جامعة دمشق
كلية الهندسة الميكانيكية والكهربائية
قسم الهندسة الطبية

نظام تشخيص آلي لمرض السكري باستخدام خوارزميات التصنيف.

An Automated Diagnosis System for Diabetes using Classification Algorithms

إعداد: م. منى صالح

إشراف: أ.د.م. رشا مسعود

الملخص

يعتبر مرض السكري من الأمراض المزمنة التي تؤدي إلى ارتفاع مستوى السكر في الدم، وعدم القدرة للسيطرة على هذه المستويات تؤدي إلى أضرار في العديد من أجهزة الجسم ولا سيما القلب والأوعية الدموية والأعصاب، وانطلاقا من أهمية الكشف المبكر عن هذا المرض كمحاولة للتقليل من مخاطره أو تأخير الإصابة به، وتحسين الحالة الصحية العامة للناس، والتخطيط لحياتهم مستقبلا، عمل هذا البحث على استخدام خوارزميتي الانحدار اللوجيستي والغابة العشوائية لبناء نماذج التصنيف وذلك باستخدام قاعدة البيانات الهندية (PID)، وموازنة عدد العينات باستخدام تقنيات التوزين: الزيادة العشوائية للبيانات(ROS)، الانقاص العشوائي للبيانات(RUS)، توليد بيانات اصطناعية جديدة(SMOTE).كما تم بناء قاعدة بيانات محلية اعتمادا على نتائج الدراسة المقطعية، وتم استخلاص سمات التنبؤ بمرض السكري من النوع الثاني من خلال تطبيق الاختبارات الإحصائية، ثم تم تحديد أفضل سمات للتنبؤ بمرض السكري باستخدام خوارزمية تقليل السمات العودية(RFE) وتدريب واختبار نماذج التصنيف بعد ضبط معاملاتها الفائقة، ثم تم تقييم أداء هذه النماذج للحصول على النموذج الأفضل.

تفوق نموذج الغابة العشوائية على الانحدار اللوجيستي عند التنبؤ باستخدام قاعدة البيانات PIDعند عدم موازنة البيانات، كما كان أداء نموذج الغابة العشوائية على البيانات المحلية أفضل.

القسم العملي

تم العمل وفق المنهجية الموضحة في الشكل(1):

الشكل(1) منهجية البحث

النتائج والمناقشة

الجدول (1) أداء نماذج التصنيف باستخدام قاعدتي البيانات المحلية والهندية

Data set	Classification Algorithms	Accuracy	F1_score
PID	LR	86%	82%
PID	LR +ROS	84%	81%
PID	LR+SMOTE	84%	81%
PID	LR+RUS	81%	79%
PID	RF	87%	84%
PID	RF +ROS	85%	82%
PID	RF+SMOTE	82%	79%
PID	RF+RUS	82%	81%
Syrian	RF+RFE	97%	97%

الجدول (2) مقارنة أفضل نموذج مع الدراسات المرجعية

Ref	Data set	Classification Algorithms	Accuracy	F1
Study[3]	PID	LR	83.1%	72.3%
Study[4]	PID	RF+SMOTE	80.7%	_
Study[5]	PID	RF +ROS	85.53%	83.06%
Study[6]	NHANCES	RF +MLR	94.15%	96.3%
Study[7]	Vanderbilt	LR+ **univariate FS**	89%	80%
Study[8]	Bangladesh	RF+**IG**	80.2%	68.7 %
Our study	PID	**RF**	**87%**	**84%**
	PID	**RF +ROS**	**85%**	**82%**
	Syrian dataset	**RF +RFE**	**97%**	**97%**

القسم النظري

مرض السكري مرض مزمن يحدث عندما يعجز البنكرياس عن إنتاج الأنسولين بكمية كافية، أو أن الكمية المفرزة منه كبيرة، ولكن هناك مقاومة من قبل أنسجة وخلايا الجسم تعوق وظيفة الأنسولين، وفي كلتا الحالتين يكون الغلوكوز غير قادر على دخول الخلايا، مما يؤدي إلى تراكمه في الدم وإمكانية ظهوره في البول، مع ازدياد تراكم السكر في الدم بدلا من دخوله خلايا الجسم، قد يؤدي إلى حدوث أضرار وخيمة على العديد من أجهزة الجسم، ولاسيما الأوعية الدموية الدقيقة في شبكة العين. [1]

تعلم الآلة: هو أحد فروع الذكاء الاصطناعي الذي يركز على تطوير الأنظمة التي تمكن أجهزة الحاسوب من التعلم والتكيف وتحسين أدائها من خلال التجربة، ويعتمد على البيانات حيث يتم استخراج المعلومات الهامة والتعلم منها مما يعطي الحاسوب قدرة على اتخاذ القرارات والتنبؤ بنتائج مستقبلية، ويوجد عدة تعاريف لتعلم الآلة، حيث يمكن تعريفه بأنه مجال الدراسة الذي يمنح الحواسيب القدرة على التعلم دون برمجتها صراحة. [2]

النتائج والمناقشة

تفوق نموذج الغابة العشوائية على الانحدار اللوجيستي عند التنبؤ بمرض السكري باستخدام قاعدة البيانات PIDعند عدم موازنة البيانات، حيث كانت الدقة 87%، ودرجة F1تساوي 84%، أما بالنسبة للمساحة تحت منحني الخواص فقد بلغت 0.81.

كان أداء نموذج الغابة العشوائية على البيانات المحلية أفضل، حيث حقق أفضل معايير أداء عند أفضل ست وعشر سمات للتنبؤ بمرض السكري، حيث كانت الدقة 97%، ودرجة F1تساوي97%، بينما المساحة تحت منحني الخواص كانت تساوي0.96.

المراجع

1. عادل، بورغدة ورضا، شريدي. (2021). "أثر النشاط البدني الرياضي المكيف في استقرار نسبة السكر في الدم ووظائف الجسم لدى مرضي السكري النوع الثاني". رسالة ماجستير، جامعة محمد بوضياف، معهد علوم وتقنيات النشاط البدني والرياضي، الجزائر

2. حداد، فاطمة الزهرة. (2023)." دراسة إحصائية لعوامل خطر الإصابة بداء السكري من النوع الثاني" مجلة روافد، المجلد(7)، العدد(2)

3- Alzain, M.A., Alzain, H.M., (2024).Using Machine Learning Techniques to Predict Type 2 Diabetes. *Scientific Journal of Faculty of Education*. Misurata University Libya.

4- Mushtaq, Z., Ali ،S., et al, (2022).Voting Classification-Based Diabetes Mellitus Prediction Using Hyper tuned Machine-Learning Techniques". *Hindawi Mobile Information Systems*. Article ID6521532, https://doi.org/10.1155/2022/6521532

5- Talukder, Md, Al., Islam, Md. M., et al, (2024).Toward reliable diabetes prediction: Innovations in data engineering and machine learning applications. *DIGITAL HEALTH*. Volume 10.

6- Maniruzzaman, Md., Ahammed, B., et al., (2020).classification and prediction of diabetes disease using machine learning paradigm. *Health Information and System*.

7- Viswanatha, V., Ramachandra, A., et al. (2023). Diabetes Prediction Using Machine Learning Approach. Vol. 10 .IS 8. http://stradresearch.org

8- Kakoly, I.j., Hoque, Md., Hassan, Naj.,(2023). Data-Driven Diabetes Risk Factor Prediction Using Machine Learning Algorithms with Feature Selection Technique