

1

2 استخدام تعلم الآلة في بناء نموذج للتنبؤ المبكر بمرض الزهايمر

3 م. جورج رياشي^{1*}، أ.د. رشا مسعود²

4 ¹ باحث أكاديمي في درجة الماجستير، قسم الهندسة الطبية، كلية الهندسة الميكانيكية و الكهربائية،

5 جامعة دمشق، البريد الإلكتروني (george.riashi@damascusuniversity.edu.sy)

6 ² أستاذ في قسم الهندسة الطبية، كلية الهندسة الميكانيكية و الكهربائية، جامعة دمشق،

7 (rasha.massoud@damascusuniversity.edu.sy)

8

9

10

الملخص:

11

تاريخ الايداع

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 وتثبيتها وموازنتها وتقسيمها إلى 80% كمجموعة تدريب و20% كمجموعة اختبار تم ضبط المحددات الفائقة للنموذجين وتدريبهما واختبارهما.

12

تاريخ القبول

13



14

حقوق النشر: جامعة دمشق

–سورية، يحتفظ المؤلفون

بحقوق النشر بموجب CC BY-NC-SA

18

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.

19

20

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.

23

24

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

25

26

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

MSc. George Riashi^{1*}, Prof. Rasha Massoud². 31

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

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



Received:
Accepted:

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

Abstract: 40

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

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

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

Keywords: Alzheimer's Disease, Machine Learning, AdaBoost, K-Nearest Neighbors. 57-59

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

1. المقدمة:

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

2. الدراسات المرجعية:

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

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
66 والذاكرة و التفكير، وكما أن نسبة عدد كبار السن الذين
67 أعمارهم فوق 60 عام في العالم سيتضاعف من 12% إلى
68 22% وذلك بين (2015-2050)، ويشكل الزهايمر تحدياً
69 كبيراً للمجتمع والعائلة والأفراد أنفسهم، ويعتقد أن التغيرات
70 الدماغية التي تسبب مشاكل في التفكير واللغة والذاكرة واللغة
71 تبدأ قبل 20 عام أو أكثر من 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 وتتمو 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].

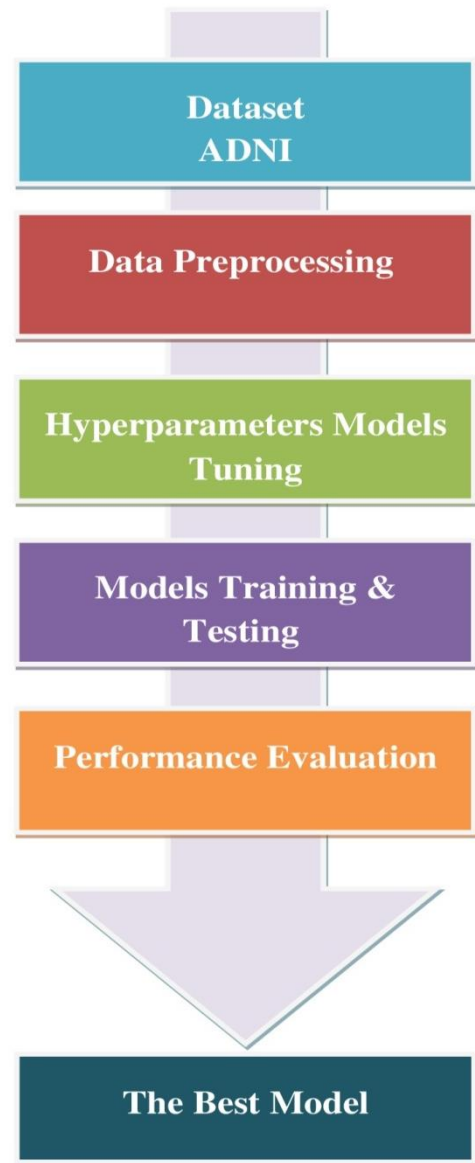
79 تسعى الدراسات إلى إيجاد طرق للكشف المبكر عن الإصابة
80 بمرض الزهايمر أو خطر الإصابة به وخصوصاً مع عدم
81 ظهور الأعراض إلا بعد الإصابة بسنوات عديدة، ويؤكد
82 الباحثون العاملون على إيجاد الأدوية أهمية الكشف المبكر لما
83 في ذلك إتاحة للمرضى باختيار المشاركة بالتجارب السريرية،
84 وكونه قد يكون مفتاح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

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

116 3. النهج:

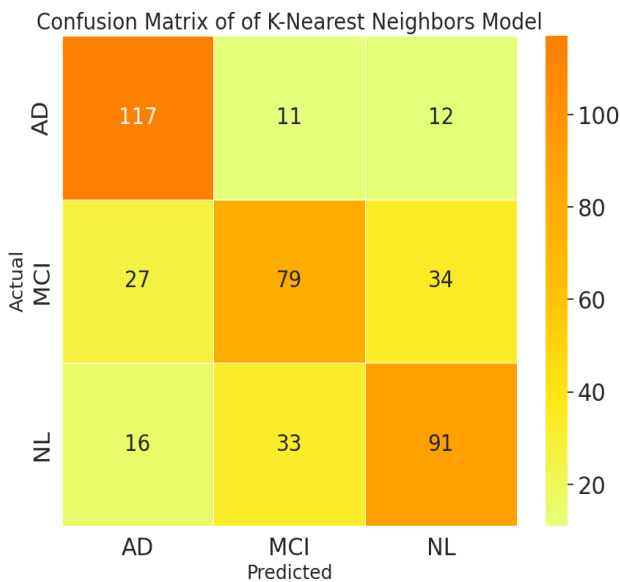
124 حيث أولاً تم العمل على قاعدة البيانات التي تم جمعها بواسطة
 125 Alzheimer's Disease Neuroimaging) ADNI (Initiative)، وهي مشروع بحثي مبتكر لتحديد علامات حيوية
 126 ودراسة تطور مرض الزهايمر، وتشمل عشرات المواقع البحثية
 127 في الولايات المتحدة وكندا، ويمكن زيارة
 128 [adni.loni.usc.edu] للمزيد من المعلومات حول ADNI.
 129
 130 ثانياً معالجة البيانات لإعدادها وتهيئتها للمرحلة التالية حيث
 131 في 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
 134 الخصائص وحذف الأعمدة التي تحوي أكثر من نصفها قيم
 135 مفقودة كاملاً.
 136 بعد ذلك تم معالجة القيم الشاذة بطريقة الربيعات (الربيع الأول
 137 والرابع الثالث)، وأيضاً تحويل الخصائص ذات القيم الاسمية
 138 إلى قيم رقمية، وبالإضافة إلى موازنة عدد العينات في صفوف
 139 الخرج الثلاثة لهذه الدراسة (Alzheimer's Disease (AD)،
 140 Mild Cognitive Impairment (MCI)، وذلك لمنع تحيز
 141 النموذج للصف ذو العينات الأكبر، في النهاية تم تقسيم
 142 البيانات التي تم التوصل إليها إلى مجموعة التدريب ومجموعة
 143 الاختبار بنسبة 80% من البيانات للتدريب و20% من
 144 البيانات للاختبار.
 145 ثالثاً تم عملية ضبط المحددات الفائقة Hyperparameters
 146 لكل من النموذجين اللذين تم استخدامهما في هذه الدراسة
 147 وهما: النموذج الأول KNN (K-Nearest Neighbors)
 148 والنموذج الثاني AB (AdaBoost) وذلك باستخدام طريقة
 149 البحث الشبكي Grid Search مع التحقق المتقاطع Cross
 150 Validation وذلك من أجل الوصول إلى النموذج الأكثر
 151 كفاءة والأفضل أداء، في البداية تم ضبط المحددات الفائقة
 152 لنموذج KNN.

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 بداية مصفوفة الارتباك لنموذج KNN الموضحة في
 184 الشكل(2) تظهر أن النموذج نجح في تصنيف 287 حالة
 185 ولكنه أخطأ بعدة حالات في كل من الفئات الثلاثة حيث كان
 186 هناك 23 حالة من عينات اختبار AD حددت بشكل خاطئ
 187 (11 حالة على أنها MCI و12 حالة على أنها NL) وكذلك
 188 49 حالة من NL حددها النموذج خاطئاً (16 حالة على أنها
 189 AD و33 حالة على أنها MCI) بينما أخطأ النموذج في 61
 190 حالة من عينات اختبار MCI (27 حالة على أنها AD و34
 191 حالة على أنها NL)، وذلك كون نموذج KNN يعتمد على
 192 البعد و MCI المرحلة المتوسطة بين AD و NL وتتقارب
 193 ببعض قيم خصائصها منهما.



الشكل(2) مصفوفة الارتباك لنموذج KNN

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

153 حيث تم ضبط n-neighbors وهو يمثل عدد الجيران الذي
 154 سيتم أخذهم بعين الاعتبار عند اتخاذ القرار لتصنيف عينة
 155 جديدة وذلك بحساب بعد العينة الجديدة المراد تصنيفها عنهم،
 156 وحيث تم اختيار n-neighbors من ضمن نطاق القيم التالية
 157 (10 - 1).

158 وكما تم ضبط Weights وهو يحدد الأوزان المعطاة للجيران
 159 الأقرب، وقد تم اختيار Weights من بين خيارين: uniform
 160 الأوزان متساوية، و distance متغيرة بحسب البعد.
 161 بالإضافة إلى ضبط P وهو معامل المسافة الذي يحدد الطريقة
 162 المتبعة لقياس بعد العينة الجديدة المراد اختبارها عن العينات
 163 الجيران، حيث تم اختيار P من بين خيارين: 1 والذي يمثل
 164 Manhattan، و2 والذي يمثل Euclidean.

165 وبعد ذلك تم ضبط المحددات الفائقة لنموذج AB، حيث تم
 166 ضبط n_estimators وهو يمثل عدد النماذج الضعيفة
 167 Weak Learners التي سيتم دمجها للوصول للنموذج
 168 النهائي، وتم اختيار n_estimators من ضمن النطاق
 169 (10 - 200).

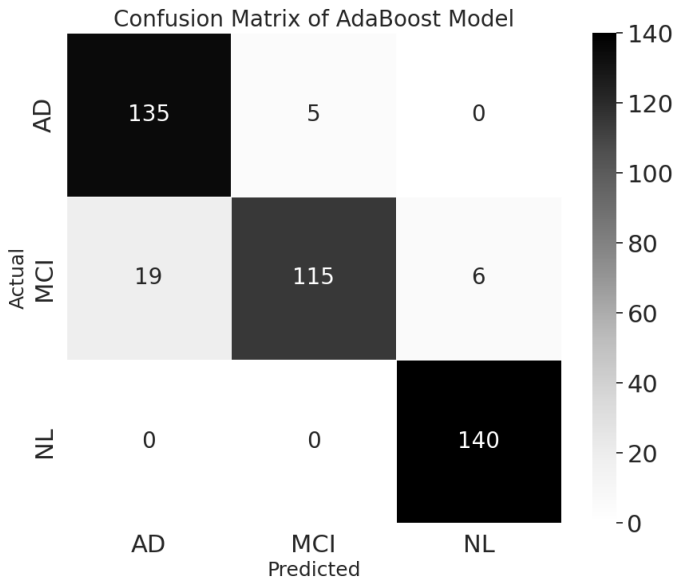
170 بالإضافة إلى ضبط learning_rate معدل التعلم وهي تمثل
 171 معدل تعلم النموذج الذي يحدد مدى تأثير كل نموذج على
 172 النموذج النهائي، وقد تم اختيار learning_rate من ضمن
 173 النطاق (0 - 1).

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

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

4. النتائج والمناقشة:

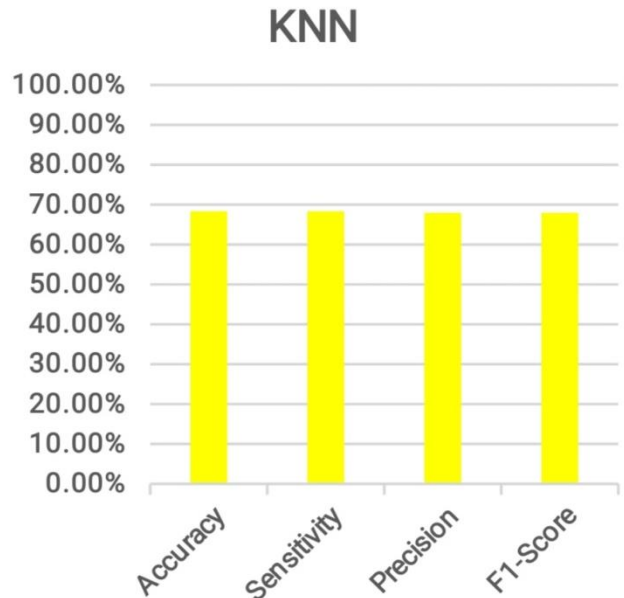
181 تم إيجاد كل من مصفوفة الارتباك ومعايير التقييم لكل من
 182 نموذج KNN ونموذج AB.



الشكل (4) مصفوفة الارتباك لنموذج AB

212

213



الشكل (3) معايير تقييم نموذج KNN

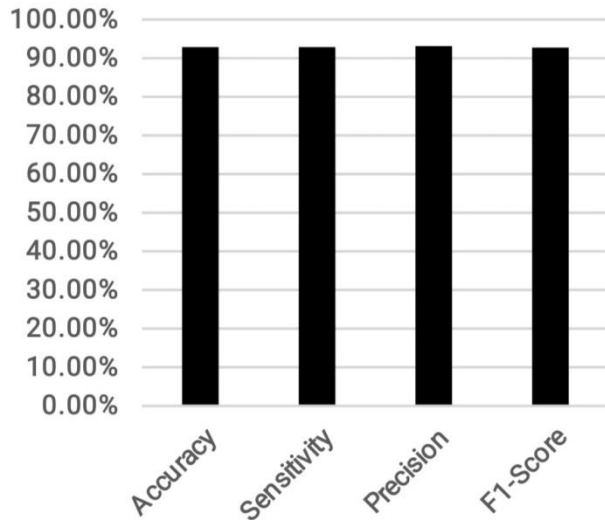
202

203

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

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

AB



الشكل (5) معايير تقييم نموذج KNN

220

221

239 حيث تم تدريب واختبار نموذجين بعد ضبط المحددات الفائقة
 240 لكليهما وذلك باستخدام كل من (K-NearestNeighbour)
 241 و(AdaBoost).
 242 بلغت قيمة الدقة لنموذج KNN نسبة 68.33% وكانت قيمة
 243 الحساسية تساوي 68.33% فيما بلغت قيمة الدقة لنموذج AB
 244 نسبة 92.86% وكانت قيمة الحساسية تساوي 92.86%.
 245 بالمقارنة بينهما أظهر نموذج AB كفاءة أعلى وتوقفاً بالأداء
 246 على نموذج KNN وبهذا قدمت الدراسة أداة لمساعدة الطبيب
 247 في اتخاذ القرار.
 248 مستقبلاً يمكن العمل على رفع أداء وكفاءة النموذج وتجربة
 249 تقنيات تعلم الآلة أخرى والمقارنة فيما بينها.
 250

251 توافر البيانات

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

254

255 شكر وتقدير

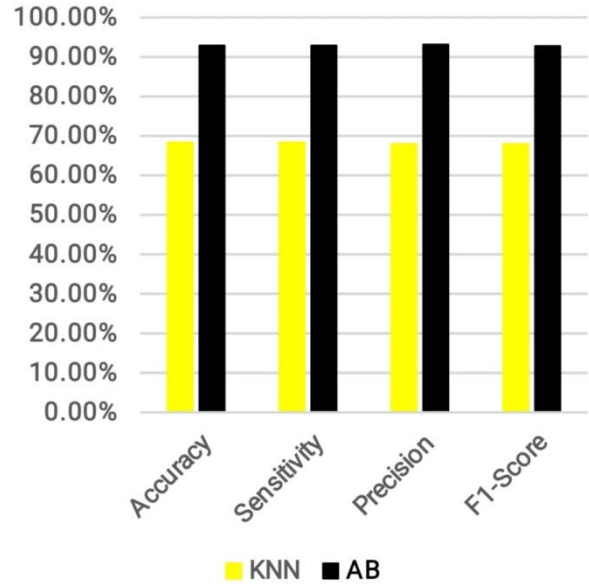
256 الشكر لمبادرة Alzheimer's Disease (ADNI)
 257 Neuroimaging Initiative.

258

259 6. المراجع:

1. Alzheimer's Association. 2023 Alzheimer's 260
 Disease Facts and Figures. Alzheimers Dement 261
 2023;19(4). DOI 10.1002/alz.13016. 262

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



227 الشكل (6) مقارنة بين نموذج AB ونموذج KNN

228

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

235 5. الاستنتاج والخاتمة:

236 عملت هذه الدراسة على إيجاد نموذج باستخدام تقنيات تعلم
 237 الآلة لتشخيص كل من الأشخاص ذوي الإدراك الطبيعي
 238 ومرضى 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. 304 305 306 307
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. 263 264 265 266 267 268 269
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. 270 271 272 273 274 275 276
4. R. Perneczky (Ed.). (2018). *Biomarkers for Preclinical Alzheimer's Disease*, Vol. 137 of *Neuromethods*, Humana Press, New York, NY. 277 278 279
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. 280 281 282 283 284
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. 285 286 287 288
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. 289 290 291 292 293
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. 294 295 296 297 298 299
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). 300 301 302 303