

## استخدام تعلم الآلة في بناء نموذج للتنبؤ المبكر بمرض الزهايمر

م. جورج رياشي<sup>1\*</sup>، أ.د. رشا مسعود<sup>2</sup>

<sup>1</sup> باحث أكاديمي في درجة الماجستير، قسم الهندسة الطبية، كلية الهندسة الميكانيكية و الكهربائية،  
جامعة دمشق، البريد الإلكتروني (george.riashi@damascusuniversity.edu.sy)

<sup>2</sup> أستاذ في قسم الهندسة الطبية، كلية الهندسة الميكانيكية و الكهربائية، جامعة دمشق،  
(rasha.massoud@damascusuniversity.edu.sy)

### الملخص:

يعد مرض الزهايمر من الأمراض التنكسية العصبية والعامل الشائع للخرف عند كبار السن وانطلاقاً من كون ظهور أعراضه السريرية الظاهرة في وقت متأخر وعدم وجود علاج له فتم العمل في هذه الدراسة على التنبؤ المبكر بمرض الزهايمر من خلال الكشف عن مرحلة ضعف الإدراك المعتدل التي يمر بها أغلب مرضى الزهايمر حيث تم العمل على قاعدة بيانات ADNI فبعد معالجتها وتهيئتها وموازنتها وتقسيمها إلى 80% كمجموعة تدريب و20% كمجموعة اختبار تم ضبط المحددات الفائقة للنموذجين وتدريبهما واختبارهما.

عملت هذه الدراسة على نموذجين من نماذج تعلم الآلة: النموذج الأول K-Nearest Neighbors والنموذج الثاني AdaBoost لتصنيف ثلاث حالات: مرض الزهايمر، ضعف الإدراك المعتدل، والإدراك الطبيعي.

أظهر نموذج AB كفاءة أعلى من نموذج KNN حيث تفوق بالأداء وبلغت قيمة الدقة لنموذج AB نسبة 92.86% وكانت قيمة الحساسية تساوي 92.86%، وهذه نتائج جيدة للمساعدة في دعم قرار الطبيب.

الكلمات المفتاحية: الزهايمر، تعلم الآلة، AdaBoost، K-Nearest Neighbors.

11 تاريخ الايداع

12 تاريخ القبول



15 حقوق النشر: جامعة دمشق –

سورية، يحتفظ المؤلفون بحقوق

النشر بموجب CC BY-NC-SA

18

## Using Machine Learning To Build a Model For Early Prediction Of Alzheimer's Disease 29 30

MSc. George Riashi<sup>1\*</sup>, Prof. Rasha Massoud<sup>2</sup>. 31

<sup>1</sup>Academic Researcher, Master's Degree at the Department of Biomedical 32  
Engineering, Faculty of Mechanical and Electrical Engineering, Damascus 33  
University, george.riashi@damascusuniversity.edu.sy 34

<sup>2</sup>Professor at the Department of Biomedical Engineering, Faculty of Mechanical 35  
and Electrical Engineering, Damascus University, 36  
rasha.massoud@damascusuniversity.edu.sy 37



Received:  
Accepted:

**Abstract:** 40

Alzheimer's disease is a neurodegenerative disease and a common factor of 41  
dementia in the elderly, and based on the fact that the appearance of its apparent 42  
clinical symptoms is late and there is no treatment for it, work was done in this 43  
study on the early prediction of Alzheimer's disease by detecting the stage of mild 44  
cognitive impairment experienced by most Alzheimer's patients. This study was 45  
used the ADNI database. After preprocessing, preparation, balancing and division 46  
into 80% as a training set and 20% as a test set, Two models were Hyperparameters 47  
tuned, trained and tested. 48

Two models of Machine Learning were worked on in this study: the first model 49  
KNN (K-Nearest Neighbors) and the second model AB (AdaBoost) to classify 50  
three cases: Normal Cognition (NL), Alzheimer's Disease (AD) and mild 51  
Cognitive improvement (MCI). 52

The AB model showed higher efficiency than the KNN model, where it 53  
outperformed, the Accuracy value of the AB model was a ratio 92.86%, and the 54  
Sensitivity value was equal to 92.86%, and these are good results to help support 55  
the doctor's decision. 56

Keywords: Alzheimer's Disease, Machine Learning, AdaBoost, K-Nearest 58  
Neighbors. 59

**Copyright:**Damascus  
University- Syria, The  
authors retain the copyright  
under a  
**CC BY- NC-SA**

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

## 61 1. المقدمة:

62 يعد مرض الزهايمر Alzheimer's Disease العامل الشائع  
 63 المسبب للخرف عند كبار السن، وهو من الأمراض التنكسية  
 64 العصبية التي تزداد سوءاً مع الوقت[1]، ويكون تأثيره المباشر  
 65 على الدماغ والإدراك حيث يسبب تدهور السلوك اليومي والذاكرة  
 66 و التفكير، وكما أن نسبة عدد كبار السن الذين أعمارهم فوق  
 67 60 عام في العالم سيتضاعف من 12% إلى 22% وذلك بين  
 68 (2015-2050)، ويشكل الزهايمر تحدياً كبيراً للمجتمع والعائلة  
 69 والأفراد أنفسهم، ويعتقد أن التغيرات الدماغية التي تسبب مشاكل  
 70 في التفكير واللغة والذاكرة واللغة تبدأ قبل 20 عام أو أكثر من  
 71 ظهور الأعراض[1,2].

72 تبدأ أعراض الزهايمر الأولية عادةً بفقدان الذاكرة القصيرة الأجل  
 73 تكرار الأسئلة ونسيان الأشياء اليومية ومع مرور الوقت يزداد  
 74 تدهور الذاكرة وتنمو الصعوبة في أداء المهام البسيطة، وعلى  
 75 الرغم من عدم وجود علاج شاف لهذا المرض إلا أن هناك  
 76 بعض العلاجات والرعاية الطبية المقدمة التي تحسن من جودة  
 77 حياة مرضى الزهايمر، وكما يوجد عدة أدوية قيد التطوير  
 78 لمعالجته[3].

79 تسعى الدراسات إلى إيجاد طرق للكشف المبكر عن الإصابة  
 80 بمرض الزهايمر أو خطر الإصابة به وخصوصاً مع عدم ظهور  
 81 الأعراض إلا بعد الإصابة بسنوات عديدة، ويؤكد الباحثون  
 82 العاملون على إيجاد الأدوية أهمية الكشف المبكر لما في ذلك  
 83 إتاحة للمرضى باختيار المشاركة بالتجارب السريرية، وكونه قد  
 84 يكون مفتاح للعلاج الناجح[4,5].

85 وهناك مراحل لتطور مرض الزهايمر، حيث يمر مرضى  
 86 الزهايمر بمرحلة ضعف الإدراك المعتدل، وهي مرحلة حرجة  
 87 لإمكانية تحويلها إلى الزهايمر بعد عدة سنوات[6].

88 يجب التعامل مع التشخيص في هذه المرحلة بعناية حيث أن  
89 10%-15% من مرضى ضعف الإدراك المعتدل يتحولون إلى  
90 مرضى الزهايمر [7].

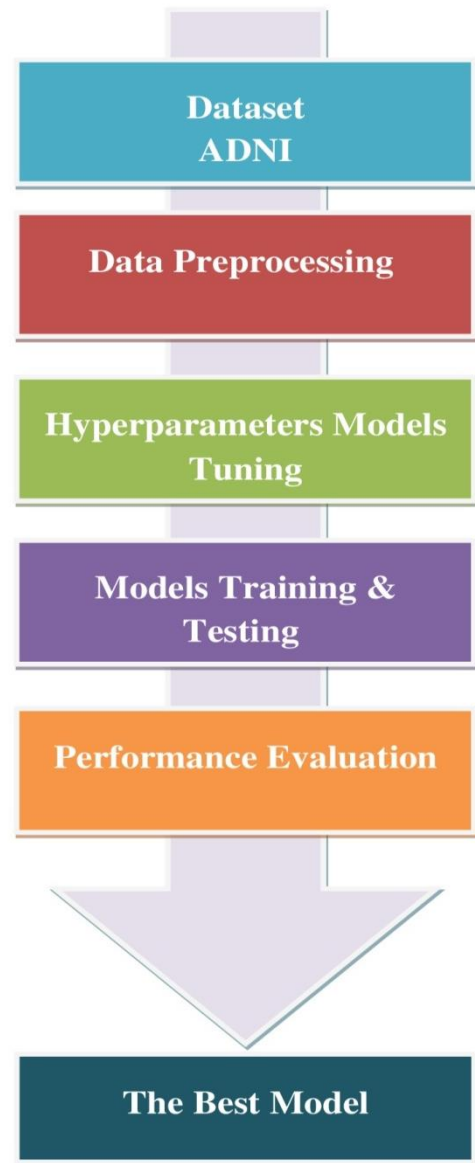
## 91. 2. الدراسات المرجعية:

- 92 تعددت الدراسات التي تعمل على تسخير تقنيات تعلم الآلة  
93 لتطوير منهجيات ذات أداء عالي للوصول إلى التشخيص المبكر  
94 لمرض الزهايمر، فحيث استخدم Ji Hwan وزملاؤه بيانات  
95 خدمة التأمين الصحي الوطنية الكورية NHIS-NSC لاختبار  
96 إمكانية نماذج التعلم الآلة للتنبؤ بحدوث مرض الزهايمر في  
97 المستقبل واكتشافه في وقت مبكر، وذلك بالاعتماد على  
98 المؤشرات الحيوية في السجلات الطبية، وحيث تم استخدام  
99 خوارزميات LR، RF، SVM للتصنيف والتنبؤ بإحدى  
100 الحالات: Probable AD، Definite AD، وأفضل دقة كانت  
101 لخوارزمية RF [8].
- 102 وعمل Muhammad وزملاؤه على تصنيف مراحل الزهايمر،  
103 واستخلاص الخصائص الأكثر تميزاً لكل مرحلة من مرحله،  
104 وذلك انطلاقاً من كون الكشف المبكر عن مراحل مرض الزهايمر  
105 وتصنيفها مفيداً جداً في علاج أعراض المرض، واستخدموا  
106 خوارزميات تعلم الآلة التالية: KNN، DT، Rule Induction،  
107 Naive Bayes، GLM، DL، وذلك باستخدام بيانات من  
108 ADNI، وأظهرت خوارزمية GLM أفضل نتائج بدقة حوالي  
109 88.24% [9].
- 110 وحرص Garam وزملاؤه على التنبؤ والكشف عن الزهايمر في  
111 مرحلة مبكرة قبل ظهور الأعراض باستخدام البيانات ADNI  
112 ونهج multi-modal deep learning حيث تم تدريب عدة  
113 نماذج تعلم عميق على مجموعة خصائص معينة ودمج نتائجها  
114 في نموذج نهائي والذي أعطى الدقة الأعلى من النماذج  
115 المنفصلة وكانت حوالي 82% [10].

## 3. النهج: 116

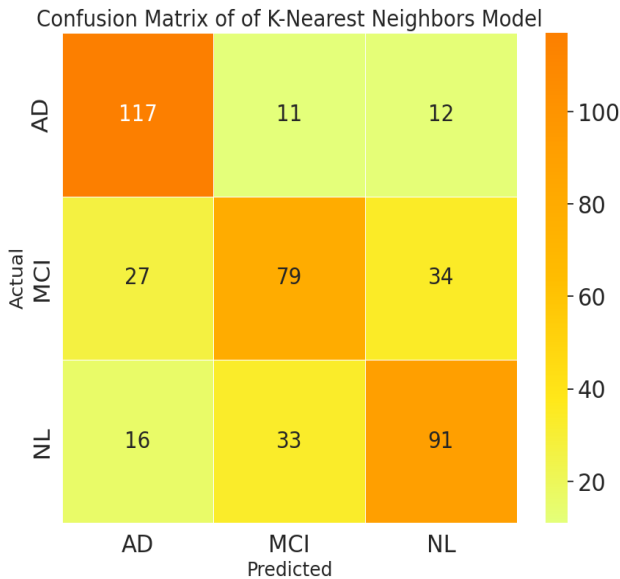
124 حيث أولاً تم العمل على قاعدة البيانات التي تم جمعها بواسطة  
 Alzheimer's Disease Neuroimaging ) ADNI 125  
 126 (Initiative)، وهي مشروع بحثي مبتكر لتحديد علامات حيوية  
 127 ودراسة تطور مرض الزهايمر، وتشمل عشرات المواقع البحثية  
 128 في الولايات المتحدة وكندا، ويمكن زيارة  
 129 [adni.loni.usc.edu] للمزيد من المعلومات حول ADNI.  
 130 ثانياً معالجة البيانات لإعدادها وتهيئتها للمرحلة التالية حيث في  
 131 البداية تم التعامل مع القيم المفقودة وكون كان عددها كبير  
 132 وبيانات طبية فلم يتم تعبئة هذه القيم حرصاً على المحافظة  
 133 عليها كما هي فتم معالجتها بالحذف كقيم في بعض الخصائص  
 134 وحذف الأعمدة التي تحوي أكثر من نصفها قيم مفقودة كاملاً.  
 135 بعد ذلك تم معالجة القيم الشاذة بطريقة الربعيات (الربع الأول  
 136 والربع الثالث)، وأيضاً تحويل الخصائص ذات القيم الاسمية  
 137 إلى قيم رقمية، وبالإضافة إلى موازنة عدد العينات في صفوف  
 138 الخرج الثلاثة لهذه الدراسة (Alzheimer's Disease (AD)،  
 139 Mild Cognitive Impairment (MCI)، وذلك لمنع تحيز  
 140 النموذج للصف ذو العينات الأكبر، في النهاية تم تقسيم البيانات  
 141 التي تم التوصل إليها إلى مجموعة التدريب ومجموعة الاختبار  
 142 بنسبة 80% من البيانات للتدريب و20% من البيانات  
 143 للاختبار.  
 144 ثالثاً تم عملية ضبط المحددات الفائقة Hyperparameters  
 145 لكل من النموذجين اللذين تم استخدامهما في هذه الدراسة وهما:  
 146 النموذج الأول (K-Nearest Neighbors) KNN والنموذج  
 147 الثاني (AdaBoost) AB وذلك باستخدام طريقة البحث الشبكي  
 148 Grid Search مع التحقق المتقاطع Cross Validation  
 149 وذلك من أجل الوصول إلى النموذج الأكثر كفاءة والأفضل  
 150 أداءً، في البداية تم ضبط المحددات الفائقة لنموذج KNN.  
 151 حيث تم ضبط n-neighbors وهو يمثل عدد الجيران الذي  
 152 سيتم أخذهم بعين الاعتبار عند اتخاذ القرار لتصنيف عينة جديدة  
 153 وذلك بحساب بعد العينة الجديدة المراد تصنيفها عنهم، وحيث

117 تم استخدام اللغة البرمجية Python وتم العمل على Google  
 118 Colab والعمل ووفق المخطط الموضح في الشكل (1) في هذه  
 119 الدراسة للوصول إلى أفضل نموذج لتحديد الأشخاص ذوي  
 120 الإدراك الطبيعي ومرضى ضعف الإدراك المعتدل ومرضى  
 121 الزهايمر.



الشكل (1) مخطط العمل

183 عينات اختبار AD حددت بشكل خاطئ (11 حالة على أنها  
 184 MCI و12 حالة على أنها NL) وكذلك 49 حالة من NL  
 185 حددها النموذج خاطئاً (16 حالة على أنها AD و33 حالة على  
 186 أنها MCI) بينما أخطأ النموذج في 61 حالة من عينات اختبار  
 187 MCI (27 حالة على أنها AD و34 حالة على أنها NL)،  
 188 وذلك كون نموذج KNN يعتمد على البعد و MCI المرحلة  
 189 المتوسطة بين AD و NL وتتقارب ببعض قيم خصائصها  
 190 منهما.



الشكل (2) مصفوفة الارتباك لنموذج KNN

193 تم إيجاد كل من الدقة، الحساسية، الإحكام، والدرجة F1 لنموذج  
 194 KNN ككل ويوضح الشكل (3) قيم هذه المعايير، وتم حساب  
 195 قيم هذه المقاييس كنسبة مئوية حيث كانت قيمة الدقة تساوي  
 196 68.33% وأما قيمة الحساسية فكانت 68.33%، وبالنسبة إلى  
 197 قيمة الإحكام فكانت 67.93%، وكانت قيمة درجة F1 تساوي  
 198 67.93%.

154 تم اختيار n-neighbors من ضمن نطاق القيم التالية (1-  
 155 10).

156 وكما تم ضبط Weights وهو يحدد الأوزان المعطاة للجيران  
 157 الأقرب، وقد تم اختيار Weights من بين خيارين: uniform  
 158 الأوزان متساوية، و distance متغيرة بحسب البعد.

159 بالإضافة إلى ضبط P وهو معامل المسافة الذي يحدد الطريقة  
 160 المتبعة لقياس بعد العينة الجديدة المراد اختبارها عن العينات  
 161 الجيران، حيث تم اختيار P من بين خيارين: I والذي يمثل  
 162 Manhattan، و2 والذي يمثل Euclidean.

163 وبعد ذلك تم ضبط المحددات الفائقة لنموذج AB، حيث تم  
 164 ضبط n\_estimators وهو يمثل عدد النماذج الضعيفة  
 165 Weak Learners التي سيتم دمجها للوصول للنموذج النهائي،  
 166 وتم اختيار n\_estimators من ضمن النطاق (10-200).

167 بالإضافة إلى ضبط learning\_rate معدل التعلم وهي تمثل  
 168 معدل تعلم النموذج الذي يحدد مدى تأثير كل نموذج على  
 169 النموذج النهائي، وقد تم اختيار learning\_rate من ضمن  
 170 النطاق (0-1).

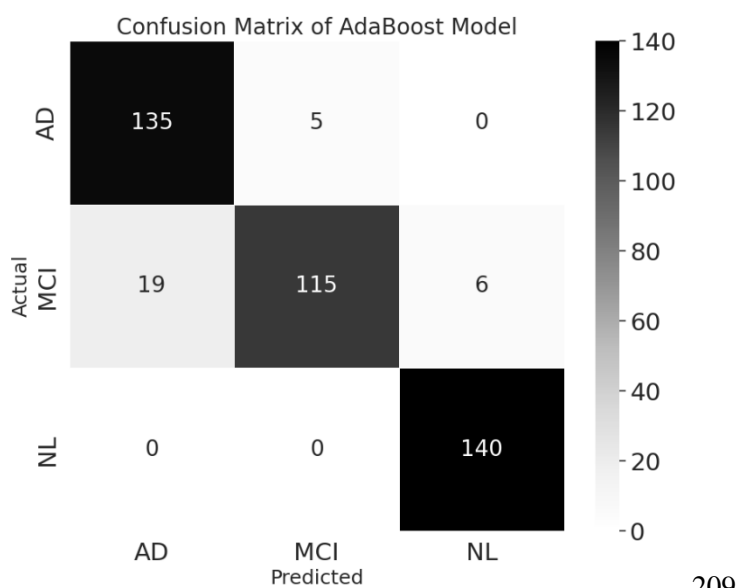
171 رابعاً بعد ذلك تم بناء كل من نموذج KNN ونموذج AB بحسب  
 172 أفضل القيم التي تم اختيارها بواسطة البحث الشبكي والتحقق  
 173 المتقاطع ثم تم تدريب كلا النموذجين واختبارهما.

174 خامساً تم تقييم أداء كل من نموذج KNN ونموذج AB لتحديد  
 175 النموذج الأكثر كفاءة وذلك بإيجاد قيمة كل من الدقة، الحساسية،  
 176 الإحكام، ودرجة F1 لكل من النموذجين.

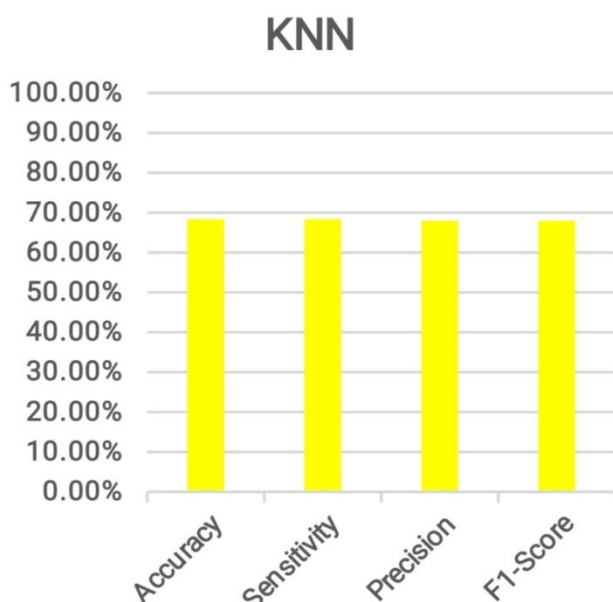
#### 4. النتائج والمناقشة:

178 تم إيجاد كل من مصفوفة الارتباك ومعايير التقييم لكل من نموذج  
 179 KNN ونموذج AB.

180 بداية مصفوفة الارتباك لنموذج KNN الموضحة في الشكل (2)  
 181 تظهر أن النموذج نجح في تصنيف 287 حالة ولكنه أخطأ بعدة  
 182 حالات في كل من الفئات الثلاثة حيث كان هناك 23 حالة من



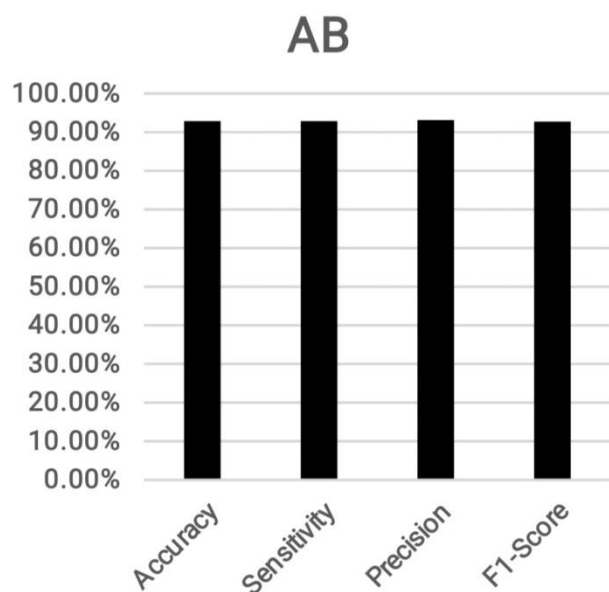
الشكل (4) مصفوفة الارتباك لنموذج AB 210



الشكل (3) معايير تقييم لنموذج KNN 199

211 وكذلك تم إيجاد كل من الدقة، الحساسية، الإحكام، والدرجة F1  
 212 لنموذج AB ككل ويوضح الشكل (5) قيم هذه المعايير، وتم  
 213 حساب قيم هذه المقاييس كنسبة مئوية حيث كانت قيمة الدقة  
 214 تساوي 92.86% وأما قيمة الحساسية فكانت 92.86%،  
 215 وبالنسبة إلى قيمة الإحكام فكانت 93.13%، وكانت قيمة درجة  
 216 F1 تساوي 92.73%.

201 وأما بالنسبة إلى مصفوفة الارتباك لنموذج AB الموضحة في  
 202 الشكل (4) تظهر أن النموذج نجح في تصنيف عدد أكبر من  
 203 الحالات بشكل صحيح (390 حالة) ولم يخطئ بأي حالة من  
 204 عينات اختبار NL ولكنه أخطأ بعدة حالات في كل من AD  
 205 وMCI حيث كان هناك 5 حالة من عينات اختبار AD حددت  
 206 بشكل خاطئ على أنها MCI وكذلك 25 حالة من عينات MCI  
 207 حددها النموذج خاطئاً (19 حالة على أنها AD و6 حالة على  
 208 أنها NL).



الشكل (5) معايير تقييم لنموذج KNN 217

218

236 حيث تم تدريب واختبار نموذجين بعد ضبط المحددات الفائقة  
 237 لكليهما وذلك باستخدام كل من (K-NearestNeighbour)  
 238 و(AdaBoost).  
 239 بلغت قيمة الدقة لنموذج KNN نسبة 68.33% وكانت قيمة  
 240 الحساسية تساوي 68.33% فيما بلغت قيمة الدقة لنموذج AB  
 241 نسبة 92.86% وكانت قيمة الحساسية تساوي 92.86%.  
 242 بالمقارنة بينهما أظهر نموذج AB كفاءة أعلى وتوقفاً بالأداء  
 243 على نموذج KNN وبهذا قدمت الدراسة أداة لمساعدة الطبيب  
 244 في اتخاذ القرار.  
 245 مستقبلاً يمكن العمل على رفع أداء وكفاءة النموذج وتجربة  
 246 تقنيات تعلم الآلة أخرى والمقارنة فيما بينها.

247

## 248 توافر البيانات

249 مجموعة البيانات متاحة على  
 250 [\[https://adni.loni.usc.edu/\]](https://adni.loni.usc.edu/)

251

## 252 شكر وتقدير

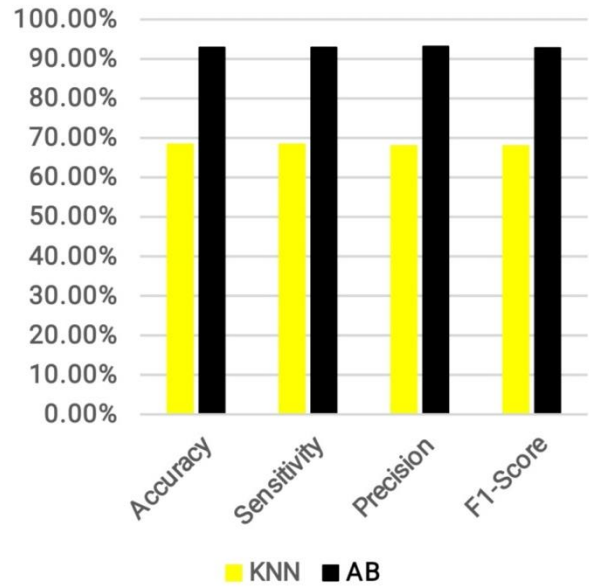
253 الشكر لمبادرة Alzheimer's Disease (ADNI)  
 254 Neuroimaging Initiative.

255

## 256 6. المراجع:

1. Alzheimer's Association. 2023 Alzheimer's 257  
 Disease Facts and Figures. Alzheimers Dement 258  
 2023;19(4). DOI 10.1002/alz.13016. 259

219 بالمقارنة بين النموذج الأول (K-Nearest Neighbors)  
 220 والنموذج الثاني (AdaBoost) كما هو موضح في الشكل (6)،  
 221 تفوق نموذج AB على نموذج KNN ذلك من الممكن كون  
 222 خوارزمية AB تعمل على تدريب نماذج ضعيفة على خطأ  
 223 بعضها للوصول إلى النموذج النهائي.



224 الشكل (6) مقارنة بين نموذج AB ونموذج KNN

224

225

226 تميزت هذه الدراسة عن الدراسات المذكورة في الدراسات  
 227 المرجعية من خلال ضبط المحددات الفائقة للنماذج وموازنة  
 228 عينات التدريب والاختبار، حيث كانت نتائج نموذج KNN  
 229 مساوية لبعض تلك الدراسات وأفضل من بعض الدراسات  
 230 الأخرى، في حين كانت نتائج نموذج AB أفضل من نتائج تلك  
 231 الدراسات.

## 232 5. الاستنتاج والخاتمة:

233 عملت هذه الدراسة على إيجاد نموذج باستخدام تقنيات تعلم الآلة  
 234 لتشخيص كل من الأشخاص ذوي الإدراك الطبيعي ومرضى  
 235 الزهايمر ومرضى ضعف الإدراك المعتدل.



10. Lee, G., Nho, K., Kang, B., Sohn, K. A., & Kim, D. (2019). Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Scientific reports*, 9(1), 1952. 300-303
2. Quiroz, Y. T., Zetterberg, H., Reiman, E. M., Chen, Y., Su, Y., Fox-Fuller, J. T., & Lopera, F. (2020). Plasma neurofilament light chain in the presenilin 1 E280A autosomal dominant Alzheimer's disease kindred: a cross-sectional and longitudinal cohort study. *The Lancet Neurology*, 19(6), 513-521. 260-266
3. Geerts, H., Dacks, P. A., Devanarayan, V., Haas, M., Khachaturian, Z. S., Gordon, M. F., ... & Brain Health Modeling Initiative. (2016). Big data to smart data in Alzheimer's disease: The brain health modeling initiative to foster actionable knowledge. *Alzheimer's & Dementia*, 12(9), 1014-1021. 267-273
4. R. Perneczky (Ed.). (2018). *Biomarkers for Preclinical Alzheimer's Disease*, Vol. 137 of *Neuromethods*, Humana Press, New York, NY. 274-276
5. Vemuri, P., Lesnick, T. G., Przybelski, S. A., Knopman, D. S., Lowe, V. J., Graff-Radford, J., ... & Jack Jr, C. R. (2017). Age, vascular health, and Alzheimer disease biomarkers in an elderly sample. *Annals of neurology*, 82(5), 706-718. 277-281
6. Basheer, S., Bhatia, S., & Sakri, S. B. (2021). Computational modeling of dementia prediction using deep neural network: analysis on OASIS dataset. *IEEE access*, 9, 42449-42462. 282-285
7. Bondi, M. W., Edmonds, E. C., & Salmon, D. P. (2017). Alzheimer's disease: past, present, and future. *Journal of the International Neuropsychological Society*, 23(9-10), 818-831. 287-289
8. Park, J. H., Cho, H. E., Kim, J. H., Wall, M. M., Stern, Y., Lim, H., Yoo, S., Kim, H., & Cha, J. (2020). Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *NPJ digital medicine*, 3(1), 46. 290-295
9. Shahbaz, M., Ali, S., Guergachi, A., Niazi, A., & Umer, A. (2019, July). Classification of Alzheimer's Disease using Machine Learning Techniques. In *Data* (pp. 296-303). 296-299