نظام تصنيف الأمراض باستخدام خوارزميات التنقيب عن البيانات وتعلم الآلة

مقدمة من

م/ امنية حسنى محمد السيد

تحت إشراف

أ.د./ رانيا احمد ابو السعود

استاذ الاشارات الرقمية بقسم الهندسة الكهربية

كلية الهندسة – جامعة الفيوم

د./ اسلام عيد على محمد المغربى

مدرس بقسم نظم المعلومات

كلية الحاسبات والمعلومات – جامعة الفيوم

فبراير 2024

ملخص الرسالة

يعد استخراج البيانات الطبية ميدانا متناميا يهدف إلى توفير معلومات طبية موثوق بها وقائمة على االدلة لألطباء والباحثين. ويمثل التصنيف وتجهيز البيانات المسبقة استراتيجيتين رئيسيتين في هذا الميدان، مع تحسين النموذج من حيث النوعية والموثوقية. ويؤثر مرض السكري، وهو مرض استقالبي واسع االنتشار، على أكثر من 422 مليون شخص على الصعيد العالمي، وال سيما في البلدان المنخفضة الدخل أو المتوسطة الدخل، مما يتسبب في وفاة نحو 1.5 مليون شخص سنويا. وتهدف هذه الدراسة إلى وضع نموذج تصنيفي يحدد بدقة مرض السكري في المرضى الذين يستخدمون بيانات تشخيصية. وتقدم الدراسة البحثية إطارا جديدا يجمع بين التعلم العميق ونهج اختيار السمات المعيارية لتحديد ما إذا كان االفراد مصابين بالسكري من النوع 2. وتُقيِّم الدراسة فعالية أساليب الفرز الذاتي العميق واالختيار التقليدي للمميزات، مما يؤدي إلى االداء االمثل. وتستخدم الدراسة مجموعة بيانات پيما للسكري الهندي ألغراض االختبار
المعياري. يتم تقييم فعالية المنهجية الهجينة من خالل مقارنة تأثيرها على مختلف خوارزميات التصنيف، مثل الشبكات العصبية، XGBoost، نايفي بايس، K-القرب الجديد جار، والتكديس. ويُجرى تحليل إحصائي لسمات المدخالت لتقييم مدى تقلبها وأهميتها. وتُطبَّق تفسيرات المضاف المشابه إلبراز أهم الميزات التي يتم الحصول عليها من خالل نهج اختيار خاصية معامل ارتباط البيرسونر. وحققت التقنية الهجينة معدل دقة قدره 83.7 في المائة، مؤكدة أهمية هذه السمات في تحديد النتيجة النهائية. وتبلغ قياسات دقة النظام المقترح ومقاييس استرجاعه 83.9 في المائة و83.7 في المائة على التوالي. وتُظهر دقة النموذج العالية ومعدالته العالية للتذكير قدرته
على الحد بفعالية من اإليجابيات الكاذبة والسلبيات الزائفة، بما يكفل التشخيص الدقيق والعالج المناسب للمرضى.