

## Article

# The Application of Intelligent Data Models for Dementia Classification

Rabah AlShboul <sup>1,\*</sup>, Fadi Thabtah <sup>2</sup>, Alexander James Walter Scott <sup>3</sup> and Yun Wang <sup>3</sup><sup>1</sup> Computer Science Department, Prince Hussein Bin Abdullah College for Information Technology, Al Al-Bayt University, Mafraq 25113, Jordan<sup>2</sup> ASD Tests, Auckland 0610, New Zealand<sup>3</sup> Manukau Institute of Technology, Auckland 0481, New Zealand

\* Correspondence: rabahshboul@aabu.edu.jo

**Abstract: Background and Objective:** Dementia is a broad term for a complex range of conditions that affect the brain, such as Alzheimer's disease (AD). Dementia affects a lot of people in the elderly community, hence there is a huge demand to better understand this condition by using cost effective and quick methods, such as neuropsychological tests, since pathological assessments are invasive and demand expensive resources. One of the promising initiatives that deals with dementia and Mild Cognitive Impairment (MCI) is the Alzheimer's Disease Neuroimaging Initiative (ADNI), which includes cognitive tests, such as Clinical Dementia Rating (CDR) scores. The aim of this research is to investigate non-invasive dementia indicators, such as cognitive features, that are typically diagnosed by clinical assessment within ADNI's data to understand their effect on dementia. **Methods:** To achieve the aim, machine learning techniques have been utilized to classify patients into Cognitively Normal (CN), MCI, or having dementia, based on the sum of CDR scores (CDR-SB) besides demographic variables. Particularly, the performance of Support Vector Machine (SVM), K-nearest neighbors (KNN), Decision Trees (C4.5), Probabilistic Naïve Bayes (NB), and Rule Induction (RIPPER) is measured with respect to different evaluation measures, including specificity, sensitivity, and harmonic mean (F-measure), among others, on a large number of cases and controls from the ADNI dataset. **Results:** The results indicate competitive performance when classifying subjects from the baseline selected variables using machine learning technology. Though we observed fairly good results across all machine learning algorithms utilized, there was still variation in the performance ability, indicating that some algorithms, such as NB and C4.5, are better suited to the task of classifying dementia status based on our baseline data. **Conclusions:** Using cognitive tests, such as CDR-SB scores, with demographic attributes to pinpoint to dementia using machine learning can be seen a less invasive approach that could be good for clinical use to aid in the diagnosis of dementia. This study gives an indication that a comprehensive assessment tool, such as CDR, may be adequate in assessing and assigning a dementia class to patients, upon their visit, in order to speed further clinical procedures.



**Citation:** AlShboul, R.; Thabtah, F.; Walter Scott, A.J.; Wang, Y. The Application of Intelligent Data Models for Dementia Classification. *Appl. Sci.* **2023**, *13*, 3612. <https://doi.org/10.3390/app13063612>

Academic Editor: Keun Ho Ryu

Received: 11 January 2023

Revised: 28 February 2023

Accepted: 9 March 2023

Published: 11 March 2023

**Keywords:** Alzheimer's disease; Clinical Dementia Rating; data analytics; neuropsychological assessment; neuroimaging; machine learning; feature selection



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Dementia can be defined as a range of cognitive impairments that include a degrading of functional skills, cognitive ability, and memory (Grassi et al., 2019 [1]; Shankle, Mani, Pazzani & Smyth, 1997 [2]). AD is the leading type of dementia with 60–80% of dementia cases being a type of AD (So, Hooshyar, Park & Lim, 2017 [3]; Alzheimer's Association, 2018 [4]). AD is a progressive disease, in which dementia symptoms typically worsen over time. The Alzheimer's Association (2018) [4] believes that being able to identify patients who are in the prodromal AD stage is critical for early management and drug development.

As such, having an efficient yet accurate way of diagnosing AD status is essential to increase the chance of early diagnosis and, therefore, treatment and planning.

AD is typically diagnosed by a combination of neuropsychological and medical assessments, such as the Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) (Rosen, Mohs & Davis, 1984) [5], the Mini Mental State Exam (MMSE) (Folstein, Folstein & McHugh, 1975) [6], and Clinical Dementia Rating (CDR) (Hughes, Berg, Danziger, Coben & Martin, 1982) [7]. Information is also derived from Magnetic Resonance Imagery (MRI) (Pantano, Caramia & Pierallini, 1999) [8], Positron Emission Tomography (PET) (Iaccarino, Sala, Caminiti & Perani, 2017) [9], genetic variables (APOE4) (Qian et al., 2017) [10], and more. As diagnosis is usually multifaceted, it may require several assessments by skilled clinicians involving much time and cost. If a neuropsychological assessment, as mentioned above, was able to give a clear dementia classification accuracy on its own, this could widen the accessibility of dementia assessments to those in poorer socio-economic areas while reducing the cost and need for assessment by skilled professionals.

In this research, we have chosen to focus on the CDR-SB as there is no extensive research on the use of intelligent techniques focusing on CDR-SB scores besides it can be considered affordable assessment. In an initial experimentation, CDR-SB showed a correlation with dementia status in the outcomes of four feature selection techniques applied to the ADNI dataset including Correlation Feature Set (CFS) (Hall, 1999) [11], ReliefF (Kira & Rendell, 1992) [12], and two Wrappers (Hall & Smith, 1999) [13]. However, this could be attributed to that CDR was used as a measure to assign the diagnostic label by clinicians during the ADNI project besides other tests.

The problem we consider in this research is a classification task, whereby we investigated and measured the performance of classification algorithms on CDR-SB scores and other demographic data to predict dementia. Our research differs from much of the existing literature, which focuses on intelligent techniques applied to MRI and other clinical data. We will focus on the use of machine learning techniques on neuropsychological data to create models to predict the clinical diagnosis for patients. Particularly, we test models derived by various classification techniques using real data and compare the predictive performance of the derived models to determine the ideal model for dementia detection. The main questions we seek to answer are:

Would a model derived from using machine learning on the ADNI dataset that was based on the CDR-SB scores and demographic variables be feasible to predict dementia?

If so, which classification technique could produce highly accurate models based on accuracy, sensitivity, and specificity rates?

The potential application of machine learning, if successful, could see an increase in detection accuracy and efficiency for clinicians. The models produced can be used to aid the decision-making process during screening or diagnosis of dementia-related conditions.

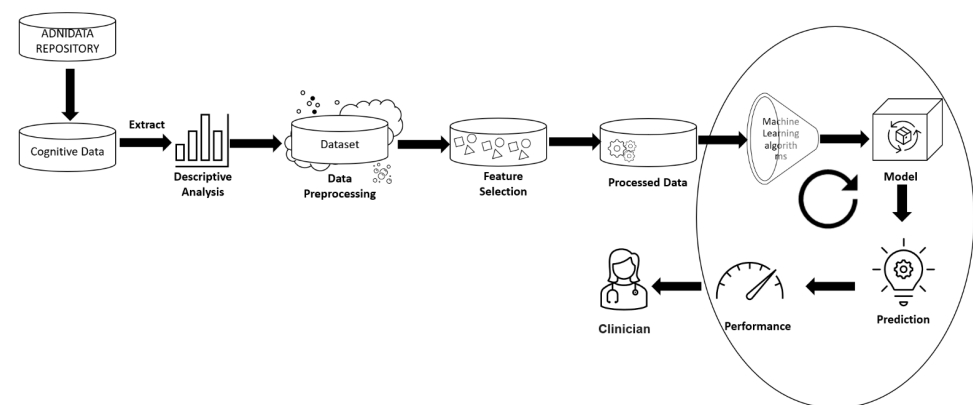
The paper sets out the methodology used in our research, followed by a literature review of research that has focused on the use of machine learning technology in predicting the dementia status of patients. The review focuses predominantly on research that has used data from the ADNI data repository. Following this review, we describe the dataset and experimental procedure used to generate our results, along with an in-depth discussion and our conclusion.

## 2. Methods

The primary variables used to derive the intelligent models in the methodology are based on the CDR method. CDR is used to rate the severity of AD in a subject, by assessing the patient's signs and symptoms in six cognitive areas. Typically, the CDR will be completed by a clinician or other specialist and follows an initial interview with the subject. During the process, the subject is scored between 0–3 across the different cognitive domains. Scores of 1, 2, and 3 indicate mild, moderate, and severe dementia, respectively, whilst 0.5 may indicate very mild dementia (Tay et al., 2015) [14]. Subsequently, the scores can be combined as the sum of each of the domains is considered with equal weightings.

This method is termed the CDR-Sum of Boxes (or CDR-SB) and assumes equal weighting among the cognitive areas (Mennella & Heering, 2015) [15].

In our study, we applied the data process described in Figure 1. The ADNI dataset was chosen as it contains both CDR scores and other neuropsychological features; additionally, its data collection was performed under a standard set of procedures to eliminate inconsistencies. Initially, a descriptive analysis was conducted to obtain more detailed statistics and information. Then, data pre-processing was performed to retain instances and variables that align within the research scope. We excluded any instance without a class (DX) and ignored any instance with missing class. We only kept the first visit data for the cases and controls.



**Figure 1.** Data process followed.

To elaborate on Figure 1, we filtered the initial ADNI dataset to retain only relevant features related to CDR and demographic variables based on the scope we have defined and our research questions. In this case, all features have been ignored other than the CDR-SB scores, age, education level, gender, ethnicity, race, exam date, and the exam site; thus, we have a dataset with 10 features (including the class, 'dx'), which is further described later. Once the dataset was prepared, we produced models using different classification algorithms. We measured the performance of our models by comparing metrics, specificity, sensitivity, precision, F1-score, and ROC area, to determine the highest performing models. Section 5 gives further details.

### 3. Results

#### 3.1. Alzheimer's Disease Neuroimaging Initiative (ADNI)

ADNI is a collection of medical centers and universities in the USA and Canada that was established to develop standardized biomarker procedures and neuroimaging techniques to better assess and diagnose subjects with AD, those with MCI, and those who are cognitively normal (CN) (Petersen et al., 2009) [16]. Commencing in 2004, under the leadership of Dr. Michael W. Weiner as a six-year study (ADNI-1) (ADNI, 2017a) [17], the primary goal was to develop and validate biomarkers to use as outcome measures in clinical trials. The purpose of these biomarkers is to aid in the (early) diagnosis and tracking of AD (ADNI, 2017a [17]; Petersen et al., 2009 [16]). The biomarkers include scores from cognitive and functional tests, demographic data, MRI, readings from PET procedures, and cerebrospinal fluid markers.

Subsequent studies (ADNI GO, ADNI-2, ADNI-3) were later initiated, where the scope was broadened to include further participants of different clinical stages (early and late mild cognitive impairment groups were introduced) (ADNI, 2017a) [17,18]. Other studies focused on an understanding of the longitudinal sequence of dementia, and investigated technologies that provide additional measures (such as the use of mass spectroscopy for the analysis of CSF markers and MRI scans to detect tau-protein tangles, included in ADNI

datasets such as Aisen, 2011 [19], Weiner et al., 2017b [20], Boutajangout, Sigurdsson & Krishnamurthy, 2011 [21].

The data used in this study was originally derived from the ADNI database (ADNI, 2017a) [17], and it is called Alzheimer’s Disease Prediction of Longitudinal Evolution (TADPOLE) dataset, which was made available in 2017 [18]. The dataset consists of 1737 unique patients with their demographic information, and the neuropsychological and medical data associated with each patient visit. Each assessment is recorded separately, supplying 12,741 dataset instances.

The dataset consists of three main subject groups (class labels): CN, MCI, and Dementia. These three groups consist of 523, 341, and 866 subjects, respectively, after pre-processing. We observed an imbalance in the class distribution with the MCI group being over twice the size of the Dementia group. The average age of a participant in each group is similar, between 73–75 years, though both the MCI ( $73.03 \pm 7.60$  years) and Dementia ( $74.93 \pm 7.81$  years) groups see a greater spread of ages in comparison to the CN group ( $74.25 \pm 5.79$  years).

Participants in the Dementia group were likely to have a lower mean education than the other groups, with the CN group having the highest education on average. Participants in the CN group were also noticeably more likely to have not be currently married at the time of the study. In total, 67.75% of CN subjects we analyzed were married, compared to 83.63% and 76.97% in the Dementia and MCI groups, respectively. CN subjects were much less likely to be carriers of the APOE 4 allele (28.60%) compared to the Dementia (65.77%) and MCI (50.23%) groups. A total of 17.27% of CN participants had a change in their diagnosis by the end of this study’s data and nearly half of the MCI group (42.01%) exhibited a change. It is important to note that for 12.88% of the MCI subjects that changed diagnosis, it was a positive change with the new diagnosis being CN.

### 3.2. Data Pre-Processing

We have excluded rows with no DX (clinical diagnosis from visit) and have kept 8904 rows. This consists of 1730 unique patients, after seven instances were removed for having no visits with a diagnosis and/or no CDR-SB data available.

As the scope of our experimentation is to predict the cognitive state of the subjects, we adjusted some of the dataset class labels that classified the level of dementia changes, as we do not observe the longitudinal progression for each participant. MCI -> Dementia and NL -> Dementia were changed to Dementia, NL -> MCI and Dementia -> MCI were changed to MCI, and MCI -> NL were changed to NL.

For our experimentations, baseline demographics, CDR-SB scores, and DX are the main features we applied, which can be seen in Table 1. We only extracted the first visit of the subjects.

**Table 1.** Features selected.

Variable	Description
SITE	ID of the site where the visit occurred
EXAMDATE	Date of the clinical examination
AGE	Age at baseline visit
PTGENDER	Sex
PTEDUCAT	Total years of education
PTETHCAT	Ethnicity
PTRACCAT	Race
PTMARRY	Marital status at baseline
CDRSB	Clinical Dementia Rating Scale Sum of Boxes
DX	Class variable

### 3.3. Experimental Settings and Evaluation Metrics

The experiments are conducted in the Waikato Environment for Knowledge Analysis (WEKA 3.8) data analytic platform. This is freely available and designed to allow easy

integration of Java code for reusability purposes within the platform. WEKA contains many supervised and unsupervised learning algorithms, dimensionality reduction, and other data pre-processing approaches and visualization methods.

The Confusion Matrix, shown in Figure 2, is adopted to evaluate the performance of the models applied in our experimentation. Performance measurements are detailed below.

1. Sensitivity =  $\frac{TP}{TP+FN}$  (1)
2. Specificity =  $\frac{TN}{TN+FP}$  (2)
3. F-measure =  $2 * \frac{Precision * Sensitivity}{Precision + Sensitivity}$  (3)
4. Precision =  $\frac{TP}{TP+FP}$  (4)
5. ROC: Receiver Operator Characteristic Curve, which is generated by plotting the TPR ( $y$ -axis) against the FPR ( $x$ -axis).

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

**Figure 2.** Confusion matrix (Narkhede, 2018) [22].

For our experiments, we are focusing on the four main performance metrics to evaluate the performance of our chosen models: sensitivity, specificity, F-measure, and ROC.

### 3.4. Classification Methods and Implementation

Several machine learning techniques were explored in this study based on 10-fold cross-validation (Mosteller & Tukey, 1968) [23]. Naïve Bayes (NB) is probabilistic algorithm that is often used in classification tasks (Islam, Wu, Ahmadi & Sid-Ahmed, 2007) [24]. It uses Bayes's theorem to assign a class label to an instance given the probabilities of the other features in the sample (Zhang, 2004) [25]. The NB classifier has been shown to be successful in practice, often utilized in classification alongside other more complex and sophisticated techniques (Rish, 2001) [26].

Repeated Incremental Pruning to Produce Error Reduction (RIPPER) is a rule-based classifier, which was proposed as an optimized version of IREP (Cohen, 1995) [27]. Classes are examined in increasing size and there is an initial set of rules that is generated for the class using incremental reduced error. RIPPER then proceeds by treating the examples in the training data as a class and finding a set of rules that will cover all members of the respective class. Once rules are derived, then RIPPER eliminates rules that have high misclassifications or cover a minimal number of instances in the dataset.

SVMs are a set of supervised learning methods used for classification as well as regression and outlier detection. SVMs can perform binary and multi-class classification tasks. On the other hand, K-nearest neighbors (KNN) is a type of instance learning algorithm. This classification classifies test data by considering the k-nearest neighbor's majority vote within the nearest neighbors of a data point (Aha, Kibler & Albert, 1991) [28]. We used  $k = 5$  [29,30]; when using  $k = 5$ , then more neighboring data instances' class labels will be utilized to assign the class label of the test data during the process of cross validation. This indeed reduces any sort of bias for the class assignment process since more neighbors are used, hence ensuring a collective decision-making.

Lastly, C4.5 (J48 in WEKA) is a Decision Tree algorithm that produces tree-based classifiers (Quinlan, 1993) [31]. This algorithm is created to predict test data by learning decision rules through the features present in the dataset and finding rules that can apply to all points in a subset. It is capable of processing both binary and multi-class classification datasets. For all algorithms, we used the default hyper parameters.



In our experimentation, we measured the performance of five machine learning algorithms to predict the dementia class of subjects from the ADNI TADPOLE challenge dataset (ADNI, 2017b) [18]. The algorithms were used on a subset of this data, which included demographic and CDR-SB variables to predict the class of each patient.

The performance metrics produced for each algorithm are shown in Table 2. NB, C4.5, and RIPPER resulted in the highest performance, as seen across all measures, with all results between 88.4–95.5%. C4.5 arguably performed the best, with the highest result across all metrics excluding ROC. The highest classification accuracy was by the C4.5 algorithm, with 89.7%. This result was 15% higher than SVM (linear) accuracy. The NB and RIPPER algorithms showed similarly competitive results within  $\pm 2\%$  when compared to C4.5 across all metrics as shown in Table 2.

**Table 2.** Experiment results.

Model	Accuracy	Sensitivity	Specificity	F-Measure
SVM-Liblinear	74.7%	74.7%	82.3%	74.8%
KNN k = 5	76.8%	76.8%	83.9%	76.7%
Naïve Bayes	88.4%	88.4%	91.7%	88.4%
C4.5-Decision Tree	89.7%	89.7%	93.2%	89.8%
RIPPER	89.0%	89.0%	92.3%	89.1%

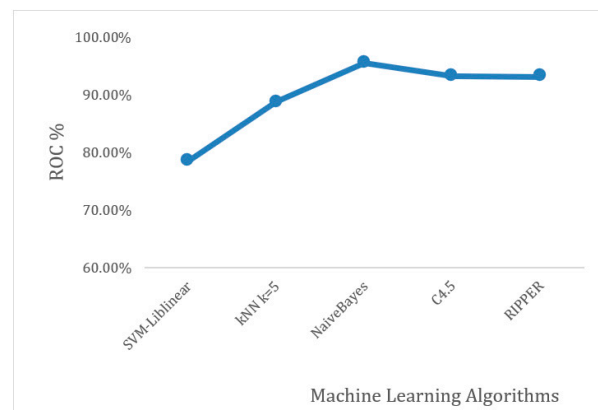
Sensitivity measures the proportion of positive results predicted correctly (Equation (1)). Sensitivity is negatively impacted when results that should be positive are predicted as negative, or a False Negative (FN). Our results indicate that C4.5 was able to correctly identify these results with the smallest proportion of FNs (0/341 dementia subjects were classified as CN and 43/866 MCI patients were classified as CN), with a sensitivity score of 89.7%. RIPPER and NB showed similar performance in regard to sensitivity, albeit a 0.7% and 1.3% lower sensitivity rate than C4.5, respectively. SVM (linear) and KNN derived lower sensitivities of 74.7% (1/341 Dementia and 61/866 MCI classed as CN) and 76.8% (7/341 Dementia and 145/866 MCI classed as CN), respectively.

Specificity measures the models' ability to correctly identify negative results. As such, it is impacted by False Positive (FP) results, as annotated in Equation (2), above. The considered machine learning algorithms demonstrated somewhat of an increase in this metric, compared to sensitivity. In the practical application of dementia classification, this high specificity indicates the strong ability of the algorithm to predict healthy or cognitive control individuals as being healthy. Though this is important, sensitivity may arguably be more important in our scenario as a FP can be corrected by subsequent testing, whereas a FN may result in no further testing and, therefore, the patient missing a diagnosis.

The C4.5 algorithm was able to correctly identify negative results, with a specificity of 93.2% (43/523 CN classified as MCI and 79/866 MCI classified as Dementia). NB and RIPPER performed well within 2% of this result. SVM (linear) and KNN performed the poorest, though markedly higher than the sensitivity results obtained, with results of 82.3% and 83.9%, respectively. These scores are a result of more FP predictions; SVM gave a total of 133/523 FPs on CN subjects, seven of which were FPs as dementia, and 121/866 MCI subjects classed as dementia. KNN had results of 102/523 FPs for CN subjects, three of which were classified as dementia, the other as MCI, and also had 46/866 MCI subjects classed as dementia.

The F-measure is the harmonic mean between sensitivity and precision, with the equation shown above. Precision (Equation (4)) is calculated as the number of TPs over all predicted positive results (true and false) and differs from sensitivity, which includes FN results. All models exhibit a similar F-measure score compared to model accuracy, with  $\pm 0.1\%$  across all five algorithms compared to their accuracy. As the F-measure is the harmonic mean between these two metrics, the algorithms that have better results in sensitivity and precision (NB, C4.5 and RIPPER) have F1-scores of over 10% higher than SVM linear and KNN.

The Receiver Operator Characteristic (ROC) curve is a plot of the TP rate against the FP rate (Fawcett, 2006) [32]. The area under this curve (ROC-AUC) gives a good measure of how well a model is at correctly predicting TP and TN results. Figure 3 shows the values of ROC based on the different algorithms used in the experiment. Though C4.5 saw the highest sensitivity, specificity, F1 score, and classification accuracy, NB resulted in a stronger ROC measurement of 95.5%, 2.3% higher than C4.5. The competitive results seen in this metric are promising as the ROC AUC addresses the capability of the models to correctly classify both TN and TP results and, therefore, minimize FN and FPs.



**Figure 3.** ROC area results.

It is imperative to minimize type II errors (FN) as much as possible in medical diagnosis applications, such as dementia detection. A FN classification, in a real-life situation, could see a patient be incorrectly diagnosed as not being cognitively impaired (or to a lesser degree, for example MCI versus a more advanced dementia). This can be incredibly detrimental to the patient's health as early diagnosis and disease management is critical for speeding up intervention due to MCI or AD. As expected, only a small number of misclassifications of CN (0 or 1 across all algorithms excluding KNN which had 7/341) when the patient was actually classified as AD. However, there is a larger proportion of CN misclassifications of MCI patients (for example 145/866 in KNN), as the symptoms (and therefore CDR-SB score) is much nearer to the CN subjects. Similarly, there is a minimal number of CN patients being classified as AD, but a more significant number of CN patients being classified as having MCI.

#### 4. Discussion

Sabuncu and Konukoglu (2014) [33] employed ADNI data, neuroimaging software, and machine learning algorithms to conduct an empirical study on multivariate pattern analysis (MVPA). Prior MVPA studies had focused on extracting imaging measurements or on improving prediction accuracy through statistical analysis algorithms (Batmanghelich, Taskar & Davatzikos, 2009 [34]; Cho, Seong, Jeong & Shin, 2012 [35]; Teipel et al., 2007) [36].

Sabuncu and Konukoglu (2014) [33] used MRI and neuropsychological features with three MVPA algorithms: Support Vector Machine (SVM), Relevance Voxel Machine (RVoxM), and Neighborhood Approximate Forest (NAF) (Cortes & Vapnik, 1995 [37]; Konukoglu, Glocker, Zikic & Criminisi, 2013 [38]; Sabuncu & Van Leemput, 2012 [39]). The results indicated that a stronger predictive accuracy can be observed when using an integration of MRI variables, rather than on one MRI result, which is typically univariate and gives information on one anatomical area of the brain than when analyzing them individually. This corresponds with other earlier research works that indicate some neurocognitive disorders are related to larger scale networks across multiple brain regions (Seeley et al., 2009) [40]. The authors conclude that there is no one universal method for dealing with neuroimaging data but rather five main limiting factors that influence prediction accuracy: sample size, data quality, standardized imaging measurements, standardized

prediction algorithms, and the underlying biological footprint of the disease. In addition, the study also showed that other variables, such as demographic and neuropsychological, can be predicted using MRI data.

Four supervised learning methods were tested by Aguilar et al. (2013) [41] to classify AD patients and controls, as well as to predict AD conversion from MCI. The authors used data from the AddNeuroMed cohort, which though a different initiative to ADNI, the data acquisition process was designed to be compatible with ADNI (Simmons et al., 2009, 2010) [42,43]. Orthogonal partial least squares to latent structures (OPLS), Decision Trees, artificial neural networks (ANN), and support vector machines (SVM) have been applied on MRI features, MRI + age, MRI + education, and MRI + APOE to generate models. The authors also used the feature selection algorithm method named Relief (Kira & Rendell, 1992) [12] to select the top ten features (nine areas of the brain area and cerebrospinal fluid (CSF) volumes). The models derived by all considered classification algorithms showed high specificity and sensitivity when tasked with distinguishing between AD and CN, with results between 82–85%. Similarly, 80–85% of MCI-converters were able to be identified. However, performance was much poorer when predicting MCI-non-converters, where roughly half or more (49–68.4%) were diagnosed as cognitively normal. The results indicated no significant statistical difference in the performance when all the MRI variables were used, versus the ranked features only. There was a noticeable decline in performance when age or education was included in the Decision Tree algorithm. Other than that, there was no significant influence for better or worse on the performance when age, education, or APOE were included with the MRI data. The highest accuracy and AUC scores were observed when MRI + APOE was used with ANN, and MRI + Edu with SVM and OPLS, with scores of 91–92%. Models with approximately 86% accuracy were derived when OPLS was used on an ADNI study using the same features (Westman et al., 2011) [44].

Spasov, Passamonti, Duggento, Liò, and Toschi, (2019) [45] utilized deep learning to combine MRI, demographic, neuropsychological, and APOE data focusing on patients who were likely to progress from MCI to AD within three years. The authors developed a feature extractor sub-network, which performs dual tasks—AD to CN classification, as well as a MCI to AD conversion prediction. Furthermore, since the feature representations are multi-layered, rich and complex data, such as MRI, can be handled well in the classification process (Spasov et al., 2019) [45]. The authors used four input combinations of biomarkers: clinical data + MRI, clinical + Jacobian Determinant data, clinical + atlas masked MRI, and clinical + Jacobian Determinant + MRI. The best classification performance is seen where MRI + clinical data are used in conjunction, with a median AUC of 92.50% and an accuracy of 86%. A decrease is observed when brain areas that are not classically associated with AD are removed, down to 92.20%. Further decreases in performance are observed when all variables are used, and when the MRI data only is used (no clinical data), a much lower AUC and accuracy, of 70% and 72%, respectively, is seen.

Stonnington et al. (2010) [46] studied the MCI—AD conversion, trying to predict a continuous measure, such as scores from the MMSE, Dementia Rating Scale (DRS), ADAS-Cog and RAVLT, from MRI data. Relevance Vector Regression (RVR) was applied for prediction using MRI data and measures from four cognitive tests, namely MMSE, DRS, ADAS-Cog, and RAVLT. The results of their research indicated a strong linear relationship between DRS, MMSE, and ADAS-Cog scores with grey matter (GM) segments of MRI data, but did not indicate this with the RAVLT assessment. DRS, followed by MMSE in dataset 1, and ADAS-Cog, followed by MMSE in datasets 2 and 3, provided the best predictive power when processed. The author suggested that images of the whole brain provided a better correlation with these cognitive tests, as they test multiple domains within the brain, whereas RAVLT targets memory, which is associated with the medial temporal lobe. The predictive accuracy of the models worsened when large groups of a class, for example CN or MCI, were either missing or removed. Further, the years-of-education feature was also shown to be significantly correlated to MMSE and ADAS-Cog from the ADNI dataset.



The authors demonstrated that RVR can be a useful multivariate method for investigating the relationship between MRI features and clinical scores. Their results support the use of MMSE, ADAS-Cog, and DRS in dementia applications.

Izquierdo et al. (2017) [47] experimented on cognitive test scores using MRI, PET, and previous cognitive information from an ADNI cohort. The authors showed that being able to accurately predict cognitive test scores would be useful in the diagnosis of AD as well as for assisting clinicians to manage patients. The aim of this study was to predict the scores of MMSE, CDRS, RAVLT, ADAS11, and ADAS13 tests. To do this, stochastic gradient boosting of Decision Trees was applied to a sample of over 1141 patients from the ADNI data. Specifically, the scores at the 24-month mark are predicted based on information at prior visits (6 months, 12 months, 18 months). The results of the experimentations indicated high correlation ( $\geq 0.9$ ) across all correlation measures. The authors compared the results from the gradient-boosted method against other algorithms such as Multilayer Perceptron, 10-KNN, Decision Trees, Bagging, and SVM. Gradient boosting outperformed other methods of predicting the cognitive test scores.

Miller et al. (2020) [48] used 741 ADNI-participants with blood microarray data to measure the cognitive decline related to AD in terms of CDR score. The authors used the CDR scores recorded in the last clinical assessment to categorize the data into three groups, and then applied machine learning algorithms to forecast the cognitive level of the individual using the reported blood microarray data. The results showed that one chloride intracellular channel 1 (CLIC1) probe was significant; the predictive rate achieved using the machine learning algorithm reached 87% when considering nonsignificant probes.

Li et al. (2021) [49] investigated the possibility of shortening the CRD clinical assessment method to develop an electronic CRD (eCDR) to improve accessibility, reduce evaluation time, and automate the scoring process. The authors utilized the item response theory (IRT) to assess the items of the CDR, and develop an IRT automatic CDR scoring model. The results demonstrated that the IRT model with a short CDR version is able to achieve medically accepted classification rate.

Thabtah et al. (2022a; 2022b; 2022c) [50–52] investigated elements related to cognitive and functional activities of dementia using real data subjects of the ADNI project. The authors primary objective is to identify functional and cognitive indicators that correlate with one type of dementia progression, which is AD. In Thabtah et al. (2022a), a number of assessment methods were compared and analyzed using real data to find out the cognitive domains that overlap in certain cognitive items using the DSM-5 criteria framework related to dementia. Then, the authors improved their initial approach by developing a computational intelligence process to detect functional indicators of AD advancement. The computational data process was evaluated using classification and feature selection methods on ADNI's data subjects by focusing on the elements that belong to the Functional Activity Questionnaire assessment (FAQ) method. Results showed that there are some functional elements that associate with the disease advancement.

Lastly, Thabtah et al. (2022c) [50] evaluated cognitive elements related to ADAS-13 cognitive assessment method (ADAS-Cog) using machine learning (classification algorithms). Results against the cognitive data of ADNI (ADAS-Cog sheet) revealed that the classification models derived can be used for the progression of AD screening, and outperform models derived from functional elements (ADNI-FAQ data sheet).

## 5. Conclusions

Neuropsychological assessment tools, such as the CDR, are commonplace in the diagnosis and assessment of neurocognitive disorders, including dementia. Compared to the clinical procedures (such as MRI and PET) that are often used in conjunction with these assessments, the neuropsychological assessments are relatively much less invasive, for instance, requiring a questionnaire or observational-based assessment. Being able to accurately predict dementia status in a patient based only on these types of tests (and

known demographic data, which can likely also be easily obtained) could see great clinical use and application to aid in the early diagnosis of dementia.

To investigate the issue of having a potentially accurate, quick, and non-invasive method for dementia classification, our research and experimentation focused on investigating the ability of different classification algorithms to perform the task as outlined above. This was done by applying different machine learning algorithms including NB, C4.5, RIPPER, SVM, and KNN, to the CDR-SB clinical scores and demographic data from an ADNI dataset.

Our results indicate competitive performance when classifying subjects into CN, MCI, or Dementia, using baseline demographic data and CDR-SB scores. Though we observed fairly strong results across all algorithms we utilized, there was still variation in the performance ability, indicating some algorithms are better suited to the task of classifying dementia status based on our baseline data.

NB, C4.5, and RIPPER exhibited sensitivities and specificities between 88.4–89.7% and 91.7–93.2%, respectively, being able to correctly assign CN, MCI, or AD to the majority of subjects. Though shown to be a strong contender as a classifier in many related literary works, SVM in our experimentation did not perform as well as expected. We did, however, utilize a linear SVM algorithm, which is a likely cause of this notable difference. Similarly, KNN gave results of between 10–20% less than the higher performing algorithms.

A limitation of this study was that the algorithms were only applied to one subset of ADNI data, which exhibited some class imbalance. This may have negatively affected the KNN and other algorithms. Furthermore, the ADNI dataset we have utilized (TADPOLE challenge) does not represent a wide spread of demographic variables, with many of the participants being Caucasian individuals with, on average, a higher level of education. Another possible limitation is that the CDR-SB assessment was used by clinicians with other neuropsychological tests, and the clinicians' experience to assign the class label for many of the participants of the ADNI study during each medical visit re-assessment. Thereby, the decision to whether an individual is demented or not could be affected by some of the CDR-SB's elements.

This study shows that CDR may be adequate in assessing and assigning a dementia class to patients on their early visit to speed up clinical procedures. Other medical procedures (such as MRI, PET scans) would likely confirm the diagnostic result obtained by neuropsychological tests like CDR. However, the relative expertise required to conduct a CDR assessment and the resources required to do so could see this as a reliable method for diagnosis by clinicians. Further analysis and comparisons would need to be conducted to validate and corroborate these results.

In the future, we intend to broaden the scope of this research by utilizing other neurocognitive assessment data alone and in conjunction with CDR, to see whether it can be incorporated to confer classification improvements. We did not use medical data that is readily available in ADNI in our experimentation, such as MRI and PET. The inclusion of this alters the practicality of a non-invasive, quick, and efficient method to class dementia patients in a clinical setting; however, may confer increased sensitivity and specificity in classifying dementia.

**Author Contributions:** Conceptualization, F.T.; Methodology, F.T., R.A., Y.W.; Software, Y.W.; Validation, R.A.; Formal analysis, R.A., F.T. and Y.W.; Investigation, R.A., F.T. and A.J.W.S.; Data curation, R.A., A.J.W.S. and Y.W.; Writing—review & editing, R.A., F.T. and A.J.W.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** No primary data was collected. The data used in this study was retrieved from the Alzheimer's Disease Prediction of Longitudinal Evolution (TADPOLE) dataset [18], which was made available by [17]. Access to the ADNI related datasets is contingent on adherence to the ADNI Data Use Agreement.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Grassi, M.; Rouleaux, N.; Caldirola, D.; Loewenstein, D.; Schruers, K.; Perna, G.; Dumontier, M.; Alzheimer's Disease Neuroimaging Initiative. A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion From Mild Cognitive Impairment to Alzheimer's Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures. *Front. Neurol.* **2019**, *10*, 756. [CrossRef]
- Shankle, W.R.; Mani, S.; Pazzani, M.; Smyth, P. Detecting very early stages of dementia from normal aging with Machine Learning methods. *Lect. Notes Comput. Sci.* **1997**, *1211*, 71–85. [CrossRef]
- So, A.; Hooshyar, D.; Park, K.W.; Lim, H.S. Early Diagnosis of Dementia from Clinical Data by Machine Learning Techniques. *Appl. Sci.* **2017**, *7*, 651. [CrossRef]
- Alzheimer's Association. 2018 Alzheimer's disease facts and figures. *Alzheimer. Dement.* **2018**, *14*, 367–429. [CrossRef]
- Rosen, W.G.; Mohs, R.C.; Davis, K.L. A new rating scale for Alzheimer's disease. *Am. J. Psychiatry* **1984**, *141*, 1356–1364. [CrossRef] [PubMed]
- Folstein, M.F.; Folstein, S.E.; McHugh, P.R. "Mini-Mental State". A Practical Method for Grading the Cognitive State of Patients for the Clinician. *J. Psychiatr. Res.* **1975**, *12*, 189–198. [CrossRef] [PubMed]
- Hughes, C.P.; Berg, L.; Danziger, W.L.; Coben, L.A.; Martin, R.L. A New Clinical Scale for the Staging of Dementia. *Br. J. Psychiatry* **1982**, *140*, 566–572. [CrossRef]
- Pantano, P.; Caramia, F.; Pierallini, A. The role of MRI in dementia. *Neurol. Sci.* **1999**, *20*, S250–S253. [CrossRef] [PubMed]
- Iaccarino, L.; Sala, A.; Caminiti, S.P.; Perani, D.F.L. The emerging role of PET imaging in dementia. *F1000Research* **2017**, *6*, 1830. [CrossRef]
- Qian, J.; Wolters, F.J.; Beiser, A.; Haan, M.; Ikram, M.A.; Karlawish, J.; Langbaum, J.B.; Neuhaus, J.M.; Reiman, E.M.; Roberts, J.S.; et al. APOE-related risk of mild cognitive impairment and dementia for prevention trials: An analysis of four cohorts. *PLoS Med.* **2017**, *14*, e1002254. [CrossRef]
- Hall, M.A. Correlation-Based Feature Selection for Machine Learning. 1999. Available online: <https://www.cs.waikato.ac.nz/~mhall/thesis.pdf> (accessed on 10 April 2020).
- Kira, K.; Rendell, L.A. The feature selection problem: Traditional methods and a new algorithm. In Proceedings of the 10th National Conference on Artificial Intelligence, San Jose, CA, USA, 12–16 July 1992; pp. 129–134.
- Hall, M.A.; Smith, L.A. Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. 1999. Available online: <http://new.aaai.org/Papers/FLAIRS/1999/FLAIRS99-042.pdf> (accessed on 2 November 2020).
- Tay, L.; Lim, W.S.; Chan, M.; Ali, N.; Mahanum, S.; Chew, P.; Lim, J.; Chong, M.S. New DSM-V Neurocognitive Disorders Criteria and Their Impact on Diagnostic Classifications of Mild Cognitive Impairment and Dementia in a Memory Clinic Setting. *Am. J. Geriatr. Psychiatry* **2015**, *23*, 768–779. [CrossRef] [PubMed]
- Mennella, H.; Heering, H. Dementia Assessment: Using the Clinical Dementia Rating Scale. 2015. Available online: [https://www.ebscohost.com/assets-sample-content/Dementia\\_Assessment\\_-\\_Using\\_the\\_Clinical\\_Dementia\\_Rating\\_Scale.pdf](https://www.ebscohost.com/assets-sample-content/Dementia_Assessment_-_Using_the_Clinical_Dementia_Rating_Scale.pdf) (accessed on 9 March 2020).
- Petersen, R.C.; Aisen, P.S.; Beckett, L.A.; Donohue, M.C.; Gamst, A.C.; Harvey, D.J.; Jack, C.R., Jr.; Jagust, W.J.; Shaw, L.M.; Toga, A.W.; et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology* **2010**, *74*, 201–209. [CrossRef] [PubMed]
- Alzheimer's Disease Neuroimaging Initiative. 2017. Available online: <http://adni.loni.usc.edu/about/> (accessed on 16 December 2021).
- Alzheimer's Disease Neuroimaging Initiative. TADPOLE Dataset 2017. Available online: <http://adni.loni.usc.edu/tadpole-challenge-dataset-available/> (accessed on 22 March 2020).
- Aisen, P.S. Q&A: The Alzheimer's disease neuroimaging initiative. *BMC Med.* **2011**, *9*, 101. [CrossRef]
- Weiner, M.W.; Veitch, D.P.; Aisen, P.S.; Beckett, L.A.; Cairns, N.J.; Green, R.C.; Harvey, D.; Jack, C.R., Jr.; Jagust, W.; Morris, J.C.; et al. The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimer. Dement.* **2017**, *13*, 561–571. [CrossRef]
- Boutajangout, A.; Sigurdsson, E.M.; Krishnamurthy, P.K. Tau as a Therapeutic Target for Alzheimer's Disease. *Curr. Alzheimer Res.* **2011**, *8*, 666–677. [CrossRef] [PubMed]
- Narkhede, S. Understanding Confusion Matrix. 2018. Available online: <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62> (accessed on 23 May 2022).
- Mosteller, F.; Tukey, J.W. Data analysis, including statistics. *Handb. Soc. Psychol.* **1968**, *2*, 80–203.
- Islam, M.J.; Wu, Q.M.J.; Ahmadi, M.; SidAhmed, M.A. Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers. *J. Conver. Inf. Technol.* **2010**, *5*, 133–137. [CrossRef]
- Zhang, H. The optimality of naive Bayes. *Fac. Comput. Sci.* **2004**, *1*, 3.
- Rish, I. An empirical study of the naive Bayes classifier. In Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, Seattle, WA, USA, 4–10 August 2001; Volume 3, pp. 41–46.
- Cohen, W.W. Fast effective rule induction. In *Machine Learning Proceedings*; Morgan Kaufmann: San Francisco, CA, USA, 1995; pp. 115–123.
- Aha, D.W.; Kibler, D.; Albert, M.K. Instance-based learning algorithms. *Mach. Learn.* **1991**, *6*, 37–66. [CrossRef]
- Abdelhamid, N.; Padmavathy, A.; Peebles, D.; Thabtah, F.; Goulder-Horobin, D. Data Imbalance in Autism Pre-Diagnosis Classification Systems: An Experimental Study. *J. Inf. Knowl. Manag.* **2020**, *19*, 2040014. [CrossRef]

30. Shen, X.; Wang, G.; Kwan, R.Y.-C.; Choi, K.-S. Using Dual Neural Network Architecture to Detect the Risk of Dementia With Community Health Data: Algorithm Development and Validation Study. *JMIR Public Health Surveill.* **2020**, *8*, e19870. [[CrossRef](#)] [[PubMed](#)]
31. Quinlan, J.R. *C4. 5: Programs for Machine Learning*; Morgan Kaufmann: San Francisco, CA, USA, 1993; pp. 155–164.
32. Fawcett, T. An Introduction to ROC analysis. *Pattern Recogn. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
33. Sabuncu, M.R.; Konukoglu, E.; Alzheimer's Disease Neuroimaging Initiative. Clinical Prediction from Structural Brain MRI Scans: A Large-Scale Empirical Study. *Neuroinformatics* **2014**, *13*, 31–46. [[CrossRef](#)]
34. Batmanghelich, N.; Taskar, B.; Davatzikos, C. A General and Unifying Framework for Feature Construction. In *Information Processing in Medical Imaging: 21st International Conference*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 423–434. [[CrossRef](#)]
35. Cho, Y.; Seong, J.-K.; Jeong, Y.; Shin, S.Y. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *Neuroimage* **2011**, *59*, 2217–2230. [[CrossRef](#)] [[PubMed](#)]
36. Teipel, S.J.; Born, C.; Ewers, M.; Bokde, A.L.; Reiser, M.F.; Möller, H.-J.; Hampel, H. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *Neuroimage* **2007**, *38*, 13–24. [[CrossRef](#)] [[PubMed](#)]
37. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
38. Konukoglu, E.; Glocker, B.; Zikic, D.; Criminisi, A. Neighbourhood approximation using randomized forests. *Med. Imag. Anal.* **2013**, *17*, 790–804. [[CrossRef](#)] [[PubMed](#)]
39. Sabuncu, M.R.; Van Leemput, K. The Relevance Voxel Machine (RVoxM): A Self-Tuning Bayesian Model for Informative Image-Based Prediction. *IEEE Trans. Med. Imag.* **2012**, *31*, 2290–2306. [[CrossRef](#)]
40. Seeley, W.W.; Crawford, R.K.; Zhou, J.; Miller, B.L.; Greicius, M.D. Neurodegenerative Diseases Target Large-Scale Human Brain Networks. *Neuron* **2009**, *62*, 42–52. [[CrossRef](#)]
41. Aguilar, C.; Westman, E.; Muehlboeck, J.-S.; Mecocci, P.; Vellas, B.; Tsolaki, M.; Kłoszewska, I.; Soininen, H.; Lovestone, S.; Spenger, C.; et al. Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. *Psychiatry Res.* **2013**, *212*, 89–98. [[CrossRef](#)]
42. Simmons, A.; Westman, E.; Muehlboeck, S.; Mecocci, P.; Vellas, B.; Tsolaki, M.; Kłoszewska, I.; Wahlund, L.-O.; Soininen, H.; Lovestone, S.; et al. MRI Measures of Alzheimer's Disease and the AddNeuroMed Study. *Ann. N. Y. Acad. Sci.* **2009**, *1180*, 47–55. [[CrossRef](#)]
43. Simmons, A.; Westman, E.; Muehlboeck, S.; Mecocci, P.; Vellas, B.; Tsolaki, M.; Kłoszewska, I.; Wahlund, L.-O.; Soininen, H.; Lovestone, S.; et al. The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease: Experience from the first 24 months. *Int. J. Geriatr. Psychiatry* **2010**, *26*, 75–82. [[CrossRef](#)] [[PubMed](#)]
44. Westman, E.; Simmons, A.; Muehlboeck, J.-S.; Mecocci, P.; Vellas, B.; Tsolaki, M.; Kłoszewska, I.; Soininen, H.; Weiner, M.W.; Lovestone, S.; et al. AddNeuroMed and ADNI: Similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. *Neuroimage* **2011**, *58*, 818–828. [[CrossRef](#)] [[PubMed](#)]
45. Spasov, S.; Passamonti, L.; Duggento, A.; Liò, P.; Toschi, N. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *Neuroimage* **2019**, *189*, 276–287. [[CrossRef](#)] [[PubMed](#)]
46. Stonnington, C.M.; Chu, C.; Klöppel, S.; Jack, C.R.; Ashburner, J.; Frackowiak, R.S. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *Neuroimage* **2010**, *51*, 1405–1413. [[CrossRef](#)] [[PubMed](#)]
47. Izquierdo, W.; Martin, H.; Cabrerizo, M.; Barker, W.W.; Loewenstein, D.A.; Duara, R.; Adjouadi, M.; Initiative, A.D.N. [P2–422]: Predicting cognitive test scores in alzheimer's patients using multimodal longitudinal data. *Alzheimer. Dement.* **2017**, *13*, 796–797. [[CrossRef](#)]
48. Miller, J.; Kauwe, J. Predicting Clinical Dementia Rating Using Blood RNA Levels. *Genes* **2020**, *11*, 706. [[CrossRef](#)]
49. Li, Y.; Xiong, C.; Aschenbrenner, A.J.; Chang, C.; Weiner, M.W.; Nosheny, R.L.; Mungas, D.; Bateman, R.J.; Hassenstab, J.; Moulder, K.L.; et al. Item response theory analysis of the Clinical Dementia Rating. *Alzheimer. Dement.* **2020**, *17*, 534–542. [[CrossRef](#)] [[PubMed](#)]
50. Thabtah, F.; Ong, S.; Peebles, D. Detection of dementia progression from functional activities data using machine learning techniques. *Intell. Decis. Technol.* **2022**, *16*, 615–630. [[CrossRef](#)]
51. Thabtah, F.; Spencer, R.; Peebles, D. Common dementia screening procedures: DSM-5 fulfilment and mapping to cognitive domains. *Int. J. Behav. Health. Res.* **2022**, *8*, 104–120. [[CrossRef](#)]
52. Thabtah, F.; Ong, S.; Peebles, D. Examining Cognitive Factors for Alzheimer's Disease Progression Using Computational Intelligence. *Healthcare* **2022**, *10*, 2045. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.