


Article

Assessment for Alzheimer's Disease Advancement Using Classification Models with Rules

Fadi Thabtah¹ and David Peebles^{2,*} ¹ ASDTests, Auckland 0610, New Zealand; fadi@asdtests.com² Department of Psychology, University of Huddersfield, Huddersfield HD1 3DH, UK

* Correspondence: d.peebles@hud.ac.uk

Abstract: Pre-diagnosis of common dementia conditions such as Alzheimer's disease (AD) in the initial stages is crucial to help in early intervention, treatment plan design, disease management, and for providing quicker healthcare access. Current assessments are often stressful, invasive, and unavailable in most countries worldwide. In addition, many cognitive assessments are time-consuming and rarely cover all cognitive domains involved in dementia diagnosis. Therefore, the design and implementation of an intelligent method for dementia signs of progression from a few cognitive items in a manner that is accessible, easy, affordable, quick to perform, and does not require special and expensive resources is desirable. This paper investigates the issue of dementia progression by proposing a new classification algorithm called Alzheimer's Disease Class Rules (AD-CR). The AD-CR algorithm learns models from the distinctive feature subsets that contain rules with low overlapping among their cognitive items yet are easily interpreted by clinicians during clinical assessment. An empirical evaluation of the Disease Neuroimaging Initiative data repository (ADNI) datasets shows that the AD-CR algorithm offers good performance (accuracy, sensitivity, etc.) when compared with other machine learning algorithms. The AD-CR algorithm was superior in comparison to the other algorithms overall since it reached a performance above 92%, 92.38% accuracy, 91.30% sensitivity, and 93.50% specificity when processing data subsets with cognitive and demographic attributes.

Keywords: Alzheimer's disease (AD); classification; dementia; machine learning; neuropsychological assessments

**Citation:** Thabtah, F.; Peebles, D.Assessment for Alzheimer's Disease Advancement Using Classification Models with Rules. *Appl. Sci.* **2023**, *13*, 12152. <https://doi.org/10.3390/app132212152>

Academic Editor: Alexander N. Pisarchik

Received: 23 September 2023

Revised: 25 October 2023

Accepted: 1 November 2023

Published: 8 November 2023



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 is a neurodegenerative condition mostly occurring in the elderly; it is characterised by difficulties in memory, orientation, and communication, psychological changes, and impairments in activities of daily living [1]. There are over 50 million people worldwide living with dementia, and this condition significantly impacts society and the economy [2]. In the United Kingdom in 2019, there were an estimated 885,000 people with dementia, and this is projected to surge by 80% by 2040 [3]. With this growing prevalence of dementia and the associated costs, research initiatives have become fundamental [4].

While pathological assessments of Alzheimer's disease (AD) diagnosis, such as using biological markers (biomarkers), and magnetic resonance imaging (MRI) [5] can be used to predict the disease, they are also time- and cost-intensive, stressful, and provide results requiring laboratory study and professional personnel who may not be available [6–8]. Therefore, cognitive assessments used for prodromal dementia such as the Alzheimer's Disease Assessment Scale-Cognitive 13 (ADAS-Cog-13), Mini Mental State Examination (MMSE), and others [9,10] are useful since they can screen for signs of early impairment. These assessments are easier to carry out than pathological methods and show acceptable performance with reference to validity, sensitivity, and specificity [11]. However, identifying a few cognitive items that can be signs of the progression of the disease and can assist

in early intervention is still a challenge [12]. Additionally, very few research studies have measured the progression of dementia using cognitive features with a data drive methodology, i.e., [6,8,13–15].

While Battista et al. [8] used machine learning to assess cognitive measures, their study focused on diagnosing AD and not the progression of the disease, which is more challenging, and it did not consider the criteria defined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) framework [16]. Moreover, Thabtah et al. [6] investigated functional elements that may have an influence on dementia advancement using real data from the ADNI data project, and no cognitive elements were considered in the evaluation. In addition, AISHboul et al. [15] studied real data related to the Clinical Dementia Rating-Sum of Boxes (CDR-SB) assessment method in the ADNI project to reveal dementia indicators and understand their impact on dementia diagnosis. The authors have not considered symptoms of dementia progression. Mapping between the assessed cognitive items and DSM-5's criteria would have been helpful for the clinicians' understanding of linking the machine learning results and the actual diagnosis of the disease.

To fill the gap, the aims of this research are:

- (1) To improve current neuropsychological method performances in terms of dementia detection rate using machine learning;
- (2) To identify the few cognitive items that can be symptoms of the advancement of dementia experimentally;
- (3) To use a classification algorithm with the ability to provide clinicians with a useful and easy-to-understand knowledge derived from real data related to the dementia stages.

The scope of our research is limited to neuropsychological assessments—results related to other pathological procedures, neuroimaging, or biomarkers are excluded. To achieve this aim, we present an enhanced version algorithm called the Alzheimer's Disease Class Rules (AD-CR), which extends a rule-based classification method [17] to model cognitive activities from real data subjects related to the ADNI data repository [18]. In this research, we attempt to offer the following benefits to clinicians:

- (1) The ability to predict the signs of dementia progression objectively using sets of features and machine learning to analyse the data subjects;
- (2) To disseminate a rule-based classification model comprised of rules on the association of the cognitive items to the DSM-5 framework as a toolkit that can be used during prodromal dementia assessment;
- (3) To measure the classification accuracy of dementia progression when using models derived from real data subjects by classification algorithms;
- (4) To discover a few cognitive items that are signs of dementia progression.

The research questions that this research paper is attempting to answer are:

- A. How can we derive predictive models from dementia data that are competitive in performance with reference to sensitivity, accuracy and specificity rates?
- B. Can a few cognitive items be used as symptoms of dementia progression using machine learning?

The AD-CR algorithm builds models with rules that are (a) used to detect possible progression of dementia and (b) exploited by clinicians for decision-making. The AD-CR algorithm was evaluated on datasets that integrate multiple features obtained from dementia cognitive assessments and demographics within the ADNI data repository. The empirical evaluation contains a feature assessment phase to assess feature–feature and feature–class (medical diagnosis) significance. We sought cognitive features that could directly be signs of dementia progression and any of their correlations.

The AD-CR algorithm is one of the rare data-driven models that have attempted to not only capture dementia progression but also to map cognitive items into their degenerative areas as defined in the DSM-5 framework. We seek to identify a few cognitive items while maintaining a high DSM-5 domain coverage, if possible. The expected findings

carry significant practical implications in the DSM-5 mapping of degenerative dementia domains. For instance, if few cognitive items are identified and, when processed by the machine learning algorithms, including the AD-CR, the models derived maintain acceptable performance in medical settings, such findings can be significant due to the limited resources available.

This paper is structured such that Section 2 introduces the cognitive assessment we used in this study and reviews recent works related to rule-based models for dementia detection. The AD-CR algorithm is discussed in Section 3. The data and their features are highlighted in Section 4, and Section 5 presents the results and analysis. Lastly, we conclude in Section 6.

2. Literature Review

In this section, we initially introduce the cognitive assessment used in this research and then provide relevant recent literature on cognitive research related to rule-based classification for predicting dementia based solely on cognitive assessments or a combination of cognitive assessments with other pathological assessments. Neuroimaging or neuropathological dementia prediction systems are out of the scope of this article. Moreover, systems that use non-rule-based classification algorithms like deep learning or deep ANN are not part of this research scope.

ADAS-Cog is a medical assessment designed to measure the level of a patient's cognitive dysfunction [9]. While its use for monitoring pre-dementia and Mild Cognitive Impairment (MCI) has been criticised [19], it is generally accepted as one of the commonly used procedures for assessing dementia. ADAS-Cog is typically administered by a professional and consists of several tasks: word recall, naming objects and fingers, commands, constructional praxis, ideational praxis, orientation, word recognition, language, comprehension of spoken language, word finding, and remembering test instructions. ADAS-Cog usually takes an hour to complete in a clinical setting and produces a score between 0 and 70, with 70 indicating the most severe cognitive dysfunction.

In addition to the 11-question version, some variations exist. The ADAS-Cog 13 additionally contains a delayed word recall section, a maze or number cancellation section, and is scored between 0 and 85 [20]. Monllau et al. [21] tested the ADAS-Cog's ability to diagnose dementia on a sample of 451 subjects (of which there were 254 control subjects with normal cognition, 86 with MCI and 111 with AD). They found that the best cut-off score for describing AD was ≥ 12 , which had a sensitivity of 89.19% and a specificity of 88.53%. For our research, we are interested in the findings of the ADAS-Cog-13 (Mohs et al., 1997)—the dataset retrieved from the ADNI uses this version.

Thabtah et al. [22] investigated the problem of dementia detection when using the Everyday Cognition (ECog) test based on real data subjects collected from ADNI (TAD-POLE) [23]. In particular, data subjects (diagnosed as CN, MCI, and AD) who participated in ECog's two versions, the patient and study partner, were analysed using machine learning algorithms with a focus on rule-based classifiers. The results obtained using the machine learning algorithms showed that when the input dataset was balanced prior to the training phase, the classification performance in terms of accuracy was enhanced for the models derived. Moreover, rule-based classifiers such as RIPPER, PART and C4.5-Rules, besides random forest, derived competitive classifiers for dementia detection.

Das et al. [24] introduced a rule-based machine learning approach called the Sparse High-Order Interaction Model with Rejection Option (SHIMR) to predict dementia. The authors utilised data subjects from the ADNI dataset with some proteomics features as well as cerebrospinal fluid-related features collected from 141 data subjects. A comparative analysis of the data subjects was conducted using the C4.5 algorithm [25], and the rule-based approach (SHIMR). Results derived based on classification accuracy revealed that the SHIMR approach was superior to C4.5 by deriving classifiers with 84% accuracy on the data subjects.

Bang et al. [26] applied machine learning algorithms to clinical data to help physicians diagnose dementia more accurately. The authors applied their model to the CREDOS study consisting of data gathered by 37 universities in Korea from 2005–2013. The model recognises the need for a diagnostic process that is made up of four steps or modules. First, the module: the kScale variable selection method is used because it provides flexibility by verifying various results from several methods. Next, a classification model is built to determine dementia symptoms. This model uses the variable used by the proposed model as input and with the CDR-SB score as the class label. Machine learning algorithms, including ANN, Decision Tree, and support vector machine (SVM), are used to build the models. A descriptive analysis was then performed to describe the process of the classification of patients based on the classification model learnt by the classification algorithms. In this descriptive analysis, Decision Tree outcomes are transformed into easy-to-interpret rules. Finally, the visualisation model helps to seek useful characteristics of instances that are clustered in a descriptive step. It uses contrasting colours so that patients with dementia and other diagnostic indicators can be distinguished immediately.

Weakley et al. [27] used classification techniques for two datasets, each consisting of participants who were classified as exhibiting ‘normal ageing’ to dementia. The first dataset consisted of 310 participants who were diagnosed according to clinical diagnosis criteria; the second dataset contained 272 participants who were diagnosed according to CDR scores: 0 for normal ageing, 0.5 for MCI, and 1 and above for AD. Both datasets initially contained 27 variables related to various cognitive test scores, but this number was reduced using a wrapper feature selection method. To evaluate the generated models, 5-fold cross-validation was used. The main comments from the authors are that the classification techniques used showed no statistically different ability in classifying the data, with the most difficult (least successfully classified) group being the middle group, alternatively MCI, or CDR = 0.5.

Jammeh et al. [28] applied machine learning techniques to identify dementia from the National Health Service (NHS) data. A total of 26,483 data instances of individuals aged over 65 years were collected from 18 general practitioner surgeries. Machine learning techniques such as random forest, SVM, Naive Bayes, and linear regression were utilised to build classification models that were then evaluated using cross-validation testing methods with $n = 10$ -fold in terms of specificity, accuracy, and sensitivity, among others. The dementia classification models derived by the machine learning algorithms pinpointed that the model derived by the Naïve Bayes algorithm produced the highest performance. The classification model of Naïve Bayes was able to detect 295 instances of dementia who had not received a clinical diagnosis.

Thabtah et al. [6,29] studied real data related to functional elements of dementia from the ADNI project to determine which functional items may be signs of disease advancement. In [6], multiple dementia diagnosis methods were studied, and then a comparative analysis based on dissimilar criteria was carried out to seek a method that may cover more cognitive domains according to the DSM-5 framework. Later on, Thabtah et al. [29] utilised a computational intelligence approach to derive impactful functional elements related to dementia progression. The authors used a number of classification techniques with feature selection within a data-driven approach that was implemented and tested on real cases and controls from ADNI’s project and that had taken the Functional Activity Questionnaire assessment (FAQ). The results showed that there are a few functional items that can be observed during disease advancement. A mapping to these results were attempted by the authors and others in [30]. Table 1 depicts the summary of the relevant research conducted using specific criteria.

Table 1. Summary of the recent literature on rule-based classification models for dementia prediction and progression.

Reference	Classification Algorithms	Features	Dataset	Medical Assessment Method	Aim
[27]	Decision Trees (C4.5), Naïve Bayes, and Logistic Regression	Cognitive items, Functional activities measured by the Instrumental Activities of Daily Living (IADL), and items related to Geriatric Depression Scale (GDS) method	Two datasets: Clinical diagnosis and CDR approved by the Washington State University Institutional Review Board	Clinical Dementia Rating (CDR) and clinical diagnosis	To identify a few clinical indicators to classify individuals as demented, MCI or Cognitively Normal using machine-learning techniques
[26]	ANN, Decision Tree, and support vector machine (SVM)	Cognitive, functional, and demographic items	CREDOS dataset; data collected by 37 universities in Korea from 2005–2013	Clinical Dementia Rating-Sum of Boxes (CDR-SB)	To identify clinical measures that can help classify data subjects with dementia levels, if any
[28]	SVM, Random Forest, Naïve Bayes, and Logistic Regression	Pathological, cognitive, behavioural, and demographic items	Primary care data from NHS Devon (now part of Northern, Eastern and Western Devon Clinical Commissioning Group)	Multiple pathological and behavioural assessments	To implement a system that is able to detect dementia during clinical evaluation
[24]	SVM, Random Forest, Decision Tree	Biomarkers, cognitive tests' total scores, and demographic items	Biomarkers Consortium Plasma Proteomics Project RBM multiplex data, and ADNI-Merge common cognitive tests' scores	MMSE score, cerebrospinal fluid (CSF) and plasma protein features like tau, amyloid- β ($A\beta$) and phosphorylated tau (p-tau) proteins	To design an affordable dementia diagnosis system using a data-driven process
[22]	RIPPER, PART, Random Forest and Decision Tree (C4.5)	eCOG cognitive items	ADNI	Everyday Cognition (eCOG) test (patient and informant versions)	Compare rule-based classification on dementia prediction
[6]	Decision Trees (C4.5), Bayesian Network, and Logistic Regression	ADAS-13 Cognitive items, demographics attributes	ADAS-13 sheet + ADNI	ADAS-13 cognitive test	Assessing cognitive items in the problem of dementia progression using machine learning techniques
[29]	Decision Trees (C4.5), Bayesian Network, and Logistic Regression	FAQ Functional items, demographics attributes	FAQ sheet + ADNI	FAQ test	Assessing functional elements for dementia progression using a data-driven process with rule-based algorithms and non-rule-based algorithms
[31]	Decision Tree, Random Forest and the local interpretable model-agnostic explanations (LIME)	Demographic information, medical history, physical examination findings, laboratory results, and imaging studies	OPTIMA (Oxford Project to Investigate Memory and Ageing) dataset	Demographic characteristics, YES/NO questions related to health and well-being, rating scales, medical history, physical examinations, neuropsychological assessments, and performance of cognitive tests	Assisting clinicians in the early identification and diagnosis of dementia by providing useful and accessible machine learning models
[32]	DS-ANFIS, C4.5, SVM, Random Forest, SGERD, SLAVE2, QuickRules	Demographic, clinical variables	Open Access Series of Imaging Studies (OASIS) repository	Demographic, Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR), and Global Deterioration Scale (GDS)	Developing a fuzzy logic-based automated system for the diagnosis of dementia

3. The Algorithm

In this section, we discuss the classification algorithm (AD-CR) that consists of If-Then rules and is constructed using a class association rules approach, then used for prediction. The AD-CR algorithm extends other covering classification algorithms, such as the Rules Machine Learning (RML) algorithm of [17], by creating less overfitted classifiers that, during the training phase, do not keep adding items into the rules to maximise its expected accuracy. Instead, the AD-CR algorithm generates the rule when it reaches a confidence level above the user's specified confidence, similar to association rule mining algorithms. More details are provided in the following sub-sections. In the next two sub-sections, we explain how the algorithm works.

3.1. Terms

The AD-CR, is based on a classification approach devised from association rule mining called class association rules. In AD-CR, the classification problem's dataset is represented as distinct items along with their corresponding values. Each distinct item is associated with the largest target class label in the dataset with which the item has occurred, and such <item, class> representation is called a 'rule_item'. Below are the main definitions related to the AD-CR using D as an input classification dataset:

1. Data Observation: A collection of features with their values plus a target class value represented as $([(O_1, v_1), (O_2, v_2), (O_3, v_3) \dots, (O_k, v_k)], Class_k)$;
2. Training Dataset D : A combination of data observations, each associated with a target class c ;
3. Feature in D : An attribute that relates to the individual undergoing the screening process of dementia, such as age, gender, visit code, etc. The feature can be categorical (linked with a predefined set of values) or continuous (numeric or decimal);
4. Target Class in D : An attribute that represents the progression of dementia stage presented in a multi-class categorical form (0,1,-1). We limit the problem to progression (1) or no progression (0);
5. 1-Rule_Item (1- RI_k) in D : Is represented as $([(O_k, v_k)], Class_k)$;
6. Support of the RI_k , i.e., $\text{supp}(RI_k)$: Calculated from D as $\frac{|[(O_1, v_1), (O_2, v_2), (O_3, v_3) \dots, (O_k, v_k)], Class_k|}{|D|}$. When $\text{Supp}(RI_k) \geq \text{min_supp_threshold}$, the RI is considered frequent;
7. Minimum Support Threshold: Denoted as min_supp and is employed to differentiate between frequent and infrequent RI s;
8. Confidence of the rule_item (RI_k), i.e., $\text{Conf}(RI_k)$: Calculated from D as $\frac{|[(O_1, v_1), (O_2, v_2), (O_3, v_3) \dots, (O_k, v_k)], Class_k|}{|[(O_1, v_1), (O_2, v_2), (O_3, v_3) \dots, (O_k, v_k)]|}$. When $\text{Conf}(RI_k) \geq \text{min_conf_threshold}$, the RI is considered a potential rule;
9. Minimum Confidence Threshold: Denoted as min_conf and employed to measure the strength of a rule generated from a RI ;
10. Potential Rule: Takes the form $(I_1 \wedge I_2 \wedge \dots \wedge I_k) \rightarrow \text{Class}$;
11. Test Dataset: A combination of data observations, each is associated with a true class c .

3.2. Learning Phase

The rules algorithm pseudocode is depicted in Algorithm 1. The inputs are the classification dataset the min_supp , and min_conf (See terms 6–9 above). The AD-CR algorithm deals with both categorical and continuous attribute values in the training dataset; any missing attribute values are removed prior to the training phase since the application data we are dealing with is medical. The min_supp threshold is employed to determine rule_items that have sufficient frequency in the training data and is mainly used to pinpoint the best frequent rule_item in any iteration. In doing so, only the best rule_item in terms of frequency is chosen each iteration by the AD-CR algorithm to start building a new rule or to append into the current rule's body. The algorithm keeps merging attribute values into the rule until the current rule passes the min_conf threshold; when this happens, the current rule is added to the candidate rules list. The confidence of the rule is calculated according to definition 8 and can be considered a performance indicator that reflects the rule's position during the learning phase.

In discovering the rules (first phase), the AD-CR algorithm iterates over the training data to discover the best one rule_item (1-rule_item) in terms of support. It should be noted that only the largest class label in the training data occurring with the attribute value of the rule_item is considered when counting the rule_item's support. Once the 1-rule item is identified, the AD-CR algorithm starts building the first potential rule as $\text{Best_Item} \rightarrow C$. The algorithm evaluates the current rule's confidence. If the current rule has confidence larger than the min_conf threshold, then it will be added to the candidate rule list, and all data examples associated with it will be removed from the training dataset.

However, if the current rule's confidence is less than the min-conf threshold, then the learning algorithm isolates its training examples into a data structure: DS. The learning algorithm then checks whether adding the best frequent item of DS into the current rule will improve its confidence value. If this check yields true, then the algorithm appends the frequent item found into the current rule's body and repeats the same process of potentially adding frequent items into the rule's body from DS until the current rule's confidence passes the min-conf threshold. When this occurs, the rule will be generated and added to the candidate rule list. And all the rule's data examples will be removed from the original training dataset whenever the rule is generated, ensuring that these examples are only used once during the training phase. The algorithm then starts creating the second potential rule from the updated training dataset after removing the first rule's data examples and repeats the same process described above until either the training dataset empties or no more potential rules can be discovered. Once all candidate rules are generated, they are sorted based on confidence and support values.

Algorithm 1. The AD-CR Algorithm.

Input: A classification dataset D, the min-supp and min-conf thresholds

Output: Candidate Rule (CR): If-Then interpretable rules

```

1.   For each rule_item (RI), i.e.  $([(Attribute, Attribute_{value_k}], Class)$  in CD do
2.     BRI  $\leftarrow$  RI with Max (supp (RI))
3.     if conf (BRI)  $\geq$  min_conf
4.       begin
5.         DS  $\leftarrow$  DataExamples'of(BRI)
6.         CR  $\leftarrow$  BRI //adding the new rule into the candidate rule set
7.         D  $\leftarrow$  D - DS // amending the original data by removing the generated
rule's data
8.       end //if
9.     else
10.      begin
11.        DS  $\leftarrow$  DataExamples'of(BRI) // identifying data subsets that belong to BRI
12.        CRF = Candidate_Rule (DS, BRI) // a method that repetitively adds new
items if any to the current BRI to reach an accepted confidence level
13.        if conf (CRF)  $\geq$  min_conf
14.          begin
15.            UDS  $\leftarrow$  Update (DS, CRF)
16.            CR  $\leftarrow$  CRF //adding the new rule into the candidate rule set
17.            D  $\leftarrow$  D - UDS // amending the original data by removing the generated
rule's data
18.          end
19.        end //if
20.      end // else
21.    end for
22.  Repeat 1–16
23.  Exit when CD' is empty/checked
24.  Produce the CR list
25.  end
26.  Order CDs by the confidence and support values

```

One potential advantage of the training algorithm of AD-CR is that fewer rules are formed due to the assurance that whenever a rule is added to the candidate rule list, all related data examples are discarded. The training process also ensures that a training data example is restricted to one rule only, and its associated data examples cannot be considered for generating other rules, therefore cutting down the search space of potential rules. This may result in a concise set of classification models that can be easily controlled and used by the clinician.

Another possible advantage of the training algorithm is that it does not seek rules with 100% accuracy as found in classic covering classification algorithms or for specific rules as found in enhanced rules algorithms. AD-CR permits the rules to be produced even when the rule's accuracy is not perfect at a good confidence level, thus reducing the chance of any overfitted predictive models. It should be noted that models that are generated by the AD-CT will be used for predicting the class of test data.

3.3. Classification Phase

The AD-CR algorithm proposes a simple yet influential classification method that uses the most suitable rule to assign an appropriate class label to the test data during the classification phase. A rule used to assign the class normally meets two conditions:

1. It has the best ranking among all other rules in terms of confidence and support values;
2. The attribute values in its body are contained within the test data, thus ensuring the similarity of the attributes' values.

During the classification phase, when a test data example is to be classified, the AD-CR seeks in the final set the rule that matches the test data's attribute values, allocates its class to the test data example, and then moves to the next test data example and so forth. However, if there is no rule in the final rules set that fully matches the test data example, then the AD-CR algorithm searches for partially matching rules; such rules have at least one attribute data value similar to the test data example. The AC-DR algorithm then allocates the class linked with more partially matching rules to the test data example. This procedure ensures more than one rule can be employed to assign the class label to the test data, unlike other covering algorithms, which employ just the first partial matching rule. In cases when no rules partially or fully match the test data example, the algorithm uses the default rule one, which denotes the class label with most of the training data examples.

Using just one rule for classifying test data examples is not only a simple approach, but it also provides good predictive power, as seen later in the experimental analysis section. In addition, the approach considers other rules that partially match the test data examples when no fully matching single rule is available rather than invoking the default class label. This reduces the number of arbitrary classifications and, thus, misclassifications.

We summarise the primary characteristics of the AD-CR algorithm:

1. Only rules with significant frequency and confidence are formed;
2. Fewer rules are formed; thus smaller in size models are produced;
3. Unlike association rule mining, no rules share data examples, thus reducing the search space of items and potential rules;
4. Rules can be associated with some degree of error to minimise overfitting;
5. A simple and effective classification method is used in the prediction phase.

4. Data, Features and Pre-Processing

The datasets used in this research were obtained from ADNI. After careful investigation of the ADNI data repository, we identified the required features and combined multiple datasets. In particular, ADNI-Merge and ADAS-Cog-sheet cover the scope of the research project, and they include patients' cognitive information as well as visits. The ADNI-Merge dataset has a cohort of 2260 participants, in which each participant is being monitored on a six-monthly basis to track and study their AD progression, with multiple observations per patient at different points in time. The ADAS-Cog-sheet dataset comprises cognitive tasks that assess learning and memory, language production, language comprehension, constructional praxis, ideational praxis, and orientation. The ADAS-Cog sheet dataset details the patient's score in each cognitive task and the total scores attained during the assessment. It contained 6770 observations and 121 attributes. We are only interested in the cognitive tasks and their associated scores.

There are two common attributes within each dataset that form the basis of the data integration of ANDI-Merge and ADAS-Cog sheet: the patient ID (RID) and the visit code

(VISCODE), which will act as a reference for the merger. The aim of the merging process is to capture individual cognitive items together from their respective datasets and the diagnostic class (DX) from ADNI-Merge for each visit per patient. There were instances where a patient observation in the ADNI-Merge dataset was not integrated due to the ADAS-Cog-sheet dataset not having a corresponding RID and visit code. This could be due to the assessments not being performed for the patient during a visit for various reasons; thus, no merging occurs, resulting in fewer observations in the newly merged dataset.

The merged dataset (ADNI-Merge-ADAS-Cog' dataset) contained 14,627 data observations and attributes related to patients' visits, cognitive tests, memory tests, functional questionnaires, genetics, demographics, and biomarkers, among others. There is one target class in the ADNI-Merge dataset, which is the diagnosis of the last examination visit (DX). There are 4243 data subjects with missing values for the DX attribute in the ADNI-Merge dataset. The age of the participants ranges from 54.4 to 94.4, and the average age is 73. The age distribution of all participants—most are between 70 and 80 years of age. The frequency of the participants' medical visits with at least one per participant and up to 22 visits. To be exact, most of the participants had two medical visits followed by one, five, and seven medical visits, respectively.

As the main element of our research is the progression of the disease, this attribute does not exist in ADNI datasets; hence, we used the diagnosis (DX), visit code, and patient ID (RID) attributes from the ADNI-Merge dataset to create it, and we named the new class 'DX Progress' as described in Algorithm 2. Initially, we created an attribute called 'DX Digit' (Line #3) to encode the three possible diagnoses (CN:1, MCI:2, and AD:3); this attribute will help us assign the appropriate values to the DX Progress. The 'DX Digit' is filled based on the current diagnosis attribute (DX) in the original dataset (ADNI-Merge). The process starts by iterating over the data subjects after they are ordered by patient number (RID), and then iterating over the patients' visits. We always set the new attribute 'DX Progress' value to '0' (no progression) for each patient's first visit. The 'DX Progress' captures the change of diagnosis in a patient, establishing a new target class attribute.

Algorithm 2. Modelling Process of the Data.

Input: D: a dataset of all patients' information and visits

Output: D': A dataset with the new target variable 'Dx Progress'

```

1.   D' = D
2.   for each rid in D' do
3.     D'.dx digit' = 0
4.       for each viscode2 in D do
5.         D'.dx digit' = d
6.         if (dn = dn-1)
7.           D'.dx progress' = 0
8.         elseif (dn > dn-1)
9.           D'.dx progress' = 1
10.        else
11.          D'.dx progress' = -1
12.        end
13.    end
14.  remove all data instances where 'dx progress' = -1

```

The 'DX Progress' attribute will model the changes of the 'DX digit' from the matching patient ID and their subsequent visit with three possible class values. When there is a progression of diagnosis from CN '1' to MCI '2', or MCI '2' to Dementia '3', we labelled the change as '1' in the 'DX Progress' attribute (Lines 8–9). If there was no progression, it was labelled as '0'; regression as '-1' (Lines 6–7). Once the new class ('DX Progress') was derived, we removed instances that had been assigned with regression '-1' (Lines 11), as

we focus only on classes that are either ‘1’ for progression or ‘0’ for no progression, with only two class values remaining.

A brief analysis is conducted after creating the new class label, and it shows that the new class attribute is imbalanced as it is linked with a high number of no progression (‘0’) classes versus progression (‘1’). Proceeding with an imbalanced dataset to learn classification models produces skewed results that favour the majority target class and ignore the minority class. To deal with the imbalanced data situation, we oversampled the data by synthesised data observations of the low-frequency class from existing data samples. We arbitrarily generated new observations of the minority class to move the number of minority class items closer to the majority class. Table 2 shows the general statistics of the newly merged dataset before and after data balancing.

Table 2. General Statistics after Data Pre-processing and Data Balancing.

Dataset Name	# of Patients before Sampling	# of Data Observations (Visits)	DX Progress—Class Distribution before Data Balancing	DX Progress—Class Distribution after Data Balancing
ADNI-Merge-ADAS-Cog dataset	1710	6330	Total observations: 6330 ‘0’: 6020 (majority 95%) ‘1’: 310 (5%)	Total observations: 11,943 ‘0’: 6020 (50.40%) ‘1’: 5923 (49.60%)

5. Empirical Results and Discussion

5.1. Experimental Settings

All experiments were conducted on a computer with an Intel® Core™ i7-6200U 2.8 Ghz with 8GB RAM, on a Windows 10 Home, 64-bit. The hyperparameters of all methods and classification algorithms remained unchanged. Moreover, all experiments were conducted using open-sourced software—WEKA 3.8 and Python 3.9—where all platforms have extensive data pre-processing, statistical and graphical tools, as well as machine learning algorithms for data analysis [33,34].

Using Python’s Seaborn library, we assessed the feature–feature correlation using Pearson Correlation as the default method to generate a correlation matrix of the data’s features as a vector of integers to reduce independent attributes’ correlations. For the implementation of the AD-AC algorithm, we used Java and integrated the algorithm within WEKA version 3.8.4. The reason for selecting Java is that WEKA is implemented in Java, and many of the functions used for rule generation and pruning can be re-engineered to design and implement a new algorithm. Ten-fold cross-validation was used during the experiments as a measure of testing to ensure less biased results.

We applied SMOTE to sample the minority class labels in the dataset. SMOTE is a data sampling technique that adjusts the class distribution by taking the entire dataset as input, increasing the minority class using K-nearest neighbours (KNN) [35]. For feature selection, we used Pearson Correlation to reduce correlations among the features, then we used the Information Gain (IG) feature selection method to show class-feature correlations. Further, to measure the performance of the models derived by the classification algorithms, we used accuracy, sensitivity, and specificity, as shown in Equations (1)–(3), respectively. Sensitivity is the measure of the proportion of actual positive cases predicted as positive. Specificity is the measure of how well a test can identify the true negatives, whilst accuracy is the measure of the correct classification of the instances based on models and measures.

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \quad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

where,

TP (True Positive) = The model predicts a positive outcome among those with the positive class;

FP (False Positive) = The model predicts a positive outcome among those with the negative class;

TN (True Negative) = The model predicts a negative outcome among those with the negative class;

FN (False Negative) = The model predicts a negative outcome among those with the positive class;

To measure the effectiveness of the AD-CR algorithm, we used seven classification algorithms: Logistic Regression (LR), Multilayer Perceptron (MLP), Sequential Minimal Optimization (SMO), K-Nearest Neighbour (KNN; $k = 5$), Naïve Bayes, Ripple-Down Rule learner (Ridor), and Non-nested generalised exemplars (Nnge) [36–42]. The reasons for choosing these classification methods are due to the following factors:

- (1) The dissimilarities in the learning mechanisms used by these algorithms;
- (2) The different types of classifier formats they offer;
- (3) Many of these methods have been used in medical-related research;
- (4) To assist in obtaining a general conclusion wherein we can compare with the AD-CR algorithms in terms of performance measures.

MLP is a type of neural network that implements a feed-forward mechanism in the way it models the problem; the structure of the MLP comprises three layers: input, output, and hidden. The input layer consists of a set of neurons that represent the features in the training dataset, and it receives the data input that requires processing. The output layer performs the task of predicting the class label when the task is classification-based on computations made in the hidden layer. The algorithm keeps adjusting the model derived by amending the weights of the neurons until it reaches an acceptable performance level.

LR is a statistical algorithm that, in its simplest form, can describe the relationship between two features of data. LR uses a logistic function, as shown in Equation (4), to model a class label with two possible values. Unlike linear regression, LR's range is restricted between 0 and 1, and it does not necessitate a linear relationship between the independent variables and the class label since it uses a nonlinear log conversion. In addition, LR employs a conditional probability loss function called the 'maximum likelihood estimation'. When the probability is less than 0.50, the test data's class label will be predicted as 1; otherwise, 0.

$$\text{Logistic function} = \frac{1}{1 + e^{-x}} \quad (4)$$

NB is a probabilistic-based algorithm that uses Bayes' theorem to develop a feature-based assumption. The algorithm assumes that the target class is independent of all other features in the dataset, and it computes the likelihood of each class given test data based on the probabilities of the attributes' values of the test data within the training dataset. The algorithm assigns the class with the largest likelihood to the test data.

KNN is an instance-based learning algorithm in which, for test data to be classified, the algorithm utilises the nearest neighbour's class information to assign the test data the appropriate class. The algorithm does not learn a model from the training dataset and then uses that model for predicting the class label as in conventional classification algorithms; rather, the algorithm employs the training dataset to make the class assignment. The selected K-nearest neighbours are often determined by KNN using distance functions, such as Manhattan distance or Euclidean distance, in which the algorithm selects the closest points to the test data point to make the prediction. In classifying test data, the algorithm assigns a class label that belongs to the largest group of neighbours.

Nnge is a generalisation of instance-based learning algorithms with an incremental function in which it utilises non-nested generalised hyperrectangles that can be represented as simple If-Then rules. Every time a new data instance is inserted into the training dataset, Nnge forms a hyperrectangle by integrating the new data instance with a group of

neighbours with a similar class label. Nnge disallows hyperrectangles to overlap by using post-pruning based on heuristics. The algorithm employs a modified Euclidean distance function that processes the features, hyperrectangles, and weights.

Ridor is a rule-based induction algorithm that initially produces a default rule and then all possible exceptions for that rule with the lowest expected error rates. Exceptions of the default rule are other rules that forecast the class labels, which are dissimilar with that of the default rule. Afterward, the algorithm finds the ideal exceptions for each produced exception and repeats the process until it reaches the best performance. Ridor expands the search of exceptions like decision tree expansion.

SMO is a SVM type of algorithm disseminated to deal with an important issue that appears during the learning phase of SVM known as a quadratic programming problem. The SMO is a repetitive algorithm that reduces the problem into a set of optimisation tasks and solves each in an analytical manner.

In all classification experiments, we used the implementation of these algorithms in WEKA without amending the algorithms' hyperparameters. In addition, the minimum support threshold for the AD-CR algorithm was optimised to be between 0.025% and 2% to ensure the generation of rules based on the experimental analysis conducted in previous research studies [43,44]. On the other hand, the minimum confidence threshold has less impact on the performance, and it has been set to 50%. This also aligns with previous research studies.

5.2. Results and Discussion

5.2.1. Feature Assessment

Multiple sets of experiments have been conducted using two feature selection methods against the 'ADNI-Merge-ADAS' dataset to evaluate and identify the cognitive items that can be signs of dementia progression. The analysis criteria were to ascertain potential effective subsets of cognitive features that could be symptoms of dementia and their association with the DSM-5 diagnostic areas related to dementia. The first subset of features in each component contains all medical test items to serve as a baseline for performance comparison against other subsets. The feature assessment of the items was based on dissimilar criteria, including:

1. High-ranked features derived using the scores calculated by the IG method;
2. The similarity of features based on feature–feature assessment where a low intercorrelation is preferred.

Each experiment derived unique subsets of features using the approaches summarised in Table 3. Unique subsets of cognitive items were derived (Table 4) using the methods and criteria described in Table 3. 'Cog-subset2' was derived based on the feature–feature Pearson correlation excluding the class label. Based on the correlation figures, 'word recall', 'word-finding', and 'language comprehension' had the highest correlation against other features. For example, 'word recall' and 'word delay' had a strong correlation of 0.77, both having the same influence on the diagnostic class attribute. Accordingly, one of the items can be removed; in this case, 'word recall' has been identified as having a larger mean absolute correlation than the other cognitive items and is thus removed.

Table 3. Summary of the Methods Used to Derive Each Neuropsychological Subset.

Analysis Approach Used	Feature Subset
All Items in ADAS-Cog13	1
Pearson correlation subset	2
IG subset	3

Table 4. Summary of the Cognitive Items for Each Data Subset.

Criteria Used	Items Description	Subset
-	All Cog items	Cog-subset1
Remove highly correlated items based on the feature–feature Pearson correlation	COMMAND, CONSTRUCT, DELAYWORD, NAMING, IDEATIONAL, ORIENT, WORDRECOG, RMBRTESTINSTR, SPOKENLG, NUMBERCANCEL	Cog-subset2
Feature to class correlation scores of IG method	WORDRECALL, DELAYWORD, WORDRECOG	Cog-subset3
Cluster analysis based on the drop score %	WORDRECALL, DELAYWORD, WORDRECOG, ORIENT, COMMAND, WORDFIND	Cog-subset4

We analysed the features to observe groups of unique features based on the ranking. ‘Cog-subset3’ and ‘Cog-subset4’ are derived from analysis after ranking the scores computed by the IG method (Table 5). ‘Cog-subset3’, covers the two DSM-5 cognitive domains of learning & memory, and language. Cog-subset4’ covers mainly learning & memory, language, and perceptual motor function cognitive domains.

Table 5. Cognitive Items with Ranked Scores Derived by the IG Method.

IG Score	Feature
0.135	WORDRECALL
0.084	DELAYWORD
0.062	WORDRECOG
0.049	COMMAND
0.044	ORIENT
0.042	WORDFIND
0.038	IDEATIONAL
0.032	NAMING
0.032	CONSTRUCT
0.028	LANGUAGE
0.025	SPOKENLG
0.021	RMBRTESTINSTR
0.021	NUMBERCANCEL

The feature selection methods used identified three items that occurred the greatest number of times within each of the derived subsets: ‘word recall’, ‘delayed word recall’, and ‘word recognition’, signalling their significance as symptoms of prodromal dementia. These three items have the common association of retrieving words that require a patient to read, remember, recall, and recognise, and thus taps into their learning and memory, and language cognitive domains. During an assessment, a clinician can pay attention to the performance of these tasks by the patient and determine whether there are signs of disease advancement and if the patient needs early intervention.

5.2.2. Classification Assessment

Table 6 (left side) illustrates the performance of the classification algorithms against each cognitive subset using only the derived features and the class. The results from the feature assessment analysis showed the superiority of the AD-CR algorithm. Specifically, the models derived by the AD-CR algorithm all have acceptable performance in terms of

sensitivity, specificity, and predictive accuracy, even when only three items were processed ('Cog-subset3').

Table 6. Performance of the Classification Methods from Different Subsets of the Cognitive Items.

Including Demographic Features			Excluding Demographic Features			Algorithm	Subset
Specificity%	Sensitivity%	Accuracy%	Specificity%	Sensitivity%	Accuracy%		
79.30	87.10	83.15	72.90	76.90	74.88	LR	Cog-subset1 (baseline)
86.40	87.70	87.03	77.80	83.60	80.64	MLP	
76.70	88.60	82.59	70.70	77.00	73.82	SMO	
84.10	93.50	88.78	82.30	92.10	87.13	KNN	
35.18	94.50	70.54	34.30	94.40	64.07	NB	
94.50	86.10	90.32	91.10	79.70	85.44	Ridor	
92.80	91.60	92.22	88.40	86.90	87.64	Nnge	
93.50	91.30	92.38	89.10	86.90	88.00	AD-CR	
87.30	93.30	90.26	72.70	69.90	71.28	LR	Cog-subset2
86.30	87.70	86.97	77.60	76.90	77.29	MLP	
75.40	88.30	81.81	76.30	63.00	69.72	SMO	
84.30	92.90	88.52	81.50	89.80	85.64	KNN	
48.80	93.20	70.81	33.30	93.50	63.15	NB	
92.10	86.70	89.41	88.50	90.20	89.32	Ridor	
92.50	91.80	92.16	84.70	82.60	83.68	Nnge	
93.00	90.80	91.90	86.60	83.80	85.21	AD-CR	
73.30	80.90	77.03	63.60	58.60	61.11	LR	Cog-subset3
82.30	85.30	83.78	60.30	64.60	62.42	MLP	
70.30	84.20	77.20	64.90	58.90	61.96	SMO	
86.30	91.90	89.10	76.80	88.90	82.81	KNN	
71.30	85.90	78.51	57.30	61.90	59.59	NB	
94.20	82.90	88.59	81.50	77.30	79.46	Ridor	
92.00	91.50	91.77	77.20	76.80	77.01	Nnge	
92.90	89.50	91.25	77.40	83.80	80.53	AD-CR	
87.60	93.10	90.35	74.80	84.10	79.40	LR	Cog-subset4
84.20	87.80	85.97	65.90	84.20	74.10	MLP	
73.60	88.10	80.80	69.60	66.90	68.26	SMO	
86.60%	92.70	89.61	81.60	90.30	85.91	KNN	
65.00	90.50	77.63	44.80	93.20	68.78	NB	
92.90%	86.30	89.63	90.10	75.80	83.06	Ridor	
92.30	92.00	92.12	83.20	83.50	83.34	Nnge	
93.20	90.70%	91.93	84.20	85.40	84.77	AD-CR	

Moreover, the AD-CR algorithm was superior to the other algorithms when processing the baseline and 'Cog-subset2' in terms of predictive accuracy. For instance, it was able to produce predictive models from the baseline subset with higher accuracy than these of LR, MLP, SMO, KNN, NB, Ridor, and Nnge with 13.12%, 7.36%, 14.18%, 0.87%, 23.93%, 2.56%, and 0.36%, respectively. In fact, the AD-CR algorithm was able to derive a model from just three cognitive items of the 'Cog-subset3' (word recall, delayed word recall, and

word recognition) with an 83.80% sensitivity rate, i.e., the model works relatively well. This rate is just 3.10% less than when the same algorithm processed the baseline cognitive items (Cog-subset1). It seems that using only three items from the ADAS can indeed help clinicians in detecting possible signs of disease progression, at least when using models generated by the AD-CR and other classification algorithms like KNN. These three items also commonly appear in the other derived subsets, underlining their importance during the dementia progression.

Furthermore, the results revealed that albeit the KNN algorithm derived predictive models that were not the most accurate from the complete cognitive items and 'Cog-subset2', it was able to derive models from 'Cog-subset3' and 'Cog-subset4' that were slightly better than the AD-CR. This supports the view that the cognitive items in these subsets are important for detecting any possible signs of disease advancement. Specifically, evaluating 'orientation', 'command', and 'word finding' in addition to the items of 'Cog-subset3' resulted in predictive models with a 3.10% and 4.24% increase in accuracy for the KNN and the AD-CR algorithms, respectively, when compared with models derived by the same algorithms from just three items ('Cog-subset3'). In addition, the Ridor and Nnge algorithms, which produce classifiers with rules, were able to produce classification models for dementia progression that were good in terms of accuracy from all cognitive subsets except 'Cog-subset3'. This suggests that rule-based classification is an appropriate machine-learning approach for the problem of predicting the disease's progression.

We investigated the confusion matrix results to understand how these classification models behave in terms of performance. For instance, for the AD-CR model produced from 'Cog-subset1', out of the 5923 positive instances that were supposed to have progression, the AD-CR algorithm was able to correctly predict 5147 instances (true positive) and misclassified 776 instances as 'no progression' (false negative), when these instances were in fact having progression. The ability of the model to predict progression correctly is significant, especially in the medical field.

While using the AD-CR algorithm, the models derived from 'Cog-subset1' and 'Cog-subset2' produced the best accuracy measure; surprisingly, the NB probabilistic algorithm derived models from these two cognitive subsets with the highest sensitivity. NB algorithm derived models from the 'Cog-subset1' and 'Cog-subset2' with 93.50% and 94.40%, respectively—superior to the remaining classification algorithms, at least on the sensitivity metric. Nevertheless, the specificity rates derived by the NB algorithm from these datasets were unacceptable at 33.30% and 34.30%, respectively. In other words, the NB algorithm can detect disease advancement but with a high number of false positives. For example, by analysing the confusion matrices of the NB model against 'Cog-subset1', we discovered that out of 5923 positive instances, only 333 instances were misclassified by the NB algorithm as 'no progression'. However, there were 3957 false positives out of 6020 cases, which contributed to the low specificity rate.

The algorithm was able to balance between specificity and sensitivity rates across all models derived from the distinct cognitive feature subsets. Specifically, the AD-CR derived models with sensitivity and specificity rates of 86.90%, 89.10%, and 83.80%, 86.60%, respectively, from the 'Cog-subset1' and 'Cog-subset2'. When comparing the sensitivity rates of the models produced by the AD-CR algorithm, it is evident that the model derived from the 'Cog-subset1' had the highest; the difference in percentages of AD-CR models across all other measures against 'Cog-subset3' is not significant, where only its specificity (77.40%) is below 80.00%. The fact that 'Cog-subset3', when processed by the AD-CR algorithm, can produce good results with only three cognitive items means less assessment is required. Furthermore, the models derived from 'Cog-subset3' achieved this despite this subset of data covering only two DSM-5-prescribed cognitive domains, in comparison to 'Cog-subset2' and 'Cog-subset4', which cover more cognitive domains, respectively. Interestingly, the overall performance of the classification techniques produced similar results as other validation studies where the accuracy ranged from 82–99.6%, sensitivity from 58–74% and specificity from 91–98% [45,46].

Reducing the number of items in assessments related to dementia conditions, such as AD, can enable diagnosticians to evaluate specific elements related to a patient's condition based on the state of the disease, the patient's medical history, and the characteristics. Focus can then be on the cognitive elements detected as symptoms for the disease progression for individualised intervention and management plans to be designed to suit the patient and their family members. This will have positive impacts on their lives as well as on the healthcare system.

We investigated the impact of adding demographic features (age, gender, level of education, race category, and marital status) to the cognitive items (right side of Table 5). The performance across all classification algorithms improved immediately, particularly evident in the SMO, LR, and MLP algorithms, with noticeable improvement across all performance measures. For example, the improvement in accuracy, sensitivity, and specificity rates of the LR models after considering demographics with cognitive items of 'Cog-subset2' increased 18.98%, 23.40%, and 14.60%, respectively, when compared with the model derived from 'Cog-subset2' without demographics. Overall, the performance metrics' results improved for all the classification algorithms besides AD-CR when including demographics in the cognitive subsets.

While the AD-CR algorithm also improved, though not as significantly (at least when processing 'Cog-subset1'), it was still superior in comparison to the other algorithms overall as it further elevated, reaching a performance above 92%, 92.38% accuracy, 91.30% sensitivity, and 93.50% specificity. The fact that the AD-CR algorithm derived classification models from all distinctive cognitive subsets with predictive accuracy above 90.00% is evidence that this algorithm is suitable for detecting dementia stage progression. The AD-CR algorithm produced models from all distinctive cognitive subsets with sensitivity and specificity rates of over 90.00%—except a specificity of 89.50% from 'Cog-subset3', which contains only three cognitive items plus demographics.

In general, the Nnge algorithm derived competitive classification models on most cognitive subsets and above 90.00% accuracy rate. The classification models' performance derived by the AD-CR and Nnge algorithms from just six features ('Cog-subset4'), particularly their accuracy, sensitivity, and specificity, are close to those derived by the same algorithm from the complete cognitive items set ('Cog-subset1'). One notable result was observed by the models produced by the LR algorithm from 'Cog-subset4' including demographics in which the accuracy, sensitivity, and specificity rates improved by 7.20%, 6.00%, and 8.30%, respectively, when compared with those derived by the same algorithm from 'Cog-subset1' with demographics. The same pattern was also observed in the accuracy and specificity rates of the NB algorithm, which improved by 7.09% and 29.82%, respectively, making the items in 'Cog-subset4' significant, especially when investigated with demographic attributes such as age and gender. In fact, age was the highest-ranked feature among demographics that correlated with the class label, followed by gender and education level, at least using the feature assessment results.

5.2.3. Strengths and Weaknesses

Overall, the models produced from 'Cog-subset4', which contains (WORDRECALL, DELAYWORD, WORDRECOG, ORIENT, COMMAND, WORDFIND) in addition to demographic attributes by the machine learning algorithms show good performance and save time by only requiring assessment in 6 activities. While the AD-CR algorithm may not have produced the best results across all measures, its overall results are above 90%, falling within the performance range of other ADAS-cog validation studies [45,46] and required the use of fewer features; thus, it can be classified as a good performing model.

The model of the AD-CR algorithm is equipped with interpretable classification rules that can easily be understood by clinicians and novice users. These models offer the patient and their family answers as to which components have been signs of dementia progression. This is aligned with the patient's right outlined in the General Data Protection Regulation (GDPR), particularly the section on decision-making using automated algorithmic methods

and the “right for an explanation [47,48]. The GDPR requires that for any data collected from a subject (the individual undergoing medical screening) in an automated decision process (the screening process using machine learning), the subject should have the right to be given the rationale behind the decision-making process. Consequently, having an intelligent and easy-to-understand screening system for the patients and their family members, besides clinicians, is valuable. The system can provide explanations to the different stakeholders and useful information in answering questions such as:

1. What are the cognitive features that relate to the progression of AD?
2. Why does the output of the screening show no progression or potential progression?
3. Why is the patient being screened for AD, MCI, or CN?
4. What further assessment can be made based on the outcome?

In addition, the AD-CR rules can be used by the clinicians during the clinical session to identify factors that may impact the disease or the disease advancement. Yet the evaluation of these rules by a specialised neuropathologist will need to be assessed in a separate research work.

One of the limitations of this study is that it only evaluates the items of the ADAS-13 cognitive assessment method. Considering more than one cognitive assessment method would allow the data-driven classification model to cover a wider scope in terms of cognitive domains related to Neurocognitive Disorder, as defined in the DSM-5 framework. In addition, the study is limited to disease progression—it does not consider the passage of time. Considering the time elapsed between two successive dementia levels may help clinicians to develop specific intervention plans that are more appropriate to patients and their dementia level. Thereby, in the near future, we will investigate the differences in the disease progression between prodromal dementia, and other dementia levels (mild dementia, etc.). Lastly, we would like to discuss the rules generated by the proposed rule-based algorithm with a pathologist to seek his/her feedback on how these rules can be exploited to enhance the intervention and the diagnostic process of dementia.

6. Conclusions

The primary dementia condition is Alzheimer’s disease (AD), which is typically diagnosed by a specialised clinician based on a set of criteria, including the scores of cognitive assessments that are designed to measure the patient’s cognitive abilities besides other pathological assessments. One of the challenging problems related to dementia is determining the symptoms of the disease advances and the cognitive items that may be used as symptoms of such an advancement. This research investigated this problem based on real data related to neuropsychological assessment items by using machine learning to identify models with few cognitive items that influence the progression of AD—this to produce fast and accurate classifications. We proposed an enhanced covering classification algorithm called an Alzheimer’s Disease Class Rules (AD-CR) that derives models that can be exploited by clinicians during the dementia screening process. The AD-CR algorithm is used to predict the progression of the disease for individuals undertaking a neuropsychological assessment based on models derived from real data. These classification models indicate associations among assessed items. They can thus be utilised as a digital information sheet to guide clinicians in diagnostic-related decisions.

Empirical results obtained by dissimilar classification techniques against datasets related to cognitive items from the ADNI repository revealed that the AD-CR algorithm produced classification models that are competitive in terms of accuracy, specificity, and sensitivity rates. Moreover, the models derived by the classification algorithms are highly competitive to ANN, SVM, statistical, rule induction, and probabilistic approaches. The results demonstrated that few cognitive items may be used for screening dementia progression as models derived by the AD-CR algorithm from 3 to 6 cognitive items are predictive. After analysing the classification models derived, there were some associations, albeit low, between cognitive items that can be captured during the progression of AD. For example, word recall, word delay, word-finding, and language comprehension. These correlations

can be utilised within neuropsychological methods to retain fewer cognitive items with the least overlapping, which is beneficial for clinicians conducting clinical assessments for dementia progression.

Analysing the classification models in regard to the impact of demographics attributes revealed that age contributes the most critical factor followed by gender and then education level at least when using the datasets, and methods considered. When age and education level were combined with cognitive items, the classification models derived by the machine learning algorithms improved. In the near future, we will study specific dementia sub-groups for 3–5 years to analyse if unique cognitive features for each sub-group can be found.

Author Contributions: Conceptualization, F.T. and D.P.; Methodology, F.T.; Formal analysis, F.T.; Investigation, F.T.; Writing—original draft, F.T. and D.P.; Writing—review and editing, F.T. and D.P.; Supervision, D.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). For up-to-date information, see www.adni-info.org for further details on data.

Acknowledgments: Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

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

References

1. Kim, H.-J.; Min, J.-Y.; Min, K.-B. The association between longest-held lifetime occupation and late-life cognitive impairment: Korean longitudinal study of aging (2006–2016). *Int. J. Environ. Res. Public Health* **2020**, *17*, 6270. [[CrossRef](#)] [[PubMed](#)]
2. World Health Organization. Dementia. Available online: <https://www.who.int/news-room/fact-sheets/detail/dementia> (accessed on 15 January 2022).
3. Wittenberg, R.; Hu, B.; Barraza-Araiza, L.; Rehill, A. *Projections of Older People with Dementia and Costs of Dementia Care in the United Kingdom, 2019–2040*; The London School of Economics and Political Science, Care Policy and Evaluation Centre: London, UK, 2019. Available online: https://www.alzheimers.org.uk/sites/default/files/2019-11/cpec_report_november_2019.pdf (accessed on 21 December 2022).
4. Pickett, J.; Bird, C.; Ballard, C.; Banerjee, S.; Brayne, C.; Cowan, K.; Clare, L.; Comas-Herrera, A.; Corner, L.; Daley, S.; et al. A roadmap to advance dementia research in prevention, diagnosis, intervention, and care by 2025. *Int. J. Geriatr. Psychiatry* **2018**, *33*, 900–906. [[CrossRef](#)] [[PubMed](#)]
5. Alghamedy, F.H.; Shafiq, M.; Liu, L.; Yasin, A.; Khan, R.A.; Mohammed, H.S. Machine Learning-Based Multimodel Computing for Medical Imaging for Classification and Detection of Alzheimer Disease. *Comput. Intell. Neurosci.* **2022**, *2022*, 9211477. [[CrossRef](#)] [[PubMed](#)]
6. Thabtah, F.; Ong, S.; Peebles, D. Examining Cognitive Factors for Alzheimer’s Disease Progression Using Computational Intelligence. *Healthcare* **2022**, *10*, 2045. [[CrossRef](#)]
7. Zhu, F.; Li, X.; Haipeng, T.; He, Z.; Zhang, C.; Hung, G.-U.; Chiu, P.-Y.; Zhou, W. Machine learning for the preliminary diagnosis of dementia. *Sci. Program.* **2020**, *2020*, 5629090. [[CrossRef](#)]
8. Battista, P.; Salvatore, C.; Castiglioni, I. Optimizing neuropsychological assessments for cognitive, behavioral, and functional impairment classification: A machine learning study. *Behav. Neurol.* **2017**, *2017*, 1850909. [[CrossRef](#)]
9. Rosen, W.; Mohs, R.; Davis, K. A new rating scale for Alzheimer’s disease. *Am. J. Psychiatry* **1984**, *141*, 1356–1364. [[CrossRef](#)]
10. Folstein, M.; Folstein, S.E.; McHugh, P. “In-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **1975**, *12*, 189–198. [[CrossRef](#)]

11. Pereira, T.; Ferreira, F.; Cardoso, S.; Silva, D.; de Mendonca, A.; Guerreiro, M.; Madeira, S. Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease: A feature selection ensemble combining stability and predictability. *BMC Med. Inform. Decis. Mak.* **2018**, *18*, 137. [CrossRef]
12. Wessels, A.; Siemers, E.; Yu, P.; Andersen, S.; Holdridge, K.; Sims, J.; Sundell, K.; Stern, Y.; Rentz, D.M.; Dubois, B.; et al. A combined measure of cognition and function for clinical trials: The integrated Alzheimer's Disease Rating Scale (iADRS). *J. Prev. Alzheimers Dis.* **2015**, *2*, 227–241. [CrossRef]
13. Jutten, R.J.; Harrison, J.E.; Brunner, A.J.; Vreeswijk, R.; Deelen RA, J.; de Jong, F.J.; Opmeer, E.M.; Ritchie, C.W.; Aleman, A.; Scheltens, P.; et al. The Cognitive-Functional Composite is sensitive to clinical progression in early dementia: Longitudinal findings from the Catch-Cog study cohort. *Alzheimer's Dement. Transl. Res. Clin. Interv.* **2020**, *6*, e12020. [CrossRef] [PubMed]
14. Shahbaz, M.; Niazi, A.; Ali, S.; Guergachi, A.; Umer, A. Classification of Alzheimer's disease using machine learning techniques. In Proceedings of the 8th International Conference on Data Science, Technology and Applications, Toyama, Japan, 7–12 July 2019; pp. 296–303. [CrossRef]
15. AlShboul, R.; Thabtah, F.; Walter Scott, A.J.; Wang, Y. The Application of Intelligent Data Models for Dementia Classification. *Appl. Sci.* **2023**, *13*, 3612. [CrossRef]
16. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. 2013. Available online: <https://www.psychiatry.org/psychiatrists/practice/dsm> (accessed on 11 March 2022).
17. Thabtah, F.; Peebles, D. A new machine learning model based on induction of rules for autism detection. *Health Inform. J.* **2020**, *26*, 264–286. [CrossRef] [PubMed]
18. Alzheimer's Disease Neuroimaging Initiative [ADNI]. 2021. Available online: <http://adni.loni.usc.edu> (accessed on 15 May 2021).
19. Kueper, J.; Speechley, M.; Montero-Odasso, M. The Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog): Modifications and responsiveness in pre-dementia populations. A Narrative Review. *J. Alzheimers Dis.* **2018**, *63*, 423–444. [CrossRef]
20. Mohs, R.C.; Knopman, D.; Petersen, R.C.; Ferris, S.H.; Ernesto, C.; Grundman, M.; Sano, M.; Bieliauskas, L.; Geldmacher, D.; Clark, C.; et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Society. *Alzheimer Dis. Assoc. Disord.* **1997**, *11*, S13–S21. [CrossRef]
21. Monllau, A.; Pena-Casanova, J.; Blesa, R.; Aguilar, M.; Bohm, P.; Sol, J.M.; Hernandez, G. Diagnostic value and functional correlations of the ADAS-Cog scale in Alzheimer's disease: Data on NORMACODEM project. *Neurologia* **2007**, *22*, 493–501.
22. Thabtah, F.; Spencer, R.; Ye, Y. The correlation of everyday cognition test scores and the progression of Alzheimer's disease: A data analytics study. *Health Inf. Sci. Syst.* **2020**, *8*, 24. [CrossRef]
23. Marinescu, R.V.; Oxtoby, N.P.; Young, A.L.; Bron, E.E.; Toga, A.W.; Weiner, M.W.; Fox, N.C.; Golland, P.; Klein, S.; Alexander, D.C. TADPOLE challenge: Accurate Alzheimer's disease prediction through crowdsourced forecasting of future data. In *Predictive Intelligence in Medicine*; Springer: Cham, Switzerland, 2019; Volume 11843, pp. 1–10. [CrossRef]
24. Das, D.; Ito, J.; Kadowaki, T.; Tsuda, K. An interpretable machine learning model for diagnosis of Alzheimer's disease. *PeerJ* **2019**, *7*, e6543. [CrossRef]
25. Quinlan, J. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: Burlington, MA, USA, 1993.
26. Bang, S.; Son, S.; Roh, H.; Lee, J.; Bae, S.; Lee, K.; Hong, C.; Shin, H. Quad-phased data mining modeling for dementia diagnosis. *BMC Med. Inform. Decis. Mak.* **2017**, *17*, 60. [CrossRef]
27. Weakley, A.; Williams, J.A.; Schmitter-Edgecombe, M.; Cook, D.J. Neuropsychological test selection for cognitive impairment classification: A machine learning approach. *J. Clin. Exp. Neuropsychol.* **2015**, *37*, 899–916. [CrossRef]
28. Jammeh, E.A.; Carroll, C.B.; Pearson, S.W.; Escudero, J.; Anastasiou, A.; Zhao, P.; Chenore, T.; Zajicek, J.; Ifeachor, E. Machine-learning based identification of undiagnosed dementia in primary care: A feasibility study. *BJGP Open* **2018**, *2*, bjgpopen18X101589. [CrossRef]
29. 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]
30. Thabtah, F.; Spencer, R.; Peebles, D. Common dementia screening procedures: DSM-5 fulfilment and mapping to cognitive domains. *Int. J. Behav. Healthc. Res.* **2022**, *8*, 104–120. [CrossRef]
31. Vyas, A.; Aisopos, F.; Vidal, M.E.; Garrard, P.; Paliouras, G. Identifying the presence and severity of dementia by applying interpretable machine learning techniques on structured clinical records. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 271. [CrossRef] [PubMed]
32. Chen, T.; Su, P.; Shen, Y.; Chen, L.; Mahmud, M.; Zhao, Y.; Antoniou, G. A dominant set-informed interpretable fuzzy system for automated diagnosis of dementia. *Front. Neurosci.* **2022**, *16*, 867664. [CrossRef] [PubMed]
33. Hall MFrank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. The WEKA Data Mining Software: An Update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [CrossRef]
34. Van Rossum, G.; Drake, F.L., Jr. *Python Reference Manual*; Centrum Voor Wiskunde en Informatica Amsterdam: Amsterdam, The Netherlands, 1995.

35. Chawla, N.; Bowyer, K.; Hall, L.; Kegelmeyer, P. SMOTE: Synthetic Minority Over-sampling Technique. In *International Conference of Knowledge Based Computer Systems*; National Center for Software Technology: Mumbai, India; Allied Press: Dunedin, New Zealand, 2000; pp. 46–57.
36. le Cessie, S.; van Houwelingen, J.C. Ridge estimators in logistic regression. *Appl. Stat.* **1992**, *41*, 191–201. [[CrossRef](#)]
37. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*; MIT Press: Cambridge, MA, USA, 1986; Volume 1, pp. 318–362.
38. Platt, J. Fast training of SVM using sequential optimization. In *Advances in Kernel Methods—Support Vector Learning*; Scholkopf, B., Burges, C., Smola, A., Eds.; MIT Press: Cambridge, MA, USA, 1998; pp. 185–208.
39. Aha, D.; Kibler, D.; Albert, M.K. Instance-based learning algorithms. *Mach. Learn.* **1991**, *6*, 37–66. [[CrossRef](#)]
40. John, G.H.; Langley, P. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Montreal, QC, Canada, 18–20 August 1995; pp. 338–345.
41. Gaines, B.R.; Compton, P. Induction of ripple-down rules applied to modeling large databases. *J. Intell. Inf. Syst.* **1995**, *5*, 211–228. [[CrossRef](#)]
42. Martin, B. *Instance-Based Learning: Nearest Neighbour with Generalisation*; Working Paper; Department of Computer Science, University of Waikato: Hamilton, New Zealand, 1995.
43. Mattiev, J.; Kavšek, B. Coverage-Based Classification Using Association Rule Mining. *Appl. Sci.* **2020**, *10*, 7013. [[CrossRef](#)]
44. Abdelhamid, N.; Ayes, A.; Thabtah, F. Phishing detection based Associative Classification data mining. *Expert Syst. Appl.* **2014**, *41*, 5948–5959. [[CrossRef](#)]
45. Nogueira, J.; Freitas, S.; Duro, D.; Almeida, J.; Santana, I. Validation study of the Alzheimer’s disease assessment scale—Cognitive subscale (ADAS-Cog) for the Portuguese patients with mild cognitive impairment and Alzheimer’s disease. *Clin. Neuropsychol.* **2018**, *32*, 46–59. [[CrossRef](#)] [[PubMed](#)]
46. Yang, H.; Cheng, Z.; Li, Z.; Jiang, Y.; Zhao, J.; Wu, Y.; Gu, S.; Xu, H. Validation study of the Alzheimer’s Disease Assessment Scale-Cognitive Subscale for people with mild cognitive impairment and Alzheimer’s disease in Chinese communities. *Int. J. Geriatr. Psychiatry* **2019**, *34*, 1658–1666. [[CrossRef](#)] [[PubMed](#)]
47. General Data Protection Regulation (GDPR). General Data Protection Regulation (GDPR)—Final Text Neatly Arranged. Available online: <https://gdpr-info.eu/> (accessed on 5 May 2022).
48. Goodman, B.; Flaxman, S. European Union regulations on algorithmic decision-making and a “right to explanation”. *AI Mag.* **2016**, *38*, 50–57. [[CrossRef](#)]

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.