Morphometric MRI as a diagnostic biomarker of frontotemporal dementia: A systematic review to determine clinical applicability

Jillian McCarthy, D. Louis Collins, Simon Ducharme

McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada

ARTICLE INFO

Keywords:
Frontotemporal dementia Classification MRI Morphometric analysis Machine learning Diagnostic biomarker

ABSTRACT

Frontotemporal dementia (FTD) is difficult to diagnose, due to its heterogeneous nature and overlap in symptoms with primary psychiatric disorders. Brain MRI for atrophy is a key biomarker but lacks sensitivity in the early stage. Morphometric MRI-based measures and machine learning techniques are a promising tool to improve diagnostic accuracy. Our aim was to review the current state of the literature using morphometric MRI to classify FTD and assess its applicability for clinical practice. A search was completed using Pubmed and PsychInfo of studies which conducted a classification of subjects with FTD from non-FTD (controls or another disorder) using morphometric MRI metrics on an individual level, using single or combined approaches. 28 relevant articles were included and systematically reviewed following PRISMA guidelines. The studies were categorized based on the type of FTD subjects included and the group(s) against which they were classified. Studies varied considerably in subject selection, MRI methodology, and classification approach, and results are highly heterogeneous. Overall many studies indicate good diagnostic accuracy, with higher performance when differentiating FTD from controls (highest result was accuracy of 100%) than other dementias (highest result was AUC of 0.874). Very few machine learning algorithms have been tested in prospective replication. In conclusion, morphometric MRI with machine learning shows potential as an early diagnostic biomarker of FTD, however studies which use rigorous methodology and validate findings in an independent real-life cohort are necessary before this method can be recommended for use clinically.

1. Introduction

Frontotemporal dementia (FTD) is one of the most common forms of early onset dementia, occurring with similar frequency to Alzheimer’s Disease (AD) in people under the age of 65 (Onyike and Diehl-Schmid, 2013). This heterogeneous disorder most often presents with combinations of personality changes such as apathy, loss of empathy, and disinhibition (behavioral variant – bvFTD) (Rascovsky et al., 2011) or language deficits (primary progressive aphasia – PPA). PPA is further divided into three variants - semantic (svPPA), nonfluent (nfvPPA) and logopenic (lvPPA) (Gorno-Tempini et al., 2011). The pathology underlying frontotemporal lobar degeneration is equally heterogeneous and involves abnormal accumulation of proteins including microtubule-associated protein tau, transactive response DNA-binding protein with molecular weight 43 kDa (TDP-43), and fused in sarcoma (FUS) protein, while the lvPPA clinical syndrome is most commonly associated with...
AD pathology (Rademakers et al., 2012).

The diagnosis of FTD currently poses a significant challenge for clinicians as the presenting symptoms overlap considerably with other diseases including primary psychiatric disorders and other dementias (Ducharme et al., 2015). This is especially true of bvFTD. Evidence suggests as many as 50% of people with bvFTD are initially diagnosed with a psychiatric disorder (Woolley et al., 2011). As well, significant memory impairment can exist in bvFTD, comparable to that seen in AD (Bertoux et al., 2014; Mansoor et al., 2015).

The most common imaging method currently used in clinical practice is structural MRI, which is insufficiently sensitive for early stage diagnosis of FTD given that atrophy can be very subtle at the disease onset. Indeed, in a mixed neuropsychiatric population that is representative of clinical practice, a standard MRI with visual review had insufficient sensitivity (70%) to identify cases with bvFTD, while the usual alternative of [18F]FDG-PET had poor specificity (68%) (Vijverberg et al., 2016). This can lead to erroneous or significantly delayed diagnosis, causing prolonged periods of uncertainty for patients and their families. The development of improved diagnostic biomarkers for the early detection of FTD is critical to ensure patients are getting the appropriate care as well as for the accurate identification of patients for clinical trials. Improving MRI methods is ideal given that MRI is already part of standard practice and there are currently no validated molecular biomarkers for FTD diagnosis. AD cerebral spinal fluid (CSF) and PET amyloid tracers can be used in the differential diagnosis of FTD from AD, as FTD will likely be negative for these (Meeter et al., 2017), however FTD-specific CSF biomarkers or tau tracers are not available.

There has been considerable interest in automated morphometric analysis of MRI, most commonly assessing gray matter (GM) atrophy and, in recent years, white matter (WM) integrity using diffusion tensor imaging (DTI). Techniques such as voxel-based morphometry (VBM) and cortical thickness have demonstrated specific patterns of frontal and temporal GM atrophy on a group level (Meeter et al., 2017). These patterns differ from those seen in other dementias (such as hippocampal atrophy found in AD). BvFTD is associated with atrophy primarily in the frontal lobe, insula, anterior cingulate cortex and basal ganglia (Meeter et al., 2017; Pan et al., 2012; Schroeter et al., 2014). PPA is primarily associated with left-sided atrophy (language dominant hemisphere) in the initial disease stages; nfvPPA with inferior frontal and insular atrophy, svPPA with anterior temporal atrophy, and lvPPA with posterior temporal and parietal atrophy (Bisenius et al., 2016; Meeter et al., 2017; Mesulam et al., 2009; Rogalski et al., 2014). WM changes have a more widespread distribution and likely precede GM atrophy (Lam et al., 2014; Mahoney et al., 2014; Meeter et al., 2017).

A high discriminative power is needed to differentiate between diseases on an individual level, in order to be useful in clinical practice. However, with improving methods of morphometric analysis and the use of multivariate statistics and machine learning methods, it is becoming increasingly feasible to improve diagnosis at the individual level. An extensive body of literature exists classifying AD in this way. An extensive body of literature exists classifying AD in this way.

The diagnosis of FTD currently poses a significant challenge for clinicians as the presenting symptoms overlap considerably with other diseases including primary psychiatric disorders and other dementias (Ducharme et al., 2015). This is especially true of bvFTD. Evidence suggests as many as 50% of people with bvFTD are initially diagnosed with a psychiatric disorder (Woolley et al., 2011). As well, significant memory impairment can exist in bvFTD, comparable to that seen in AD (Bertoux et al., 2014; Mansoor et al., 2015).

Eleven studies conducted a binary classification of FTD or specifically bvFTD from a control group. Seventeen studies conducted a binary classification of FTD or specifically bvFTD from AD. Six studies conducted a multi-class classification to differentiate FTD, AD and controls, while four studies conducted a multi-class classification between various dementia types and controls. Four studies conducted classifications of PPA; two studies differentiated PPA subtypes from each other and controls. One study classified PPA from controls. One study differentiated FTD subtypes (bvFTD and PPA) from a combined group of all other subtypes and AD. Results are summarized in Tables 1–5. Accuracy, sensitivity, specificity, and/or area under the receiver operating characteristic curve (AUC) are reported, if provided. In cases where raw numbers were reported, applicable performance measures were calculated from these numbers. In this paper we consider performance of 90% or greater as high, 70–90% as moderate, and < 70% as low.

Studies varied considerably in methodology. The majority of studies looked at changes in GM structure, most commonly using VBM to assess features derived from structural MRI, either alone or in combination. In the case of studies which conducted classifications based on MRI morphometry alone and in combination with other methods, only those results pertaining to MRI morphometry were included in this review. Reference lists of included articles were also manually searched to identify other relevant articles. The risk of bias and applicability of each included study was assessed with the QUADAS-2 tool (Whiting et al., 2011).

3. Results

The search produced 151 articles. Of these, 25 relevant articles were identified. Cross-reference list searches of each relevant article yielded three additional papers, resulting in a total of 28 papers for inclusion in this review (Fig. 1).

3.1. Study characteristics

Eleven studies conducted a binary classification of FTD or specifically bvFTD from a control group. Seventeen studies conducted a binary classification of FTD or specifically bvFTD from AD. Six studies conducted a multi-class classification to differentiate FTD, AD and controls, while four studies conducted a multi-class classification between various dementia types and controls. Four studies conducted classifications of PPA; two studies differentiated PPA subtypes from each other and controls. One study classified PPA from controls. One study differentiated FTD subtypes (bvFTD and PPA) from a combined group of all other subtypes and AD. Results are summarized in Tables 1–5. Accuracy, sensitivity, specificity, and/or area under the receiver operating characteristic curve (AUC) are reported, if provided. In cases where raw numbers were reported, applicable performance measures were calculated from these numbers. In this paper we consider performance of 90% or greater as high, 70–90% as moderate, and < 70% as low.

Studies varied considerably in methodology. The majority of studies looked at changes in GM structure, most commonly using VBM to assess either GM concentration or volume. WM integrity was commonly assessed using DTI measures. Studies used a variety of whole-brain and region of interest (ROI) based approaches, including a priori selection of ROIs and the use of ROIs that showed significant differences in group-level comparison. Studies also varied widely in classification methods. Machine learning classification techniques were utilized by most studies, the most common being support vector machines (SVM). Most studies used a k-fold cross validation (CV) approach, most commonly with a leave-one-out CV strategy. Only one study used independent subject data (from a different cohort) in a separate testing set (Klöppel et al., 2015).

Almost all studies used a clinically defined diagnosis as the reference standard. Six studies (Chow et al., 2008; Frings et al., 2014; Mahoney et al., 2014; Meyer et al., 2017; Muñoz-Ruiz et al., 2012; Wang et al., 2016) included a subset of patients with pathologically confirmed diagnosis or those with a known genetic mutation consistent with FTD. Three studies (Klöppel et al., 2008b; Lehmann et al., 2010; Vemuri et al., 2011) used pathologically defined dementia diagnosis as the gold standard. Two studies (McMillan et al., 2014; McMillan et al., 2012) grouped subjects as AD or FTD based on the presence or absence of CSF biomarkers consistent with AD. Studies also varied considerably
in disease severity. Studies report a variety of methods for evaluating disease severity (Mini Mental State Exam, Clinical Dementia Rating, disease duration) making comparison difficult. Four studies used a control group consisting in part or entirely of those with subjective cognitive decline (Dukart et al., 2011; Koikkalainen et al., 2016; Möller et al., 2016; Tong et al., 2017). All others consisted of healthy, cognitively normal subjects. Studies also varied widely in their exclusion criteria. Some studies included FTD with concurrent motor symptoms while others excluded these subjects.

3.2. bvFTD vs Controls

Five studies classified bvFTD from a control group (Chow et al., 2008; Mahoney et al., 2014; Meyer et al., 2017; Möller et al., 2016; Raamana et al., 2014) (Table 1 and Fig. 2a). In general studies could distinguish FTD from controls with moderate to high accuracy, although results are heterogeneous. Two studies measured GM concentration with VBM using a SVM classifier. Meyer et al. (2017) achieved highest accuracy, sensitivity and specificity when using a ROI approach (frontal and temporal lobes – 84.6%, 80.7% and 88.5%, respectively), while Möller et al. (2016) reported low sensitivity (60%) but high specificity (98%) with a whole-brain approach. Mahoney et al. (2014) achieved moderate results using radial diffusivity from DTI. The highest result was reported by Raamana et al. (2014) using surface displacements of the left lateral ventricle as inputs to a SVM, using a train/test approach (AUC of 0.938, sensitivity of 100% and specificity of 88%) The result was somewhat lower when using leave-one-out CV (AUC of 0.826, sensitivity of 79, specificity of 87). These results contrast with this study's reported results for other regions (right lateral ventricle and left and right hippocampus) in which sensitivity is low. None of the studies classifying the bvFTD subtype from controls looked at different MRI metrics in combination.

3.3. FTD vs controls

Six studies classified a combined group of FTD clinical subtypes from a control group (Table 2 and Fig. 2b), again with overall moderate to high accuracy (Bron et al., 2017; Davatzikos et al., 2008; Du et al., 2007; Dukart et al., 2011; Muñoz-Ruiz et al., 2012; Zhang et al., 2013). Davatzikos et al. (2008) reported 100% accuracy when using GM and WM volumetric features derived from principle component analysis as inputs to an SVM, however this study was small (FTD n = 12) and may not have used a completely independent test set. Very high results were also reported by Bron et al. (2017) when using GM, WM, or supratentorial brain volume with an SVM (AUC 0.95–0.96). This study did not report sensitivity and specificity numbers. In contrast, Zhang et al. (2013) reported poor results using GM or WM volumes and logistic regression in a ROI approach extracted from group differences, but achieved best results using radial diffusivity (accuracy, sensitivity, specificity, and AUC of 81.4%, 80.7%, 80.5%, 0.877, respectively). Two other studies reported moderately high results using various measures of GM structure alone (tensor-based morphometry, volumetry, VBM, cortical thickness) (Du et al., 2007; Muñoz-Ruiz et al., 2012). Only one study (Bron et al., 2017) assessed a multimodal approach (WM volume and fractional anisotropy), which achieved a similar result to that by WM volume alone (AUC 0.95).

3.4. bvFTD vs AD

Six studies classified bvFTD from AD (Canu et al., 2017; Frings et al., 2014; Mahoney et al., 2014; Möller et al., 2016; Raamana et al., 2014; Wang et al., 2016) (Table 1 and Fig. 2c). In general, results indicate that this is a much harder task than distinguishing from controls and results are highly variable. Canu et al. (2017) achieved moderately high results using cortical thickness in a random forest approach to
distinguish bvFTD from AD (accuracy, sensitivity, and specificity of 82%, 80%, and 87% respectively). These results were not majorly improved when combined with DTI measures. No other study looked at the accuracy of combined MRI metrics. Other studies reported low to moderate accuracy in classifying bvFTD from AD using a range of single metrics including DTI, GM concentration, volumetry, and surface displacements (Frings et al., 2014; Mahoney et al., 2014; Möller et al., 2016; Raamana et al., 2014; Wang et al., 2016).

### 3.5. FTD vs AD

Eleven studies classified FTD (combined clinical subtypes, pathological subtypes, or CSF-defined) from AD (Bron et al., 2017; Davatzikos et al., 2008; Du et al., 2007; Dukart et al., 2011; Klöppel et al., 2015; Klöppel et al., 2008b; Lehmann et al., 2010; McMillan et al., 2014; McMillan et al., 2012; Muñoz-Ruiz et al., 2012; Whitwell et al., 2011) (Table 2 and Fig. 2d). Again, results are highly variable. McMillan et al. (2012) reported highest accuracy when using a combination of GM density and fractional anisotropy (sensitivity, specificity and AUC of 87%, 83%, and 0.93 respectively) when distinguishing CSF-defined FTD and AD using regression, although this study did not use an independent testing set. McMillan et al. (2014) also reported moderately high sensitivity, specificity and AUC (89%, 89%, and 0.874 respectively) to classify CSF-defined FTD and AD when using a combination of cortical thickness and fractional anisotropy in a data-driven approach. In contrast Klöppel et al. (2008b) reported similar numbers using GM volume alone, in a whole-brain approach with an SVM (accuracy, sensitivity, and specificity of 89.2%, 94.7%, and 83.3% respectively), while Whitwell et al. (2011) reported high AUC (0.93) using GM volumes of the temporoparietal cortex and hippocampus. Other studies again reported low to moderate accuracy in classifying FTD from AD with a range of different metrics (Bron et al., 2017; Davatzikos et al., 2008; Du et al., 2007; Dukart et al., 2011; Klöppel et al., 2015; Lehmann et al., 2010; Muñoz-Ruiz et al., 2012).
3.6. Multi-class classifications

Several studies attempted a multi-class classification with varying accuracy. Six studies included a three-way classification between FTD, AD, and controls (Bron et al., 2017; Dukart et al., 2011; Kuceyeski et al., 2012; Möller et al., 2015; Raamana et al., 2014; Wang et al., 2016) (Table 3). Kuceyeski et al. (2012) reported the highest accuracy using radial diffusivity, with accuracy and sensitivity of 89.09% and 97.3% but lower specificity (72.22%) using linear discriminant analysis. Results were similar using the LoCo metric, a measurement of the amount of structural network disruption incurred by a GM region for a particular pattern of WM integrity loss (accuracy, sensitivity, and specificity of 87.27%, 91.89%, 77.78% respectively). Four studies conducted a multi-class classification between various dementias and controls.

Table 2: Classifications of FTD vs Controls or AD.

| Name          | Sample | Classification | Measure                        | ROIs                           | vs Controls Acc | vs AD Acc | vs Controls SS | vs AD SS | vs Controls SP AUC | vs AD SP AUC |
|---------------|--------|----------------|--------------------------------|--------------------------------|-----------------|-----------|-----------------|-----------|---------------------|--------------|
| Bron et al., 2017 | 33 FTD | SVM            | VBM-GM volume                   | Whole-brain                    | 0.95            | 0.78      |                 |           |                     |              |
|               | 24 AD  | 4-fold CV      | VBM-WM volume                   |                                | 0.96            | 0.76      |                 |           |                     |              |
|               | 34C    |                | VBM-Supratentorial brain volume |                               | 0.95            | 0.72      |                 |           |                     |              |
|               |        |                | DTI-FA                          |                               |                 |           |                 |           |                     |              |
|               |        |                | VBM-WM volume + DTI-FA           |                               | 0.91            | 0.80      |                 |           | 0.95                | 0.81         |
| Davatzikos et al., 2008 | 12 FTD | SVM            | RAVENS-GM and WM volume         | PCA                            | 100             | 84.3      |                 |           |                     |              |
|               | 37 AD  | LOOCV          |                                 |                                |                 |           |                 |           |                     |              |
|               | 12 C   |                |                                 | Volume                         | 75              | 70.9      |                 |           |                     |              |
| Du et al., 2007 | 19 FTD | Logistic       |                                  | Volume                         | Frontal         | 89        |                 |           |                     |              |
|               |        | regression     |                                 |                                |                 |           |                 |           |                     |              |
|               | 22 AD  | LOOCV          |                                 | Parietal                        | 81              | 79        |                 |           |                     |              |
|               | 23 C   |                |                                 | Temporal                        | 85              |           |                 |           |                     |              |
|               |        |                |                                 | Cortical thickness              | Frontal         | 88        |                 |           |                     |              |
|               |        |                |                                 | Parietal                        | 82              | 82        |                 |           |                     |              |
|               |        |                |                                 | Temporal                        | 85              |           |                 |           |                     |              |
| Dukart et al., 2011 | 14 FTD | SVM            | GM                              | Whole-brain                    | 77.8            | 80        |                 |           |                     |              |
|               | 21 AD  | LOOCV          | WM                              | Whole-brain                     | 77.8            | 74.3      |                 |           |                     |              |
|               | 13C    |                | GM                              | ROI (a priori)                  | 85.2            | 60        |                 |           |                     |              |
| Kloppel et al., 2008b | 19 FTD | SVM            | GM volume                       | Whole-brain                    | 89.2            | 94.7      | 83.3            |           |                     |              |
| Kloppel et al., 2015 | 12 FTD | SVM            | VBM-GM volume                   | Whole-brain                    | 89.2            | 94.7      | 83.3            |           |                     |              |
| Lehmann et al., 2010 | 23 FTD | SVM            | Cortical Thickness              | Whole-brain                    | 79.4            | 91.3      | 54.5            | 0.87     |                     |              |
| McMillan et al., 2012 | 38 FTD | Logistic       | GM density                      | Parietal                        | 82              | 79        | 88.3            |           |                     |              |
|               | 29 AD  | regression     |                                 | Temporal                        | 97              | 69        | 79.2            |           |                     |              |
|               |        |                |                                 | Cortical thickness              | Frontal         | 99        | 59              | 79.95    |                     |              |
| McMillan et al., 2014 | 72 FTD | Linear         | Cortical thickness              | Data-driven                     | 89              | 81        | 77.8            |           |                     |              |
|               | 21 AD  | regression     |                                 | Anatomical                      | 100             | 54        | 80.2            |           |                     |              |
|               |        | Train/test     |                                 | Volume                          | 65              | 100       | 82.0            |           |                     |              |
|               |        |                |                                 | Global GM                       | 65              | 100       | 82.0            |           |                     |              |
|               |        |                |                                 | Global ventricles               | 100             | 65        | 82.6            |           |                     |              |
|               |        |                |                                 | Data-driven                     | 100             | 46        | 80.8            |           |                     |              |
|               |        |                |                                 | Anatomical                      | 56              | 76        | 64.9            |           |                     |              |
|               |        |                |                                 | Data-driven                     | 89              | 89        | 87.4            |           |                     |              |
|               |        |                |                                 | Anatomical                      | 78              | 70        | 62              |           |                     |              |
| Muñoz-Ruiz et al., 2012 | 37 FTD | Regression     | Volume                          | Hippocampus                     | 83              | 80        | 84              | 55        | 55                  | 55           |
|               | 46 AD  | Train/Test     |                                 |                                |                 |           |                 |           |                     |              |
|               | 26 C   |                |                                 | TBM                             | 82              | 90        | 77              | 62        | 67                  | 76           |
|               |        |                |                                 | VBM-GM concentration            | 83              | 91        | 77              | 72        | 76                  | 67           |
|               |        |                |                                 | VBM-GM volume                   | 85              | 89        | 82              | 69        | 71                  | 66           |
| Whitwell et al., 2011 | 14 FTD | Logistic       | GM volume                       | Temporoparietal cortex          | 65.7            | 80.1      | 48.7            | 0.665    |                     |              |
|               | 14 AD  | regression     |                                 | Hippocampus                     | 89              | 89        | 68              |           |                     |              |
| Zhang et al., 2013 | 25 FTD | Logistic       | VBM-GM volume                   | ROI1 (B frontotemporal, anterior callosal) | 65.7            | 80.1      | 48.7            | 0.665    |                     |              |
|               | 19 C   | regression     |                                 | ROI2 (l. temporal)              | 63.9            | 77.0      | 46.6            | 0.722    |                     |              |
|               |        | 4-fold CV      |                                 | ROI3 (l. dorsal frontal)        | 45.7            | 74.2      | 5.4             | 0.566    |                     |              |
|               |        |                |                                 | VBM-WM volume                   | 59.2            | 77.2      | 34.6            | 0.627    |                     |              |
|               |        |                |                                 | ROI2                            | 58.1            | 71.5      | 36.4            | 0.657    |                     |              |
|               |        |                |                                 | ROI3                            | 47.4            | 79.8      | 5.3             | 0.606    |                     |              |
|               |        |                |                                 | DTI-RD                          | 76.0            | 79.9      | 72.3            | 0.853    |                     |              |
|               |        |                |                                 | ROI2                            | 81.4            | 80.7      | 80.5            | 0.877    |                     |              |
|               |        |                |                                 | ROI3                            | 67.6            | 73.3      | 58.6            | 0.722    |                     |              |
### Table 3

**Multi-class Classifications of FTD, AD, and Controls.**

| Name            | Sample | Classification | Measure                          | ROIs                              | Acc | SS (FTD) | SP (FTD) | AUC  |
|-----------------|--------|----------------|----------------------------------|-----------------------------------|-----|----------|----------|------|
| **FTD, AD and controls** |        |                |                                  |                                   |     |          |          |      |
| Bron et al., 2017 | 33 FTD | SVM            | VBM-GM volume                    | Whole-brain                       | 0.85|          |          |      |
|                 | 24 AD  | 4-fold CV      | VBM-WM volume                    |                                   | 0.83|          |          |      |
|                 | 34 C   |                | VBM-Supratentorial brain volume  |                                   | 0.84|          |          |      |
|                 |        |                | DTI-FA                           |                                   | 0.83|          |          | 0.87 |
| Dukart et al., 2011 | 14 FTD | SVM            | GM                               | Whole-brain                       | 72.9|          |          |      |
|                 | 21 AD  | LOOCV          | WM                               |                                   | 66.7|          |          |      |
|                 | 13 C   |                | GM                               | a priori ROIs                     | 56.3|          |          |      |
| Kuceyeski et al., 2012 | 18 FTD | Linear discriminant analysis | GM volume                        | Whole-brain parcellation          | 76.36| 81.08  | 66.67   |      |
|                 |        |                | DWI-FA                           |                                   | 76.36| 72.97  | 83.33   |      |
|                 |        |                | DWI-RD                           |                                   | 89.09| 97.30  | 72.22   |      |
|                 |        |                | DWI-LD                           |                                   | 85.45| 89.19  | 77.78   |      |
|                 |        |                | Combination GM + DWI             |                                   | 83.64| 91.89  | 66.67   |      |
|                 |        |                | LoCo                             |                                   | 87.27| 91.89  | 77.78   |      |
| Möller et al., 2015 | 30 bvFTD | Discriminant function analyses LOOCV | 1st analysis: VBM-GM volume, Subcortical volumes, DWI-FA | Significant voxels/regions from paired group comparisons | 91.4| 66.7   |          |      |
|                 | 39 AD  |                |                                 |                                   |     |          |          |      |
|                 | 41 C   |                |                                 |                                   |     |          |          |      |
| Raamana et al., 2014 | 30 bvFTD | SVM            | Volumes                          |                                     |     |          |          |      |
|                 | 34 AD  |                |                                 |                                     |     |          |          |      |
|                 | 14 C   |                |                                 |                                     |     |          |          |      |
| Klöppel et al., 2015 | 12 FTD | SVM            | VBM-GM volume                    | Whole-brain                        | 50.4|          |          | 0.78 |
|                 | 122 AD |                | VBM-GM concentration             |                                     | 65.1|          |          |      |
|                 | 4 DLBL |                | TBM                              |                                     | 64.3|          |          |      |
|                 | 18 C   |                | ROI-based grading                |                                     | 56.4|          |          |      |
| Koikkalainen et al., 2016 | 92 FTD | Disease State Index (DSI)        | Volumes                           | Whole-brain parcellation           | 50.4|          |          |      |
|                 | 10-fold CV |                |                                 |                                     | 65.1|          |          |      |
|                 | 223 AD |                |                                 |                                     | 64.3|          |          |      |
|                 | 47 DLBL|                |                                 |                                     | 56.4|          |          |      |
| Tong et al., 2017 | 92 FTD | SVM            | Volumes                          | Whole-brain parcellation           | 50.4|          |          |      |
|                 | 10-fold CV |                |                                 |                                     | 65.1|          |          |      |
|                 | 219 AD |                |                                 |                                     | 64.3|          |          |      |
|                 | 47 DLBL|                |                                 |                                     | 56.4|          |          |      |
| Vemuri et al., 2011 | 47 FTD | SVM            | VBM-GM volume                    | Whole-brain                        | 84.4|          |          |      |
|                 | 48 AD  |                | VBM-Supratentorial brain volume  |                                     | 93.8|          |          |      |
|                 | 20 DLBL|                | DTI-FA                           |                                     |      |          |          |      |
|                 | 21 C   |                | GM                               |                                     |      |          |          |      |

### Table 4

**Multi-class Classifications of Dementia.**

**Multi dementia types**

| Name            | Sample | Classification | Measures                          | ROIs                              | Acc | SS (FTD) | SP (FTD) | AUC  |
|-----------------|--------|----------------|----------------------------------|-----------------------------------|-----|----------|----------|------|
| Möller et al., 2015 | 12 FTD | SVM            | VBM-GM volume                    | Whole-brain                        | 50.4|          |          | 0.78 |
|                 | 122 AD |                | VBM-GM concentration             |                                     | 65.1|          |          |      |
|                 | 4 DLBL |                | TBM                              |                                     | 64.3|          |          |      |
|                 | 18 C   |                | ROI-based grading                |                                     | 56.4|          |          |      |
| Koikkalainen et al., 2016 | 92 FTD | Disease State Index (DSI)        | Volumes                           | Whole-brain parcellation           | 50.4|          |          |      |
|                 | 10-fold CV |                |                                 |                                     | 65.1|          |          |      |
|                 | 223 AD |                |                                 |                                     | 64.3|          |          |      |
|                 | 47 DLBL|                |                                 |                                     | 56.4|          |          |      |
| Tong et al., 2017 | 92 FTD | SVM            | VBM-GM volume                    | Whole-brain                        | 50.4|          |          | 0.78 |
|                 | 10-fold CV |                |                                 |                                     | 65.1|          |          |      |
|                 | 219 AD |                |                                 |                                     | 64.3|          |          |      |
| Vemuri et al., 2011 | 47 FTD | SVM            | VBM-GM volume                    | Whole-brain                        | 50.4|          |          | 0.78 |
|                 | 48 AD  |                | VBM-Supratentorial brain volume  |                                     | 65.1|          |          |      |
|                 | 20 DLBL|                | DTI-FA                           |                                     | 64.3|          |          |      |
|                 | 21 C   |                | GM                               |                                     | 56.4|          |          |      |
Table 5
PPA classifications.

| Name                  | Sample      | Classification | Measures          | ROIs                                      | Acc  | SS  | SP  | AUC | Acc  | SS  | SP  | AUC | Acc  | SS  | SP  | AUC |
|-----------------------|-------------|----------------|-------------------|-------------------------------------------|------|-----|-----|-----|------|-----|-----|-----|------|-----|-----|-----|
| Bisenius et al., 2017 | 16 nfvPPA   | SVM            | VBM-GM density    | Whole-brain ROI (a priori from meta-analyses) | 91   | 88  | 94  | 0.94| 95   | 91  | 100 | 0.95| 97   | 94  | 100 | 0.97|
|                       | 17 lvPPA    | LOOCV          |                   |                                           | 84   | 81  | 88  | 0.90| 82   | 82  | 82  | 0.91| 100  | 100 | 1   | 1   |
|                       | 11 lvPPA    |                |                   |                                           | 82   | 82  | 82  | 0.91| 100  | 100 | 1   | 1   |      |     |     |     |
| Wilson et al., 2009   | 32 svPPA    | SVM            | GM volume         | PCA                                       | 89.1 | 87.5| 90.6| 0.94| 100  | 100 | 1   | 1   | 100  | 100 | 1   | 1   |
|                       | 38 svPPA    | 2-level CV     |                   |                                           |      |     |     |     |      |     |     |     |      |     |     |     |
|                       | 16 lvPPA    | 115 C          |                   |                                           |      |     |     |     |      |     |     |     |      |     |     |     |
| Bisenius et al., 2017 | 16 nfvPPA   | SVM            | VBM-GM density    | Whole-brain ROI (a priori from meta-analyses) | 78   | 81  | 75  | 0.88| 95   | 100 | 91  | 0.93| 55   | 64  | 45  | 0.59|
|                       | 17 lvPPA    | LOOCV          |                   |                                           | 78   | 81  | 75  | 0.87| 95   | 100 | 91  | 0.91| 64   | 73  | 55  | 0.64|
| Wilson et al., 2009   | 32 svPPA    | SVM            | GM volume         | PCA                                       | 89.1 | 84.4| 93.8| 0.964| 93.8 | 93.8| 93.8| 0.984| 81.3 | 81.3| 81.3| 0.879|
|                       | 38 svPPA    | 2-level CV     |                   |                                           |      |     |     |     |      |     |     |     |      |     |     |     |
|                       | 16 lvPPA    | 115 C          |                   |                                           |      |     |     |     |      |     |     |     |      |     |     |     |
| Chow et al., 2008     | 14 PPA      | Logistic       | Volumes           | L anterior temporal                       | 90.9 |     |     |     | 78.6 |     |     | 96.7 |      |     |     |     |
|                       | 30 C        | regression     |                   |                                           |      |     |     |     |      |     |     |     |      |     |     |     |
| Tahmasian et al., 2016| 11 bvFTD    | SVM            | VBM-GM volume     | A priori based on the NDH                 | 72.5 | 45.4| 82.7|     | 92.5 | 50  | 97.5|     | 82.5 | 0   | 94.2|     |
|                       | 4 svPPA     | LOOCV          |                   |                                           |      |     |     |     |      |     |     |     |      |     |     |     |
|                       | 5 nfvPPA    |                |                   |                                           |      |     |     |     |      |     |     |     |      |     |     |     |
|                       | 20 AD       |                |                   |                                           |      |     |     |     |      |     |     |     |      |     |     |     |

Fig. 2. Visual representation of the classification accuracy for the different comparisons (for studies which conducted more than one classification, the best result is shown). a) behavioral variant frontotemporal dementia (bvFTD) vs Controls. b) Frontotemporal dementia (any subtype - FTD) vs Controls. c) bvFTD vs AD. d) FTD (any subtype) vs AD.
Vemuri et al., 2011) reported moderate sensitivity (84.4%) and high specificity (93.8%) for FTD classification versus all others using whole brain GM density approach and a novel classification approach (referred to as differential-STAND), however they did not have a completely independent test set. Results were considerably lower for other studies (Klöppel et al., 2015; Koikkalainen et al., 2016; Tong et al., 2017).

### 3.7. PPA subtypes

Four studies included classifications of PPA (Bisenius et al., 2017; Chow et al., 2008; Tahmasian et al., 2016; Wilson et al., 2009) (Table 5). Two studies classified each PPA subtype against controls using SVM of GM atrophy, with moderate to high accuracy across studies (accuracy ranged from 84 to 100%) (Bisenius et al., 2017; Wilson et al., 2009). Both studies also classified subtypes against each other, with varying results. Wilson et al. (2009) reported highest accuracy, sensitivity, and specificity (89.1%, 84.4%, 93.8% respectively, AUC of 0.964) to distinguish svPPA from nfvPPA using GM volume and a principal component analysis approach. Results were very high for both studies for lvPPA vs svPPA, while Wilson et al. (2009) achieved highest results for lvPPA vs nfvPPA (accuracy, sensitivity, specificity, AUC of 81.3%, 81.3%, 81.3% and 0.879 respectively). Tahmasian et al. (2016) classified each FTD subtype against a group of all others and AD using GM volume and SVM, resulting in high specificity (97.5% and 94.2%) but very poor sensitivity (50% and 0%) for both svPPA and nfvPPA vs others, while Chow et al. (2008) combined svPPA and nfvPPA subtypes together in a classification from a control group, achieving moderate sensitivity (78.6%) and high specificity (96.7%).

### 3.8. Risk of bias assessment

The results of the QUADAS-2 evaluation are given in Table 6. The patient selection domain was rated as high risk of bias in six studies that had inappropriate exclusion criteria (e.g. exclusion of subjects with abnormalities on structural MRI other than atrophy, such as WM hyperintensities) combined with a case-control design. The index test was rated as high risk of bias in eight studies which did not use separate testing data or used all data to perform ROI selection or dimensionality reduction prior to classification. Two studies were given an unclear risk of bias on this domain. One study was rated as having applicability concerns on the index test domain as it only looked at the overall accuracy of multi-class classification of dementia types.

### 4. Discussion

This systematic review provides a summary of studies attempting to classify FTD from non-FTD via morphometric MRI data with the aim to determine its potential for use as a diagnostic aid in clinical practice. Studies included in this review are highly heterogeneous in terms of subject selection, MRI methodology and classification methods, complicating the comparison of accuracy of results. However, overall studies report good levels of accuracy (see Table 7 for a summary of the best performance for each classification), indicating the potential value of MRI morphometry in the diagnosis of FTD. FTD could be diagnosed with high accuracy from control groups, with many studies finding accuracies of over 80% or 90% with good sensitivity and specificity. However, most studies include subjects with well characterized patients in which there is significant atrophy, and therefore the added benefit of morphometry is uncertain. Results distinguishing FTD from AD were somewhat poorer. This is unsurprising given that minimal atrophy is expected in control subjects and that there exists overlap in atrophy patterns between FTD and AD (De Souza et al., 2013). Studies which conducted multi-class classifications did not all report specific sensitivity and specificity values for FTD, although Vemuri et al. (2011) reported good sensitivity and specificity (84.4% and 93.8%) in distinguishing FTD from other dementias. Only four studies specifically classified PPAs, generally with moderate to high accuracy. No studies attempted to distinguish bvFTD patients from

---

**Table 6** QUADAS-2 Evaluation.

| Study                  | Patient selection | Index test | Reference standard | Flow and timing | Patient selection | Index test | Reference standard |
|------------------------|-------------------|------------|--------------------|-----------------|-------------------|------------|--------------------|
| Bisenius et al., 2017  | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Brod et al., 2017      | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Canu et al., 2017      | High              | Low        | Low                | Low             | Low               | Low        | Low                |
| Chow et al., 2008      | Low               | High       | Low                | Low             | Low               | Low        | Low                |
| Davatzikos et al., 2008| Low               | Unclear    | Low                | Low             | Low               | Low        | Low                |
| Du et al., 2017        | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Dukart et al., 2011    | High              | Low        | Low                | Low             | Low               | Low        | Low                |
| Frings et al., 2014    | Low               | High       | Low                | Low             | Low               | Low        | Low                |
| Kloppel et al., 2008a,b| Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Kloppel et al., 2015   | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Koikkalainen et al., 2016| Low            | Low        | Low                | Low             | Low               | Low        | Low                |
| Kuceyveski et al., 2012| Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Lehmann et al., 2010   | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Mahoney et al., 2014   | Low               | High       | Low                | Low             | Low               | Low        | Low                |
| McMillan et al., 2012  | Low               | High       | Low                | Low             | Low               | Low        | Low                |
| McMillan et al., 2014  | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Meyer et al., 2017     | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Möller et al., 2015    | High              | Low        | Low                | Low             | Low               | Low        | Low                |
| Möller et al., 2016    | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Muñoz-Ruiz et al., 2012| High              | Unclear    | Low                | Low             | Low               | Low        | Low                |
| Raamana et al., 2014   | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Tahmasian et al., 2016 | High              | Low        | Low                | Low             | Low               | Low        | Low                |
| Tong et al., 2017      | Low               | Low        | Low                | Low             | Low               | High       | Low                |
| Vemuri et al., 2011    | Low               | High       | Low                | Low             | Low               | Low        | Low                |
| Wang et al., 2016      | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Whithwell et al., 2011 | Low               | High       | Low                | Low             | Low               | Low        | Low                |
| Wilson et al., 2009    | Low               | Low        | Low                | Low             | Low               | Low        | Low                |
| Zhang et al., 2013     | High              | Low        | Low                | Low             | Low               | Low        | Low                |
those with psychiatric disorders, and these two disorders have been shown to be difficult to distinguish clinically (Woolley et al., 2011). However, it is likely that this distinction will be similar to that of control subjects as no atrophy is expected in most psychiatric disorders other than severe and persistent mental illness, such as schizophrenia with chronic psychotropic treatment, that have been linked to subtle volume loss over time (Andreasen et al., 2011).

Most studies have looked at GM atrophy. Fewer studies have used DTI measures, proving mixed results but with some studies suggesting DTI may be more sensitive in the early stages of the disease (Kuceyeski et al., 2012; Zhang et al., 2013). Most studies included in this review only looked at single MRI measures. Hypothetically a multimodal approach combining various MRI modalities such as GM structure and WM integrity should produce more accurate classification than a single modality, as these modalities should provide complimentary information about different aspects of the disease. This is supported by some studies (McMillan et al., 2014; McMillan et al., 2012) while others found no improvement when adding white matter to cortical metrics (Brom et al., 2017; Klöppel et al., 2008a). These differences are likely due to differing patient groups and methodology.

This review focuses on morphometric MRI measures as the majority of studies in this area have focused on morphometry, however a few recent studies have looked at the added benefit of arterial spin labeling MRI or functional MRI (Bron et al., 2017; Tahmasian et al., 2016). This may provide additional discriminative power and is feasible given that these are all MRI sequences that can be performed in the same session.

4.1. Comparison to visual MRI reading

Currently, FTD diagnosis is usually assisted via visual reading of MRI scans with or without semi-structured visual rating scales in clinical practice. It is therefore important that an effective MRI morphometry-based classification tool improves on current practices.

Klöppel et al. (2008a) found that radiologists with different levels of experience varied widely in their ability to distinguish pathologically defined FTD from AD on visual reading of MRI (ranges for accuracy, sensitivity, and specificity were 56.2–83.8%, 55.6–83.8%, and 57.9–90.0% respectively) and generally performed poorer than an SVM classifier of GM volume on the same cohort (Klöppel et al., 2008b). Accuracy was positively correlated with the radiologist’s level of experience. Koikkalainen et al. (2016) reported much poorer results (overall accuracy of 46.6%, with a sensitivity of 50% for FTD versus others) when using a disease state index classifier on multiple visual rating scales in the multi-class classification of dementia types compared to their morphometric results. In a mixed neuropsychiatric population, visual reading of baseline MRIs by neuroradiologists using visual rating scales reported high specificity (93%) but only moderate sensitivity (70%) in distinguishing bvFTD from non-bvFTD, using clinical diagnosis at two-year follow-up as the gold standard (Vijverberg et al., 2016).

In a cohort of pathologically defined dementia (Harper et al., 2016), unstructured visual assessment by experienced raters resulted in moderate sensitivity (82%) and high specificity (99%) in distinguishing FTD from controls, while moderate sensitivity (74%) and specificity (81%) was achieved when distinguishing FTD from AD. These results are comparable with many of the results obtained from morphometry studies. Semi-structured visual rating scales were found to provide comparatively high sensitivity and specificity in distinguishing FTD from controls (82% and 89% using the medial temporal lobe atrophy (MTA) scale, and 89% and 97% when using an SVM on the results of multiple visual rating scales). Visual rating scales resulted in moderate specificity (81% for an orbito-frontal scale, and 89% and 97% when using an SVM on the results of multiple visual rating scales) but low sensitivity (55% and 56%) when distinguishing FTD from AD.

Overall the results from visual radiologists’ review appear generally poorer than the best reported results from MRI morphometry studies, indicating the potential usefulness of automated MRI morphometry for improving diagnosis of FTD. However, it is not proven at this point if morphometry outperforms semi-structures visual rating scales (Chow et al., 2011; Harper et al., 2016). It is possible that morphometric
approaches could improve diagnostic accuracy in settings where clinicians have less experience in identifying FTD neuroradiological features. (Klöppel et al., 2008a).

4.2. Single-subject approach to structural MRI

While there has been major improvement in automated structural MRI processing pipelines over the years, there remain significant methodological challenges to its application at the single-subject level. One of the main limitations to the clinical validity of such methods is the variability with regards to different sites, scanners and repeated image acquisitions. This variability leads to inconsistencies in measurements that reduce the accuracy of diagnostic classifications based on subtle differences in atrophy or other morphometric measures (Potvin et al., 2017). While a comparison of the performance of the different currently available processing pipelines is beyond the scope of this paper, the ideal MRI processing pipeline must perform robust registration and tissue contrast normalization to achieve precise cortical and subcortical segmentation across different scanners. It should further be able to perform intra-subject registration to measure subtle brain changes over time. Being able to compare subjects to a large database of healthy controls across ages, sex and education level is also of significant benefit (Potvin et al., 2017).

4.3. Limitations

Studies included in this review are highly heterogeneous in terms of population demographics and methodology. These issues are similar to those regarding the diagnostic classification of AD (Falahati et al., 2014; Rathore et al., 2017).

Studies varied considerably on the subjects they included. Studies using small homogenous samples may result in the overfitting of data. A major issue with studies is the inclusion of well-characterized subjects that tend to be at a later disease stage and therefore may find higher accuracy because brain changes are more substantial and easier to differentiate. Ideally studies need to include patients in the earliest stages of the disease when diagnoses are ambiguous, such as the naturalistic symptom-based inclusion approach taken by the Late-Onset Frontal lobe study (Krudop et al., 2014). Many studies grouped FTD subtypes in vivo (Perry et al., 2017). Morphometry could also improve clinical stage prediction and to build models that could predict pathological subtypes in vivo (Perry et al., 2017). Morphometry could also improve practice by identifying data-driven subtypes with clinically relevant differences in symptom profile or prognosis (Ranasinghe et al., 2016). The methodology needs to be feasible for use in clinical practice; a straight-forward process that is not time consuming and is easy to interpret is needed, and it needs to be applicable across scanner types and centers. This type of method may be especially helpful for those clinicians with less experience diagnosing FTD, such as community hospitals and primary care physicians that do not have easy access to specialty FTD clinics. In addition to leading to earlier diagnosis and improved prognosis clinically, morphometric biomarkers could potentially improve patient selection and reduce required sample sizes in clinical trials (Pankov et al., 2016), which would accelerate drug discovery.

5. Conclusions

Automated morphometric MRI has potential to improve the diagnosis and prognosis of early stage FTD in clinical practice. Current evidence provides good support for its ongoing development. The inclusion of 3D-T1 MRI sequences in clinical imaging protocols would facilitate the development of these tools, and eventually the integration of these methods in practice. However, more studies that use rigorous methodology and prospectively validate findings in independent real-
