ملخص رسالة ماجستير بعنوان

نظام تشخيص آلي لمرض السكري باستخدام خوارزميات التصنيف.

اسم الطالب

م. منى صالح

المشرف

أ.د.م رشا مسعود

القسم والاختصاص

الهندسة الطبية

الملخـــــص

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 بمرض السكري باستخدام قاعدة البيانات PID عند 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.

Master's thesis summary entitled

An Automated Diagnosis System for Diabetes using Classification Algorithms

Student Name

Eng. Mona Saleh

Supervisor

Prof. Eng. Rasha Massoud

Department

Biomedical Engineering

Summary

Diabetes mellitus is a chronic disease that elevates blood glucose levels and, if uncontrolled, can damage various body systems, particularly the heart, blood vessels, and nerves. Early diagnosis is challenging due to overlapping symptoms with other diseases.

This research focuses on building predictive models for early detection of type 2 diabetes using machine learning algorithms: Random Forest and Logistic Regression.

Pima Indian Diabetes dataset (PID) was utilized and preprocessed to address missing and outlier values, normalize features, and balance class distributions using: random over sampler, random under sampler, and SMOTE techniques. Hyper parameters were tuned using grid search and cross-validation, and models were trained, tested, and evaluated to determine the optimal one.

Additionally, a local dataset was built by the result of cross sectional study. Feature selection was conducted using statistical tests (Man-Whitney, T-test, Chi-Square) and Recursive Feature Elimination (RFE).The model was retrained on this dataset and evaluated.

Results showed that the Random Forest model outperformed Logistic Regression on the PID dataset (accuracy: 87%, F1-score: 84%, AUC: 0.81). On the local dataset, Random Forest achieved even better performance (accuracy: 97%, F1-score: 97%, AUC: 0.96).