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 22, 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

Motor neuron disease (MND) is a fatal neurodegenerative condition affecting both upper (UMN) and lower motor neurons (LMN). It encompasses various subtypes, notably amyotrophic lateral sclerosis (ALS), which constitutes a 90% of cases [**?**]. Other subtypes include Progressive Spinal Muscular Atrophy (PMA), Primary Lateral Sclerosis (PLS), and Progressive Bulbar Palsy (PBP), characterised by varying degrees of UMN and LMN involvement. Distinguishing between these subtypes during diagnosis is crucial due to their distinct prognoses and clinical presentations. For instance, PLS tends to have a more favourable prognosis compared to ALS due to less respiratory involvement, offering the possibility of a normal lifespan [2]. Despite advancements, there is currently no cure for ALS, with an average survival time of 2 to 4 years post-diagnosis, although a minority of patients survive beyond 10 years [3, 4, 5]. Incidence rates of ALS vary globally, with European populations and males experiencing higher rates [6, 7].

It is unknown exactly what causes MND, but genetics play a large role. Historically, ALS has been categorised into familial ALS (fALS) and sporadic ALS (sALS), distinguished by whether there is a family history of ALS. However, with 10% of sALS patients exhibiting fALS-associated mutations and a genetic contribution estimated at 61%, there is a blurred boundary between sALS and fALS causes, prompting a reclassification into “genetically confirmed“ and “non-genetically confirmed“ ALS [8, 9, 10]. As of 2022, there have been over 40 genes associated with ALS [4], the most common of which is the C9orf72 hexanucleotide repeat

expansion, which occurs in 5–15% of fALS cases [11]. 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 [12, 13].

The primary symptom of MND is progressive motor loss in voluntary muscles, sparing involuntary movements like pupillary responses [11]. 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 [14]. Ongoing research delves into the incomplete understanding of the neurodegenerative mechanism, particularly focusing on TDP-43 aggregation in the brain, a common feature in nearly all ALS cases [15].

Recently, there's been heightened attention on the cognitive and behavioural aspects of MND, affecting 35 to 50% of ALS patients, with factors like genetics and symptom onset site influencing their occurrence [16, 17]. ALS-FTD, occurring in 15% of cases, is considered a spectrum, with reciprocal links observed, such as 12.5% of behavioural-variant FTD patients progressing to ALS [18]. The most common behavioural symptoms affecting 10% of ALS patients are apathy and loss of sympathy [19]. Other common symptoms are issues with language fluency, social cognition, and executive function [20], although long-term and spatial memory are usually spared [21]. 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 El Escorial criteria, revised in 2015, categorize patients based on progressive weakness spread, aiding in diagnosis [40]. 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 [19]. 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 [22].

Disease progression is assessed using the ALSFRS-R (revised amyotrophic lateral sclerosis functional rating scale), a questionnaire evaluating patients' ability to perform disease-related functions such as swallowing or mobility, with scores ranging from 4 (normal function) to 0 (loss of function), yielding a total score ranging from 48 to 0 [23]. While often employed as the primary outcome measure in clinical trials, there is debate over its appropriateness due to its broad domain coverage, prompting suggestions to focus on domain-specific scores like limb movement or bulbar function [24, 25]. Additionally, MND patients may be staged using systems such as the King's Clinical Staging [26] or the ALS Milano-Torino Staging [27], though their widespread clinical adoption remains limited [10, 28].

Common therapies for MND, such as gastrostomy, weight maintenance, and non-invasive ventilation, are standard approaches [29]. In the UK, Riluzole is the primary treatment, potentially offering benefits in advanced stages, with a demonstrated increase in median survival by 3 months [30, 31]. Outside the UK, Edaravone shows promise in slowing disease progression but faces scrutiny regarding trial criteria and safety [33].

Imaging, such as brain magnetic resonance imaging (MRI) and spinal MRI, is currently used in the clinical pathway for MND diagnosis to rule out mimicking diseases. Although not used for prognosis clinically, neuroimaging-derived measures have been shown to be associated with patient survival and progression [63, 65]. Machine learning (ML) is a growing area of research in medical imaging, and its application to neuroimaging been shown to be useful for prognosis in other diseases [**?**]. Moreover, ML has been applied to clinical data for MND prognosis, with some success [**?**, 34].

The work presented in this thesis explores the creation of a machine learning model for MND prognosis prediction using clinical and imaging data from the

diagnostic appointment. The combination of clinical and imaging data is known as multimodal data fusion, and is a growing area of machine learning research, in the medical imaging field and beyond [1, **?**].

1.1 Motivation

Prognosis is the prediction of the course of a disease, and is important for both patients and clinicians. For patients, an accurate prognosis can help them to plan their lives and feel empowered to make decisions about their care [**?**, **?**]. For clinical use, prognosis can be used to stratify patients for clinical trials in order to investigate how different treatments affect different patient groups [**?**]. Additionally, a predicted disease course could be used as an outcome measure for clinical trials [**?**].

Imaging from the diagnostic appointment is a potentially valuable source of data for prognosis prediction. Applying imaging for this purpose would not require any extra data collection, as it is already being taken for differential diagnosis. Morever, a baseline scan is easier for patients to undergo than a scan once their disease has progressed, due to the physical limitations of progressed MND .

Since MND is multifactorial and complex, investigating the value of all available data for prognosis is a logical step. Clinical prognostic factors for MND are well established [36]. Combining these known clinical factors with imaging data could provide a more accurate prognosis than using clinical or imaging factors alone, because these ML approaches are designed to find complex patterns in large datasets that traditional statistical methods may miss.

1.2 Project Aims

The aims of this project are as follows:

Aim 1: To identify factors that are associated with MND prognosis and assess the value of imaging in survival analysis without machine learning.

Aim 2: To explore methods already developed for multimodal data fusion in the literature.

Aim 3: To apply and compare multimodal data fusion methods to MND prognosis prediction with clinical and imaging-derived features.

Aim 4: To investigate the added value of different imaging modalities and preprocessing methods, such as subregion segmentation, texture analysis, and whole brain analysis, in multimodal data fusion for MND prognosis prediction.

Aim 5: To develop the optimal multimodal data fusion model further by adding the capability to include more data modalities, such as fluid biomarkers and natural language processing (NLP) from radiological reports.

Aim 6: To assess how the performance of the proposed prognostic model compares to existing models and clinical prognostic factors, and to assess its suitability to different prognostic tasks, such as stratifying patients for clinical trials or predicting treatment need.

1.3 Upgrade Thesis Outline

Chapter 2 provides a literature review of multimodal prognostic factors in MND, machine learning for MND prognosis, and multimodal methods applied to MND. Chapter 3 presents the results of Aim 1, identifying survival factors in our dataset through Cox proportional hazards models. Chapter 4 describes multimodal data fusion in more detail and outlines the development and design of a Python package, Fusilli, for training, evaluating, and comparing multimodal data fusion methods, and presents the results of Aim 2. Chapter 5 presents the results of Aim 3, applying Fusilli to MND prognosis prediction with clinical and imaging data. Finally, Chapter 6 summarises the findings of this thesis and outlines future work.

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 [34, 35]. 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 [36]. 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 [36]. 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 [36] 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) [36] and another study showed a delay of less than one year indicates shorter survival (HR=3.43) [37].

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 [36]. A short interval between the first motor onset and the next site involvement is also associated with shorter survival [38], 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 [36, 39].

Both a smaller ALSFRS-R at diagnosis and a faster rate of ALSFRS-R decline are associated with short survival [36]. 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 [36].

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“ [40]. “Definite“ ALS patients have been found to progress faster than “probable“ or “possible“ patients in the meta-analysis and a large multi-centre study [36, 41]. However, in a large cohort of 1,809 Chinese patients, there was no significant relationship between El Escorial and survival [37].

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) [36], also supported by smaller meta-analysis [42] and a large population study [37]. 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 [43] and after diagnosis [44]. 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 [45].

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 [36]. 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 [46]. 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 developed to target specific genetic mutations [47]. 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) [48], backed up by other large cohort study [49]. 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) [48].

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 [36]. 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 [50], and the concentration of NfL plateaus around a year after symptom onset [51, 50, 52].

Higher baseline NfL concentration is associated with shorter survival, concluded from consensus of over 20 studies [53]. 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 [54].

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 [55]. Furthermore, many studies correlate brain imaging measures with ALSFRS-R, which is a measure dominated by the effects of lower motor neuron degeneration [55]. 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 [56, 57]. 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 [57]. 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) [58]. Conversely, two separate studies found that lower overall FA is associated with faster progression [59, 60].

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 [61, 62].

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 [63, 64, 65, 66, 67]. 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 [68, 66]. 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 [69].

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 ($N < 50$) [70, 71]. 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 [72]. 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 [73].

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 [73, 74]. Additionally, basal ganglia gray matter atrophy in the left caudate and right putamen is associated with faster progression [59, 75]. Initial smaller basal ganglia and amygdala volumes predict shorter survival, although significance diminishes with age of onset adjustment [76]. Furthermore, texture changes in the basal ganglia and hippocampus, detected via DTI analysis, are notable in short-term survivors [77].

The GM volumes of subcortical structures have been associated with cognitive impairment in other diseases [78]. 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 [36].

Hippocampus

Studies have reported no associations between hippocampal volume [79, 73] or FA [63] 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 [80], and Stoppel and colleagues found that increased hippocampal activation in resting-state fMRI is associated with lower ALSFRS-R [81], 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 [36]. Fast disease progression has been associated with GM atrophy [59], decreased FA [59, 67], and decrease functional connectivity [58] in the frontotemporal lobe. ALSFRS-R itself has no association with frontal areas FA [63] but has been correlated with reduced functional connectivity in the left sensorimotor cortex [82].

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 [76].

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 [83], and no correlation between brain stem FA and ALSFRS-R was found in a study of 253 patients [63]. Although, in a smaller cohort of 60 ALS patients, baseline medulla oblongata volume significantly predicted short versus long survival [84].

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 [85]. 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 [86].

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 better 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 [87]. 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 [88]. 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 [68]. 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 [89].

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 [90, 83]. "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 [91]. 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 [41]. 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 progression 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 [92]. 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 [93]. 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 [92], and a post-challenge study found highest performance with ensembles of classical ML models [94]. Some unexpected findings emerged, such as a high variability in individual ALSFRSr scores being a strong predictor of progression rate [95] 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 [96].

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 [91]. 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 [97, 98]. 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 [99] or on deterioration pattern [100]. The success of these stratification approaches suggests that the heterogeneity 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 [101]. 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 [102]. The 4 resulting papers all found that it was comparatively simple to predict the risk of clinical events, but not the timings [103, 104, 105, 106]. Other attempts at predicting time to treatment have focused on only predicting time to NIV [107, 108]. 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, incorpo-

rating clinical and biological logic into their model [109]. 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 [110]. 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) [41]. 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 [111] and Parkinson's disease [112], 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 [113], and to classify patients into neuropathological disease stages using DTI features of ALS-associated tracts [114]. 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 [115].

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 [116]. 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 [117]. 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 [118], and a later extension added simulated TDP-43 accumulation levels [119]. 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

Survival Analysis with Multimodal Baseline Clinical and Neuroimaging Data

3.1 Introduction

Many of the clinical and imaging features that were found to be associated with survival in Chapter 2 were found using the Cox proportional hazards (CPH) model. CPH is a survival analysis model that is used to investigate the relationship between the time to an event and the factors that may influence it.

As a first step in my investigation into the predictive power of clinical and imaging features in MND, I used a CPH model to investigate the relationship between the time to death and the clinical and neuroimaging features that were extracted from the ALS Biomarkers Study and Ospedale San Raffaele MND cohorts.

Previously, Querin and colleagues used a multivariable CPH to compare the predictive power of clinical and spinal cord imaging features in ALS in a cohort of 49 ALS patients, and concluded that spinal MRI measures were more predictive than clinical measures [115]. In this chapter, univariable, unimodal multivariable, and multimodal multivariable CPH models were used to investigate the relationship between the time to death and the clinical and imaging features in a larger cohort of 125 MND patients.

3.2 Data

Data from two studies was used for this survival analysis: ALS Biomarkers Study and Opsedale San Raffaele. Both of these datasets contain clinical information on MND patients and structural imaging conducted during their disease course.

The outcome of interest in this analysis is time to death, censored by date of censorship if recorded in the dataset or date of last data update if not. The Ospedale San Raffaele cohort defines their endpoint as death or tracheostomy, but the ALS Biomarkers Study only defines their endpoint as death.

The clinical features in this study are the patient's sex, baseline ALSFRS-R, diagnostic delay, age at diagnosis, site of onset (categorised as bulbar and non-bulbar), signs of FTD, and their MND subtype (categorised as ALS and non-ALS). These features were chosen for their clinical relevance to survival in MND and their availability in both datasets.

Patients were excluded if they did not have a T1- or T2-weighted MRI within 12 months before or after an MND diagnosis. Regional brain volumes were extracted from the MRI using SynthSeg [**?**], a modality-agnostic deep-learning segmentation tool. A modality agnostic tool was chosen to overcome the inconsistency in MRI protocols within the ALS Biomarkers Study and between the ALS Biomarkers Study and Ospedale San Raffaele's MND cohort. The dimensionality of the 33 regions was reduced to 12 by summing right and left regions and choosing regions relevant to MND pathology. The remaining regions were z-score normalised.

The demographics and clinical characteristics of the patients who were included in the analysis are shown in Table 3.1. Out of the 125 patients, 21 were censored and 104 died during the study period. Censored patients were those who were still alive at the end of the study period (or at the time of data download: 5th March 2024) or who were lost to follow-up before the end of the study period.

Table 3.1: Demographics and clinical characteristics of the patients included in the analysis. ALSFRSr: ALS Functional Rating Scale-Revised.

Variable	
n	125
Sex, n (%)	
Female	55 (44.0)
Male	70 (56.0)
ALSFRSr, mean (SD)	38.9 (6.4)
Survival (months), mean (SD)	34.3 (27.3)
Diagnostic Delay (months), mean (SD)	13.6 (13.2)
Age at Diagnosis (years), mean (SD)	62.6 (11.4)
Site of Onset, n (%)	
Non-bulbar	91 (72.8)
Bulbar	34 (27.2)
Frontotemporal Dementia, n (%)	
Not present	87 (69.6)
Present	38 (30.4)
MND Type, n (%)	
Not ALS	21 (16.8)
ALS	104 (83.2)
Outcome, n (%)	
Censored	21 (16.8)
Died	104 (83.2)

3.3 Methods

A CPH model is expressed as

$$h(t) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n), \tag{3.1}$$

where t is the survival time, $h(t)$ is the hazard function, X are the variables investigated, $h_0(t)$ is the baseline hazard if all the variables were 0. The hazard ratios (HRs) are represented as β, with β_1 being the hazard ratio for the variable X_1. A HR larger than 1 indicates the variable is associated with a poorer prognosis, or increased risk of the event, which is death in this CPH application. The assumption of proportional hazards is that each individual in the analysis has the same hazard function, but the hazard functions are scaled by a constant factor, which is not dependent on time. The Python package, lifelines, was used to implement the CPH models and to test the proportional hazards assumption [**?**]. Since the aim of this analysis was to see how

multimodal data affects survival analysis of MND, four models were run:

1. **Univariable model**: each feature was tested individually to see if it was associated with survival without adjusting for other features.
2. **Clinical multivariable model**: multivariable model with only clinical features.
3. **Imaging multivariable model**: multivariable model with only imaging features.
4. **Multimodal multivariable model**: multivariable model with both clinical and imaging features.

The CPH models estimated the hazard ratios and their 95% confidence intervals for each variable, with significance level set at $p < 0.05$.

The quality of models fit to the data was assessed using the concordance index, which is a measure of how well the model predicts the order of the survival times, and the Akaike Information Criterion (AIC), which is a measure of the model's goodness of fit balanced with its complexity.

3.4 Results

3.4.1 Univariable

Table 3.2 shows the HRs, confidence intervals, and significance p values of the features in the univariable CPH, multivariable clinical CPH, multivariable imaging CPH, and multivariable multimodal CPH. The significantly harmful univariable factors are an older age of diagnosis (HR=1.48), co-presence of FTD (HR=1.58), having ALS MND (HR=2.40), and larger volumes in CSF (HR=1.40) and lateral ventricles (HR=1.58). Significantly protective factors include higher baseline ALSFRS-R (HR=0.70), a longer diagnostic delay (0.77), and larger volumes in the brain stem (HR=0.66), hippocampus (HR=0.59), amygdala (HR=0.59), thalamus (HR=0.78), caudate (HR=0.77), putamen (HR=0.72), pallidum (HR=0.75) and cerebellum cortex (HR=0.71). The only features included that did not significantly affect survival were sex, bulbar site of onset, cerebral white matter volume, cerebellum white matter volume, and cerebral cortex volume.

Table 3.2: Hazard ratios of survival risk in patients with motor neuron disease for univariable and three multivariable Cox proportional hazards regressions: clinical only, imaging-features only, and clinical and imaging features together (multimodal). Acronyms: FTD - frontotemporal dementia, ALSFRS-R - revised amytrophic lateral sclerosis functional rating scale, CSF - cerebrospinal fluid.

	Univariable		Multivariable					
			Clinical		Imaging		Multimodal	
Variable	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*
Clinical								
Sex								
Female	1.00, Ref	-	1.00, Ref	-			1.00, Ref	-
Male	0.91 (0.62–1.34)	0.6417	1.13 (0.72 – 1.77)	0.5995			1.11 (0.62 – 1.99)	0.7315
ALSFRS-R	0.70 (0.60–0.82)	<**0.0001**	0.62 (0.51 – 0.76)	<**0.0001**			0.68 (0.53 – 0.88)	**0.0034**
Diagnostic delay, mo	0.77 (0.60–0.99)	**0.0409**	0.79 (0.59 – 1.06)	0.1165			0.86 (0.63 – 1.18)	0.3500
Age at diagnosis, yr	1.48 (1.29–1.84)	**0.0005**	1.53 (1.23 – 1.92)	**0.0002**			1.19 (0.84 – 1.67)	0.3265
Site of onset								
Non-bulbar	1.00, Ref	-	1.00, Ref	-			1.00, Ref	-
Bulbar	1.36 (0.88–2.10)	0.1605	0.90 (0.55 – 1.47)	0.6695			0.94 (0.54 – 1.64)	0.8280
FTD								
No	1.00, Ref	-	1.00, Ref	-			1.00, Ref	-
Yes	1.58 (1.04–2.41)	**0.0337**	1.46 (0.91 – 2.35)	0.1152			1.20 (0.69 – 2.08)	0.5177
MND Subtype								
Non-ALS	1.00, Ref	-	1.00, Ref	-			1.00, Ref	-
ALS	2.40 (1.36–4.23)	**0.0026**	1.96 (1.04 – 3.71)	**0.0384**			2.32 (1.10 – 4.88)	**0.0264**
Volumes								
Brain stem	0.66 (0.53 – 0.83)	**0.0003**			0.64 (0.41 – 1.01)	0.0557	0.64 (0.38 – 1.08)	0.0958
CSF	1.40 (1.16 – 1.70)	**0.0005**			1.10 (0.79 – 1.54)	0.5794	0.93 (0.61 – 1.43)	0.7370
Lateral ventricles	1.58 (1.32 – 1.89)	<**0.0001**			1.52 (1.06 – 2.17)	**0.0214**	1.47 (0.96 – 2.24)	0.0731
Hippocampus	0.59 (0.47 – 0.73)	<**0.0001**			0.96 (0.58 – 1.61)	0.8811	1.18 (0.67 – 2.08)	0.5663
Amygdala	0.59 (0.47 – 0.73)	<**0.0001**			0.65 (0.42 – 1.00)	0.0502	0.68 (0.43 – 1.07)	0.0927
Thalamus	0.78 (0.64 – 0.96)	**0.0173**			1.30 (0.79 – 2.14)	0.2931	1.14 (0.67 – 1.93)	0.6229
Caudate	0.77 (0.64 – 0.93)	**0.0052**			0.62 (0.41 – 0.94)	**0.0245**	0.69 (0.43 – 1.11)	0.1253
Putamen	0.72 (0.60 – 0.87)	**0.0007**			1.30 (0.71 – 2.37)	0.3884	0.88 (0.47 – 1.66)	0.6932
Pallidum	0.75 (0.61 – 0.91)	**0.0041**			0.75 (0.50 – 1.14)	0.1819	0.87 (0.57 – 1.32)	0.5007
Cerebral white matter	0.99 (0.82 – 1.20)	0.9339			2.35 (1.16 – 4.77)	**0.0181**	2.43 (1.16 – 5.10)	**0.0185**
Cerebellum white matter	0.83 (0.67 – 1.02)	0.0775			0.97 (0.56 – 1.66)	0.9054	1.14 (0.63 – 2.05)	0.6714
Cerebellum cortex	0.71 (0.57 – 0.88)	**0.0016**			0.79 (0.52 – 1.21)	0.2888	0.82 (0.51 – 1.32)	0.4135
Cerebral cortex	0.99 (0.82 – 1.21)	0.9462			0.90 (0.45 – 1.83)	0.7769	0.87 (0.41 – 1.84)	0.7118

Table 3.3: Univariable-screened hazard ratios of survival risk in patients with motor neuron disease for three multivariable Cox proportional hazards regressions: clinical only, imaging-features only, and clinical and imaging features together (multimodal). The features included are significant in univariable Cox regressions. Acronyms: FTD - frontotemporal dementia, ALSFRS-R - revised amytrophic lateral sclerosis functional rating scale, CSF - cerebrospinal fluid.

	Univariable (Significant Only)		Multivariable with Univariable Screening					
			Clinical		Imaging		Multimodal	
Variable	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*
Clinical								
ALSFRS-R	0.70 (0.60–0.82)	<**0.0001**	0.64 (0.53 – 0.77)	<**0.0001**			0.69 (0.55 – 0.88)	**0.0026**
Diagnostic delay, mo	0.77 (0.60–0.99)	**0.0409**	0.81 (0.61 – 1.07)	0.1385			0.83 (0.61 – 1.13)	0.2346
Age at diagnosis, yr	1.48 (1.29–1.84)	**0.0005**	1.52 (1.21 – 1.9)	**0.0003**			1.03 (0.74 – 1.42)	0.8646
FTD								
No	1.00, Ref	-	1.00, Ref	-			1.00, Ref	-
Yes	1.58 (1.04–2.41)	**0.0337**	1.42 (0.89 – 2.26)	0.1429			1.17 (0.69 – 2.01)	0.5561
MND Subtype								
Non-ALS	1.00, Ref	-	1.00, Ref	-			1.00, Ref	-
ALS	2.40 (1.36–4.23)	**0.0026**	2.01 (1.09 – 3.71)	**0.0262**			2.11 (1.03 – 4.32)	**0.0410**
Volumes								
Brain stem	0.66 (0.53 – 0.83)	**0.0003**			0.67 (0.47 – 0.94)	**0.0218**	0.74 (0.51 – 1.06)	0.1041
CSF	1.40 (1.16 – 1.70)	**0.0005**			1.33 (1.02 – 1.73)	**0.0332**	1.24 (0.92 – 1.68)	0.1562
Lateral ventricles	1.58 (1.32 – 1.89)	<**0.0001**			1.63 (1.14 – 2.32)	**0.0072**	1.56 (1.03 – 2.36)	**0.0374**
Hippocampus	0.59 (0.47 – 0.73)	<**0.0001**			1.10 (0.68 – 1.78)	0.7018	1.38 (0.81 – 2.37)	0.2371
Amygdala	0.59 (0.47 – 0.73)	<**0.0001**			0.65 (0.42 – 0.99)	**0.0473**	0.62 (0.39 – 0.97)	**0.0381**
Thalamus	0.78 (0.64 – 0.96)	**0.0173**			1.74 (1.11 – 2.71)	**0.0149**	1.38 (0.86 – 2.23)	0.1844
Caudate	0.77 (0.64 – 0.93)	**0.0052**			0.64 (0.43 – 0.97)	**0.0339**	0.73 (0.46 – 1.14)	0.1609
Putamen	0.72 (0.60 – 0.87)	**0.0007**			1.55 (0.87 – 2.73)	0.1338	1.19 (0.67 – 2.11)	0.5556
Pallidum	0.75 (0.61 – 0.91)	**0.0041**			0.77 (0.52 – 1.13)	0.1779	0.84 (0.58 – 1.22)	0.363
Cerebellum cortex	0.71 (0.57 – 0.88)	**0.0016**			0.81 (0.58 – 1.11)	0.1900	0.86 (0.60 – 1.23)	0.4093

Table 3.4: Metrics assessing the Cox proportional hazards models' fits: c-index (concordance index) and AIC (Akaike Information Criterion).

			Fit metrics	
Model	**Factors Included**	**Number of Factors**	c-index	AIC
Clinical	All variables	7	0.73	799.66
	Univariable-screened	5	0.73	796.40
Imaging	All variables	13	0.75	794.90
	Univariable-screened	10	0.74	799.60
Multimodal	All variables	20	**0.78**	**787.53**
	Univariable-screened	15	0.77	789.11

3.4.2 Multivariable

3.4.2.1 Clinical

When the clinical features were input into a multivariable CPH, baseline ALSFRS-R was a significant protective factor (HR=0.62), and ALS MND and older age at diagnosis were significantly harmful (HRs of 1.96 and 1.53). Diagnostic delay and co-presence of FTD were no longer significant in the multivariable CPH.

The proportional hazards assumptions were broken by age at diagnosis ($p = 0.002$) and diagnostic delay ($p = 0.038$). In an effort to correct the broken assumptions, another multivariable CPH was fit with only the univariably-significant clinical factors. Table 3.3 shows the results from the univariable-screened CPH models. This fixed the broken proportional hazards assumptions and the same factors remained significant: high baseline ALSFRS-R (HR=0.64), age at diagnosis (HR=1.52), and ALS MND (HR=2.01).

3.4.2.2 Imaging

The multivariable imaging CPH resulted in three significant survival factors: lateral ventricles (harmful, HR=1.52), cerebral white matter (harmful, HR=2.35), and caudate (protective, HR=0.62). However, the cerebral white matter and CSF variables broke the CPH assumptions ($p = 0.0325$ and 0.005 respectively).

When only the univariably-significant imaging factors are input into a multivariable CPH, more factors were significantly associated with survival, shown in Table 3.3. Higher volumes of the brain stem (HR=0.67), amygdala (HR=0.65), and

caudate (HR=0.64) were protective, and higher volumes of the CSF (HR=1.33), lateral ventricles (HR=1.63), and thalamus (HR=1.74) were harmful.

3.4.2.3 Multimodal

Only three factors were significant in the multimodal multivariable CPH: baseline ALSFRS-R (HR=0.68), ALS MND (HR=2.32), and cerebral white matter volume (HR=2.43). The proportional hazards assumption was broken by sex ($p = 0.0069$), bulbar site of onset ($p = 0.0298$), and cerebral white matter ($p = 0.0497$).

Screening input variables by their univariable significance resulted in a CPH that had no broken assumptions and four significant survival factors: two clinical and two imaging. Higher baseline ALSFRS-R (HR=0.69) and larger amygdala volume (HR=0.62) were significantly protective, and ALS MND (HR=2.11) and larger lateral ventricle volume (HR=1.56) were significantly harmful.

Table 3.4 shows the metrics of model fit for the multivariable CPH models. The multimodal multivariable models resulted in the best fit metrics, both when considering all the models and also when considering only the univariable-screened models and the "all variable" models.

3.5 Discussion

The univariably-significant clinical factors in this analysis were consistent with the existing literature. However, having a bulbar site of onset was not significantly associated with survival, which was surprising, as bulbar onset is often associated with a poorer prognosis in MND and was found to be significant in a meta-analysis of survival factors [36]. 27.2% of the patients in this analysis had a bulbar site of onset, which is consistent with the general MND population [10], so it is unlikely that the lack of significance was due to a small sample size. Our analysis categorised patients as having a bulbar or non-bulbar site of onset, but a respiratory site of onset is associated with a poorer prognosis than bulbar onset [36], so it is possible that the lack of significance was due to the categorisation of the site of onset. Future work could investigate the relationship between the site of onset and survival in MND by using a more granular categorisation of the site of onset, such as bulbar, respiratory,

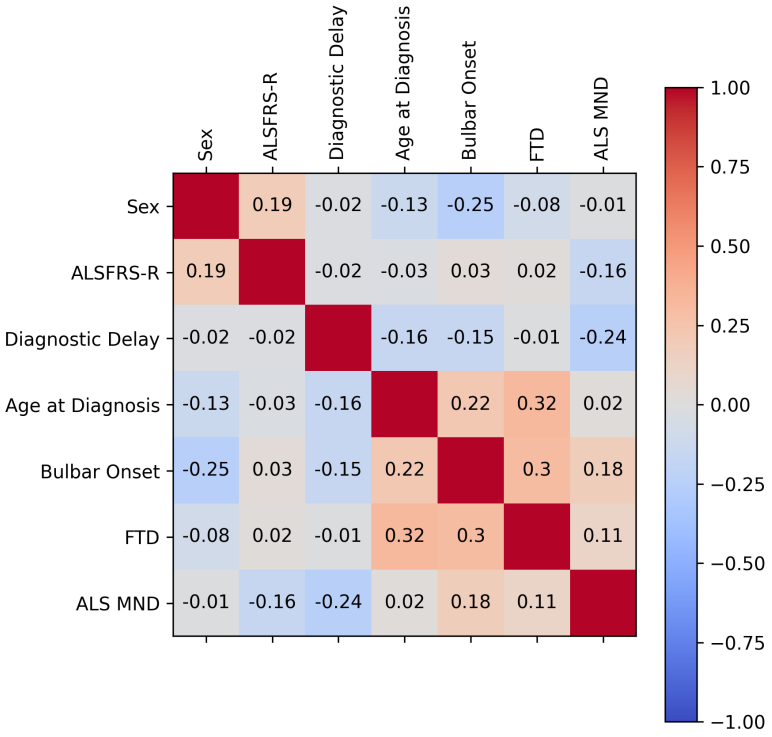

Figure 3.1: Correlation between the clinical features input into the Cox proportional hazards survival model, as calculated by Pearson's correlation coefficient.

and limb onset.

However, in general, a univariable model is not the ideal model for survival analysis of a multi-factorial disease like MND, as it does not account for the effects of other variables on the outcome.

The multivariable models all resulted in broken proportional hazards assumptions, which were fixed by screening the input variables by their univariable significance. This approach is not without controversy, as it is possible that features that are not significant univariably are significant in the multivariable model. However, we did not find this to be the case in our analysis, since the screened imaging and multimodal models resulted in more significant features than the unscreened models.

The multivariable imaging model resulted in some clinically unintuitive results, such as reported significantly harmful effects of larger cerebral white matter volume and larger volume of the thalamus. In the univariable analysis, larger thalamus volume was significantly protective, and larger cerebral white matter volume was not

significantly associated with survival. Moreover, previous work suggests that atrophy in the thalamus suggests faster disease progression, although this is not necessarily associated with survival [59, 73]. By its nature, the regional brain volumes are highly correlated, which increases the risk of collinearity in the model. Collinearity can lead to unstable estimates of the coefficients, and possibly explain a variable effect switching from harmful to protective or vice versa.

One way to mitigate the effect of collinearity in the model is to use dimensionality reduction techniques, such as principal component analysis (PCA), to reduce the number of features in the model. Westeneng and colleagues used this technique to group the brain into larger regions based on their principal components before applying to a survival analysis model [76].

Diagnostic delay and FTD lost significance in the multivariable clinical model (both screened and unscreened) when age at diagnosis was included. This lost significance could be explained by the correlations in the clinical features, shown in Figure 3.1. In multivariable models, the features that are correlated with other features are less likely to be significant, because the effect of the correlated features is already accounted for by the significant features. Figure 3.1 shows that age at diagnosis is the most positively correlated with FTD, which could explain why FTD lost significance in the clinical model when age at diagnosis was included, and age at diagnosis remained significant. Moreover, diagnostic delay is the most negatively correlated with ALS MND, which resulted in diagnostic delay losing significance, even though it was significant univariably.

The presence of the amygdala in the final multimodal significant features is interesting, since brain regions which are more commonly associated with MND, such as the brain stem, were not significant. This could, again, be explained by the correlations between the features included in the model. Figure 7.1 in the appendices shows that the most negative correlation between FTD and a brain region is with the amygdala (coefficient of -0.40), which could explain why FTD lost significance in the clinical model when the amygdala was included. Moreover, the brain stem is highly correlated with the amygdala (coefficient of 0.60), which could explain why

the brain stem lost significance in the multimodal model, even though it is implicated in lower motor neuron degeneration in MND .

Another example of this is that the age at diagnosis was not significant in the multimodal model, even though it was significant in the clinical model. Inspecting the correlations between the clinical features and the brain volumes found that age at diagnosis is correlated with lateral ventricle volume (coefficient of 0.49), which could explain why age at diagnosis lost significance in the multimodal model when lateral ventricle volume was included.

The absence of these clinical variables in the imaging multivariable CPH may have led to the model picking up on brain structure not directly influencing survival, but markers of clinical features that are influencing survival. For example, older age at diagnosis, which is associated with brain atrophy, and FTD, which is associated with limbic atrophy leading to three limbic structures being found as significant: thalamus, amygdala, and caudate. The multimodal CPH mitigated the effect of the model picking up on brain structure not directly influencing survival, because we included the clinical features that are influencing survival into the model as well. However, the trade-off was that the collinearity within the brain volumes but also between the brain volumes and the clinical features was high, which could mean that individual volumes are picked out as significant over clinical features and vice versa, even if both are influencing survival.

Finally the fit statistics in Table 3.4 show that the fit improved with multimodal features. Although adding more features to a survival model is likely to increase its fit metrics, AIC also considers the number of features in the model in its calculation, so it can be concluded that the multimodal features are adding more information to the model than the clinical or imaging features alone.

3.5.1 Limitations

The counter-intuitive results in the imaging multivariable model were likely due to limitations in this analysis. Firstly, the sample size was relatively small ($n = 125$), which could mean that the results were due to cohort-specific factors. Moreover, the data came from two studies and multiple sites, and no harmonisation was per-

formed to account for site-specific factors, although an argument could be made that removing site-specific factors could remove important information about the disease. Further work could be done to investigate the effect of the sites on the results, such as by including site as a variable in the model.

Secondly, the scans were not generally taken at the same time as diagnosis, which could mean that the brain volumes are not representative of the patient's brain at the time when the clinical features were recorded. MND is a fast-progressing disease and allowing for 12 months either side of diagnosis for including the scan is a wide window. However, this wide window was a trade-off between sample size and data quality. In the future, we aim to increase the sample size, which would allow us to look at scans taken closer to diagnosis to check the consistency of the results.

The collinearity of the brain volumes introduced instability into the model. The summing of the left and right regions was done to mitigate this, but this could have also introduced information loss if the left and right regions were not equally important to survival. Future work to mitigate this further would be to implement dimensionality reduction techniques before inputting the brain volumes into the model.

Finally, the univariable screening of the features could have introduced bias into the model, as it is possible that features that are not significant univariably are significant in the multivariable model. Although we did not find this to be the case, there are other ways to attempt to fix the proportional hazards assumption, such as using time-dependent covariates or stratifying the model by a variable, and there are other ways to screen the features, such as using a LASSO regression to select the most important features.

3.6 Conclusion

In this chapter, we have shown that both brain region volumes and clinical features were significantly associated with survival in MND, and that including both data modalities in a survival model increased the model's fit metrics. The clinical features that were significantly associated with survival were consistent with the existing

literature, but some of the imaging features were not, until the clinical features were included in the model. This suggests that incorporating imaging-derived features with clinical features is a valuable approach to survival analysis in MND, and that the multimodal model is more informative than the clinical or imaging models alone.

The issues encountered during the analysis were likely due to the collinearity of the brain volumes and the correlations between the brain volumes and the clinical features. Future work will aim to address these issues by extending this analysis to include dimensionality reduction and improved clinical characterisation of the patients, as well as increased sample size.

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

Multimodal data fusion is the process of combining data from different sources to make predictions or decisions, often through the use of deep learning. The goal of combining different modalities is to improve the performance of a model by leveraging the relevant information from each modality and fusing them in a way that improves the model's performance. There are many research fields where multimodal data fusion is used, such as in agriculture to predict crop yields and detect diseases [120, 121], in disaster management to analyse response scenarios from audio and social media posts [122], and in robotics to help direct the robots with multiple sensors [123]. Moreover, the types of models used in multimodal data fusion can vary a lot, from geometric deep learning to relatively simple neural network architectures [1].

My PhD investigates multimodal data fusion for predicting MND prognosis, an area with minimal deep learning research [96, 110] and only one model specially-created for deep-learning based multimodal data fusion [118]. Despite systematic reviews on the topic attempting qualitative comparison between models [1, 124, 125, 126], a lack of quantitative comparison necessitates experimentation with many models from other fields to identify the most effective for MND prognosis prediction.

However, acquring a diverse range of models for experimentation is made difficult by the use of variable terminology and the common absence of maintained, quality code available in studies.

A way to address these problems is to create a curated collection of models for somebody interested in multimodal data fusion to consult. As far as I am aware, there are three Python packages that house collections of deep learning based data fusion models: "Multi-view-AE" [127], "CCA-Zoo" [128], and "pytorch-widedeep" [129]. However, each of these packages only includes models with specific frameworks (autoencoders, CCA, and Google's "wide and deep" models, respectively), which limits the variety of models available for comparison.

Therefore, I aimed to develop a Python package for training and comparing multimodal data fusion models with any architecture. This Python package is named Fusilli, as a portmanteau of "fuse easily". Fusilli works by taking the user's multimodal data and training it on a variety of models, and then comparing the models' performances.

4.2 Development and Implementation

4.2.1 Software Design Choices

Before developing Fusilli, the following design goals were set to ensure the package would be useful for a wide range of users and tasks, as well as for my own research.

Modularity: Fusilli should be modular, meaning that the various functionalities within the package should be independent of each other. This would allow for easy addition of new models in the future and easy adjustments to the package's functionality. Testing, which is the process of writing code to check that the package

works as expected, would also be easier with a modular design, and so the package would be more reliable.

Beginner-friendly and expert-friendly: Fusilli should be beginner-friendly, with users able to compare the different models without needing expertise in deep learning or Python programming. This would make it different from other similar packages, which require the user to set up their own experiments.

On the other hand, Fusilli should also be expert-friendly, with users who are more capable being able to change the training parameters, modify the models, and access the trained models for further experiments.

Wide applicability: Fusilli should include a wide variety of models, to ensure that the best model for a given task can be found. Moreover, this would ensure that the package is useful for a wide range of users, as different users may have different requirements for model architectures based on their task and data.

Support for two modalities: The models in Fusilli will support data fusion between either two types of tabular data (e.g. clinical data and brain region volumes data) or between an image and tabular data (e.g. MRI images and clinical data). Again, for wide applicability, Fusilli should be able to handle two-dimensional and three-dimensional images, and tabular data of any size.

Support for different prediction tasks: The prediction tasks that Fusilli should support are regression and classification. From the literature, these are the most common tasks for multimodal data fusion, and are the tasks that I am interested in for my own research.

4.2.2 Implementation

Fusilli was implemented in Python, using the PyTorch and PyTorch Lightning libraries for deep learning. This is because PyTorch is a popular library for deep learning, and is known for its flexibility and ease of use.
The user must specify:

- **Data**: The user's data, which must be in the form of a .csv file for tabular data,

and .pt files for images.

- **Task**: The task that the user wants to perform, which can be either regression or classification (binary or multiclass).
- **Models**: The models that the user wants to compare, which can be any of the models included in Fusilli.
- **Output**: The output directory for the trained models and the results of the experiments.

The user can also specify experiment specifics, which are set to default values if not specified. These are:

- Training and validation data splits.
- Maximum number of epochs to train for, including early stopping.
- Any model hyperparameters or architectures that the user wants to modify.
- Batch size.

After the user has specified these, the user calls two functions in Fusilli to train the models:

- `prepare_fusion_data()`: This function prepares the data for training, including splitting it into training and validation sets and running any model-specific data preparation steps.
- `train_and_save_models()`: This function trains the models on the user's data, and outputs the trained models.

Finally, the user can call functions to evaluate a single model or compare the models, which will output the performance of the models on the user's validation data or external test data. The evaluation figures are saved in the output directory, and the user can also access the trained models for further analysis.

4.2.3 Fusion Methods

A literature search was conducted to find models that could be included in Fusilli. The search criteria aimed to return papers that mentioned machine learning, multimodality, and image and tabular data, with variants of these terms used to capture

Table 4.1: Descriptions of data fusion model categories used in Fusilli, first proposed by Cui and colleagues [1].

Category	Description
Operation-based	Models that fuse data based on operations such as concatenation, addition, or multiplication. This can be done at any point in the model, such as at the input, hidden layers, or output.
Attention-based	Models that use attention mechanisms to weight the importance of different modalities.
Graph-based	Models with a graph structure, such as graph convolutional networks.
Subspace-based	Models that project the data into a joint lower-dimensional space, possibly with deep learning methods such as autoencoders.
Tensor-based	Models that use tensor operations to fuse the data to capture inter- and intra-modality correlations.

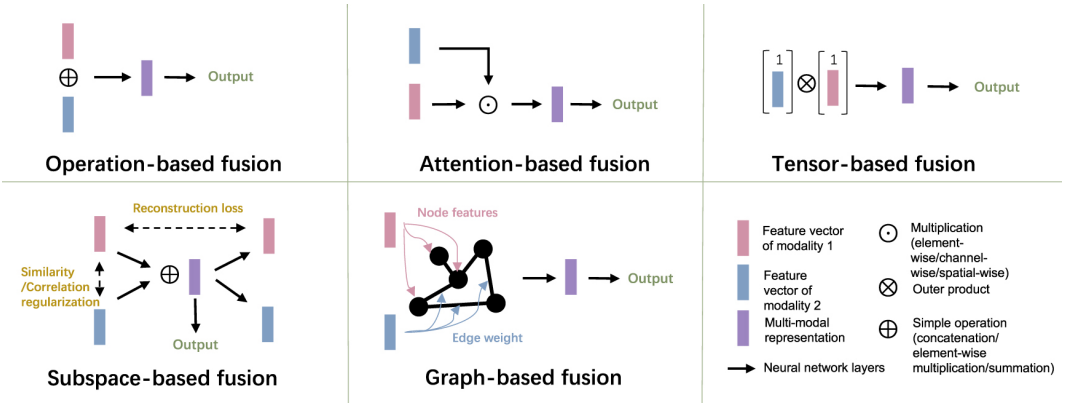

Figure 4.1: Cui and colleagues' diagram of the architecture-based taxonomy of multimodal data fusion models used in Fusilli [1].

a wide range of papers. The resulting papers were checked for relevance and papers were discarded which use the same model as another paper, which happened frequently.

The models were categorised based on the taxonomy defined by Cui and colleagues in their review on data fusion methods for diagnosis and prognosis [1]. Figure 4.1 is a diagram taken from the review, which shows the general architecture differences between the categories, and Table 4.1 includes descriptions of the categories.

Another common categorisation is "early", "intermediate", and "late" fusion,

which refers to the point in the model where the data is fused. However, this categorisation is too simplistic for the variety of models architectures in Fusilli, and so the architecture-based taxonomy was chosen.

More models were found in the literature than were included in Fusilli due to the time constraints of the project, but the models included in Fusilli were chosen based on their popularity, availability of code, and the variety of architectures they represent. Moreover, unimodal benchmarks were included in Fusilli to allow for comparison between the multimodal models and unimodal models.

Table 4.2: A list of the models included in Fusilli v1.2.3, categorised by their fusion method and modalities. References are included where applicable, although the Fusilli implementation is not a direct copy of the referenced model, and may have been modified to fit the package's requirements. **Acronyms**: GNN = Graph Neural Network, MCVAE = Multi-Channel Variational Autoencoder.

Model name (and reference where applicable)	**Fusion**	**Modalities**
Tabular1 uni-modal	Unimodal	Tabular Only
Tabular2 uni-modal	Unimodal	Tabular Only
Image unimodal	Unimodal	Image Only
Activation function map fusion [130]	Operation	Tabular-tabular
Activation function and tabular self-attention [130]	Operation	Tabular-tabular
Concatenating tabular data	Operation	Tabular-tabular
Concatenating tabular feature maps [131]	Operation	Tabular-tabular
Tabular decision	Operation	Tabular-tabular
Channel-wise multiplication net (tabular) [132]	Attention	Tabular-tabular
Tabular Crossmodal multi-head attention [133]	Attention	Tabular-tabular
Attention-weighted GNN [134]	Graph	Tabular-tabular
Edge Correlation GNN	Graph	Tabular-tabular
MCVAE Tabular [135]	Subspace	Tabular-tabular
Concatenating tabular data with image feature maps [136]	Operation	Tabular-image
Concatenating tabular and image feature maps [131]	Operation	Tabular-image
Image decision fusion	Operation	Tabular-image
Channel-wise Image attention [132]	Attention	Tabular-image
Crossmodal multi-head attention [133]	Attention	Tabular-image
Trained Together Latent Image + Tabular Data [137]	Subspace	Tabular-image
Pretrained Latent Image + Tabular Data [137]	Subspace	Tabular-image
Denoising tabular autoencoder with image maps [138]	Subspace	Tabular-image

4.3 Results

Fusilli version 1.0.0 was published on 30th November 2023, and has since been updated to version 1.2.3. Currently, Fusilli is under review for publication in the Journal of Open Source Software (JOSS). The package has been well-received on GitHub, having been favourited by 138 users and specially cloned by 12 users for their own modifications, which is a lot of attention for a package that has only been available for a few months. Moreover, I wrote an article on Medium about Fusilli, which has been read by 243 people, and two articles have been written about Fusilli by other people.

Fusilli is documented using Sphinx, a Python documentation generator. Figure 4.2 shows an overview of the documentation, which includes tutorials for getting started, guidance for more advanced users, example Python scripts which run the models on simple data, a guide for people to contribute their own models, and the source code documentations. The page "Fusion Model Explanations", shown in Figure 4.2B, has a diagram and a text explanation for each model in Fusilli, which is useful for users to understand and compare the models' architectures with a standardised diagram style.

Figures 4.3A, B, and C show the figures output by Fusilli after training, validating, and comparing the models respectively. The format of the figures depends on the prediction task and the type of training used, with the figures shown being for a binary classification task with 5-fold cross validation. If the user chooses to use train-test split training, the comparison figure is a bar chart instead of a violin plot. Additionally, if the user trains a regression model, the performance evaluation will be a scatter plot of the predicted values against the true values.

4.4 Discussion

Overall, Fusilli achieves the goals set out at the beginning of the project. It has had interest from both beginners and experts, and there has been interest from people in different fields, such as thermal imaging. Additionally, a hackathon project was run during the 2023 Centre for Medical Imaging Hackathon, where participants aimed

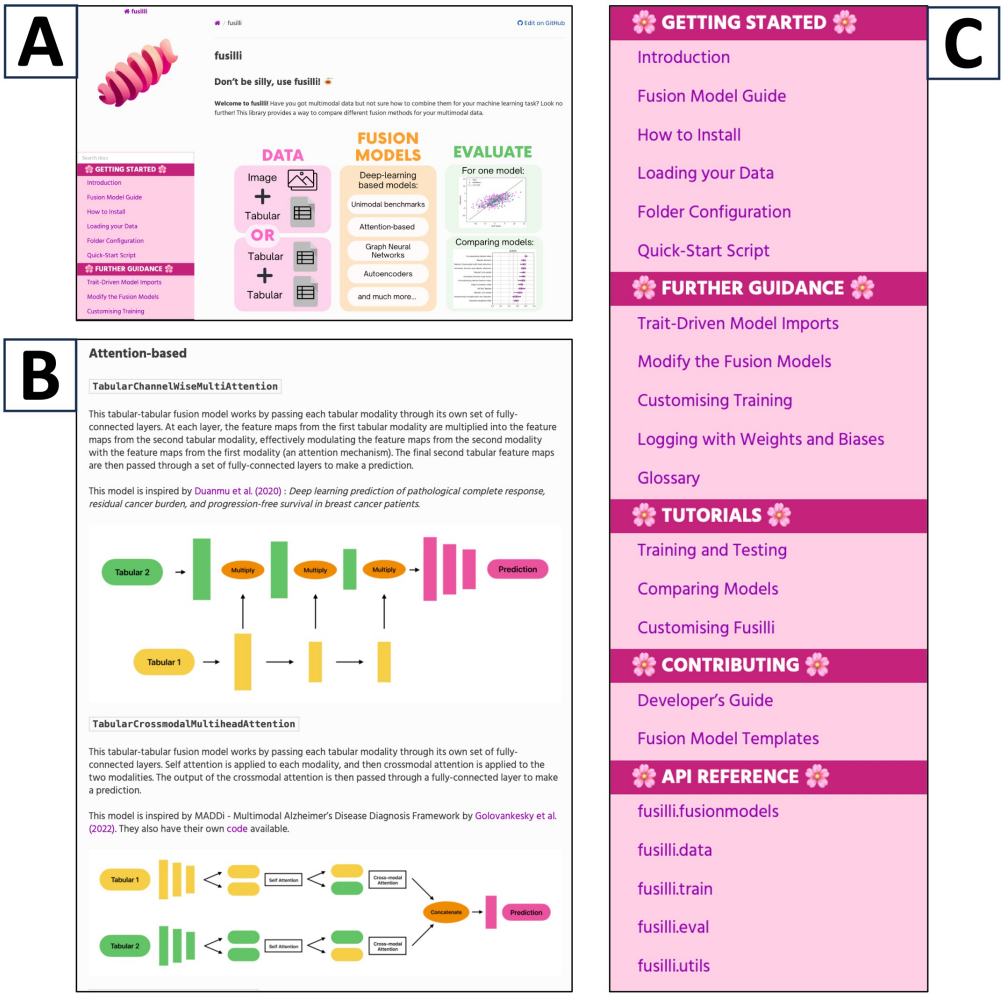

Figure 4.2: Overview of the Fusilli documentation.
A: The home page of the documentation, showing the logo and the diagram explaining Fusilli's purpose.
B: Two examples of the fusion model explanations, complete with diagram and description with reference to source paper.
C: The sidebar of the Fusilli documentation showing all the pages available.

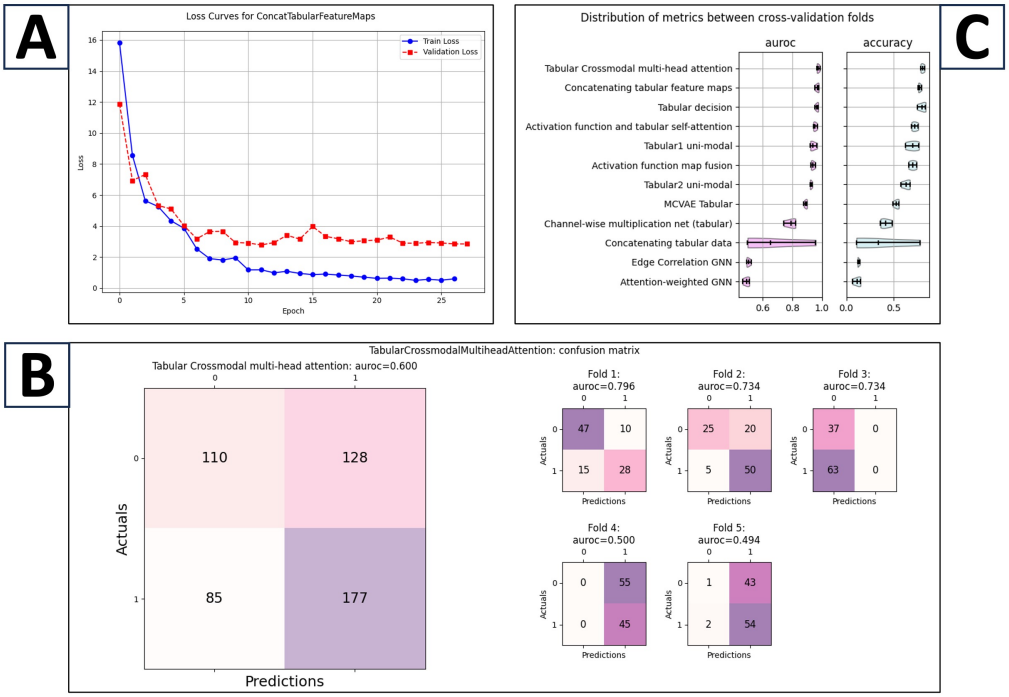

Figure 4.3: Fusion model training, evaluation, and comparison figures output by Fusilli.
A: The training and validation loss curves for an individual model.
B: Performance evaluation of a model which was trained on a binary task with 5-fold cross validation. Both the validation performances of each fold and the overall performance from the aggregated folds are shown.
C: Comparison of the models' performances on the validation data. The violin plot distributions are the distributions of the fold performances when using k-fold cross validation. If train-test split training is used, the comparison figure is a bar chart.

to add new models to Fusilli, which was a success with two new models being added and more in development.

Fusilli has a number of limitations, most of which I hope to address with future updates to the software. Firstly, Fusilli only supports two modalities, which is a limitation for tasks that may benefit from more than two modalities. For example, in the context of MND prognosis prediction, using different types of MRI images (e.g. T1-weighted, T2-weighted, diffusion tensor imaging) could be beneficial, but Fusilli currently only supports one type of image alongside clinical information. Moreover, Fusilli only supports inputting image data the form of PyTorch .pt files, which requires the user to do some pre-processing of their data. This is a limitation

for users who may not be comfortable with Python, and would prefer to input their data in the form of JPEGs or NiFTI files, for example. Fusilli also only supports regression and classification tasks, and does not support other tasks such as time-to-event analysis. Finally, although the models included in Fusilli are popular and cover a wide variety of architectures, there are many more models in the literature that could be included in Fusilli, and I hope to include these.

4.5 Conclusion

In conclusion, Fusilli is a Python package for multimodal data fusion experimentation and analysis, specifically tackling the problem of lack of ways to compare models and lack of standardisation in the field. I specifically made it for my PhD in fusing different data modalities in MND prognosis prediction. The next chapter will discuss the application of Fusilli to MND prognosis prediction, using tabular-tabular fusion of clinical information and brain region volumes.

Chapter 5

Comparison of Multimodal Data Fusion Models for Predicting Survival Classification in Motor Neuron Disease

5.1 Introduction

Only one deep learning model has been used to predict survival in MND through clinical data and neuroimaging data [118]. This lack of research in multimodal data fusion in MND makes it difficult to conclude whether multimodal data is useful in predicting survival in MND . Additionally, we have shown that many different multimodal data fusion methods have been developed for other research applications. In Chapter 4, we developed a package called Fusilli to compare the performance of multimodal data fusion methods, approximately half of which are designed to combine two tabular modalities.

In this chapter, we will use Fusilli to compare the performance of different tabular-tabular multimodal data fusion methods on predicting survival in MND patients using clinical and imaging extracted features data. The aims of this work are to, firstly, assess the effect of different data fusion model architectures on prognostic performance, and secondly, to assess the value of baseline clinical and neuroimaging

data in MND prognosis prediction.

5.2 Data

The data used in this analysis is from two studies: University College London Queen's Square Institute of Neurology's ALS Biomarkers Study [**?**] and Ospedale San Raffaele's MND cohort. Both of these datasets contain clinical data from the diagnostic visit and brain MRI data.

For patients to be included in the analysis, they must have a diagnosis of MND and have an outcome of interest (death or tracheostomy). In the ALS Biomarkers Study, the outcome of interest is death, whereas in the Ospedale San Raffaele's MND cohort, the outcome of interest is death or tracheostomy. Unfortunately, the Ospedale San Raffaele's MND cohort does not specify the date of death, so we have assumed that the date of death is the date of tracheostomy.

Patients in this analysis must have non-missing data for age at diagnosis, sex, date of diagnosis, date of death, date of symptom onset, site of onset, and baseline ALSFRS-R . Additionally, patients must have a T1-weighted or T2-weighted MRI within 12 months before or after their date of diagnosis.

The final dataset contains 110 MND patients. The patients in the cohort were split into two groups based on the median survival time: short survival group (less than 24 months) and long survival group (more than 24 months).

5.2.1 Clinical Data

The clinical variables chosen to be included in the analysis are based on the variables used in the ENCALS model [41] described in Chapter 2. El Escorial criteria and FVC were not included in the analysis as they were not available in the Ospedale San Raffaele's MND cohort. Features with missing data after the inclusion criteria were applied were features of FTD and presence of C9orf72 mutation. Where these features were missing, they were assumed to be negative. Moreover, where the MND type was missing, it was assumed to be ALS, as it is the most common.

Table 5.1 shows the clinical features included in this analysis and statistical differences between the long and short survival groups. The longer survival group

Table 5.1: Differences in clinical demographics between the long and short survival groups. PRB is progression rate to baseline, calculated as the rate of decline of ALSFRS-R between symptom onset and diagnosis.
*Chi-square test, †Fisher's exact test, ‡ Two-sample t-test.

	Overall	**Short**	**Long**	**P-Value**
n	110	55	55	
Categorical, n (%)				
Sex (Male)	52 (47.3)	27 (49.1)	25 (45.5)	0.849*
Bulbar Onset	31 (28.2)	20 (36.4)	11 (20.0)	0.090*
FTD	32 (29.1)	24 (43.6)	8 (14.5)	**0.002***
C9orf72	7 (6.4)	2 (3.6)	5 (9.1)	0.438†
ALS	96 (87.3)	51 (92.7)	45 (81.8)	0.153*
Continuous, mean (SD)				
ALSFRS-R	37.5 (7.2)	36.3 (7.2)	38.7 (7.0)	0.081‡
Diagnostic Delay, mo	12.5 (12.0)	10.0 (9.8)	14.9 (13.4)	**0.031**‡
Age at Diagnosis, yr	63.2 (11.8)	69.1 (9.1)	57.3 (11.3)	<**0.001**‡
PRB (points/month)	1.4 (1.7)	2.0 (2.2)	0.9 (0.7)	**0.001**‡
Survival, mo	29.3 (23.2)	12.3 (6.4)	46.4 (21.3)	<**0.001**‡

Table 5.2: Differences in clinical demographics between the two data sites: the ALS Biomarkers Study from University College London and Ospedale San Raffaele. PRB is progression rate to baseline, calculated as the rate of decline of ALSFRS-R between symptom onset and diagnosis.
*Chi-square test, †Fisher's exact test, ‡ Two-sample t-test.

	Overall	ALS Biomarkers Study	Ospedale San Raffaele	P-Value
n	110	46	64	
Categorical, n (%)				
Sex (Male)	52 (47.3)	26 (56.5)	26 (40.6)	0.146*
Bulbar Onset	31 (28.2)	20 (43.5)	11 (17.2)	**0.005***
FTD	32 (29.1)	16 (34.8)	16 (25.0)	0.367*
C9orf72	7 (6.4)	3 (6.5)	4 (6.2)	1.000*
ALS	96 (87.3)	44 (95.7)	52 (81.2)	0.052‡
Continuous, mean (SD)				
ALSFRSr	37.5 (7.2)	34.1 (8.4)	40.0 (5.0)	<**0.001**†
Diagnostic Delay, mo	12.5 (12.0)	11.9 (10.2)	12.8 (13.2)	0.682*
Age at Diagnosis, yr	63.2 (11.8)	66.2 (12.1)	61.1 (11.1)	**0.028**‡
PRB (points/month)	1.4 (1.7)	1.9 (2.3)	1.0 (1.1)	**0.017**‡
Survival, mo	29.3 (23.2)	24.3 (26.8)	33.0 (19.6)	0.066‡

had significantly fewer patients with FTD, a longer diagnostic delay, a younger age at diagnosis, and a slower rate of decline in ALSFRS-R . These differences are consistent with the literature on factors associated with survival in MND [36].

Table 5.2 shows the group differences between the two data sites. The cohort from Ospedale San Raffaele had a significantly lower proportion of bulbar onset patients, a higher mean baseline ALSFRS-R score, a lower mean age at diagnosis, and a slower rate of decline in ALSFRS-R . These differences, in combination with survival factors suggested in the literature, suggest that the cohort from Ospedale San Raffaele have had less aggressive disease progression at diagnosis compared to the ALS Biomarkers Study cohort.

Statistically significant differences between the sites: **List them here** Why didn't we do any site-specific analysis or correction? - Wanted to see how the model would perform in a real-world setting - Not enough data to do one site

5.2.2 Imaging Data

The same segmentation pipeline used in Chapter 3 was used for this analysis also: using SynthSeg to segment MRI conducted within 12 months before or after diagnosis. SynthSeg returns the volumes of 33 regions of the brain, which, apart from intra-cranial volume, were used as features in this analysis. The region volumes were z-score normalised across the entire cohort.

The left and right volumes were summed to simplify the comparison of regional brain volumes between the long and short survival groups, but the left and right volumes were kept separate for the main analysis. Two-sample t-tests were used to compare the regional brain volumes between the long and short survival groups. The long survival group had significantly larger volumes in the cerebellum white matter and cortex, thalamus, caudate, putamen, pallidum, brain stem, hippocampus, amygdala, accumbens area, and ventral diencephalon. The short survival group had significantly larger volumes in the cerebrospinal fluid, 3rd ventricle, lateral ventricle, and inferior lateral ventricle.

5.3 Methods

Prediction task - split survival time on median. - Why did we do this? Small sample size so needed to split the data to get a reasonable number of patients in each group - Median survival time is 24 months which is a reasonable cut off for short and long survival going off of epidemiological studies of MND, 2 to 4 years after diagnosis

What models - Fusilli tabular-tabular methods - Diagrams in the appendix - table of the models included, and a brief description of each, number of parameters - Why can't we use graph based models in fusilli: currently not set up to do external test set on them - Benchmark unimodal models

Three experiments: - Using both sites together: Data split into training and testing: 20% (n=22) kept back - Train on one site, test on the other: Essex trained on Milan, Milan trained on Essex. The other site is not used in the training at all

Training set up - All the experiment's training sets used with K-fold cross validation - 5 folds - why? Small sample size so improving stability - The test data was run in each of the models' folds, then the predictions were aggregated and the performance metrics calculated on the aggregated predictions - Another method for improving stability is running repetitions of the training and evaluation process until the mean of the performance of the repetitions converges to a stable value (within 1% of the main metric: balanced accuracy) - Early stopping on the validation fold: minimising the loss of the validation fold, patience of 5, minimum change of 0.001

Metrics: - Table of metrics we're using and their definitions

5.4 Results

5.4.1 Training together

Results observations from the figure: - Figure 5.1 has the models ranked by their balanced accuracy on the test set on the left, and the corresponding validation set balanced accuracy on the right, aggregated over the 5 validation folds. Distributions that make the violin plots are over the repetitions of training on both sites. - Table 5.3 shows the mean and standard deviations of the test metrics over the repetitions of training on both sites. - Early concatenation performed the best with 70% balanced

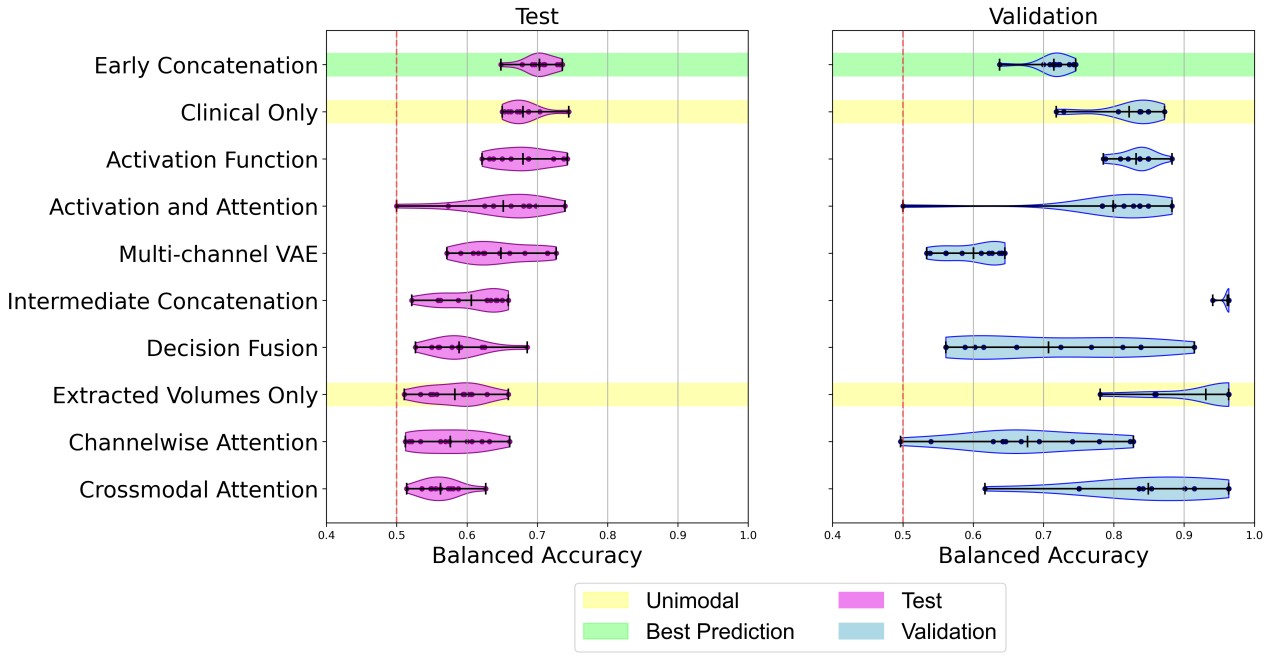

Figure 5.1: Balanced accuracies of test set and validation set when sites trained together.

Table 5.3: Means and standard deviations of test metrics over repetitions of training on both sites. Bold is the best metric model.

	Performance metric: mean (standard deviation)				
Method	Balanced Accuracy	AUROC	Precision	Recall	F1
Early Concatenation	**0.703 (0.023)**	**0.815 (0.014)**	0.846 (0.018)	0.598 (0.065)	0.698 (0.044)
Clinical Only	0.680 (0.025)	0.748 (0.052)	**0.870 (0.044)**	0.493 (0.049)	0.627 (0.040)
Activation Function	0.680 (0.039)	0.744 (0.054)	0.868 (0.048)	0.493 (0.620)	0.626 (0.057)
Activation and Attention	0.652 (0.061)	0.712 (0.093)	0.772 (0.237)	0.451 (0.158)	0.566 (0.186)
Multi-channel VAE	0.648 (0.049)	0.727 (0.048)	0.833 (0.075)	0.474 (0.084)	0.597 (0.073)
Intermediate Concatenation	0.606 (0.044)	0.663 (0.029)	0.708 (0.030)	**0.764 (0.036)**	**0.735 (0.029)**
Decision Fusion	0.589 (0.040)	0.647 (0.042)	0.737 (0.030)	0.488 (0.166)	0.567 (0.141)
Extracted Volumes Only	0.583 (0.041)	0.619 (0.038)	0.695 (0.029)	0.718 (0.101)	0.703 (0.057)
Channelwise Attention	0.576 (0.047)	0.596 (0.052)	0.730 (0.059)	0.486 (0.193)	0.558 (0.137)
Crossmodal Attention	0.562 (0.028)	0.610 (0.041)	0.681 (0.025)	0.720 (0.079)	0.697 (0.032)

accuracy and 81% AUROC, with uni-modal clinical data and the activation functions models also performing well. - The worst performing models had similar performed to each other at around 55% accuracy: the attention models, imaging-only, and decision fusion.

Validation results - Validation accuracies had a generally higher variance than the test accuracies, with the better test-performing models having validation accuracies with less spread.

Results observations from the table: - Looking at the table, recall was above 50% for only 4 out of the 10 models: imaging only, intermediate concatenation, early

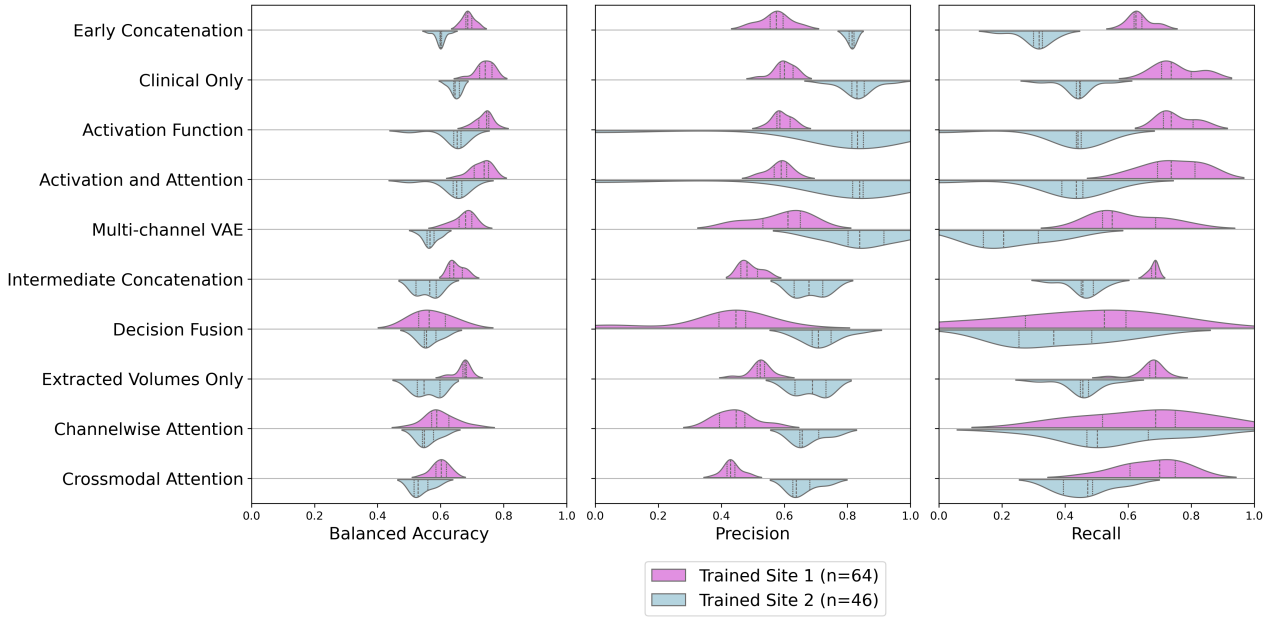

Figure 5.2: Effect of training on one site and testing on the other on the balanced accuracy, precision, and recall of the different models. Ordered by best to worst balanced accuracy of trained together.

concatenation, and cross modal attention. - Means that the methods weren't very good at accurately detecting long survivors

5.4.2 Leave-site-out analysis

- Figure 5.2 shows the balanced accuracy of the different models when trained on one site and tested on the other. - The models are ranked by their balanced accuracy when trained on both sites together. - Training on site one generally resulted in better performance than training on site two, when looking at overall balanced accuracy. - Training on ALS Biomarkers Study generally resulted in better precision, but worse recall - Similar ranking on the methods too, but early concatenation doesn't perform as well as it does when trained on both sites together.

5.5 Discussion

Only one of the multimodal models was better than the clinical only model when comparing balanced accuracy: early concatenation, which is the most simple model. Imaging only was one of the worst performing models, possibly because of the ratio of number of features to number of samples being high. Figure 5.1 validation plot

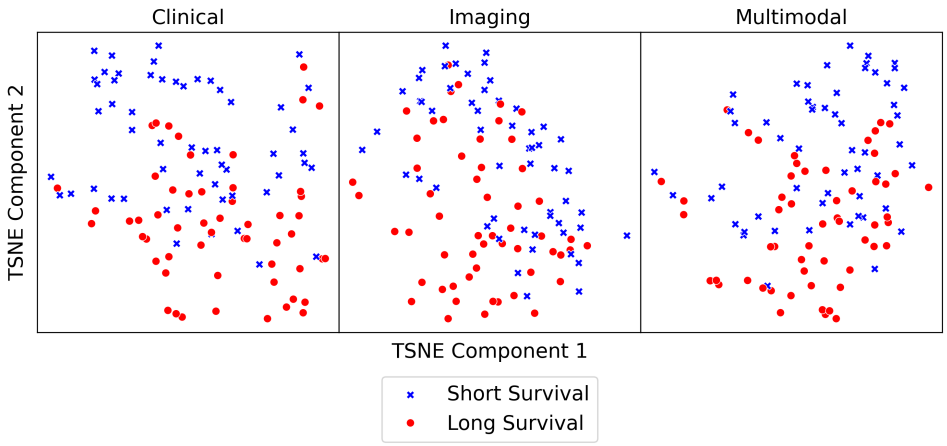

Figure 5.3: TSNE with two components of the clinical, imaging, and multimodal data, split by survival category to illustrate the separability of the data.

shows imaging overfitting on the validation set, with a much higher accuracy than the test set, probably because of the high number of features.

Generally, results aren't great: Figure 5.3 shows the separability of the data from doing TSNE. - Visually looks like clinical-only has the best separation but it's not obvious - Can't expect to get excellent results - This chapter is more of a proof of concept and early analysis, future work will look to improve the actual metrics

Overfitting on the validation set - The distributions of validation fold balanced accuracies show that intermediate concatenation and imaging-only had very higher accuracy, but much lower test set accuracy, suggesting overfitting on the validation set.

Leave-site-out: - Training on the higher sample size site (Milan) performs better than training on the lower sample size site (ALS Biomarkers Study) - Early concatenation doesn't perform as well as it does when trained on both sites together, suggesting that it may be overfitting on the Milan data - Statistically significant differences between the sites in Table 5.2 showed that the Milan patients had different clinical features to the Essex patients, which may have meant the results were ungeneralisable to ALS Biomarkers Study. - However, the overall results are similar to the trained together results, either suggesting that the models are generalisable to the ALS Biomarkers Study data or that model architecture is more important than the data used to train it.

5.5.1 Limitations

- Limitations on sample size
- Evaluating on validation set rather than a completely external test set
- 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*. However, whole image may introduce bias because further progressed patients may have worse quality scans.
- Two sites put together without harmonisation
- Using whole ALSFRS-R rather than individual components - not possible to get with Milan data
- Needs more hyperparameter tuning of the different models to see if they can be improved. Next steps would be to test different network architectures and hyperparameters to see if the results can be improved.

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

6.1.1 Cox model

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

6.1.2 Fusilli

In Chapter 4, I detailed the development of Fusilli, my Python library for easy comparison of multimodal data fusion methods. Even though Fusilli is now finished and released for anybody to use in their personal projects, for my future PhD chapters, using Fusilli will greatly speed up my analysis and experimentation.

6.1.3 Fusilli with MND data

My first experimentation with Fusilli for MND prognosis prediction is shown in Chapter 5.

6.2 Future Work

I have split the possible future work sections into roughly three chapters. Most of the work described here relies on the availability of a larger sample size of data from MND cohorts. Currently, my supervisory team and I are liaising with other researchers for access to more data, as well as applying for access to MRI scans from UCLH (University College London Hospitals), which correspond to subjects in ALS Biomarkers Study.

Moreover, I am undertaking a research visit to Charité - Universitätsmedizin Berlin in April and May of 2024, where I will be working on extending Fusilli's capabilities and applying Fusilli to multimodal medical data on mental health and addiction disorders. I will be making connections and collaborating with researchers at Charité, and while I am there, I hope to visit MND clinics in Berlin to improve my understanding of other areas of MND clinical research.

6.2.1 Applying Fusilli to multimodal medical data: MND and other applications

The next step in my PhD is to continue to apply Fusilli to MND data, especially with larger sample sizes and more tuning of hyperparameters. I will also apply Fusilli to other multimodal medical datasets to round out the development of the library, by showing what Fusilli can do in multiple applications.

Illustrating Fusilli capabilities with multimodal medical data

Although Chapter 4 showed the whole development journey of Fusilli, a valuable addition to the narrative would be to show how Fusilli can be used in practice. My plan is to use Fusilli "out-of-the-box" on two multimodal medical datasets: one on neurodegenerative diseases and one on critical care chest X-rays.

In terms of feasibility, I already have access to ADNI (Alzheimer's Disease Neuroimaging Initiative) and PPMI (Parkinson's Progression Markers Initiative) data for neurodegenerative diseases, and MIMIC CXR (Medical Information Mart for Intensive Care - Chest X-ray) data for critical care.

To finish this analysis off, I will first need to download the data, for which

I will ask for assistance from colleagues at Centre for Medical Image Computing (CMIC) who have used these datasets before. Secondly, I will need to preprocess the data by dealing with data missingness in clinical data and running preprocessing pipelines on the imaging data: segmentation and registration for MRI data, and preprocessing for chest X-ray data. Next, I will choose a task to apply Fusilli to, such as prognosis prediction or disease classification, and I will seek clinical advice on the most appropriate task for each dataset. Finally, I will run the Fusilli pipeline on the data, and analyse the results.

Extending work on applying Fusilli to MND prognosis prediction

In Chapter 5, I showed the first application of Fusilli to MND data. The next step is to extend this work by using more data from collaborations and the ALS Biomarkers Study, and by fine-tuning the parameters and architectures of the fusion models. I plan to do this alongside applying non-MND data to Fusilli, as described above.

- Extending the work in Chapter 5
- Doing this alongside project 1
- We're getting more data from collaborations and the ALS Biomarkers Study
- Also getting D50 measurements from collaboration with Jena, possibly allowing us to use mri from further away from diagnosis if we can estimate the alsfrsr at the scan date
- ALS Biomarkers study team are running analysis to get more NfL measurements from samples they still have
- Fine tuning parameters and architectures of the fusion models as well
- Motivation: fluid biomarkers are more accessible than MRI etc.
- Feasibility: ALS Biomarkers Study etc.

6.2.2 Effect of MRI preprocessing on Fusilli prognosis prediction

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

6.2.3 Final project options

Sensitive analysis of Fusilli prognosis prediction on varying prognosis definitions

Looking at different prognosis outcomes, such as survival, ALSFRS-R decline, or other measures of disease progression, could provide a more nuanced understanding of the performance of Fusilli.

Adding spinal cord MRI data to Fusilli prognosis prediction

Dependent on data availability

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

6.2.3.1 Adding radiological report derived features to Fusilli prognosis prediction

Dependent on data availability

6.3 Timeline

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

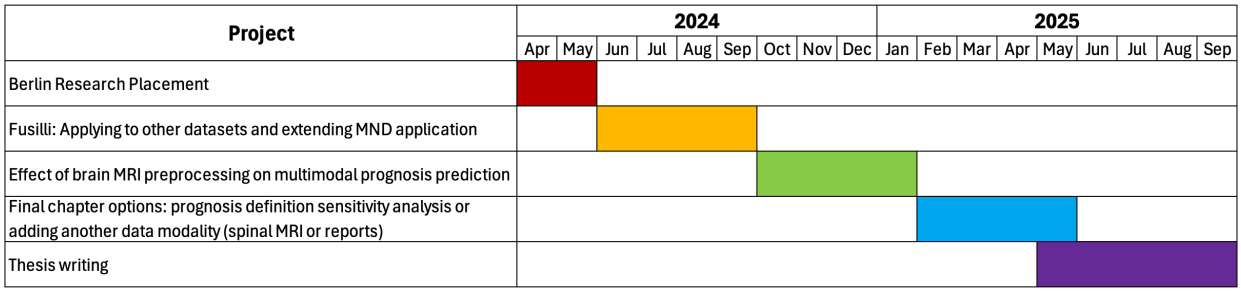

Figure 6.1: Gantt chart showing the timeline for the remaining work in my PhD

Appendix A

Multimodal Correlations

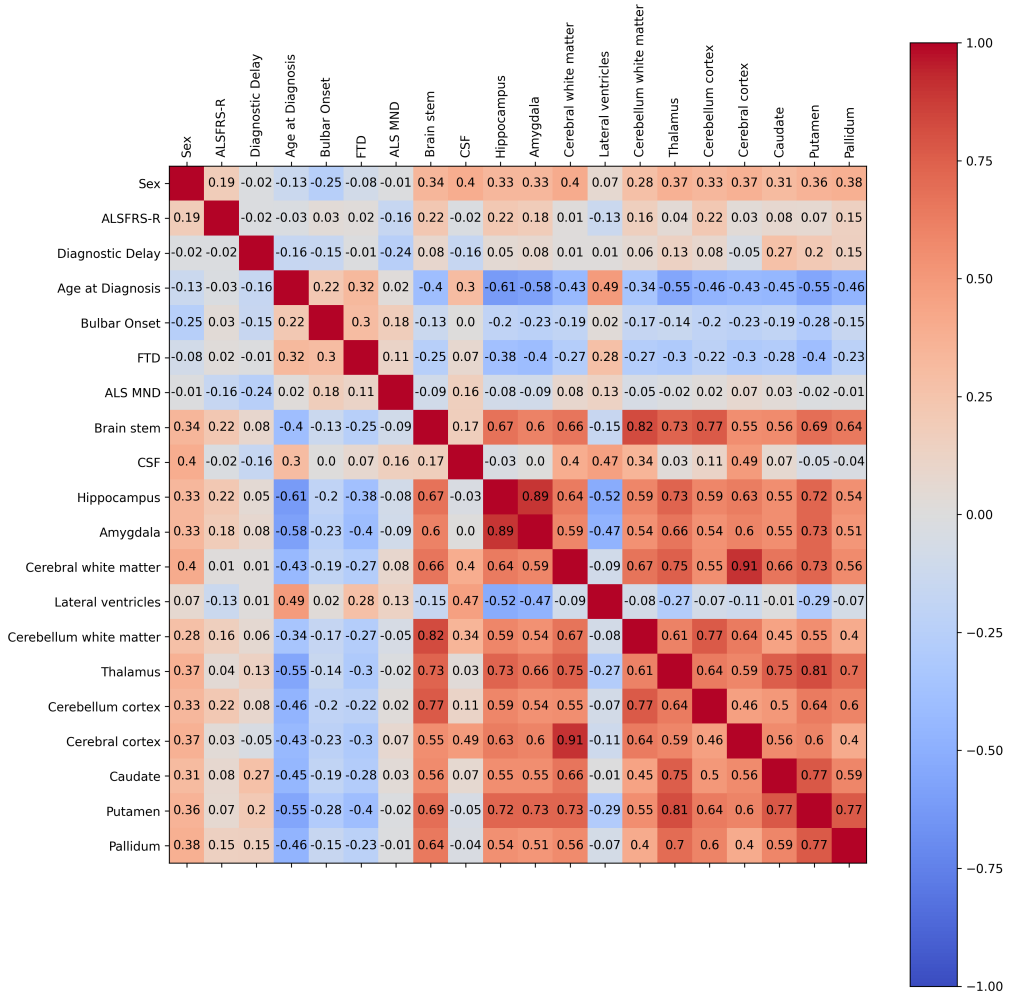

Figure A.1: Correlations of all of the features included in the multimodal multivariable Cox proportional hazards model in Chapter 3.

Appendix B

Fusion Model Architectures

Bibliography

[1] Can Cui, Haichun Yang, Yaohong Wang, Shilin Zhao, Zuhayr Asad, Lori A. Coburn, Keith T. Wilson, Bennett A. Landman, and Yuankai Huo. Deep Multi-modal Fusion of Image and Non-image Data in Disease Diagnosis and Prognosis: A Review, March 2022. arXiv:2203.15588 [cs].

[2] Jeffrey M. Statland, Richard J. Barohn, Mazen M. Dimachkie, Mary Kay Floeter, and Hiroshi Mitsumoto. Primary Lateral Sclerosis. *Neurologic clinics*, 33(4):749–760, November 2015.

[3] Bart Swinnen and Wim Robberecht. The phenotypic variability of amyotrophic lateral sclerosis. *Nature Reviews Neurology*, 10(11):661–670, November 2014. Number: 11 Publisher: Nature Publishing Group.

[4] Stephen A Goutman, Orla Hardiman, Ammar Al-Chalabi, Adriano Chió, Masha G Savelieff, Matthew C. Kiernan, and Eva L Feldman. Recent advances in the diagnosis and prognosis of ALS. *The Lancet. Neurology*, 21(5):480–493, May 2022.

[5] Elisabetta Pupillo, Paolo Messina, Giancarlo Logroscino, Ettore Beghi, and the SLALOM Group. Long-term survival in amyotrophic lateral sclerosis: A population-based study. *Annals of Neurology*, 75(2):287–297, 2014. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ana.24096.

[6] Benoît Marin, Farid Boumédiene, Giancarlo Logroscino, Philippe Couratier, Marie-Claude Babron, Anne Louise Leutenegger, Massimilano Copetti, Pierre-Marie Preux, and Ettore Beghi. Variation in worldwide incidence of amy-

otrophic lateral sclerosis: a meta-analysis. *International Journal of Epidemiology*, 46(1):57–74, February 2017.

[7] Andrea Fontana, Benoit Marin, Jaime Luna, Ettore Beghi, Giancarlo Logroscino, Farid Boumédiene, Pierre-Marie Preux, Philippe Couratier, and Massimilano Copetti. Time-trend evolution and determinants of sex ratio in Amyotrophic Lateral Sclerosis: a dose-response meta-analysis. *Journal of Neurology*, 268(8):2973–2984, August 2021.

[8] Martha F. Hanby, Kirsten M. Scott, William Scotton, Lokesh Wijesekera, Thomas Mole, Catherine E. Ellis, P. Nigel Leigh, Christopher E. Shaw, and Ammar Al-Chalabi. The risk to relatives of patients with sporadic amyotrophic lateral sclerosis. *Brain*, 134(12):3451–3454, December 2011.

[9] A. Al-Chalabi, F. Fang, M. F. Hanby, P. N. Leigh, C. E. Shaw, W. Ye, and F. Rijsdijk. An estimate of amyotrophic lateral sclerosis heritability using twin data. *Journal of Neurology, Neurosurgery, and Psychiatry*, 81(12):1324–1326, December 2010.

[10] Eva L Feldman, Stephen A Goutman, Susanne Petri, Letizia Mazzini, Masha G Savelieff, Pamela J Shaw, and Gen Sobue. Amyotrophic lateral sclerosis. *The Lancet*, 400(10360):1363–1380, October 2022.

[11] Michael A van Es, Orla Hardiman, Adriano Chio, Ammar Al-Chalabi, R Jeroen Pasterkamp, Jan H Veldink, and Leonard H van den Berg. Amyotrophic lateral sclerosis. *The Lancet*, 390(10107):2084–2098, November 2017.

[12] Kyla A. McKay, Kelsi A. Smith, Lidija Smertinaite, Fang Fang, Caroline Ingre, and Fabian Taube. Military service and related risk factors for amyotrophic lateral sclerosis. *Acta Neurologica Scandinavica*, 143(1):39–50, 2021. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ane.13345.

[13] Eleonora Lacorte, Luigina Ferrigno, Emanuele Leoncini, Massimo Corbo, Stefania Boccia, and Nicola Vanacore. Physical activity, and physical activity

related to sports, leisure and occupational activity as risk factors for ALS: A systematic review. *Neuroscience & Biobehavioral Reviews*, 66:61–79, July 2016.

[14] Renée Walhout, Esther Verstraete, Martijn P. van den Heuvel, Jan H. Veldink, and Leonard H. van den Berg. Patterns of symptom development in patients with motor neuron disease. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 19(1-2):21–28, January 2018. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21678421.2017.1386688.

[15] Anna M. Blokhuis, Ewout J. N. Groen, Max Koppers, Leonard H. van den Berg, and R. Jeroen Pasterkamp. Protein aggregation in amyotrophic lateral sclerosis. *Acta Neuropathologica*, 125(6):777–794, 2013.

[16] Tianmi Yang, Yanbing Hou, Chunyu Li, Bei Cao, Yangfan Cheng, Qianqian Wei, Lingyu Zhang, and Huifang Shang. Risk factors for cognitive impairment in amyotrophic lateral sclerosis: a systematic review and meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, 92(7):688–693, July 2021. Publisher: BMJ Publishing Group Ltd Section: Cognitive neurology.

[17] Adriano Chiò, Cristina Moglia, Antonio Canosa, Umberto Manera, Fabrizio D'Ovidio, Rosario Vasta, Maurizio Grassano, Maura Brunetti, Marco Barberis, Lucia Corrado, Sandra D'Alfonso, Barbara Iazzolino, Laura Peotta, Maria Francesca Sarnelli, Valentina Solara, Jean Pierre Zucchetti, Fabiola De Marchi, Letizia Mazzini, Gabriele Mora, and Andrea Calvo. ALS phenotype is influenced by age, sex, and genetics: A population-based study. *Neurology*, 94(8):e802–e810, February 2020.

[18] Michael J. Strong, Sharon Abrahams, Laura H. Goldstein, Susan Woolley, Paula Mclaughlin, Julie Snowden, Eneida Mioshi, Angie Roberts-South, Michael Benatar, Tibor HortobáGyi, Jeffrey Rosenfeld, Vincenzo Silani, Paul G Ince, and Martin R. Turner. Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnos-

tic criteria. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 18(3-4):153–174, April 2017. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21678421.2016.1267768.

[19] Sharon Abrahams, Judith Newton, Elaine Niven, Jennifer Foley, and Thomas H. Bak. Screening for cognition and behaviour changes in ALS. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*, 15(1-2):9–14, March 2014.

[20] Emma Beeldman, Joost Raaphorst, Michelle Klein Twennaar, Marianne de Visser, Ben A. Schmand, and Rob J. de Haan. The cognitive profile of ALS: a systematic review and meta-analysis update. *Journal of Neurology, Neurosurgery, and Psychiatry*, 87(6):611–619, June 2016.

[21] Christopher Crockford, Judith Newton, Katie Lonergan, Theresa Chiwera, Tom Booth, Siddharthan Chandran, Shuna Colville, Mark Heverin, Iain Mays, Suvankar Pal, Niall Pender, Marta Pinto-Grau, Ratko Radakovic, Christopher E. Shaw, Laura Stephenson, Robert Swingler, Alice Vajda, Ammar Al-Chalabi, Orla Hardiman, and Sharon Abrahams. ALS-specific cognitive and behavior changes associated with advancing disease stage in ALS. *Neurology*, 91(15):e1370–e1380, October 2018.

[22] John Douglas Mitchell, Pauline Callagher, Joyce Gardham, Catriona Mitchell, Mandy Dixon, Robert Addison-Jones, Wendy Bennett, and Mary R. O'Brien. Timelines in the diagnostic evaluation of people with suspected amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND) – a 20-year review: Can we do better? *Amyotrophic Lateral Sclerosis*, 11(6):537–541, December 2010. Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/17482968.2010.495158.

[23] J. M. Cedarbaum, N. Stambler, E. Malta, C. Fuller, D. Hilt, B. Thurmond, and A. Nakanishi. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group

(Phase III). *Journal of the Neurological Sciences*, 169(1-2):13–21, October 1999.

[24] Ruben P. A. van Eijk, Adriaan D. de Jongh, Stavros Nikolakopoulos, Christopher J. McDermott, Marinus J. C. Eijkemans, Kit C. B. Roes, and Leonard H. van den Berg. An old friend who has overstayed their welcome: the ALSFRS-R total score as primary endpoint for ALS clinical trials. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*, 22(3-4):300–307, May 2021.

[25] James Rooney, Tom Burke, Alice Vajda, Mark Heverin, and Orla Hardiman. What does the ALSFRS-R really measure? A longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 88(5):381–385, May 2017.

[26] Jose C. Roche, Ricardo Rojas-Garcia, Kirsten M. Scott, William Scotton, Catherine E. Ellis, Rachel Burman, Lokesh Wijesekera, Martin R. Turner, P. Nigel Leigh, Christopher E. Shaw, and Ammar Al-Chalabi. A proposed staging system for amyotrophic lateral sclerosis. *Brain: A Journal of Neurology*, 135(Pt 3):847–852, March 2012.

[27] Adriano Chiò, Edward R. Hammond, Gabriele Mora, Virginio Bonito, and Graziella Filippini. Development and evaluation of a clinical staging system for amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 86(1):38–44, January 2015.

[28] Ton Fang, Ahmad Al Khleifat, Daniel R Stahl, Claudia Lazo La Torre, Caroline Murphy, Carolyn Young, Pamela J Shaw, P Nigel Leigh, and Ammar Al-Chalabi. Comparison of the King's and MiToS staging systems for ALS. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 18(3-4):227–232, April 2017. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21678421.2016.1265565.

[29] Stephen C. Bourke, Mark Tomlinson, Tim L. Williams, Robert E. Bullock, Pamela J. Shaw, and G. John Gibson. Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: a randomised controlled trial. *The Lancet. Neurology*, 5(2):140–147, February 2006.

[30] Robert G. Miller, J. D. Mitchell, and Dan H. Moore. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *The Cochrane Database of Systematic Reviews*, 2012(3):CD001447, March 2012.

[31] Michael Hinchcliffe and Alan Smith. Riluzole: real-world evidence supports significant extension of median survival times in patients with amyotrophic lateral sclerosis. *Degenerative Neurological and Neuromuscular Disease*, 7:61–70, 2017.

[32] Jinsy A. Andrews, Carlayne E. Jackson, Terry D. Heiman-Patterson, Paolo Bettica, Benjamin Rix Brooks, and Erik P. Pioro. Real-world evidence of riluzole effectiveness in treating amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 21(7-8):509–518, October 2020. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21678421.2020.1771734.

[33] Simon Witzel, André Maier, Robert Steinbach, Julian Grosskreutz, Jan C. Koch, Anastasia Sarikidi, Susanne Petri, René Günther, Joachim Wolf, Andreas Hermann, Johannes Prudlo, Isabell Cordts, Paul Lingor, Wolfgang N. Löscher, Zacharias Kohl, Tim Hagenacker, Christian Ruckes, Birgit Koch, Susanne Spittel, Kornelia Günther, Sebastian Michels, Johannes Dorst, Thomas Meyer, Albert C. Ludolph, and German Motor Neuron Disease Network (MND-NET). Safety and Effectiveness of Long-term Intravenous Administration of Edaravone for Treatment of Patients With Amyotrophic Lateral Sclerosis. *JAMA neurology*, 79(2):121–130, February 2022.

[34] Fabiano Papaiz, Mario Emílio Teixeira Dourado, Ricardo Alexsandro de Medeiros Valentim, Antonio Higor Freire de Morais, and Joel Perdiz Arrais. Machine Learning Solutions Applied to Amyotrophic Lateral Sclerosis Prognosis: A Review. *Frontiers in Computer Science*, 4, 2022.

[35] Erica Tavazzi, Enrico Longato, Martina Vettoretti, Helena Aidos, Isotta Trescato, Chiara Roversi, Andreia S. Martins, Eduardo N. Castanho, Ruben Branco, Diogo F. Soares, Alessandro Guazzo, Giovanni Birolo, Daniele Pala, Pietro Bosoni, Adriano Chiò, Umberto Manera, Mamede de Carvalho, Bruno Miranda, Marta Gromicho, Inês Alves, Riccardo Bellazzi, Arianna Dagliati, Piero Fariselli, Sara C. Madeira, and Barbara Di Camillo. Artificial intelligence and statistical methods for stratification and prediction of progression in amyotrophic lateral sclerosis: A systematic review. *Artificial Intelligence in Medicine*, 142:102588, August 2023.

[36] Wei-Ming Su, Yang-Fan Cheng, Zheng Jiang, Qing-Qing Duan, Tian-Mi Yang, Hui-Fang Shang, and Yong-Ping Chen. Predictors of survival in patients with amyotrophic lateral sclerosis: A large meta-analysis. *eBioMedicine*, 74:103732, December 2021.

[37] Ming Gao, Na Liu, Xue-Mei Li, Liu-Wen Chao, Hong-Qi Lin, Yan Wang, Yan Sun, Chen Huang, Xiao-Gang Li, and Min Deng. Epidemiology and factors predicting survival of amyotrophic lateral sclerosis in a large Chinese cohort. *Chinese Medical Journal*, 134(18):2231–2236, September 2021.

[38] Chieko Fujimura-Kiyono, Fumiharu Kimura, Simon Ishida, Hideto Nakajima, Takafumi Hosokawa, Masakazu Sugino, and Toshiaki Hanafusa. Onset and spreading patterns of lower motor neuron involvements predict survival in sporadic amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 82(11):1244–1249, November 2011.

[39] M. Elamin, J. Phukan, P. Bede, N. Jordan, S. Byrne, N. Pender, and O. Hardiman. Executive dysfunction is a negative prognostic indicator in patients with ALS without dementia. *Neurology*, 76(14):1263–1269, April 2011.

[40] Albert Ludolph, Vivian Drory, Orla Hardiman, Imaharu Nakano, John Ravits, Wim Robberecht, and Jeremy Shefner. A revision of the El Escorial criteria - 2015. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 16(5-6):291–292, August 2015. Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/21678421.2015.1049183.

[41] Henk Jan Westeneng, Thomas P. A. Debray, Anne E. Visser, Ruben P A van Eijk, James Rooney, Andrea Calvo, Sarah Martin, Christopher J McDermott, Alexander G. Thompson, Susana Pinto, Xenia Kobeleva, Angela Rosenbohm, Beatrice Stubendorff, Helma Sommer, Bas M. Middelkoop, Bas Middelkoop, Annelot M. Dekker, Annelot M. Dekker, Joke J.F.A. van Vugt, Joke J.F.A. van Vugt, Wouter van Rheenen, Wouter van Rheenen, Alice Vajda, Mark Heverin, Mbombe Kazoka, Hannah Hollinger, Marta Gromicho, Sonja Körner, Thomas M. Ringer, A. Rödiger, A. Gunkel, Christopher Shaw, Annelien L. Bredenoord, Annelien L. Bredenoord, Michael A. van Es, Michael A van Es, Philippe Corcia, Philippe Couratier, Markus Weber, Julian Grosskreutz, Albert C. Ludolph, Susanne Petri, Mamede de Carvalho, Philip Van Damme, Kevin Talbot, Martin R Turner, Pamela J. Shaw, Ammar Al-Chalabi, Adriano Chiò, Orla Hardiman, Karel G.M. Moons, Karel G. M. Moons, Jan H. Veldink, Leonard H. van den Berg, and Leonard H. van den Berg. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurology*, 17(5):423–433, May 2018. MAG ID: 2794654744.

[42] Efthimios Dardiotis, Vasileios Siokas, Maria Sokratous, Zisis Tsouris, Athina-Maria Aloizou, Desponia Florou, Metaxia Dastamani, Alexios-Fotios A. Mentis, and Alexandros G. Brotis. Body mass index and survival from amyotrophic

lateral sclerosis: A meta-analysis. *Neurology Clinical Practice*, 8(5):437–444, October 2018.

[43] S.A. Goutman, J. Boss, G. Iyer, H. Habra, M.G. Savelieff, A. Karnovsky, B. Mukherjee, and E.L. Feldman. Body mass index associates with amyotrophic lateral sclerosis survival and metabolomic profiles. *Muscle and Nerve*, 67(3):208–216, 2023.

[44] Ali Jawaid, Santosh B. Murthy, Andrew M. Wilson, Salah U. Qureshi, Moath J. Amro, Michael Wheaton, Ericka Simpson, Yadollah Harati, Adriana M. Strutt, Michele K. York, and Paul E. Schulz. A decrease in body mass index is associated with faster progression of motor symptoms and shorter survival in ALS. *Amyotrophic Lateral Sclerosis*, 11(6):542–548, December 2010. Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/17482968.2010.482592.

[45] Eva Lindauer, Luc Dupuis, Hans-Peter Müller, Heiko Neumann, Albert C. Ludolph, and Jan Kassubek. Adipose Tissue Distribution Predicts Survival in Amyotrophic Lateral Sclerosis. *PLoS ONE*, 8(6):e67783, June 2013.

[46] Marc G Weisskopf, Joseph Levy, Aisha S Dickerson, Sabrina Paganoni, and Maya Leventer-Roberts. Statin Medications and Amyotrophic Lateral Sclerosis Incidence and Mortality. *American Journal of Epidemiology*, 191(7):1248–1257, June 2022.

[47] EFNS Task Force on Diagnosis and Management of Amyotrophic Lateral Sclerosis:, Peter M. Andersen, Sharon Abrahams, Gian D. Borasio, Mamede de Carvalho, Adriano Chio, Philip Van Damme, Orla Hardiman, Katja Kollewe, Karen E. Morrison, Susanne Petri, Pierre-Francois Pradat, Vincenzo Silani, Barbara Tomik, Maria Wasner, and Markus Weber. EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS)–revised report of an EFNS task force. *European Journal of Neurology*, 19(3):360–375, March 2012.

[48] Wei-Ming Su, Xiao-Jing Gu, Qing-Qing Duan, Zheng Jiang, Xia Gao, Hui-Fang Shang, and Yong-Ping Chen. Genetic factors for survival in amyotrophic lateral sclerosis: an integrated approach combining a systematic review, pairwise and network meta-analysis. *BMC Medicine*, 20(1):209, June 2022.

[49] Adriano Chiò, Cristina Moglia, Antonio Canosa, Umberto Manera, Maurizio Grassano, Rosario Vasta, Francesca Palumbo, Salvatore Gallone, Maura Brunetti, Marco Barberis, Fabiola De Marchi, Clifton Dalgard, Ruth Chia, Gabriele Mora, Barbara Iazzolino, Laura Peotta, Bryan J. Traynor, Lucia Corrado, Sandra D'Alfonso, Letizia Mazzini, and Andrea Calvo. Association of Copresence of Pathogenic Variants Related to Amyotrophic Lateral Sclerosis and Prognosis. *Neurology*, 101(1), July 2023.

[50] Michael Benatar, Lanyu Zhang, Lily Wang, Volkan Granit, Jeffrey Statland, Richard Barohn, Andrea Swenson, John Ravits, Carlayne Jackson, Ted M. Burns, Jaya Trivedi, Erik P. Pioro, James Caress, Jonathan Katz, Jacob L. McCauley, Rosa Rademakers, Andrea Malaspina, Lyle W. Ostrow, and Joanne Wuu. Validation of serum neurofilaments as prognostic and potential pharmacodynamic biomarkers for ALS. *Neurology*, 95(1):e59–e69, July 2020.

[51] Michael Benatar, Joanne Wuu, Vittoria Lombardi, Andreas Jeromin, Robert Bowser, Peter M. Andersen, and Andrea Malaspina. Neurofilaments in presymptomatic ALS and the impact of genotype. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*, 20(7-8):538–548, November 2019.

[52] Alexander G. Thompson, Elizabeth Gray, Nick Verber, Yoana Bobeva, Vittoria Lombardi, Stephanie R. Shepheard, Ozlem Yildiz, Emily Feneberg, Lucy Farrimond, Thanuja Dharmadasa, Pamela Gray, Evan C. Edmond, Jakub Scaber, Delia Gagliardi, Janine Kirby, Thomas M. Jenkins, Pietro Fratta, Christopher J. McDermott, Sanjay G. Manohar, Kevin Talbot, Andrea Malaspina, Pamela J. Shaw, and Martin R. Turner. Multicentre appraisal of amyotrophic lateral sclerosis biofluid biomarkers shows primacy of blood neurofilament light chain. *Brain Communications*, 4(1):fcac029, 2022.

[53] Katherine E. Irwin, Udit Sheth, Philip C. Wong, and Tania F. Gendron. Fluid biomarkers for amyotrophic lateral sclerosis: a review. *Molecular Neurodegeneration*, 19(1):9, January 2024.

[54] Marie Dreger, Robert Steinbach, Nayana Gaur, Klara Metzner, Beatrice Stubendorff, Otto W. Witte, and Julian Grosskreutz. Cerebrospinal Fluid Neurofilament Light Chain (NfL) Predicts Disease Aggressiveness in Amyotrophic Lateral Sclerosis: An Application of the D50 Disease Progression Model. *Frontiers in Neuroscience*, 15:651651, April 2021.

[55] Peter Bede and Orla Hardiman. Lessons of ALS imaging: Pitfalls and future directions — A critical review. *NeuroImage: Clinical*, 4:436–443, January 2014.

[56] Mohamed Mounir El Mendili, Annie Verschueren, Jean-Philippe Ranjeva, Maxime Guye, Shahram Attarian, Wafaa Zaaraoui, and Aude-Marie Grapperon. Association between brain and upper cervical spinal cord atrophy assessed by MRI and disease aggressiveness in amyotrophic lateral sclerosis. *Neuroradiology*, 65(9):1395–1403, September 2023.

[57] Peter Bede and Orla Hardiman. Longitudinal structural changes in ALS: a three time-point imaging study of white and gray matter degeneration. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 19(3-4):232–241, April 2018. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21678421.2017.1407795.

[58] Francesca Trojsi, Federica Di Nardo, Mattia Siciliano, Giuseppina Caiazzo, Carla Passaniti, Giulia D'Alvano, Dario Ricciardi, Antonio Russo, Alvino Bisecco, Luigi Lavorgna, Simona Bonavita, Mario Cirillo, Fabrizio Esposito, and Gioacchino Tedeschi. Resting state functional MRI brain signatures of fast disease progression in amyotrophic lateral sclerosis: a retrospective study. *Amyotrophic Lateral Sclerosis and Frontotemporal Degenera-*

tion, 22(1-2):117–126, January 2021. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21678421.2020.1813306.

[59] Joe Senda, Naoki Atsuta, Hirohisa Watanabe, Epifanio Bagarinao, Kazunori Imai, Daichi Yokoi, Yuichi Riku, Michihito Masuda, Ryoichi Nakamura, Hazuki Watanabe, Mizuki Ito, Masahisa Katsuno, Shinji Naganawa, and Gen Sobue. Structural MRI correlates of amyotrophic lateral sclerosis progression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 88(11):901–907, November 2017.

[60] Dobri Baldaranov, Andrei Khomenko, Ines Kobor, Ulrich Bogdahn, Martin Gorges, Jan Kassubek, and Hans-Peter Müller. Longitudinal Diffusion Tensor Imaging-Based Assessment of Tract Alterations: An Application to Amyotrophic Lateral Sclerosis. *Frontiers in Human Neuroscience*, 11, 2017.

[61] Giovanni Rizzo, Anna Federica Marliani, Stella Battaglia, Luca Albini Riccioli, Silvia De Pasqua, Veria Vacchiano, Rossella Infante, Patrizia Avoni, Vincenzo Donadio, Massimiliano Passaretti, Ilaria Bartolomei, Fabrizio Salvi, Rocco Liguori, and on behalf of the BoReALS group. Diagnostic and Prognostic Value of Conventional Brain MRI in the Clinical Work-Up of Patients with Amyotrophic Lateral Sclerosis. *Journal of Clinical Medicine*, 9(8):2538, August 2020. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

[62] Marie Catherine Boll, Oscar René Marrufo Meléndez, Camilo Rios, Jesus Maciel Zenil, and Yara de Alba. Is the Hypointensity in Motor Cortex the Hallmark of Amyotrophic Lateral Sclerosis? *Canadian Journal of Neurological Sciences*, 46(2):166–173, March 2019.

[63] Hans-Peter Müller, Martin R. Turner, Julian Grosskreutz, Sharon Abrahams, Peter Bede, Varan Govind, Johannes Prudlo, Albert C. Ludolph, Massimo Filippi, Jan Kassubek, and Neuroimaging Society in ALS (NiSALS) DTI Study Group. A large-scale multicentre cerebral diffusion tensor imaging

study in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 87(6):570–579, June 2016.

[64] Haining Li, Qiuli Zhang, Qianqian Duan, Jiaoting Jin, Fangfang Hu, Jingxia Dang, and Ming Zhang. Brainstem Involvement in Amyotrophic Lateral Sclerosis: A Combined Structural and Diffusion Tensor MRI Analysis. *Frontiers in Neuroscience*, 15, 2021.

[65] Federica Agosta, Elisabetta Pagani, Melissa Petrolini, Maria P. Sormani, Domenico Caputo, Michele Perini, Alessandro Prelle, Fabrizio Salvi, and Massimo Filippi. MRI predictors of long-term evolution in amyotrophic lateral sclerosis. *The European Journal of Neuroscience*, 32(9):1490–1496, November 2010.

[66] Ricarda A. L. Menke, Sonja Körner, Nicola Filippini, Gwenaëlle Douaud, Steven Knight, Kevin Talbot, and Martin R. Turner. Widespread grey matter pathology dominates the longitudinal cerebral MRI and clinical landscape of amyotrophic lateral sclerosis. *Brain: A Journal of Neurology*, 137(Pt 9):2546–2555, September 2014.

[67] Sanjay Kalra, Hans-Peter Müller, Abdullah Ishaque, Lorne Zinman, Lawrence Korngut, Angela Genge, Christian Beaulieu, Richard Frayne, Simon J. Graham, and Jan Kassubek. A prospective harmonized multicenter DTI study of cerebral white matter degeneration in ALS. *Neurology*, 95(8):e943–e952, August 2020.

[68] G. Grolez, M. Kyheng, R. Lopes, C. Moreau, K. Timmerman, F. Auger, G. Kuchcinski, A. Duhamel, P. Jissendi-Tchofo, P. Besson, C. Laloux, M. Petrault, J. C. Devedjian, Thierry Pérez, Pierre François Pradat, L. Defebvre, R. Bordet, V. Danel-Brunaud, and D. Devos. MRI of the cervical spinal cord predicts respiratory dysfunction in ALS. *Scientific Reports*, 8:1828, January 2018.

[69] Robert Steinbach, Nayana Gaur, Annekathrin Roediger, Thomas E. Mayer, Otto W. Witte, Tino Prell, and Julian Grosskreutz. Disease aggressiveness signatures of amyotrophic lateral sclerosis in white matter tracts revealed by the D50 disease progression model. *Human Brain Mapping*, 42(3):737–752, 2021. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.25258.

[70] A. d'Ambrosio, A. Gallo, F. Trojsi, D. Corbo, F. Esposito, M. Cirillo, M. R. Monsurrò, and G. Tedeschi. Frontotemporal Cortical Thinning in Amyotrophic Lateral Sclerosis. *American Journal of Neuroradiology*, 35(2):304–310, February 2014. Publisher: American Journal of Neuroradiology Section: Brain.

[71] Esther Verstraete, Jan H. Veldink, Jeroen Hendrikse, H. Jurgen Schelhaas, Martijn P. van den Heuvel, and Leonard H. van den Berg. Structural MRI reveals cortical thinning in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 83(4):383–388, April 2012. Publisher: BMJ Publishing Group Ltd Section: ALS and FTD Special Edition.

[72] Hannelore K. van der Burgh, Henk-Jan Westeneng, Renée Walhout, Kevin van Veenhuijzen, Harold H. G. Tan, Jil M. Meier, Leonhard A. Bakker, Jeroen Hendrikse, Michael A. van Es, Jan H. Veldink, Martijn P. van den Heuvel, and Leonard H. van den Berg. Multimodal longitudinal study of structural brain involvement in amyotrophic lateral sclerosis. *Neurology*, 94(24):e2592–e2604, June 2020. Publisher: Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology Section: Article.

[73] Nora Dieckmann, Annekathrin Roediger, Tino Prell, Simon Schuster, Meret Herdick, Thomas E. Mayer, Otto W. Witte, Robert Steinbach, and Julian Grosskreutz. Cortical and subcortical grey matter atrophy in Amyotrophic Lateral Sclerosis correlates with measures of disease accumulation independent of disease aggressiveness. *NeuroImage: Clinical*, 36:103162, January 2022.

[74] Scott L.M. Johns, Abdullah Ishaque, Muhammad Khan, Yee-Hong Yang, Alan H. Wilman, and Sanjay Kalra. Quantifying changes on susceptibility weighted images in amyotrophic lateral sclerosis using MRI texture analysis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 20(5-6):396–403, July 2019. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21678421.2019.1599024.

[75] Federica Agosta, Maria Luisa Gorno-Tempini, Elisabetta Pagani, Stefania Sala, Domenico Caputo, Michele Perini, Ilaria Bartolomei, Maria Elena Fruguglietti, and Massimo Filippi. Longitudinal assessment of grey matter contraction in amyotrophic lateral sclerosis: A tensor based morphometry study. *Amyotrophic Lateral Sclerosis*, 10(3):168–174, January 2009. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17482960802603841.

[76] Henk-Jan Westeneng, Esther Verstraete, Renée Walhout, Ruben Schmidt, Jeroen Hendrikse, Jan H. Veldink, Martijn P. van den Heuvel, and Leonard H. van den Berg. Subcortical structures in amyotrophic lateral sclerosis. *Neurobiology of Aging*, 36(2):1075–1082, February 2015.

[77] Abdullah Ishaque, Dennell Mah, Peter Seres, Collin Luk, Dean Eurich, Wendy Johnston, Yee-Hong Yang, and Sanjay Kalra. Evaluating the cerebral correlates of survival in amyotrophic lateral sclerosis. *Annals of Clinical and Translational Neurology*, 5(11):1350–1361, September 2018.

[78] Hyon-Ah Yi, Christiane Möller, Nikki Dieleman, Femke H. Bouwman, Frederik Barkhof, Philip Scheltens, Wiesje M. van der Flier, and Hugo Vrenken. Relation between subcortical grey matter atrophy and conversion from mild cognitive impairment to Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 87(4):425–432, April 2016. Publisher: BMJ Publishing Group Ltd Section: Neurodegeneration.

[79] Susanne Abdulla, Judith Machts, Jörn Kaufmann, Karina Patrick, Katja Kollewe, Reinhard Dengler, Hans-Jochen Heinze, Susanne Petri, Stefan

Vielhaber, and Peter J. Nestor. Hippocampal degeneration in patients with amyotrophic lateral sclerosis. *Neurobiology of Aging*, 35(11):2639–2645, November 2014.

[80] Woo-Suk Tae, Joo Hye Sung, Seol-Hee Baek, Chan-Nyoung Lee, and Byung-Jo Kim. Shape Analysis of the Subcortical Nuclei in Amyotrophic Lateral Sclerosis without Cognitive Impairment. *Journal of Clinical Neurology*, 16(4):592–598, October 2020.

[81] Christian Michael Stoppel, Stefan Vielhaber, Cindy Eckart, Judith Machts, Jörn Kaufmann, Hans-Jochen Heinze, Katja Kollewe, Susanne Petri, Reinhard Dengler, Jens-Max Hopf, and Mircea Ariel Schoenfeld. Structural and functional hallmarks of amyotrophic lateral sclerosis progression in motor- and memory-related brain regions. *NeuroImage: Clinical*, 5:277–290, January 2014.

[82] F. Agosta, P. Valsasina, M. Absinta, N. Riva, S. Sala, A. Prelle, M. Copetti, M. Comola, G. Comi, and M. Filippi. Sensorimotor Functional Connectivity Changes in Amyotrophic Lateral Sclerosis. *Cerebral Cortex*, 21(10):2291–2298, October 2011.

[83] Robert Steinbach, Meerim Batyrbekova, Nayana Gaur, Annika Voss, Beatrice Stubendorff, Thomas E. Mayer, Christian Gaser, Otto W. Witte, Tino Prell, and Julian Grosskreutz. Applying the D50 disease progression model to gray and white matter pathology in amyotrophic lateral sclerosis. *NeuroImage: Clinical*, 25:102094, January 2020.

[84] Giammarco Milella, Alessandro Introna, Alma Ghirelli, Domenico Maria Mezzapesa, Ucci Maria, Eustachio D'Errico, Angela Fraddosio, and Isabella Laura Simone. Medulla oblongata volume as a promising predictor of survival in amyotrophic lateral sclerosis. *NeuroImage: Clinical*, 34:103015, January 2022.

[85] A. Hermann, G.N. Tarakdjian, A.G.M. Temp, E. Kasper, J. MacHts, J. Kaufmann, S. Vielhaber, J. Prudlo, J.H. Cole, S. Teipel, and M. Dyrba. Cognitive and behavioural but not motor impairment increases brain age in amyotrophic lateral sclerosis. *Brain Communications*, 4(5), 2022.

[86] S Kalra, A Vitale, N R Cashman, A Genge, and D L Arnold. Cerebral degeneration predicts survival in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 77(11):1253–1255, November 2006.

[87] M.M. El Mendili, G. Querin, P. Bede, and P.-F. Pradat. Spinal cord imaging in amyotrophic lateral sclerosis: Historical concepts—novel techniques. *Frontiers in Neurology*, 10(APR), 2019.

[88] Lucas M. T. Branco, Milena De Albuquerque, Helen Maia T. De Andrade, Felipe P. G. Bergo, Anamarli Nucci, and Marcondes C. França Jr. Spinal cord atrophy correlates with disease duration and severity in amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 15(1-2):93–97, March 2014. Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/21678421.2013.852589.

[89] H.K. van der Burgh, H.-J. Westeneng, J.M. Meier, M.A. van Es, J.H. Veldink, J. Hendrikse, M.P. van den Heuvel, and L.H. van den Berg. Cross-sectional and longitudinal assessment of the upper cervical spinal cord in motor neuron disease. *NeuroImage: Clinical*, 24, 2019.

[90] Koen Poesen, Maxim De Schaepdryver, Beatrice Stubendorff, Benjamin Gille, Petra Muckova, Sindy Wendler, Tino Prell, Thomas M. Ringer, Heidrun Rhode, Olivier Stevens, Kristl G. Claeys, Goedele Couwelier, Ann D'Hondt, Nikita Lamaire, Petra Tilkin, Dimphna Van Reijen, Sarah Gourmaud, Nadin Fedtke, Bianka Heiling, Matthias Rumpel, Annekathrin Rödiger, Anne Gunkel, Otto W. Witte, Claire Paquet, Rik Vandenberghe, Julian Grosskreutz, and Philip Van Damme. Neurofilament markers for ALS cor-

relate with extent of upper and lower motor neuron disease. *Neurology*, 88(24):2302–2309, June 2017. Publisher: Wolters Kluwer.

[91] Divya Ramamoorthy, Kristen Severson, Soumya Ghosh, Karen Sachs, Jonathan D. Glass, Christina N. Fournier, Todd M. Herrington, James D. Berry, Kenney Ng, and Ernest Fraenkel. Identifying patterns in amyotrophic lateral sclerosis progression from sparse longitudinal data. *Nature Computational Science*, 2(9):605–616, September 2022. Number: 9 Publisher: Nature Publishing Group.

[92] Robert Küffner, Neta Zach, Raquel Norel, Johann Hawe, Johann Hawe, David A. Schoenfeld, Liuxia Wang, Guang Li, Lilly Fang, Lester Mackey, Orla Hardiman, Merit Cudkowicz, Alexander Sherman, Alexander Sherman, Gökhan Ertaylan, Moritz Grosse-Wentrup, Torsten Hothorn, Jules van Ligtenberg, Jakob H. Macke, Timm Meyer, Bernhard Schölkopf, Linh Tran, Rubio Vaughan, Gustavo Stolovitzky, Melanie Leitner, Melanie Leitner, and Melanie Leitner. Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. *Nature Biotechnology*, 33(1):51–57, January 2015. MAG ID: 2124885415.

[93] N. Atassi, J. Berry, A. Shui, N. Zach, A. Sherman, E. Sinani, J. Walker, I. Katsovskiy, D. Schoenfeld, M. Cudkowicz, and M. Leitner. The PROACT database: Design, initial analyses, and predictive features. *Neurology*, 83(19):1719–1725, November 2014.

[94] Hamza Turabieh, Askar S. Afshar, Jeffery Statland, and Xing Song. Towards a Machine Learning Empowered Prognostic Model for Predicting Disease Progression for Amyotrophic Lateral Sclerosis. *AMIA Annual Symposium Proceedings*, 2023:718–725, January 2024.

[95] Torsten Hothorn and Hans H. Jung. RandomForest4Life: A Random Forest for predicting ALS disease progression. *Amyotrophic Lateral Sclerosis*, 15:444–452, August 2014. MAG ID: 2094100061.

[96] Corrado Pancotti, Giovanni Birolo, Cesare Rollo, Tiziana Sanavia, Barbara Di Camillo, Umberto Manera, Adriano Chiò, and Piero Fariselli. Deep learning methods to predict amyotrophic lateral sclerosis disease progression. *Scientific Reports*, 12(1):13738, August 2022. Number: 1 Publisher: Nature Publishing Group.

[97] Mei-Lyn Ong, Pei Fang Tan, Joanna D. Holbrook, and Joanna D. Holbrook. Predicting functional decline and survival in amyotrophic lateral sclerosis. *PLOS ONE*, 12(4), April 2017. MAG ID: 2606315231.

[98] Muzammil Arif Din Abdul Jabbar, Ling Guo, Sonakshi Nag, Yang Guo, Zachary Simmons, Erik P. Pioro, Savitha Ramasamy, and Crystal Jing Jing Yeo. Predicting amyotrophic lateral sclerosis (ALS) progression with machine learning. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 0(0):1–14, 2023. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21678421.2023.2285443.

[99] Sofia Pires, Marta Gromicho, Susana Pinto, Mamede Carvalho, and Sara C. Madeira. Predicting Non-invasive Ventilation in ALS Patients Using Stratified Disease Progression Groups. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 748–757, November 2018. ISSN: 2375-9259.

[100] Dan Halbersberg and Boaz Lerner. Temporal Modeling of Deterioration Patterns and Clustering for Disease Prediction of ALS Patients. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 62–68, December 2019.

[101] Vincent Grollemund, Gaétan Le Chat, Marie-Sonia Secchi-Buhour, François Delbot, Jean-François Pradat-Peyre, Peter Bede, and Pierre-François Pradat. Development and validation of a 1-year survival prognosis estimation model for Amyotrophic Lateral Sclerosis using manifold learning algorithm UMAP. *Scientific Reports*, 10:13378, August 2020.

[102] A. Guazzo, I. Trescato, E. Longato, E. Hazizaj, D. Dosso, G. Faggioli, G.M. Di Nunzio, G. Silvello, M. Vettoretti, E. Tavazzi, C. Roversi, P. Fariselli, S.C. Madeira, M. de Carvalho, M. Gromicho, A. Chiò, U. Manera, A. Dagliati, G. Birolo, H. Aidos, B. Di Camillo, and N. Ferro. Overview of iDPP@CLEF 2022: The Intelligent Disease Progression Prediction Challenge. volume 3180, pages 1130–1210, 2022. ISSN: 1613-0073.

[103] R. Branco, D.F. Soares, A.S. Martins, E. Auletta, E.N. Castanho, S. Nunes, F. Serrano, R.T. Sousa, C. Pesquita, S.C. Madeira, and H. Aidos. Hierarchical Modelling for ALS Prognosis: Predicting the Progression Towards Critical Events. volume 3180, pages 1211–1227, 2022.

[104] Aidan Mannion, Thierry Chevalier, Didier Schwab, and Lorraine Goeuriot. Predicting the Risk of & Time to Impairment for ALS patients: Report for the Lab on Intelligent Disease Progression Prediction at CLEF 2022. In *Conference & Labs of the Evaluation Forum (CLEF) 2022*, Bologne, Italy, September 2022.

[105] I. Trescato, A. Guazzo, E. Longato, E. Hazizaj, C. Roversi, E. Tavazzi, M. Vettoretti, and B. Di Camillo. Baseline Machine Learning Approaches To Predict Amyotrophic Lateral Sclerosis Disease Progression. volume 3180, pages 1277–1293, 2022. ISSN: 1613-0073.

[106] C. Pancotti, G. Birolo, T. Sanavia, C. Rollo, and P. Fariselli. Multi-Event Survival Prediction for Amyotrophic Lateral Sclerosis. volume 3180, pages 1269–1276, 2022. ISSN: 1613-0073.

[107] Andr V. Carreiro, Pedro M. T. Amaral, Susana Pinto, Pedro Toms, Mamede de Carvalho, and Sara C. Madeira. Prognostic models based on patient snapshots and time windows. *Journal of Biomedical Informatics*, 58:133–144, December 2015. MAG ID: 1820442240.

[108] Telma Pereira, Sofia Pires, Marta Gromicho, Susana Pinto, Mamede de Carvalho, and Sara C. Madeira. Predicting assisted ventilation in Amyotrophic

Lateral Sclerosis using a mixture of experts and conformal predictors. *arXiv: Learning*, July 2019. MAG ID: 2964426516.

[109] Erica Tavazzi, Sebastian Daberdaku, Alessandro Zandonà, Rosario Vasta, Beatrice Nefussy, Christian Lunetta, Gabriele Mora, Jessica Mandrioli, Enrico Grisan, Claudia Tarlarini, Andrea Calvo, Cristina Moglia, Vivian Drory, Marc Gotkine, Adriano Chiò, and Barbara Di Camillo. Predicting functional impairment trajectories in amyotrophic lateral sclerosis: a probabilistic, multifactorial model of disease progression. *Journal of Neurology*, 269(7):3858–3878, 2022.

[110] Marcel Müller, Marta Gromicho, Mamede de Carvalho, and Sara C. Madeira. Explainable models of disease progression in ALS: Learning from longitudinal clinical data with recurrent neural networks and deep model explanation. *Computer Methods and Programs in Biomedicine Update*, 1:100018, January 2021.

[111] R C. Petersen, P S. Aisen, L A. Beckett, M C. Donohue, A C. Gamst, D J. Harvey, C R. Jack, W J. Jagust, L M. Shaw, A W. Toga, J Q. Trojanowski, and M W. Weiner. Alzheimer's Disease Neuroimaging Initiative (ADNI). *Neurology*, 74(3):201–209, January 2010.

[112] Kenneth Marek, Sohini Chowdhury, Andrew Siderowf, Shirley Lasch, Christopher S. Coffey, Chelsea Caspell-Garcia, Tanya Simuni, Danna Jennings, Caroline M. Tanner, John Q. Trojanowski, Leslie M. Shaw, John Seibyl, Norbert Schuff, Andrew Singleton, Karl Kieburtz, Arthur W. Toga, Brit Mollenhauer, Doug Galasko, Lana M. Chahine, Daniel Weintraub, Tatiana Foroud, Duygu Tosun-Turgut, Kathleen Poston, Vanessa Arnedo, Mark Frasier, and Todd Sherer. The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*, 5(12):1460–1477, October 2018.

[113] Wenbin Li, Qianqian Wei, Yanbing Hou, Du Lei, Yuan Ai, Yuan Ai, Yuan Ai, Kun Qin, Jing Yang, Jing Yang, Jing Yang, Graham J. Kemp, Huifang Shang, and Qiyong Gong. Disruption of the white matter structural network and its correlation with baseline progression rate in patients with sporadic amyotrophic lateral sclerosis. *Translational neurodegeneration*, 2021. MAG ID: 3199144798.

[114] Anna Behler, Hans-Peter Müller, Albert C. Ludolph, Dorothée Lulé, and Jan Kassubek. A multivariate Bayesian classification algorithm for cerebral stage prediction by diffusion tensor imaging in amyotrophic lateral sclerosis. *NeuroImage: Clinical*, 35:103094, January 2022.

[115] G. Querin, M. M. El Mendili, T. Lenglet, S. Delphine, V. Marchand-Pauvert, H. Benali, and P.-F. Pradat. Spinal cord multi-parametric magnetic resonance imaging for survival prediction in amyotrophic lateral sclerosis. *European Journal of Neurology*, 24(8):1040–1046, 2017. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ene.13329.

[116] Anna Behler, Hans-Peter Müller, Kelly Del Tredici, Heiko Braak, Albert C. Ludolph, Dorothée Lulé, and Jan Kassubek. Multimodal in vivo staging in amyotrophic lateral sclerosis using artificial intelligence. *Annals of Clinical and Translational Neurology*, 9(7):1069–1079, 2022. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/acn3.51601.

[117] Virgilio Kmetzsch, Emmanuelle Becker, Dario Saracino, Daisy Rinaldi, Agnès Camuzat, Isabelle Le Ber, Olivier Colliot, and for the PREV-DEMALS study group. Disease Progression Score Estimation From Multimodal Imaging and MicroRNA Data Using Supervised Variational Autoencoders. *IEEE Journal of Biomedical and Health Informatics*, 26(12):6024–6035, December 2022. Conference Name: IEEE Journal of Biomedical and Health Informatics.

[118] Hannelore K. van der Burgh, Ruben Schmidt, Henk Jan Westeneng, Marcel A. de Reus, Leonard H. van den Berg, and Martijn P. van den Heuvel. Deep

learning predictions of survival based on MRI in amyotrophic lateral sclerosis. *NeuroImage: Clinical*, 13:361–369, January 2017. MAG ID: 2531444579.

[119] Jil M. Meier, Hannelore K. van der Burgh, Abram D. Nitert, Peter Bede, Siemon C. de Lange, Orla Hardiman, Leonard H. van den Berg, and Martijn P. van den Heuvel. Connectome-Based Propagation Model in Amyotrophic Lateral Sclerosis. *Annals of Neurology*, 87(5):725–738, May 2020.

[120] P. S. S. Gopi and M. Karthikeyan. Multimodal Machine Learning Based Crop Recommendation and Yield Prediction Model. *Intelligent Automation & Soft Computing*, 36(1):313–326, 2023.

[121] Rutuja R. Patil and Sumit Kumar. Rice-Fusion: A Multimodality Data Fusion Framework for Rice Disease Diagnosis. *IEEE Access*, 10:5207–5222, 2022.

[122] Nilani Algiriyage, Raj Prasanna, Kristin Stock, Emma E. H. Doyle, and David Johnston. Multi-source Multimodal Data and Deep Learning for Disaster Response: A Systematic Review. *SN Computer Science*, 3(1):92, November 2021.

[123] Shengshun Duan, Qiongfeng Shi, and Jun Wu. Multimodal Sensors and ML-Based Data Fusion for Advanced Robots. *Advanced Intelligent Systems*, 4(12):2200213, 2022. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aisy.202200213.

[124] Jing Gao, Peng Li, Zhikui Chen, and Jianing Zhang. A Survey on Deep Learning for Multimodal Data Fusion. *Neural Computation*, 32(5):829–864, May 2020.

[125] Sören Richard Stahlschmidt, Benjamin Ulfenborg, and Jane Synnergren. Multimodal deep learning for biomedical data fusion: a review. *Briefings in Bioinformatics*, 23(2):bbab569, March 2022.

[126] Xiaoqiang Yan, Shizhe Hu, Yiqiao Mao, Yangdong Ye, and Hui Yu. Deep multi-view learning methods: A review. *Neurocomputing*, 448:106–129, August 2021.

[127] Ana Lawry Aguila, Alejandra Jayme, Nina Montaña-Brown, Vincent Heuveline, and Andre Altmann. Multi-view-AE: A Python package for multi-view autoencoder models. *Journal of Open Source Software*, 8(85):5093, May 2023.

[128] James Chapman and Hao-Ting Wang. CCA-Zoo: A collection of Regularized, Deep Learning based, Kernel, and Probabilistic CCA methods in a scikit-learn style framework. *Journal of Open Source Software*, 6(68):3823, December 2021.

[129] Javier Rodriguez Zaurin and Pavol Mulinka. pytorch-widedeep: A flexible package for multimodal deep learning. *Journal of Open Source Software*, 8(86):5027, June 2023.

[130] Qian Chen, Min Li, Chen Chen, Panyun Zhou, Xiaoyi Lv, and Cheng Chen. MDFNet: application of multimodal fusion method based on skin image and clinical data to skin cancer classification. *Journal of Cancer Research and Clinical Oncology*, 149(7):3287–3299, July 2023.

[131] Riqiang Gao, Thomas Li, Yucheng Tang, Kaiwen Xu, Mirza Khan, Michael Kammer, Sanja L. Antic, Stephen Deppen, Yuankai Huo, Thomas A. Lasko, Kim L. Sandler, Fabien Maldonado, and Bennett A. Landman. Reducing uncertainty in cancer risk estimation for patients with indeterminate pulmonary nodules using an integrated deep learning model. *Computers in Biology and Medicine*, 150:106113, November 2022.

[132] Hongyi Duanmu, Pauline Boning Huang, Srinidhi Brahmavar, Stephanie Lin, Thomas Ren, Jun Kong, Fusheng Wang, and Tim Q. Duong. Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast

Cancer Using Deep Learning with Integrative Imaging, Molecular and Demographic Data. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, Lecture Notes in Computer Science, pages 242–252, Cham, 2020. Springer International Publishing.

[133] Michal Golovanevsky, Carsten Eickhoff, and Ritambhara Singh. Multimodal attention-based deep learning for Alzheimer's disease diagnosis. *Journal of the American Medical Informatics Association*, 29(12):2014–2022, November 2022.

[134] Kyriaki-Margarita Bintsi, Vasileios Baltatzis, Rolandos Alexandros Potamias, Alexander Hammers, and Daniel Rueckert. Multimodal brain age estimation using interpretable adaptive population-graph learning, July 2023. arXiv:2307.04639 [cs].

[135] Luigi Antelmi, Nicholas Ayache, Philippe Robert, and Marco Lorenzi. Sparse Multi-Channel Variational Autoencoder for the Joint Analysis of Heterogeneous Data. In *Proceedings of the 36th International Conference on Machine Learning*, pages 302–311. PMLR, May 2019. ISSN: 2640-3498.

[136] Weipeng Li, Jiaxin Zhuang, Ruixuan Wang, Jianguo Zhang, and Wei-Shi Zheng. Fusing Metadata and Dermoscopy Images for Skin Disease Diagnosis. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1996–2000, April 2020. ISSN: 1945-8452.

[137] Dan Zhao, Morteza Homayounfar, Zhe Zhen, Mei-Zhen Wu, Shuk Yin Yu, Kai-Hang Yiu, Varut Vardhanabhuti, George Pelekos, Lijian Jin, and Mohamad Koohi-Moghadam. A Multimodal Deep Learning Approach to Predicting Systemic Diseases from Oral Conditions. *Diagnostics*, 12(12):3192, December 2022.

[138] Rui Yan, Fa Zhang, Xiaosong Rao, Zhilong Lv, Jintao Li, Lingling Zhang, Shuang Liang, Yilin Li, Fei Ren, Chunhou Zheng, and Jun Liang. Richer fusion network for breast cancer classification based on multimodal data. *BMC Medical Informatics and Decision Making*, 21(1):134, April 2021.