Recursive Feature Elimination with Cross Validation for Alzheimer's Disease Classification using Cognitive Exam Scores

2023 Intelligent Methods, Systems, and Applications (IMSA) | 979-8-3503-3556-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/IMSA58542.2023.10217660

Christian Yaphet Freytes
Electrical & Computer Engineering
Florida International University
Miami FL, USA
cfrey001@fiu.edu

Robin Perry Mayrand
Electrical & Computer Engineering
Florida International University
Miami FL, USA
rmayr002@fiu.edu

Luana Okino Sawada
Computing & Information Sciences
Florida International University
Miami FL, USA
lokin001@fiu.edu

Thony Yan Liang
Electrical & Computer Engineering
Florida International University
Miami FL, USA
tyan001@fiu.edu

Rosie E. Curiel Cid
Department of Neuropsychology
University of Miami
Miami FL, USA
rcuriel2@med.miami.edu

Shanna Burke
Public Health and Social Work
Florida International University
Miami, United States
sburke@fiu.edu

David Loewenstein
Cognitive Neurosciences and Aging
University of Miami
Miami FL, USA
dloewenstein@med.miami.edu

Ranjan Duara
Wien Center for AD & MD
Mt Sinai Medical Center
Miami FL, USA
ranjan.duara@msmc.com

Malek Adjouadi
Electrical & Computer Engineering
Florida International University
Miami, United States
adjouadi@fiu.edu

***Abstract*—Prodromal detection of Alzheimer's Disease(AD) is a substantial challenge in the research community. Among the tools used in AD diagnosis, cognitive exams are standard in most procedures. However, the barrage of cognitive examinations is both time and resource consuming. With the use of Machine Learning, Feature Elimination (FE) can be combined with classification algorithms to determine which cognitive exams are best suited for diagnosis. Using the results of FE, it can be determined if subsections of different composite scores can be combined to create a new enhanced and exhaustive exam. This paper implements a Recursive Feature Elimination with Cross Validation (RFECV) machine learning algorithm to determine which cognitive exams perform best for AD classification tasks. Out of 119 features, an average of 16 features were selected as optimal. These optimal features average 75% Accuracy, 70% Precision, and 75% Recall and an F1 Weighted score of 71% in classification.**

***Index Terms*—Alzheimer's Disease, Machine Learning, Recursive Feature Elimination, Feature Elimination, Decision Tree, Random Forest**

This research is supported by the National Science Foundation under grants: CNS-1920182, CNS-2018611, and CNS-1551221, and with the National Institutes of Health through the P30AG066506 with the 1Florida Alzheimer's Disease Research Center (ADRC).

I. Introduction

Alzheimer's Disease is a progressive neurodegenerative disease that currently affects millions of Americans, accounting for over half of all cases of dementia, with roughly 50 million people affected worldwide [1]. For the year of 2020, the cost of Alzheimer's Disease treatment was estimated to be $305 billion and expected to increase to over $1 trillion [2].

Currently, the only definitive way to confirm Alzheimer's is via post mortem examination of brain tissue [3]. Without an autopsy, Alzheimer's diagnosis is a complex issue, but alternatives do exist. Alzheimer's diagnostics can be performed using a multitude of other procedures such as neuroimaging, biomarkers, or clinical assessments [4]. Neuroimaging, such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Computed Tomography (CT) scans also show potential for AD diagnosis [5]. However, even with these modalities, there is variability to what results represents a diagnosis of Alzheimer's Disease. For example, the thresholds agreed upon as signs of Alzheimer's pathology vary depending on the lab or center performing the evaluation [6]. Additionally, a subject might not be exhibiting signs of dementia but presents pathology in neuroimaging, which brings up concerns of whether or not such a patient should be diagnosed with AD. Due to these discrepancies, AD diagnosis

performed without autopsy are estimations of disease status based on a clinicians evaluation. Additionally, AD databases face an issue of incompleteness due to an increased attrition rate as some subjects may lack the time or resources to complete every kind of assessment.

Out of all available modalities for AD diagnosis, clinical evaluations are some of the most standard procedures. A clinical evaluation is accomplished by reviewing medical and familial history in addition to making behavioral observations of a subject. It is common for neuropsychological tests such as the Mini-Mental State Exam (MMSE) and the Montreal Cognitive Assessment (MoCA) to be included in clinical evaluations for Alzheimer's Disease. Although a standard, cognitive examinations can potentially be biased, whether it's the format or language the exam is written in, and many of these exams require more sensitivity to the prodromal stages of Alzheimer's Disease [7]–[9] . Therefore, a need exists for more comprehensive exams to precisely capture the earlier stages of Alzheimer's Disease. Considering all these challenges, developing a standardized exam that consumes fewer resources and is flexible in its diagnostic capabilities would be highly beneficial to AD research.

The current landscape of AD research incorporates various machine learning (ML) applications that show potential as tools to aid in the classification and diagnosis of AD [10], [11]. Machine learning performance, as with all computing techniques, will be dependent on the quality of its source data. In short, introducing substandard input data to ML algorithms generates substandard results. Medical data is inherently extensive and has a natural tendency to be highly dimensional which can become burdensome to the processing ability of many ML algorithms. In order to alleviate the computational stress of both the dimensionality and quality of input data, researchers employ feature optimization to include only the most pertinent information [12], [13].

One form of feature optimization is feature elimination (FE), where the weakest features given to a learning model are removed. Recursive Feature Elimination (RFE) is a popular FE algorithm due to it's simplicity, speed, and compatibility with all supervised learning estimators. RFE is a wrapper-style method that assigns importance scores to features using an iterative process. This ranking style allows for enhanced interpretability not only in the final rankings of the selected features, but of the selection process itself. RFE can be combined with cross-validation (CV) using stratified k-fold techniques to address overfitting and data incompleteness issues to simulate an independent dataset.

Machine learning research in AD faces the challenge of highly dimensional datasets burdening predictors. This resource burden could potentially be solved by implementing RFE to determine an optimal feature space. This paper incorporates feature reduction via a Recursive Feature Elimination with Cross Validation (RFECV) machine learning classification algorithm with cognitive examinations to identify the most relevant exams for AD diagnosis in order to develop a reduced and efficient feature set. Overall, RFECV was able to generate an optimal feature space of 16 features for cognitive status classification. The rest of this document is structured as follows: $Section\ II$ describes the data collection method, $Section\ III$ describes the method implemented and how results were evaluated $Section\ IV$ concludes the paper and discusses limitations and future works.

II. Data

TABLE I
SUBJECT DEMOGRAPHIC DATA

Diagnosis	Avg Age	Sex (F/M)	Total
NC	72.2	11/5	16
eMCI	72.5	48/41	89
AD	72.7	13/5	18
Late MCI	73.0	10/12	22
PreMCI Clinical	75.0	7/4	11
PreMCI NP	71.2	18/4	22
Total	72.8	107/71	178

All data used was gathered by the 1Florida Alzheimer's Disease Research Center (1FADRC) study [14]. 1FADRC collects clinical data, multiple neuroimaging modalities, and neuropsychological test scores for medical and research applications. 499 subjects between the ages of 52 and 93 were made available via the 1FADRC(Table I). Subjects enrolled in the 1FADRC may not have a full cognitive examination battery completed due to limited resources or time constraints, so filtering was used to include exams that include the largest majority of patients . After filtering, the final data set contained 178 unique subjects with each having 14 different completed neuropsychological examinations [15]. The new data set subjects were categorized into different cognitive diagnosis groups: Cognitively Normal (CN), PreMild Cognitive Impairment (pMCI) Clinical, PreMCI Neuropsychological (pNP), Early MCI (EMCI), Late MCI (LMCI), and AD.

The 14 neuropsychological exams included after filtering are, in alphabetical order:

1) Benson Complex Figure: Assesses a subject's visioconstructional and visual memory functions.

2) Clinical Dementia Rating (CDR): A composite rating scale used to quantify cognitive impairment by analyzing functional performance of a subject in different areas such as memory, orientation, and judgement.

3) Controlled Oral Word Association Test (COWA): Verbal fluency test that measures executive function via spontaneous word association.

4) Category Fluency: Evaluation of semantic memory where a subject is instructed to name different items of a given category.

5) Craft Story Recall: Assessment of a subject's ability to recall a short story after a certain amount of time.

6) Hopkins Verbal Learning Test: A list learning exam used to determine cognitive status.

7) Loewenstein-Acevedo Scales for Semantic Interference and Learning (LASSI): Cognitive stress test designed to identify early cognitive fluctuations linked to cognitive impairments.

8) Mini Mental State Exam (MMSE): Questionnaire used in clinical and research settings to assess cognitive status.

9) Montreal Cognitive Assessment (MoCA): A screening assessment used in the detection of cognitive status composed of various evaluations of cognitive functions.

10) Multilingual Naming Test (MINT): A object naming test containing items with similar levels of usage and familiarity in English, Spanish, Hebrew, and Mandarin.

11) Number Span Test: : Examination that utilizes numbers to assess working memory by having a subject count forwards and backwards.

12) Trail Making Test: An assessment of a subject's processing speed and executive function judged in terms of time.

13) Stroop Test: Exam used to assess cognitive interference via the stroop color word test.

14) Wechsler Memory Scale: A neuropsychological test composed of five distinct index scores used to measure cognitive functions of a subject.

Each exam uses unique scoring metrics and are composite scores of several subsections. All subsections' scores are made available to the RFECV algorithm. After accounting for section totals and subtotals there are 119 total input features.

III. Method

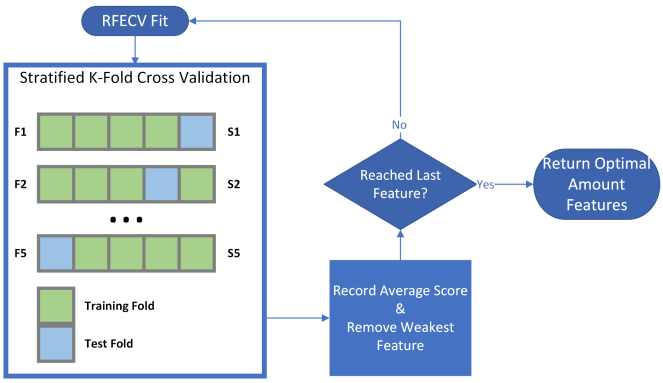

Fig. 1. Flowchart of RFECV Process

Recursive Feature Elimination with Cross Validation (RFECV) for multi-class classification was implemented via the sci-kit learn Python library [16]. The RFECV algorithm is composed of three major components: (1) stratified cross-validation (CV), (2) classifiers, and (3) Recursive Feature Elimination (RFE). After implementing RFECV, the performance of both classifiers and features is evaluated.

A. Stratified Cross Validation

Due to the highly imbalanced nature of the input data a stratified k-fold of $n = 5$ splits was used. RFECV functions by first performing cross-validation on the input data dividing it into testing and training sets prior to introducing it to a classifier. The stratified k-fold process ensures that each training and testing sets are balanced.

B. Classifiers

Two tree classifiers were implemented in the RFECV algorithm: Decision Tree (DT) and Random Forest (RF). Both classifiers were trained using cross-validation and no maximum tree depth.

1) Decision Tree: Decision trees are supervised learning algorithms used in both classification and regression analysis. Tree models predict the values of target variables via the use of discrete values called classification trees.

2) Random Forest: An ensemble learning method used for classification which is made up of multiple randomly assembled decision trees. Random Forest models for classification return the class that is selected by the most trees.

C. Recursive Feature Elimination and Classifier Training Parameters

Recursive Feature Elimination functions by executing a classifier with a target metric, such as accuracy or precision, and then removing the feature with the lowest gini impurity score (Figure 1). After removing the lowest performing feature, the classifier is refitted with the newly reduced set of features and run again. Reiterations are executed until a set minimum amount of features is achieved, in this implementation the minimum number of features is defined as $n = 1$. After generating feature rankings based on the RFE method, an optimal amount of features is determined. The optimal feature range is calculated using the confidence interval across all k-folds, meaning selecting a feature range with the lowest standard deviation and the highest metric score. When a feature makes it into the optimal feature range, it is considered a supportive feature.

IV. Results

A. RFECV Classifier Performance Evaluation

For both classifiers, the calculated amount of optimal features for each metric was recorded across all runs and their averages recorded. The performance of each RFECV classifier was calculated using four metrics: Accuracy, Precision, Recall, and F1 weighted score. These scores are calculated using the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). All scoring metrics are the calculated averages across all 5 folds.

1) Accuracy: In order to compute observational error, accuracy was selected as a metric. Accuracy evaluates how a model performs across all classes and is calculated as follows:

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

2) Precision: Precision determines how efficient a model is at predicting each individual category and is calculated as:

$$Precision = \frac{TP + TN}{TP + FP} \tag{2}$$

3) Recall: Also known as sensitivity, recall is the probability of a model providing a correct prediction and is calculated as:

$$Recall = \frac{TP + TN}{TP + FN} \tag{3}$$

4) F1 Weighted: An F1 score is the mean of precision and recall. An F1 weighted score provides the weighted mean of F1 measured in relation to class probability. An F1 Score is calculated via:

$$F1_{Score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

The F1 Weighted score is calculated for an N-class dataset as:

$$F1_{weighted} = \sum_{i=1}^{N} w_i * F1Score_i \tag{5}$$

Where,

$$w_i = \frac{\text{Number of samples in class i}}{\text{Total number of samples}} \tag{6}$$

B. RFECV Feature Performance Evaluation

Each time a feature is flagged as a support feature, it is recorded and stored. The combined total of every instance a feature was listed as supportive is then calculated as a percentage across all runs and aggregated into a single table.

C. RFECV Optimization Figures

Visualizations of the RFECV Optimization process were generated for all 4 metrics across all runs. Figures 2 and 3 are examples using two samplings of the iterations used in RFECV.

D. Classifier Performance

Out of both classifiers, Random Forest outperformed Decision Tree across all metrics except precision (Table II). Combined, the algorithms averaged 73.37% accuracy, 70.63% precision, 73.37% recall, and 71.01% F1 weighted score. Across all metrics, the average amount of optimal features for both classifiers was 16 (Table III). The largest amount of selected features was 30 for the F1 Weighted metric in DT and the smallest amount was 6 in both accuracy and recall for DT.

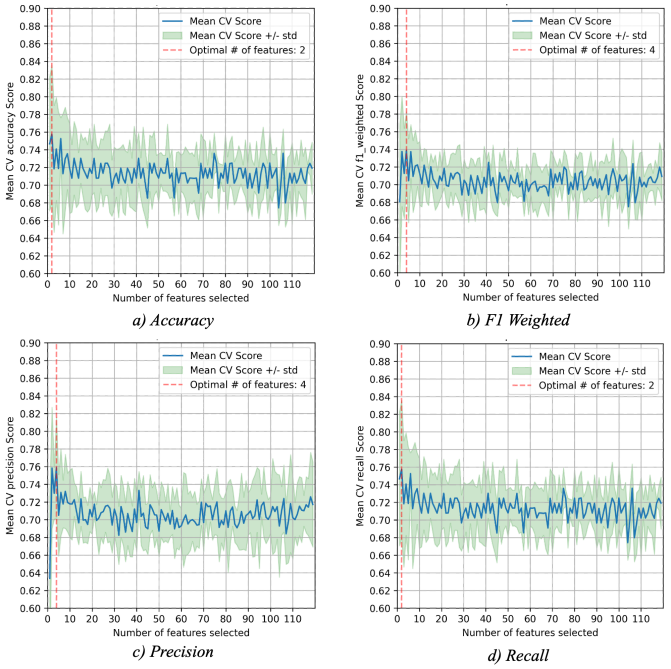

Fig. 2. Visualization of RFECV Optimization for the Decision Tree Classifier of a) Accuracy, b) F1 Weighted Score, c) Precision, and d) Recall

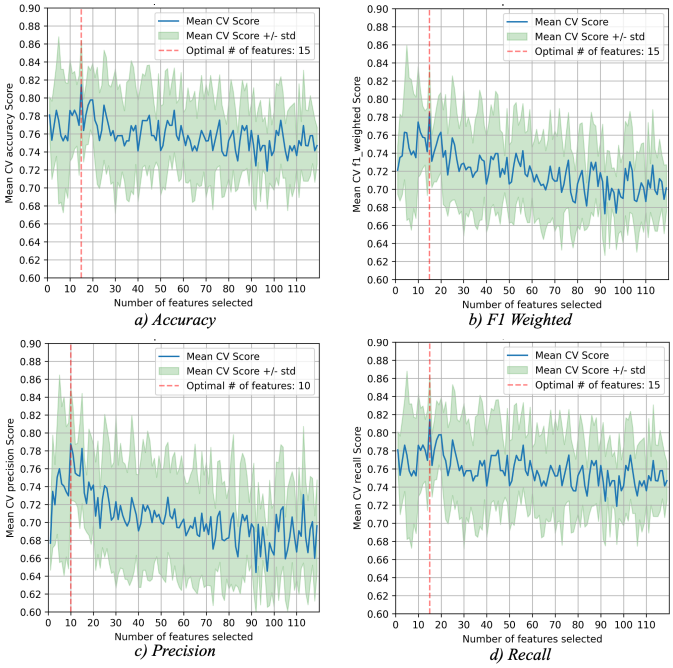

Fig. 3. Visualization of RFECV Optimization for the Random Free Classifier of a) Accuracy, b) F1 Weighted Score, c) Precision, and d) Recall

E. Feature Performance

CDRSUM, a derived score using CDR, made it into the optimal feature range in 100% of the RFECV iterations across all metrics(Table IV). The next closest overall support percentage was HVLT_DR, a subsection of HVLT, with a support

TABLE II
REFCV METRIC AVERAGES

Classifier	Overall	Accuracy	Precision	Recall	F1 Weighted
Random Forest	**73.39**%	**75.69**%	70.42%	**75.69**%	**71.78**%
Decision Tree	70.80%	71.06%	**70.84**%	71.06%	70.23%
Combined	72.10%	73.37%	70.63%	73.37%	71.01%

Combined represents the average across both classifiers for a given metric. Best model for each metric in **bold**.

TABLE III
REFCV OPTIMAL FEATURE RANGE AVERAGES

Classifier	Overall	Accuracy	Precision	Recall	F1 Weighted
Random Forest	**15**	17	**12**	17	**14**
Decision Tree	17	**6**	24	**6**	30
Combined	16	12	18	12	22

Combined represents the average across both classifiers for a given metric. Best model for each metric in **bold**.

percentile of 65.38%. Out of the top 30 supportive features, only 7 were selected at least 50% of the time. Although top supportive feature, FL_COWA_A, a subsection of the COWA exam, only made it into 5% of the time into the accuracy and recall metrics across all runs, the lowest percentile score of any top feature.

V. DISCUSSION

Overall, for both classifiers, RFECV was able to reduce the feature space from 119 features to an average of 16 optimal features across all runs and metrics. In comparing classifiers, RF outperformed DT on average across most categories except feature optimization in the accuracy and recall metrics. Additionally, many of the top 30 features are subsections of their respective composite totals. The prevalence of feature subsections indicates a potential for the construction of an enhanced exam that only utilizes these scores. With a performance average of 73% across all metrics, feature optimization using RFECV has potential viability as a solution to the resource issues faced in AD research.

A. Limitations

An important consideration for classification with this RFECV implementation is that some features introduced are part of the algorithmic diagnosis used to generate the class label. The algorithmic diagnosis is a formula used to determine to which cognitive diagnosis class a subject belongs to [17]. Therefore, some features are intrinsically correlated with the diagnosis which could introduce bias to the classifiers' predictions. Additionally, even though all features used in the algorithmic diagnosis were utilized, RFECV could only achieve a maximum performance of 73% across all metrics. This gap in predictive ability could be due to the sensitivity of the exams or the necessity for a more extensive and balanced data set.

TABLE IV
RFECV FEATURE SUPPORT PERCENTILES

Feature	Overall	Accuracy	Recall	F1 Weighted	Precision
CDRSUM	100.00%	100.00%	100.00%	100.00%	100.00%
HVLT_DR	65.38%	52.50%	52.50%	82.00%	74.50%
HVLT_IR	64.88%	54.00%	54.00%	77.50%	74.00%
MCDR_MN_MN	63.88%	54.00%	54.00%	75.00%	72.50%
CDRGLOB	58.25%	52.50%	52.50%	66.00%	62.00%
MCDR_SM_MN	58.13%	52.00%	52.00%	66.00%	62.50%
FL_WMS_LM_DR	56.00%	52.00%	52.00%	63.50%	56.50%
MEMORY	48.00%	49.00%	49.00%	48.00%	46.00%
LASSI_B_CR2	44.63%	43.50%	43.50%	48.50%	43.00%
LASSI_B_IR1	37.00%	31.50%	31.50%	47.00%	38.00%
FL_WAIS_DIGISYM	33.88%	34.00%	34.00%	38.50%	29.00%
FL_TRLS_B	28.13%	27.00%	27.00%	32.50%	26.00%
FL_TRLS_A	28.00%	25.50%	25.50%	34.00%	27.00%
TRAILB	27.88%	28.00%	28.00%	31.00%	24.50%
TRAILA	26.75%	26.50%	26.50%	30.00%	24.00%
COMMUN	26.13%	32.00%	32.00%	23.50%	17.00%
LASSI_A_IR1	25.38%	22.00%	22.00%	33.00%	24.50%
STRP_CW	24.75%	22.00%	22.00%	31.00%	24.00%
HOMEHOBB	24.25%	31.00%	31.00%	22.00%	13.00%
HVLT_RET	23.25%	22.50%	22.50%	29.00%	19.00%
STRP_W	22.63%	24.00%	24.00%	25.50%	17.00%
UDSVERLC	18.50%	6.50%	6.50%	32.00%	29.00%
LASSI_A_CR1	18.00%	5.50%	5.50%	30.50%	30.50%
FL_COWA_A	17.25%	5.00%	5.00%	28.50%	30.50%
FL_WMS_LM_IR	16.75%	14.00%	14.00%	23.50%	15.50%
MCDR_1_MN	14.25%	15.50%	15.50%	16.00%	10.00%
STRP_INT	14.13%	7.00%	7.00%	23.50%	19.00%
JUDGMENT	13.88%	18.50%	18.50%	12.00%	6.50%
STRP_C	13.25%	9.50%	9.50%	19.50%	14.50%
FL_COWA_F	11.63%	7.00%	7.00%	19.00%	13.50%

Feature input labels are defined by 1FADRC Data Element Dictionaries [15].

B. Future Work

In future feature reduction implementations, a larger and a more balanced data set should be considered as more data is collected via the 1Florida Alzheimer's Disease Research Center. Moreover, additional classification algorithms could include more supervised learning such as support vector machines or gradient boosting [18]. These classifiers can also be paired with filtering methods like principal component analysis (PCA) or independent component analysis (ICA) beforehand to aid in reducing highly correlated or noisy features. A more balanced data set should include more normal controls, more overall subjects, and more completed cognitive exams. Longitudinal analysis using optimal features should also be considered to determine whether or not the optimal feature

sets are sensitive to cognitive changes over time [19]. Another interesting application would be to determine whether or not these optimal feature sets are sensitive to the various etiologies of Alzheimer's Disease.

ACKNOWLEDGMENT

This research is supported by the National Science Foundation under grants: CNS-1920182, CNS-2018611, and CNS-1551221, and with the National Institutes of Health through the P30AG066506 with the 1Florida Alzheimer's Disease Research Center (ADRC).

REFERENCES

[1] "Dementia." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia (Accessed 2023-04-30).

[2] P. Winston Wong, "Economic Burden of Alzheimer Disease and Managed Care Considerations," vol. 26, Aug. 2020, publisher: MJH Life Sciences. [Online]. Available: https://www.ajmc.com/view/economic-burden-of-alzheimer-disease-and-managed-care-considerations (Accessed 2023-04-30).

[3] S. Khan, K. H. Barve, and M. S. Kumar, "Recent Advancements in Pathogenesis, Diagnostics and Treatment of Alzheimer's Disease," *Current Neuropharmacology*, vol. 18, no. 11, pp. 1106–1125, Nov. 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709159/ (Accessed 2023-04-30).

[4] https://www.facebook.com/NIHAging, "Alzheimer's Disease Fact Sheet." [Online]. Available: https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet (Accessed 2023-04-30).

[5] R. Perry Mayrand, C. Y. Freytes, L. O. Sawada, M. Adeyosoye, R. E. Curiel Cid, D. Lowenstein, R. Duara, and M. a. Adjouadi, "Computational Analysis of a Light-Weight SUVr Processing Technique for Neuroimaging Alzheimer's Disease."

[6] "The Challenges of Diagnosis in Alzheimer's Disease," *US Neurology.*, Mar. 2018, section: Uncategorized. [Online]. Available: https://touchneurology.com/alzheimers-disease-dementia/journal-articles/the-challenges-of-diagnosis-in-alzheimers-disease/ (Accessed 2023-06-14).

[7] D. Mungas, C. Shaw, E. Hayes-Larson, C. DeCarli, S. T. Farias, J. Olichney, H. H. Saucedo, P. Gilsanz, M. M. Glymour, R. A. Whitmer, and E. R. Mayeda, "Cognitive impairment in racially/ethnically diverse older adults: Accounting for sources of diagnostic bias," *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring*, vol. 13, no. 1, p. e12265, Dec. 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8719430/ (Accessed 2023-04-30).

[8] R. E. Curiel Cid, D. A. Loewenstein, M. Rosselli, J. A. Matias-Guiu, D. Piña, M. Adjouadi, M. Cabrerizo, R. M. Bauer, A. Chan, S. T. DeKosky, T. Golde, M. T. Greig-Custo, G. Lizarraga, A. Peñate, and R. Duara, "A cognitive stress test for prodromal Alzheimer's disease: Multiethnic generalizability," *Alzheimer's & Dementia (Amsterdam, Netherlands)*, vol. 11, pp. 550–559, Dec. 2019.

[9] D. A. Loewenstein, R. E. Curiel, S. DeKosky, R. M. Bauer, M. Rosselli, S. M. Guinjoan, M. Adjouadi, A. Peñate, W. W. Barker, S. Goenaga, T. Golde, M. T. Greig-Custo, K. S. Hanson, C. Li, G. Lizarraga, M. Marsiske, and R. Duara, "Utilizing semantic intrusions to identify amyloid positivity in mild cognitive impairment," *Neurology*, vol. 91, no. 10, pp. e976–e984, Sep. 2018.

[10] R. M. Chapman, M. Mapstone, A. P. Porsteinsson, M. N. Gardner, J. W. McCrary, E. DeGrush, L. A. Reilly, T. C. Sandoval, and M. D. Guillily, "Diagnosis of Alzheimer's Disease Using Neuropsychological Testing Improved by Multivariate Analyses," *Journal of clinical and experimental neuropsychology*, vol. 32, no. 8, pp. 793–808, Oct. 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896992/ (Accessed 2023-04-30).

[11] F. Thabtah, S. Ong, and D. Peebles, "Examining Cognitive Factors for Alzheimer's Disease Progression Using Computational Intelligence," *Healthcare (Basel, Switzerland)*, vol. 10, no. 10, p. 2045, Oct. 2022.

[12] B. Mwangi, T. S. Tian, and J. C. Soares, "A review of feature reduction techniques in neuroimaging," *Neuroinformatics*, vol. 12, no. 2, pp. 229–244, Apr. 2014.

[13] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 2663–2693, Jun. 2022. [Online]. Available: https://doi.org/10.1007/s40747-021-00637-x (Accessed 2023-04-30).

[14] "Neuroimaging Web Services Interface – Neuroimaging Web Services Interface by CATE Center at FIU." [Online]. Available: https://cate-alz.fiu.edu/ (Accessed 2023-06-14).

[15] "For Scientists," Jul. 2020. [Online]. Available: https://1floridaadrc.org/for-scientists/ (Accessed 2023-04-30).

[16] "scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation." [Online]. Available: https://scikit-learn.org/stable/ (Accessed 2023-04-30).

[17] R. Duara, D. A. Loewenstein, M. Greig, A. Acevedo, E. Potter, J. Appel, A. Raj, J. Schinka, E. Schofield, W. Barker, Y. Wu, and H. Potter, "Reliability and Validity of an Algorithm for the Diagnosis of Normal Cognition, MCI and Dementia: Implications for Multi-Center Research Studies," *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry*, vol. 18, no. 4, pp. 363–370, Apr. 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2844658/ (Accessed 2023-04-30).

[18] M. Shojaie, M. Cabrerizo, S. T. DeKosky, D. E. Vaillancourt, D. Loewenstein, R. Duara, and M. Adjouadi, "A transfer learning approach based on gradient boosting machine for diagnosis of Alzheimer's disease," *Frontiers in Aging Neuroscience*, vol. 14, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnagi.2022.966883 (Accessed 2023-06-15).

[19] S. Tabarestani, M. Aghili, M. Eslami, M. Cabrerizo, A. Barreto, N. Rishe, R. E. Curiel, D. Loewenstein, R. Duara, and M. Adjouadi, "A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study," *NeuroImage*, vol. 206, p. 116317, Feb. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811919309085 (Accessed 2023-06-15).