



جامعة افريقيا العالمية
عمادة الدراسات العليا و البحث العلمي و النشر
كلية إقرا لدراسات الحاسوب
قسم تقانة المعلومات

كشف مرض اعتلال الشبكية لمرضى السكري باستخدام
خوارزمية نظام الدعم الالي

Detecting diabetic retinopathy using Support Vector Machines

بحث تكميلي لنيل درجة الماجستير في تقنية المعلومات

اشراف :

ا.د سعد صبير

اعداد الطالب :

محمد عبد المنعم الخير سعيد

1441هـ – 2020م

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

الاستهلال

﴿ قَالَ الَّذِي عِنْدَهُ عِلْمٌ مِنَ الْكِتَابِ أَنَا آتِيكَ بِهِ قَبْلَ أَنْ يَرْتَدَّ إِلَيْكَ طَرْفُكَ فَلَمَّا رآهُ مُسْتَقِرًّا عِنْدَهُ قَالَ هَذَا مِنْ فَضْلِ رَبِّي لِيَبْلُوَنِي أَأَشْكُرُ أَمْ أَكْفُرُ وَمَنْ شَكَرَ فَإِنَّمَا يَشْكُرُ لِنَفْسِهِ وَمَنْ كَفَرَ فَإِنَّ رَبِّي غَنِيٌّ كَرِيمٌ ﴾

النمل [40]

الاهداء

الي القلب الكبير

والذي العزيز

الي القلب الناصع بالبياض

والدتي الحبيبة

الي الانسانة الغالية

اختي الحبيبة

الي جمال الحياة

اخوتي

الي بنات خالاتي لم يقصروا في تشجيعي و تحفيزي

"هند و اسلام"

الي صديقي العزيز الذي كان يقف الي جانبي دائما

صديقي العزيز "منذر عمر "

شكر و تقدير

أول مشكور هو الله عز و جل ,ثم والداي علي كل مجهوداتهم منذ ولادتي الي هذه اللحظات,أنتم كل شي احبكم في الله أشد الحب.يسرني أن أوجه شكري لكل من نصحني أو ارشدني أو وجهني أو ساهم معي في اعداد هذا البحث بايصالي للمراجع و المصادر المطلوبة في أي مرحلة من مراحل و اشكر علي وجه الخصوص استاذي الفاضل البروف سعد صبير علي مساندي و ارشادي بالنصح و التصحيح و علي اختيار العنوان و الموضوع كما ان شكري موجه لادارة كلية الحاسوب بجامعة افريقيا العالمية

المستخلص

اعتلال الشبكية السكري هو مرض وعائي في الشبكية يصيب مرضى السكري. هو السبب الأول للعمى لدى الأشخاص الذين تتراوح أعمارهم بين 20 و 64 عامًا . مرض السكري شائع للغاية ، لذلك ليس من المستغرب أن يصيب % 3.4 من السكان . الهدف من هذا البحث هو تصنيف المرضى الي اولئك الذين يعانون من اعتلال الشبكية السكري و الذين ليس لديهم المرض ، وقد استخدمنا مجموعة البيانات التي تحتوي على ميزات مستخرجة من قاعدة بيانات صور ميسيدور (messidor) . تمثل جميع الميزات إما آفة مكتشفة ، أو ميزة وصفية لجزء تشريحي أو وصف على مستوى الصورة. في هذا البحث ، تم استخدام مجموعة بيانات من قاعدة بيانات تسمى ميسيدور (messidor) و تم اسخدام خوارزمية نظام الدعم الالي (Support Vector Machines) مع الكيرنل الخطي و تمت التجربة علي اداتين مختلفتين و هما البايثون python و R programming و تم الحصول علي نتائج مختلفة حيث حققت خوارزمية نظام الدعم الالي (SVM) التي ادرجت عن طريق R علي دقة (Accuracy) %74.78 و النسبة المئوية للسجلات المصنفة حسب المصنف على أنها إيجابية وإيجابية بالفعل يسمى الضبط (precision) علي %86.66 و النسبة المئوية للسجلات الإيجابية المصنفة حسب المصنف على أنها إيجابية يسمى الاستدعاء (recall) علي %71.23 . عند استخدام خوارزمية نظام الدعم الالي (SVM) التي ادرجت عن طريق python تم الحصول علي دقة (Accuracy) %78.44 و النسبة المئوية للسجلات المصنفة حسب المصنف على أنها إيجابية وإيجابية بالفعل يسمى الضبط (precision) علي %83.33 و النسبة المئوية للسجلات الإيجابية المصنفة حسب المصنف على أنها إيجابية يسمى الاستدعاء (recall) علي %66.03. لسوء الحظ لم نحقق الدقة المطلوبة بسبب قلة البيانات و عدم معالجتها بالشكل الصحيح لذلك نوصي باستخدام بيانات اكثر و اسخدام تقنيات معالجة افضل في المستقبل.

Abstract

Diabetic retinopathy (DR) is a vascular disease of the retina which affects patients with diabetes mellitus. It is the number one cause of blindness in people between the ages of 20–64. Diabetes mellitus is extremely common, so it is not surprising that DR affects 3.4 percent of the population. The goal of this research to classify patients having diabetic retinopathy or not having, we used dataset contains features extracted from the Messidor image database. All features represent either a detected lesion, a descriptive feature of an anatomical part or an image–level descriptor. In this research we used datasets from messidor database and SVM (Support Vector Machines) algorithm was used. The experiment was done on two different tools ,namely R programming and python and different results were obtained , as SVM algorithm was included by R programming achieved an accuracy of 74.78% with precision 86.66% and recall 71.23% ,when using SVM algorithm was included by python achieved accuracy was obtained 78.44% with precision 83.33% and recall 66.03%.unfortunately ,we did not achieve accuracy due to lack of data and incorrect handling of it ,therefore ,we recommend that you use more data and better processing technologies in the future.

قائمة المحتويات

| رقم الصفحة | الموضوع | |
|------------|----------------------------|-----|
| أ | الاية | |
| ب | الاهداء | |
| ت | شكر و تقدير | |
| ث | ملخص البحث | |
| ج | Abstract | |
| | الفصل الاول : الاطار العام | |
| 1 | مقدمة | 1. |
| 1 | معلومات اساسية | 1.1 |
| 2 | مشكلة البحث | 2.1 |
| 2 | اهمية البحث | 3.1 |
| 2 | اهداف البحث | 4.1 |
| 2 | منهجية البحث | 5.1 |
| 3 | نطاق البحث | 6.1 |
| 3 | مصطلحات البحث | 7.1 |

| | | |
|----------------------------|------------------------------------|-------|
| 4 | الدراسات السابقة | 8.1 |
| 10 | هيكل البحث | 9.1 |
| الفصل الثاني الاطار النظري | | |
| 12 | اعتلال الشبكية السكري | 1.2 |
| 14 | التصنيف | 1.2.1 |
| 14 | اعتلال الشبكية السكري غير التكاثري | 1.2.2 |
| 16 | اعتلال الشبكية السكري التكاثري | 1.2.3 |
| 17 | الاعراض | 1.2.4 |
| 17 | الشاشة | 1.2.5 |
| 18 | الفحص | 1.2.6 |
| 18 | خيارات العلاج | 1.2.7 |
| 19 | مقدمة عن نظام الدعم الالي | 1.3 |
| 20 | التمهيدية: المصنف الخطي | 1.3.1 |
| 21 | تصنيف الهامش الكبير | 1.3.2 |
| 22 | مقدمة عن R | 1.4 |
| 22 | مقدمة عن لغة python | 1.5 |
| الفصل الثالث :التحليل | | |
| 25 | مقدمة | .6 |
| 25 | التحليل | 1.6 |
| 25 | لبيانات | 2.6 |

| | | |
|------------------------------------|---|------|
| 25 | الأدوات | 3.6 |
| 25 | وصف البيانات | 4.6 |
| 26 | معلومات مجموعة البيانات | 5.6 |
| الفصل الرابع: الاطار العملي | | |
| 30 | التنفيذ و الاختبار | .7 |
| 30 | فتح أداة R studio | 1.7 |
| 31 | استيراد مجموعة البيانات إلى R studio | 2.7 |
| 32 | التحقق من هيكل إطار البيانات | 3.7 |
| 33 | تحقق أعلى 6 صفوف من مجموعة البيانات | 4.7 |
| 34 | تشريح البيانات هو خطوة لتقسيم البيانات إلى مجموعة تدريب واختبار | 5.7 |
| 35 | عملية اعادة المعالجة | 6.7 |
| 36 | لمجموعة البيانات تلخيص التفاصيل | 7.7 |
| 37 | عملية توحيد (Standardize) للبيانات | 8.7 |
| 38 | عملية تحويل المتغير الهدف (Target) الي عامل متغير (factor) | 9.7 |
| 39 | تدريب نموذج ال SVM بالاستخدام الكيرنل الخطي (linear kernel) | 10.7 |
| 40 | تنبؤ مجموعة الاختبار | 11.7 |
| 41 | مصفوفة التضارب confusion matrix | 12.7 |
| 42 | فتح محرر Jupyter notebook | 13.7 |
| 43 | ادراج البيانات باستخدام مكتبة pandas | 14.7 |
| 44 | انشاء مجموعة تدريب بواسطة dropping لعمود | 15.7 |

| | | |
|---|---|------|
| | (المتغير المستهدف) | |
| 45 | انشاء مجموعة للمتغير المستهدف (target set) | 16.7 |
| 46 | تقسيم مجموعة البيانات لي مجموعة تدريب و مجموعة اختبار | 17.7 |
| 47 | انشاء مصنف SVM مع الكيرنل الخطي (linear) | 18.7 |
| 48 | تحديد دقة المصنف | 19.7 |
| الفصل الخامس : النتائج و مناقشتها | | |
| 50 | النتائج | .8 |
| 50 | النتائج التي تم الحصول عليها عن طريق R programming | 1.8 |
| 53 | النتائج التي تم الحصول عليها عن طريق Python | 2.8 |
| الفصل السادس : الخاتمة و التوصيات و قائمة المصادر و المراجع | | |
| 58 | الخاتمة | .9 |

| | | |
|----|-------------------------|-----|
| 58 | التوصيات | .10 |
| 58 | قائمة المصادر و المراجع | .11 |

فهرس الاشكال

- الشكل (2.1) يوضح عين طبيعية 13.....
- الشكل (2.2) يوضح مخطط العين العادية..... 13
- الشكل (3.2) يوضح اعتلال الشبكية السكري غير التكاثري (NPDR)..... 15.....
- الشكل (4.2) يوضح اعتلال الشبكية السكري التكاثري (PDR)..... 17.....
- الشكل (5.2) يوضح الطبقة الخطية. تقسم حدود القرار (النقاط x بحيث $w^T x + b = 0$) المستوى إلى مجموعتين حسب علامة $w^T x + b$ 20.....
- الشكل (6.2) يوضح مصنف نظام الدعم الالي الخطي..... 22.....
- الشكل (1.7) يوضح الشاشة الرئيسية..... 30.....
- الشكل (2.7) يوضح الشاشة استيراد البيانات في R studio..... 31.....
- الشكل (3.7) يوضح التحقق من هيكل اطار البيانات..... 32.....
- الشكل (4.7) يوضح أعلى 6 صفوف من مجموعة البيانات..... 33.....
- الشكل (5.7) تقسيم البيانات إلى مجموعة تدريب ومجموعة اختبار..... 34.....
- الشكل (6.7) يوضح عملية المعالجة..... 35.....
- الشكل (7.7) يوضح عملية تلخيص البيانات..... 36.....
- الشكل (8.7) يوضح عملية Standardize للبيانات..... 37.....
- الشكل (9.7) يوضح عملية تحويل المتغير الهدف..... 38.....
- الشكل (10.7) يوضح تدريب نموذج ال SVM بالاستخدام الكيرنل الخطي..... 39.....

- الشكل (11.7) يوضح تنبؤ مجموعة الاختبار..... 40.....
- الشكل (12.7) يوضح مصفوفة التضارب..... 41.....
- الشكل (13.7) يوضح الواجهة الرئيسية لمحرر Jupyter notebook 42.....
- الشكل (14.7) يوضح ادراج البيانات بواسطة مكتبة pandas..... 43.....
- الشكل (15.7) انشاء مجموعة تدريب بواسطة dropping لعمود (المتغير المستهدف)..... 44.....
- الشكل (16.7) يوضح انشاء مجموعة للمتغير المستهدف (target set)..... 45.....
- الشكل (17.7) يوضح تقسيم مجموعة البيانات لي مجموعة تدريب و مجموعة اختبار..... 46.....
- الشكل (18.7) انشاء مصنف SVM مع الكيرنل الخطي (linear)..... 47.....
- شكل (19.7) يوضح دقة المصنف..... 48.....

فهرس الجداول

- جدول (1.1) يوضح مصطلحات البحث.....3
- جدول (2.1) يوضح الدراسات السابقة.....4
- جدول (1.6) يوضح وصف مجموعات البيانات.....25
- جدول (2.6) يوضح معلومات الخصائص (Attribute).....26
- جدول (3.6) يوضح عينات مجموعات البيانات.....28
- جدول (1.8) يوضح و صف معلومات التشغيل عن طريق R programming.....50
- جدول (2.8) يوضح مصفوفة التضارب في R programming.....51
- جدول (3.8) يوضح تلخيص النتيجة النهائية عن طريق R programming.....52
- جدول (4.8): يوضح و صف معلومات التشغيل عن طريق Python.....53
- جدول (5.8) يوضح مصفوفة التضارب في python.....54
- جدول (6.8) يوضح تلخيص النتيجة النهائية عن طريق Python.....55
- جدول (7.8) يوضح مقارنة النتائج النهائية.....56

1. المقدمة

اعتلال الشبكية السكري (DR) هو مرض خطير في العين يحدث بسبب داء السكري وقد أصبح السبب الأكثر شيوعاً للعمى في العالم الحالي. بناءً على آخر التقارير بحلول عام 2030 ، يحتمل ان يرتفع معدل انتشار الوباء بنسبة 4.4% في انتشار مرض السكري على مستوى العالم. يمكن أن تتأثر رؤية المريض بمرض السكري الذي يسبب إعتام عدسة العين ، الجلوكوما ، والأهم من ذلك ، تلف الأوعية الدموية داخل العين ، وهي حالة تعرف باسم "اعتلال الشبكية السكري". تتوفر علاجات فعالة لاعتلال الشبكية السكري على الرغم من أنها تتطلب التشخيص المبكر والمراقبة المستمرة لمرضى السكري. يتم إجراء تشخيص اعتلال الشبكية السكري بواسطة تقييم صور الشبكية (قاع العين). الدرجات اليدوية لهذه الصور لتحديد شدة اعتلال الشبكية السكري بطيئة إلى حد ما وتتطلب موارد. يحدث ذلك عندما يتسبب السكري في إتلاف الأوعية الدموية الدقيقة داخل شبكية العين ، وهي الأنسجة الحساسة للضوء في الجزء الخلفي من العين. هذه الأوعية الدموية الصغيرة سوف تسرب الدم والسوائل على أشكال شبكية العين مثل تمدد الأوعية الدموية الدقيقة ، النزيف ، الإفرازات الصلبة ، بقع الصوف القطني أو الحلقات الوريدية.

1.1 معلومات اساسية

تعليم الآلة او ما يعرف باسم machine learning هو واحد من بين اشهر فروع الذكاء الاصطناعي تطور بشكل مذهل و سمي بتعليم الآلة لانه يعتمد علي الآلة , تسمح للحواسيب بامتلاك خاصية "التعلم" و هو كل طريقة تسمح لنا بانشاء نموذج حقيقي انطلاقا من معلومات موجودة او قمنا بانشائه من جديد. بشكل عام يقوم هناك مستويين من التعلم :الاستقرائي و الاستنتاجي. يقوم الاستقرائي باستنتاج قواعد و أحكام عامة عن البيانات الضخمة,بينما الاستنتاجي فهو العكس ,ينطبق من احكام عامة و يطبقها في امثلة خاصة.

2.1 مشكلة البحث

اعتلال الشبكية السكري (DR) هو السبب الأسرع نموًا للعمى حيث يوجد حوالي 415 مليون مريض بالسكري معرضون للخطر في جميع أنحاء العالم. إذا تم اكتشاف المرض يمكن علاجه ؛ إن لم يكن هذا يمكن أن يؤدي إلى العمى الذي لا رجعة فيه لسوء الحظ ، فإن الأخصائيين الطبيين القادرين على اكتشاف المرض غير متاحين في جزء كبير من العالم .

3.1 أهمية البحث

للتعرف علي دور التعليم الالي في تطوير انظمة الرعاية الصحية التعليمية و الصحة العامة و تحسين جودة الرعاية الصحية

4.1 اهداف البحث

1. استخدام خوارزمية نظام الدعم الالي للكشف عن مرض اعتلال الشبكية لمرضي السكري
2. تنفيذ الخوارزمية علي البيانات و اختبار دقتها و تقييمها
3. لتقييم دور البرمجيات الالية القائمة علي الذكاء الاصطناعي للكشف عن اعتلال الشبكية السكري

5.1 منهجية البحث

- المرحلة (1): جمع البيانات: استخدمنا مجموعة البيانات الرقمية المقدمة من Kaggle Community ، وتشمل 1151 عينة ثنائية من الصور التي تم معالجتها من اعتلال الشبكية السكري مع 21 خا صية

- المرحلة (2): استخدام أدوات : R programming و python و Orange

- المرحلة (3): إدراج مجموعة البيانات في مصنف SVM للتنبؤ بما إذا كان المريض يعاني من اعتلال الشبكية السكري ام لا (وجود 1 وغياب 0) في النهاية نحدد دقة المصنف

6.1 نطاق البحث

اسباب اختيار البحث : تركز الدراسة علي استخدام التكنولوجيا التي ظهرت مؤخرا مثل التعلم الالي و التي يمكن الاستفادة منها بشكل كبير في المجالات الطبية .

الحدود المكانية : جهاز كمبيوتر (سطح المكتب)

الحدود الزمنية: يتم اجراء هذا البحث من 2018 - 2020

7.1 مصطلحات البحث

جدول (1.1) يوضح مصطلحات البحث

| المصطلح بالانجليزي | المصطلح بالعربي |
|--|------------------------------------|
| Machine learning | التعلم الالة |
| Diabetic retinopathy | اعتلال الشبكية السكري |
| Non-proliferative diabetic retinopathy | اعتلال الشبكية السكري غير التكاثري |
| proliferative diabetic retinopathy | اعتلال الشبكية السكري التكاثري |
| Support Vector Machine | نظام الدعم الالي |
| Python language | لغة بايثون |
| R language | لغة R |

جدول (2.1) يوضح الدراسات السابقة

| المراجع | مشكلة البحث | الحل المقترح | اسلوب البحث | اهم النتائج | القصور |
|---|--|--|---|---|--|
| DIAGNOSIS OF DIABETIC RETINOPATHY MACHINE USING LEARNING TECHNIQUES” ICTACT JOURNAL ON SOFT COMPUTING, JULY 2013, VOLUME: 03, ISSUE: 04 | تشخيص مرض اعتلال الشبكية السكري باستخدام ثلاثة نماذج | ثلاثة نماذج و هي : Probabilistic Neural network (PNN) Support vector machine (SVM) Bayesian Classification | استخدام الماتلاب لمعالجة الصور و تحويلها الي صور رمادية و بعد ذلك استخراج الميزات مثل الاوعية الدموية و النزيف و الافرازات و اخيرا ادراج الميزات الي ثلاثة نماذج "عدد الصور المستخدمة 100 , 350 | وتظهر النتائج أن : PNN 87.69% Bayes Classification 90.76% SVM 95.38% | كل النماذج الثلاثة حققت نتائج افضل لكن مصنف SVM كان الافضل لذلك كان كانت المشكلة في مصنفات PNN و Bayes |

| | | | | | |
|-----------------------------------|--|---|--|--|---|
| | | للتدريب و 250 للاختبار " | | | |
| كانت المشكلة في مصنف kNN | حقق SVM حساسية 91% و خصوصية 100% و دقة التصنيف 86% بينما حقق KNN حساسية 32.80% و خصوصية 100% و دقة التصنيف 55% | استخدام الماتلاب لمعالجة الصور و بعد ذلك ازالة القرص البصري و كشف الاوعية الدموية و كشف تمدد الاوعية الاموية اخيرا ادراج الميزات الي مصنفات SVM و KNN "عدد الصور 94 " 73 للتدريب و 27 للاختبار " | SVM classifier KNN classifier | الكشف عن تمدد الاوعية الدموية و الافرازات | Diagnosis of Diabetic Retinopathy using Machine Learning" Journal of Research and Development DOI: 10.4172/2311- 3278.1000127" Department of Computer Science and Engineering, Shri Ramdeobaba College Of Engineering and Management, Nagpur, India ,2015 |

| | | | | | |
|------------------------------------|--|--|--------------------------------------|---|---|
| | | | | | |
| المشكلة كانت في مصنف KNN | حقق SVM من الخصوصية %100 و الحساسية %90 | استخدام الماتلاب لمعالجة الصور و استخدام العمليات المورفولوجية للتعرف علي الافرازات و تمدد الاوعية الدموية | SVM classifier KNN classifier | الكشف عن تمدد الاوعية الدموية و الافرازات 3 | Diabetic Retinopathy using Morphological Operations and Machine Learning”, IEEE International Advance Computing Conference (IACC) 978-1-4799-8047-5/15/\$31.00 ©2015 IEEE,2015 3 |
| استغرق تشغيل الصور التحقق من الصحة | ، تحقق شبكة CNN المقترحة حساسية قدرها %95 ودقة %75 | زيادة الطبقات التلافيفية للسماح للشبكة للتعلم ميزات اعق | الشبكات العصبية التلافيفية CNN | كان التعرف علي الميزات الصغيرة من الصور يتطلب | Convolutional Neural Networks for Diabetic Retinopathy” “International |

| | | | | | |
|--|--|--|-----------------------------------|---|---|
| <p>على الشبكة 188 ثانية</p> | | <p>(learn deeper features) في مجموعة البيانات التي تضم 80,000 صورة مستخدمة</p> | | <p>اطباء ذوي الخبرة و قد يستهلك الوقت</p> | <p>Conference On Medical Imaging Understanding and Analysis 2016, MIUA 2016, 6-8 July 2016, Loughborough, UK” ,2016 4</p> |
| <p>الحساسية يجب ان تكون اعلي لتقليل الاجابية الكاذبة في البيانات .</p> | <p>تحقيق دقة 97% و حساسية 94% و خصوصية 98%</p> | <p>استخدام 75135 من الصور من قاعدة بيانات و استخدام خوارزمية التعلم العميق</p> | <p>تطوير خوارزمية تعليم عميقة</p> | <p>تقنية تشخيصية قوية للحد من فقدان البصر</p> | <p>“Automated Identification of Diabetic Retinopathy Using Deep Learning “AMERICAN ACADEMY™ OF OPHTHALMOLOGY ,2017 5</p> |

R. Priya¹ و P. Aruna² (2013) , تشخيص اعتلال الشبكية السكري ، تم وصف ثلاثة نماذج مثل الشبكة العصبية الاحتمالية (PNN) ، ونظام الدعم الالي (SVM) و Bayes Classifier ومقارنة أدائها. يمكن تحديد مقدار المرض المنتشر في شبكية العين عن طريق استخراج ميزات شبكية العين. يتم استخراج الميزات مثل الأوعية الدموية ، نرف صورة NPDR و exudates صورة PDR من الصور الخام باستخدام تقنيات معالجة الصور وتغذيتها إلى المصنف للتصنيف. تم استخدام مجموعه 350 صورة قاع ، منها 100 تم استخدامها للتدريب و 250 صورة تم استخدامها للاختبار. تظهر النتائج التجريبية أن PNN لديها دقة تبلغ 89.6%. تصنيف Bayes Classifier هو دقة 94.4% و SVM لديها دقة 97.6%. هذا يدل على أن طراز SVM يتفوق على جميع الطرز الأخرى. كما يتم تشغيل نظامنا أيضًا على 130 صورة متوفرة من "DIARETDB0: قاعدة بيانات التقييم ومنهجية لاعتلال الشبكية السكري" وتظهر النتائج أن PNN تتمتع بدقة 87.69%. تصنيف Bayes Classification هو دقة 90.76% و SVM دقة 95.38%].

Karandikar , et al. (2015) , اكتشاف تمدد الأوعية الدموية الدقيقة في شبكية العين والإفرازات من أجل الفحص التلقائي لل DR باستخدام المصنف. لتطوير نظام فحص تلقائي للكشف عن الآفات المظلمة والآفات الساطعة في الصناديق الرقمية ، هناك حاجة إلى صور. للكشف عن تمدد الأوعية الدموية الدقيقة في شبكية العين وتتضح ، يتم التقاط الصور من مجموعة بيانات Messidor. بعد المعالجة المسبقة ، يتم إجراء العمليات المورفولوجية للعثور على الميزة ، ثم يتم استخراج الميزات مثل GLCM و Splat للتصنيف. في هذا نحقق حساسية وخصوصية 87% و 100% على التوالي مع دقة 86% [2].

Mr. JaykumarLachure¹, et al. (2015), الكشف عن microanurysm الشبكية والإفرازات للفحص التلقائي لل DR باستخدام نظام الدعم الالي SVM ومصنف KNN. لتطوير هذا النظام المقترح ، هناك حاجة إلى الكشف عن الآفات الحمراء والمشرقة في صور قاع رقمية. تمدد الأوعية الدموية الدقيقة هي أول علامة سريرية . للكشف عن تمدد الأوعية

الدموية الدقيقة في شبكية العين ، يتم التقاط صور قاع الشبكية من Messidor ، DB-reet ، dataset . بعد المعالجة المسبقة ، يتم إجراء العمليات المورفولوجية للعثور على تمديد الأوعية الدموية الدقيقة ثم يتم استخراج الميزات مثل GLCM والميزات الهيكلية للتصنيف. حقق SVM من الخصوصية 100% و الحساسية 90% [3].

Harry Pratt^{a,*} ، Frans Coenen^b et al. (2016). ، تطوير شبكة ذات بنية شبكة تلافيفية عصبية و في مجموعة البيانات التي تضم 80,000 صورة مستخدمة، تحقق شبكة CNN المقترحة حساسية قدرها 95% ودقة 75% [5].

تطوير خوارزمية تعليم عميقة و تم استخدام 75135 ، (2017) ، Rishab Gargeya et al.

من الصور من قاعدة بيانات و استخدام خوارزمية التعلم العميق تحقيق دقة 97% و حساسية 94% و خصوصية 98% [6].

أوجه التشابه :

كل الدراسات السابقة تشخص نفس المرض و تستخدم معظم الميزات مثل الإفرازات و تمدد لاوعية الدموية التي يتم استخراجها بعد عملية معالجة صور العين أوجه الاختلاف :

نلاحظ ان كل دراسة تستخدم مجموعة مختلفة من البيانات و خوارزمية تعلم الآلي مختلفة

مقارنة مع الدراسة الحالية :

الدراسة الحالية تستخدم بيانات (رقمية) تم معالجتها مسبقا و تم الحصول عليها من Kaggle Community ، وتشمل 1151 عينة ثنائية من الصور التي تم معالجتها من اعتلال الشبكية السكري مع 21 خاصية و قد تمت الدراسة علي تجربتين مختلفتين التجربة الاولى باستخدام R و التجربة الثانية باستخدام ال python و تم استخدام نفس الخوارزمية

في التجريبتين و تختلف هذه الدراسة عن الدراسات السابقة في نوع البيانات المستخدمة و الأدوات المستخدمة

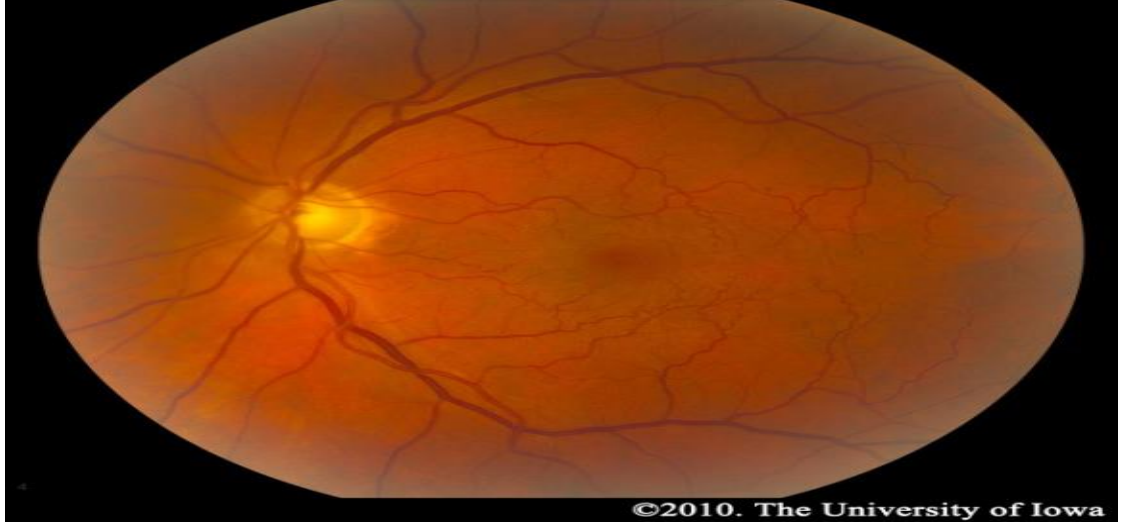
9.1 هيكل البحث

- الفصل الاول (1) الاطار العام و يشمل : المقدمة - معلومات عامة -مشكلة البحث - اهمية البحث - اهداف البحث - نطاق البحث -مصطلحات البحث - الدراسات السابقة
- الفصل الثاني (2) الاطار النظري و يشمل : مقدمة عن مرض اعتلال الشبكية السكري - مقدمة عن لغة الباحثون - مقدمة عن لغة R
- الفصل الثالث (3) التحليل و يشمل : تحليل البيانات - معلومات عن البيانات - وصف البيانات
- الفصل الرابع (4) الاطار العملي و يشمل : عمل التجارب علي البيانات و اختبار دقة المصنف
- الفصل الخامس (5) النتائج و مناقشتها
- الفصل السادس (6) يشمل الخاتمة و التوصيات و المراجع و المصادر

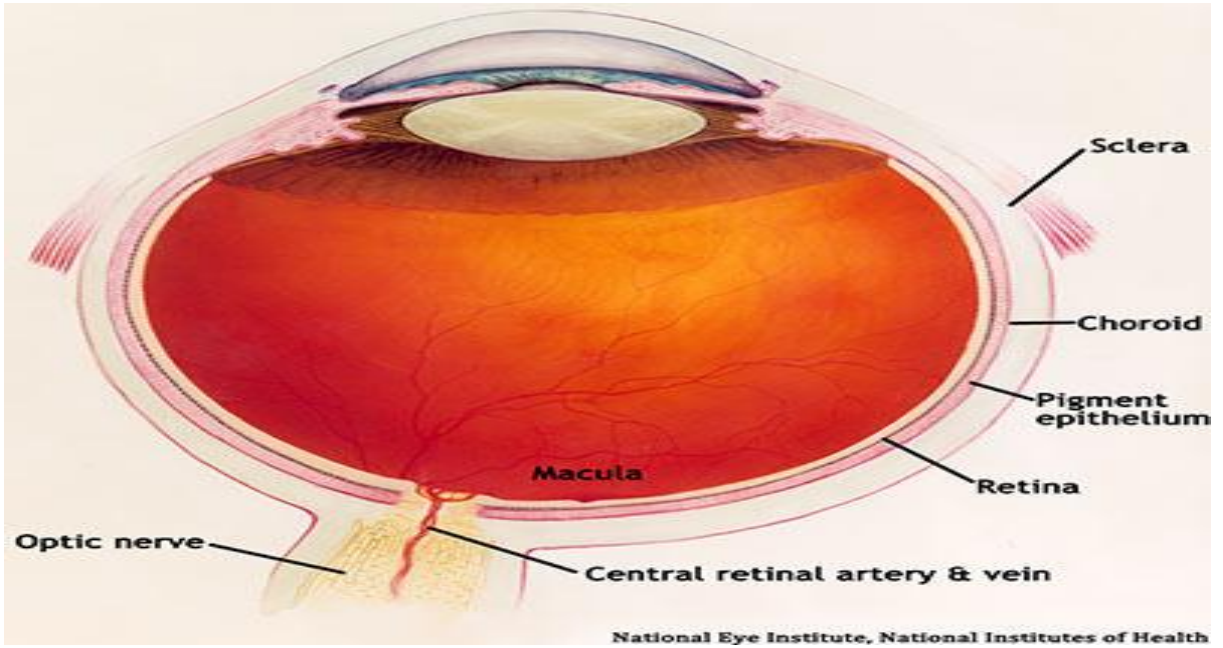
الفصل الثاني
الاطار النظري

1.2 اعتلال الشبكية السكري (DR)

اعتلال الشبكية السكري (DR) هو مرض وعائي في الشبكية يصيب مرضى السكري. إنه السبب الأول للعمى لدى الأشخاص الذين تتراوح أعمارهم بين 20 و 64 عامًا. لذلك ، يعد موضوعًا مفيدًا لجميع طلاب الطب. مرض السكري شائع للغاية ، لذلك ليس من المستغرب أن يصيب مرض اعتلال الشبكية السكري 3.4 ٪ من السكان (4.1 مليون شخص). من بين ملايين الأشخاص المصابين بالتهاب اعتلال الشبكية السكري ، يعاني ما يقرب من ربعهم من مرض يهدد الرؤية (AAO 2008). يرتبط احتمال تطوير اعتلال الشبكية السكري بفترة المرض. النوع الثاني من داء السكري لديه بداية خبيثة ويمكن أن يمر دون أن يلاحظها أحد لسنوات. نتيجة لذلك ، قد يكون لدى المرضى بالفعل اعتلال الشبكية السكري في وقت التشخيص. من ناحية أخرى ، يتم تشخيص مرضى السكري من النوع الأول في وقت مبكر من مرضهم ، وهم عادة لا يصابون باعتلال الشبكية إلا بعد سنوات من إجراء التشخيص. يزيد خطر الإصابة باعتلال الشبكية بعد البلوغ. بعد عشرين عامًا من تشخيص مرض السكري ، يظهر 80 ٪ من مرضى السكري من النوع 2 وجميع مرضى السكري من النوع 1 تقريبًا بعض علامات اعتلال الشبكية (كلاين 1984 أ ، كلاين 1984 ب). في حين أن هذه الأرقام تفتح عينيك ، يمكن لمرضى السكر تقليل مخاطر اعتلال الشبكية وإبطاء تقدم المرض بعد أن بدأ مع مراقبة الجلوكوز (DCCTRG 1993). التحكم في الجلوكوز له أيضًا فائدة إضافية تتمثل في تقليل خطر حدوث مضاعفات أخرى لمرض السكري ، لذلك من المهم أن يتم تعليم مرضى السكري حول هذا الموضوع. يعد الوقت منذ التشخيص ومدى ارتفاع السكر في الدم من أهم عوامل الاختطار بالنسبة لاعتلال الشبكية السكري ، لكن عوامل الخطر الأخرى تشمل ارتفاع ضغط الدم ، وداء شحميات الدم ، والتدخين ، واعتلال الكلية ، والحمل (AAO 2008). شبكية العين عبارة عن ورقة متعددة الطبقات تتكون من الخلايا العصبية ومستقبلات الضوء وخلايا الدعم. إنه أحد أكثر الأعضاء نشاطًا في الجسم ، وبالتالي ، فهو شديد الحساسية لنقص التروية ونقص العناصر الغذائية (Frank 2004).



الشكل (1.2) يوضح عين طبيعية



الشكل (2.2) يوضح مخطط العين العادية

الثالث الخارجي من شبكية العين إمداد الدم من المشيماء ، وهي شبكة وعائية تقع بين شبكية العين والصلبة. يتم توفير الثلثين الداخليين من شبكية العين من خلال فروع الشريان المركزي للشبكية ، والتي تأتي من الشريان البصري (الفرع الأول من الشريان السباتي الداخلي). يخرج الشريان المركزي للشبكية من العصب البصري وتتفرع فروعه مؤقتاً فوق وتحت البقعة (المنطقة الحساسة لشبكية العين المسؤولة عن الرؤية المركزية).

على الرغم من أن الفيزيولوجيا المرضية الدقيقة لأمراض الأوعية الدموية الدقيقة لمرض السكري غير معروفة ، يُعتقد أن ارتفاع السكر في الدم يتسبب في أضرار بطانية ، وفقدان انتقائي للبيسيريت ، وسماكة غشاء القاع ، وكلها تساهم في الأوعية الدموية المتسربة .

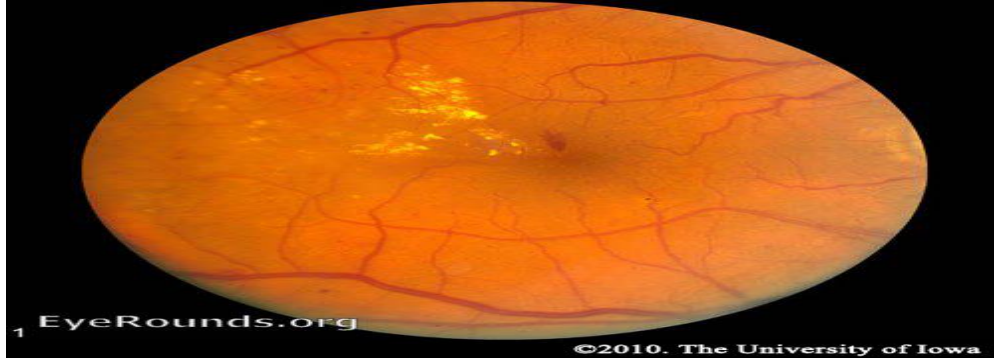
1.2.1 تصنيف

يقع اعتلال الشبكية السكري في فئتين رئيسيتين: غير التكاثري والتكاثري. تشير كلمة "التكاثري" إلى ما إذا كان هناك أو لا يوجد توسع في الأوعية الدموية (نمو غير طبيعي للأوعية الدموية). في مرض شبكية العين المبكر يسمى اعتلال الشبكية السكري غير التكاثري (NPDR). مع تقدم المرض ، قد يتطور إلى اعتلال الشبكية السكري التكاثري (PDR) ، والذي يتم تحديده من خلال وجود الأوعية الدموية وزيادة احتمال حدوث عواقب بصرية خطيرة.

1.2.2 اعتلال الشبكية السكري غير التكاثري (NPDR) Non-proliferative diabetic retinopathy

يؤدي ارتفاع السكر في الدم إلى تلف الشعيرات الدموية في الشبكية. يؤدي ذلك إلى إضعاف جدران الشعيرات الدموية ويؤدي إلى ضخ كميات صغيرة من أوعية الوعاء الدموي ، والتي تعرف باسم (Microaneurysms). تمزق الأوعية الدقيقة لتشكيل نزيف عميق داخل الشبكية ، محصوراً بالغشاء الداخلي المحدد. بسبب مظهرها الشبيه بالنقاط ، يطلق عليها نزيف "النقط". كما تصبح الأوعية الضعيفة متسربة وتسرب السوائل إلى الشبكية. يتداخل ترسب السوائل تحت البقعة ، أو الوذمة البقعة الصفراء ، مع الوظيفة الطبيعية للبقعة ، وهو سبب شائع لفقدان

البصر لدى المصابين باعتلال الشبكية السكري .تتكون هذا الرواسب من منتجات ثانوية للدهون ، ويبدو أنه رواسب صفراء شمعية تسمى الإفرازات الصلبة. مع تقدم اعتلال الشبكية السكري غير التكاثري ، تصبح الأوعية المتأثرة في النهاية معاقة. قد يتسبب هذا العائق في احتشاء طبقة الألياف العصبية ، مما يؤدي إلى ظهور بقع بيضاء ناعمة تسمى (بقع الصوف القطني) .



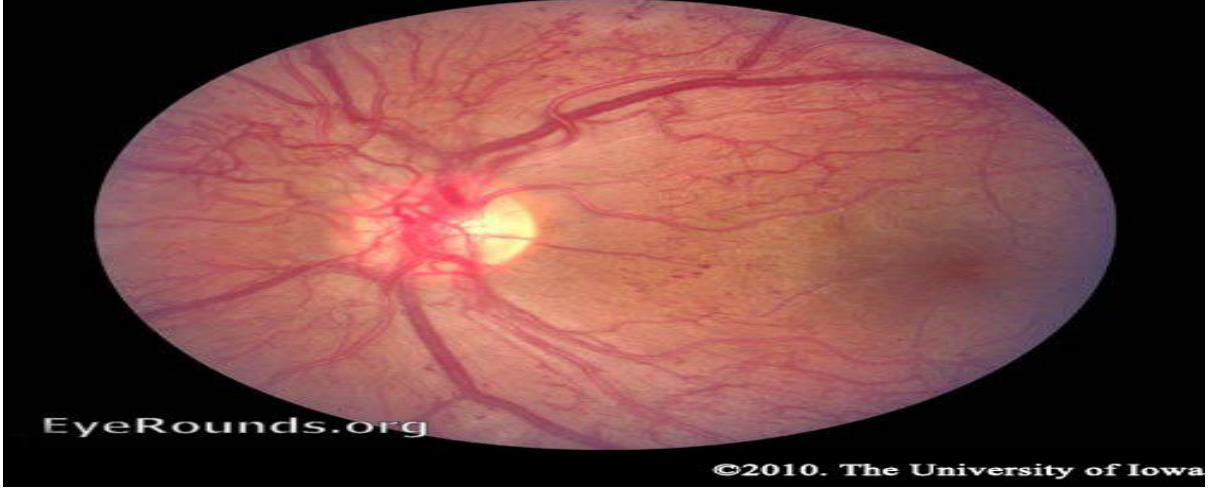
الشكل (3.2) يوضح اعتلال الشبكية السكري غير التكاثري (NPDR)

اعتلال الشبكية السكري غير التكاثر يتم تقديمه أيضًا على أساس النتائج النهائية:

- اعتلال الشبكية السكري غير التكاثري في وقت مبكر - واحد على الأقل من الاوعية الدموية (microaneurysm) موجودة في فحص الشبكية.
- اعتلال الشبكية السكري غير التكاثري المعتدل - يتميز بمتعدد في الأوعية الدموية الدقيقة ، ونزيف أو بقع الصوف القطني.
- اعتلال الشبكية السكري غير التكاثري شديد - في أشد مرحلة من مراحل اعتلال الشبكية السكري غير التكاثري ، ستجد بقع من الصوف القطني ، ، وتشوهات الأوعية الدقيقة الدقيقة .

1.2.3 اعتلال الشبكية السكري التكاثري (proliferative diabetic retinopathy) PDR

كما ذكر سابقًا ، فإن شبكية العين لديها متطلبات أيضية عالية ، لذلك مع استمرار نقص التروية ، تستجيب خلايا الشبكية بإصدار إشارات وعائية مثل عامل نمو بطانة الأوعية الدموية (VEGF). تحفز العوامل الوعائية ، مثل VEGF ، نمو الأوعية الدموية الشبكية الجديدة لتجاوز الأوعية التالفة. يشار إلى هذا باسم الأوعية الدموية. في اعتلال الشبكية السكري التكاثري ، يمتد انتشار الأوعية الليفية . قد يبدو هذا فكرة جيدة ، لكن الأوعية الجديدة تتسرب منها الهشاشة وغالبًا ما يتم توجيهها بطريقة خاطئة. حتى أنها قد تنمو خارج شبكية العين إلى الزجاجي. عندما يتقلص الجسم الزجاجي مع تقدم العمر ، فإنه يسحب على هذه الأوعية الهشة ويمكن أن يسبب تمزقها ، مما يؤدي إلى نزيف زجاجي وفقدان مفاجئ في الرؤية. قد تندب هذه الأوعية أيضًا ، وتشكل مثبتات قوية بين شبكية العين وجسم الزجاجي مما يسبب الجر على شبكية العين. هذه آلية أخرى يمكن أن يسببها اعتلال الشبكية السكري هي فقدان البصر المفاجئ. إذا لم تتم إعادة ربط شبكية العين قريبًا ، خاصةً إذا كانت البقعة ضالعة ، فقد تتعرض الرؤية لخطر دائم. على الرغم من أن تأثيرات الأوعية الدموية في اعتلال الشبكية السكري التكاثري يمكن أن تكون مدمرة ، فإن السبب الأكثر شيوعًا لفقدان البصر لدى مرضى السكري هو الوذمة البقعية. يمكن أن تحدث الوذمة البقعية في اعتلال الشبكية السكري غير التكاثري ، لكنها أكثر شيوعًا في الحالات الأكثر خطورة من اعتلال الشبكية السكري بسبب تسرب الأوعية الدموية الجديدة (Wani 2003). يمكن أن يعاني مرضى السكر أيضًا من مشاكل في العين. قد تطفو جزيئات وعائية المنشأ التي تنتجها شبكية العين من الأمام ، مما تسبب في الأوعية الدموية للقزحية. يمكن أن تنمو هذه الأوعية في زاوية الحجرة الأمامية حيث يوجد شبك شبكي ثلاثي العين . هذا يمكن أن يعرقل تدفق السائل المائي ، مما يزيد الضغط داخل العين ويسبب الجلوكوما الحادة.



الشكل (4.2) يوضح اعتلال الشبكية السكري التكاثري (PDR)

1.2.4 الأعراض SYMPTOMS

عادةً لا يعاني المرضى من الأعراض حتى وقت متأخر أثناء المرض عندما يكون العلاج غير فعال. تختلف الأعراض المتأخرة لاعتلال الشبكية السكري اعتمادًا على السبب. نزيف في الجسم الزجاجي يمكن أن يسبب فقدان مفاجئ للرؤية، وذمة البقعة الصفراء ونقص التروية هما آليتان أخريان من انخفاض الرؤية.

1.2.5 الشاشة SCREENING

يعد فحص اعتلال الشبكية السكري أمرًا لا يصدق لأن معظم المرضى لا يواجهون أي أعراض حتى المراحل المتقدمة من المرض. إذا تم التعرف عليه في وقت مبكر ، يمكن في الغالب منع الآثار الجانبية التي تهدد الرؤية من خلال الإدارة المناسبة. تختلف توصيات الفحص بالنسبة لمرضى السكري من النوع 1 والنوع 2. نظرًا لأن العديد من مرضى السكري من النوع 2 يعانون من اعتلال الشبكية في الوقت الذي يتم فيه تشخيص مرض السكري لديهم ، فمن المستحسن أن تبدأ اختبارات العين السنوية المتوسعة بعد وقت قصير من إجراء تشخيص مرض السكري. يجب أن يبدأ مرضى السكري من النوع الأول بإجراء فحوصات سنوية للعيون بعد 3-5 سنوات من تشخيص مرض السكري (AAO 2008).

اعتلال الشبكية السكري يمكن أن تتقدم بسرعة أثناء الحمل. يجب أن تخضع النساء الحوامل المصابات بداء السكري لفحص متوسّع للعين قبل الحمل ، في بداية الثلث الأول من الحمل ، ثم كل 3 أشهر حتى الولادة. بخلاف مرض السكري الموجود مسبقًا ، لا يُعَرِّضُ سكري الحمل المرضى لخطر اعتلال الشبكية السكري المتزايد ولا يحتاج هؤلاء المرضى إلى فحوصات العين المتكررة (AAO 2008).

1.2.6 الفحص EXAMINATION

يجب أن يبدأ فحص العين السليم لمرض السكري دائمًا عن طريق جمع تاريخ شامل من المريض. اسأل عن أي أعراض بصرية ومشاكل جهازية قد تؤثر على خطر الإصابة بداء اعتلال الشبكية السكري مثل الحمل ، وضغط الدم ، ومستويات الكوليسترول ، والوضع الكلوي. بالإضافة إلى ذلك ، تأكد من فحص آخر هيموغلوبين A1c الخاص بك للحصول على فكرة عن كيفية التحكم في مستوى الجلوكوز في الدم خلال الأشهر الثلاثة الماضية. (AAO 2008)

يجب أن يبدأ الفحص بحدة البصر ، قياسات الضغط داخل العين ، وفحص المصباح الشقي ، بما في ذلك الفحص الدقيق للقرحية . في حالة الاشتباه في توسع الأوعية الدموية للقرحية أو إذا كان المريض يعاني من ضغوط داخل العين ، فيجب إجراء التنظير المناعي لتقييم الزاوية القرحية. يستخدم Gonioscopy عدسة خاصة معكوسة تسمح لك بمشاهدة الزاوية والشبكة التريبقية. لا يمكن تحقيق ذلك باستخدام المصباح الشقي وحده.

1.2.7 خيارات العلاج TREATMENT OPTIONS

عادة ما يتم إدارة اعتلال الشبكية السكري غير التكاثري من خلال تحسين الصحة العامة للمريض. أفضل علاج لاعتلال الشبكية السكري هو منع تطورها وتقدمها من خلال التحكم الدقيق في مستوى الجلوكوز (DCCTRG 1993). يجب على المرضى الحفاظ على نسبة $HbA1c \leq 7\%$. كما تبين أن إدارة ضغط الدم تقلل من تطور المرض (UKPDSG 1998)

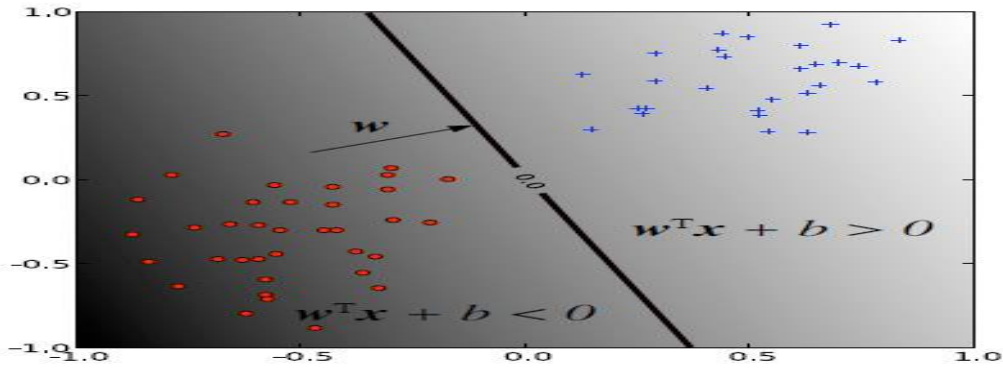
، وينبغي علي المرضى التوقف عن التدخين. يجب أن يتدخل أطباء العيون إذا كان المريض يعاني من وذمة البقعة الصفراء الكبيرة (CSME) مع اعتلال الشبكية السكري غير التكاثري . وغالبًا ما يستخدم تصوير الأوعية فلوريس ، ثم يتم علاجه مباشرةً باستخدام العلاج بالليزر. إذا كان التسرب أكثر انتشارًا ، يمكن لشبكة من حروق الليزر الخفيفة أن تبطئ الوذمة. أخيرًا ، تتوفر العديد من الخيارات الطبية خارج التسمية ، مثل الحقن داخل التريامسينولون وغيرها.

بمجرد تطویر المريض اعتلال الشبكية السكري التكاثري ، هناك العديد من طرق العلاج المتاحة. علاج الوذمة البقعية في اعتلال الشبكية السكري التكاثري يشبه علاج اعتلال الشبكية السكري غير التكاثري . لكن لدى اعتلال الشبكية السكري التكاثري أيضًا خيارات علاجية إضافية تهدف إلى ترويض نمو أوعية جديدة . الدعامة الأساسية للعلاج هي تخثر ضوئي كامل (PRP) ، حيث يتم تدمير أجزاء من شبكية العين باستخدام آلاف الحروق بالليزر أثناء تجنب البقعة. من المفترض أن هذا قد يقلل من كمية الشبكية الدماغية ، وبالتالي ، يقلل من إنتاج جزيئات الأوعية الدموية. قد يبدو العلاج شديدًا ، ولكنه يسبب في الواقع فقدانًا ضعيفًا بشكل مدهش (فرانك 1975). لقد وجد أنه فعال للغاية ، مما يقلل من خطر فقدان البصر الشديد بنسبة 50 ٪ (ETDRS 1987 ، محمد 2007)

1.3 مقدمة نظام الدعم الآلي Support Vector Machines

إن نظام الدعم الآلي هي طريقة تصنيف حديثة تم تقديمها في عام 1992 بواسطة Boser و Guyon و Vapnik. تُستخدم فئة نظام الدعم الآلي على نطاق واسع في المعلوماتية الحيوية (وغيرها من التخصصات) بسبب دقتها العالية وقدرتها على التعامل مع البيانات عالية الأبعاد مثل التعبير الجيني والمرونة في نمذجة مصادر البيانات المتنوعة. ينتمي نظام الدعم الآلي إلى الفئة العامة لطرق kernel. طريقة kernel هي خوارزمية تعتمد على البيانات فقط من خلال dot-products. عندما يكون هذا هو الحال ، يمكن استبدال dot-product بوظيفة kernel التي تحسب dot-product في مساحة ميزة ذات أبعاد عالية ربما. هذا له ميزتان: أولاً ، القدرة على إنشاء حدود قرار (decision boundaries) غير خطية باستخدام طرق

مصممة للفئة الخطية. ثانيًا ، يسمح استخدام وظائف kernel للمستخدم بتطبيق فئة على البيانات التي ليس لها تمثيل واضح لمسافة متجه الأبعاد . المثال الرئيسي لمثل هذه البيانات في المعلوماتية الحيوية هو التسلسل ، إما الحمض النووي أو البروتين ، وهيكل البروتين يتطلب استخدام نظام الدعم الآلي بشكل فعال فهم كيفية عملها. عند تدريب نظام الدعم الآلي ، يجب أن يتخذ الممارس عددًا من القرارات: كيفية معالجة البيانات مسبقًا ، ونواة الاستخدام ، وأخيرًا ، تحديد معلمات نظام الدعم الآلي والنواة.



الشكل (5.2) يوضح الطبقة الخطية. تقسم حدود القرار (النقاط x بحيث $w^T x + b = 0$) المستوى إلى مجموعتين حسب علامة $w^T x + b$.

1.3.1 التمهيدية: المصنف الخطي Preliminaries: Linear Classifier

يعتبر نظام الدعم الآلي مثالاً على فئة خطية من فئتين. يشرح هذا القسم ماذا يعني ذلك. تتكون البيانات الخاصة بمشكلة التعلم في فصلين من الكائنات التي تحمل علامة واحدة من تصنيفين مطابقين للفئتين ؛ ، نفترض أن التصنيفات هي $1+$ (أمثلة إيجابية) أو $1-$ (أمثلة سلبية). في ما يلي x يدل على متجه مع المكونات x_i تدل العلامة x_i على المتجه i^{th} في مجموعة البيانات $f(x_i; y_i) g$ ، حيث y_i هي التسمية المرتبطة x_i . تسمى الكائنات x_i بالأنماط أو الأمثلة. نفترض أن الأمثلة تنتمي إلى مجموعة X . في البداية ، نفترض أن الأمثلة عبارة عن متجهات (vectors)، ولكن بمجرد تقديم النواة ، سيتم الاسترخاء

على هذا الافتراض ، وعند هذه النقطة يمكن أن يكون أي كائن مستمر / منفصل (مثل تسلسل البروتين / الحمض النووي أو بنية البروتين). المفهوم الرئيسي المطلوب لتعريف الطبقة الخطية هي نقطة الانتاج (dot product) بين متجهين ، ، يُعرف بالمعادلة $(w^T x = \sum_i w_i x_i)$. تعتمد الفئة الخطية على دالة تمييز خطية للنموذج.

$$F(x) = w^T x + b$$

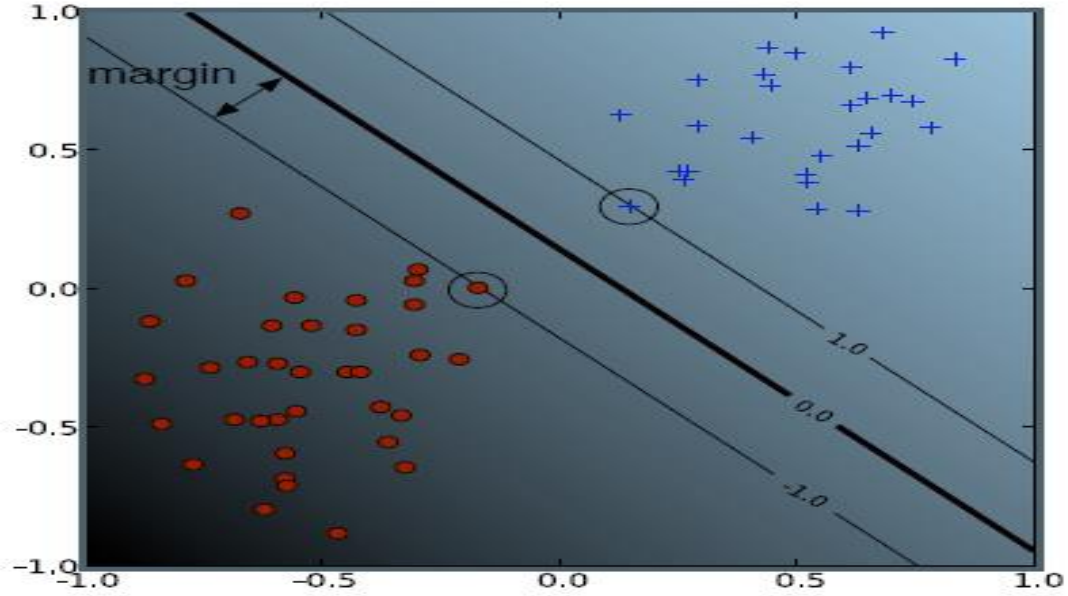
يُعرف المتجه w باسم متجه الوزن ، ويسمى b التحيز. النظر في القضية $b = 0$ أولاً. مجموعة النقاط x بحيث تكون $w^T x = 0$ كلها نقاط متعامدة مع w وتذهب إلى الأصل خط في بعدين ، الخط في ثلاثة أبعاد ، تسمى الخط الفائق (hyperplane).

$$fx: f(x) = w^T x + b = 0$$

إلى جانب الخط ذات $f(x)$ يقسم المسافة إلى قسمين: تشير علامة الدالة التمييزية المستوى الفائق التي تشغلها نقطة (انظر 1). وتسمى الحدود بين المناطق المصنفة بأنها إيجابية وسلبية حدود قرار الطبقة. يقال إن حدود القرار المعرفّة من الخطوط الفائقة خطية لأنها خطية في أمثلة المدخلات (راجع المعادلة 1). تسمى الطبقة ذات حدود القرار الخطي الطبقة الكلاسيكية.

1.3.2 تصنيف الهامش الكبير: Large Margin Classification

في ما يلي نستخدم المصطلح قابل للفصل خطياً للإشارة إلى البيانات التي توجد لها حدود قرار خطي تفصل بين الموجب والأمثلة السلبية (انظر الشكل 5.2). في البداية ، سوف نفترض بيانات مفصولة خطياً ، ونشير لاحقاً إلى كيفية التعامل مع البيانات التي لا يمكن فصلها خطياً



الشكل (6.2) يوضح مصنف نظام الدعم الآلي الخطي . نقاط البيانات الدائرية هي متجهات الدعم - الأمثلة الأقرب إلى حدود القرارات وهم يحددون الهامش الذي يتم به فصل الفئتين.

1.4 مقدمة عن R

لغة R programming هي لغة برمجة وبيئة لتحليل البيانات التفاعلية. تم إصدار النسخة الأولى من R في عام 1993 ، ونمت من لغة S المطورة في Bell Labs في أواخر السبعينيات. تتمثل القوة الرئيسية لـ R في نظامها الإيكولوجي الغني ، والذي يشمل البائعين التجاريين ، وآلاف الحزم الإضافية ، وآليات الحصول على المساعدة والعديد من أماكن النشر العلمي

1.5 لغة البايثون Python

بايثون لغة برمجة، من لغات المستوى العالي، تتميز ببساطة كتابتها وقراءتها، سهلة التعلم، تستخدم أسلوب البرمجة الكائنية (OOP) ، مفتوحة المصدر، وقابلة للتطوير. تعتبر لغة بايثون لغة تفسيرية، متعددة الأغراض وتستخدم بشكل واسع في العديد من المجالات، كبناء البرامج المستقلة باستخدام الواجهات الرسومية المعروفة وفي عمل برامج الويب، بالإضافة إلى

استخدامها كلغة برمجة نصية للتحكم في أداء بعض من أشهر البرامج المعروفة أو في بناء برامج ملحقة لها. وبشكل عام يمكن استخدام بايثون لبرمجة البرامج البسيطة للمبتدئين، ولإنجاز المشاريع الضخمة كأى لغة برمجية أخرى في نفس الوقت. غالباً ما يُنصح المبتدئين في ميدان البرمجة بتعلم هذه اللغة لأنها من بين أسرع اللغات البرمجية تعلماً.

الفصل الثالث التحليل

6. مقدمة : في هذا الفصل يتم تحليل البيانات و وصف البيانات

1.6 التحليل Analysis :

تحليل البيانات هو عملية فحص وتطهير وتحويل ونمذجة البيانات بهدف اكتشاف معلومات مفيدة وإبلاغ الاستنتاج ودعم اتخاذ القرار. في عملية التحليل أولاً تحديد البيانات التي تم جمعها من مصادر مختلفة ثم يتم التحليل باستخدام أدوات R programming و python و orange تسمى الخوارزمية المستخدمة (نظام الدعم الالي SVM)

2.6 البيانات dataset

وصف البيانات المستخدمة في البحث وكيفية معالجتها والتجربة التي أجريت

3.6 الأدوات Tools

تسمى الأدوات المستخدمة في هذا البحث R programming و python والخوارزمية تسمى خوارزمية نظام الدعم الالي (SVM)

4.6 وصف البيانات Data description

تحتوي مجموعة بيانات اعتلال الشبكية السكري على ميزات features مستخرجة من قاعدة بيانات صور messidor، 1151 رقم instance مع 21 Attribute

جدول (1.6) يوضح وصف مجموعات البيانات

| | | | | | |
|------------|-------------------------|------|-----------------|---------------------|-----------------------------|
| Life | مساحة | 1151 | عدد instance | متعددة المتغيرات | خصائص مجموعة البيانات |
| 2014-11-03 | البيانات المتبرع بها | 12 | عدد الخصائص | عدد Integer/real | الخصائص Attribute |

| | | | | | |
|-------|------------------------|-------|----------------|-------|--------------------|
| | | | Attribute | | |
| 94145 | عدد مرات تصفح الويب | N / A | القيم المفقودة | تصنيف | المهام المرتبطة |

5.6 معلومات مجموعة البيانات

تحتوي مجموعة البيانات هذه على ميزات (features) مستخرجة من مجموعة صور Messidor للتنبؤ بما إذا كانت الصورة تحتوي على علامات اعتلال الشبكية السكري أم لا. تمثل جميع الميزات إما آفة مكتشفة (detected lesion) ، أو ميزة وصفية لجزء تشريحي (anatomical part). يوصف تحليل الصورة الأساسي للطريقة واستخراج الميزات وكذلك تقنية التصنيف الخاصة بنا في Andras Hajdu ، Balint Antal: نظام قائم على المجموعة للفحص التلقائي لاعتلال الشبكية السكري ، النظم القائمة على المعرفة 60 (أبريل 2014) ، 27-20. تتوفر مجموعة الصور (Messidor) على [Web Link]

جدول (2.6) يوضح معلومات الخصائص (Attribute)

| | |
|-----|--|
| 0 | النتيجة الثنائية لتقييم الجودة. 0 = جودة سيئة 1 = جودة كافية. |
| 1 | النتيجة الثنائية للفحص المسبق ، حيث تشير 1 إلى خلل حاد في الشبكية و 0 افتقارها. |
| 7-2 | نتائج الكشف عن تمدد الأوعية الدموية الدقيقة (MA) ، تشير كل قيمة للقيمة إلى عدد MAs الموجودة في مستويات الثقة = alpha = 0.5 ، 1 ، على التوالي |

| | |
|-----------------|--|
| 15-8 | تحتوي على نفس المعلومات مثل (2-7) للافرزات. ومع ذلك ، حيث يتم تمثيل الافرازات بمجموعة من النقاط بدلاً من عدد البكسلات التي تصنع الآفات . |
| 16 | المسافة الإقليدية من مركز البقعة ووسط القرص البصري لتوفير معلومات مهمة عن حالة المريض. |
| 17 | قطر القرص البصري |
| 18 | النتيجة الثنائية للتصنيف القائم على AM / FM amplitude-modulation-frequency-modulation تعديل السعة و تشكيل التردد للتمييز بين صورة الشبكية الطبيعية و المرضية |
| الهدف Target | يحتوي علي علامات اعتلال الشبكية السكري , 1,2,3 يشير لوجود المرض حسب قاعدة بيانات Messidor و 0 يشير الي عدم وجود المرض |

جدول (3.6) يوضح عينات مجموعات البيانات

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | target |
|---|---|-----|----|----|----|----|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----|--------|
| 1 | 1 | 22 | 22 | 22 | 19 | 18 | 14 | 49.9 | 17.78 | 5.271 | 0.772 | 0.019 | 0.007 | 0.004 | 0.004 | 0.487 | 0.1 | 1 | 0 |
| 1 | 1 | 24 | 24 | 22 | 18 | 16 | 13 | 57.71 | 23.8 | 3.325 | 0.234 | 0.004 | 0.004 | 0.004 | 0.004 | 0.521 | 0.144 | 0 | 0 |
| 1 | 1 | 62 | 60 | 59 | 54 | 47 | 33 | 55.83 | 27.99 | 12.69 | 4.852 | 1.394 | 0.373 | 0.042 | 0.008 | 0.531 | 0.129 | 0 | 1 |
| 1 | 1 | 55 | 53 | 53 | 50 | 43 | 31 | 40.47 | 18.45 | 9.119 | 3.079 | 0.84 | 0.272 | 0.008 | 0.002 | 0.483 | 0.115 | 0 | 0 |
| 1 | 1 | 44 | 44 | 44 | 41 | 39 | 27 | 18.03 | 8.571 | 0.41 | 0 | 0 | 0 | 0 | 0 | 0.476 | 0.124 | 0 | 1 |
| 1 | 1 | 44 | 43 | 41 | 41 | 37 | 29 | 28.36 | 6.936 | 2.306 | 0.324 | 0 | 0 | 0 | 0 | 0.503 | 0.127 | 0 | 1 |
| 1 | 0 | 29 | 29 | 29 | 27 | 25 | 16 | 15.45 | 9.114 | 1.633 | 0 | 0 | 0 | 0 | 0 | 0.542 | 0.14 | 0 | 1 |
| 1 | 1 | 6 | 6 | 6 | 6 | 2 | 1 | 20.68 | 9.498 | 1.224 | 0.15 | 0 | 0 | 0 | 0 | 0.576 | 0.071 | 1 | 0 |
| 1 | 1 | 22 | 21 | 18 | 15 | 13 | 10 | 66.69 | 23.55 | 6.151 | 0.496 | 0 | 0 | 0 | 0 | 0.5 | 0.117 | 0 | 1 |
| 1 | 1 | 79 | 75 | 73 | 71 | 64 | 47 | 22.14 | 10.05 | 0.875 | 0.1 | 0.023 | 0 | 0 | 0 | 0.561 | 0.109 | 0 | 1 |
| 1 | 1 | 45 | 45 | 45 | 43 | 40 | 32 | 84.36 | 50.98 | 17.29 | 1.974 | 0 | 0 | 0 | 0 | 0.546 | 0.112 | 0 | 0 |
| 1 | 0 | 25 | 25 | 25 | 23 | 22 | 18 | 22.48 | 13.95 | 0.436 | 0.116 | 0 | 0 | 0 | 0 | 0.552 | 0.14 | 1 | 0 |
| 1 | 1 | 70 | 69 | 65 | 63 | 63 | 50 | 10.56 | 3.108 | 0.626 | 0.288 | 0.104 | 0.005 | 0 | 0 | 0.534 | 0.09 | 0 | 1 |
| 1 | 1 | 48 | 43 | 39 | 32 | 27 | 18 | 23.01 | 6.738 | 2.404 | 0.189 | 0.011 | 0 | 0 | 0 | 0.502 | 0.138 | 1 | 1 |
| 1 | 1 | 94 | 93 | 92 | 89 | 86 | 77 | 8.611 | 1.981 | 0.401 | 0.066 | 0 | 0 | 0 | 0 | 0.541 | 0.125 | 0 | 0 |
| 1 | 1 | 20 | 18 | 16 | 15 | 13 | 9 | 65.11 | 33.12 | 8.785 | 0.674 | 0.052 | 0.003 | 1E-03 | 1E-03 | 0.569 | 0.09 | 1 | 0 |
| 1 | 1 | 105 | 95 | 81 | 66 | 46 | 32 | 123.1 | 70.57 | 37.41 | 19.94 | 14.79 | 6.115 | 2.346 | 1.002 | 0.524 | 0.134 | 1 | 1 |

الفصل الرابع

الاطار العملي

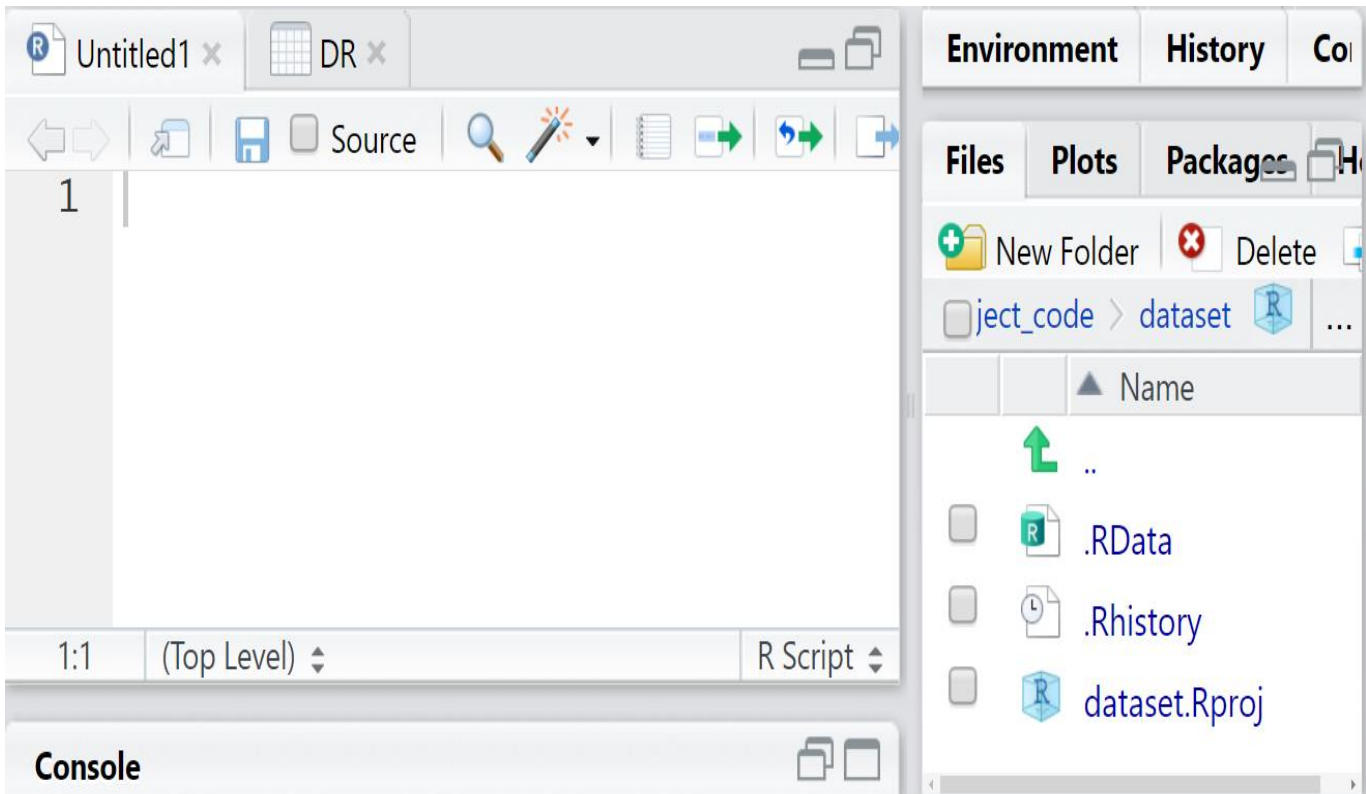
7. التنفيذ و الاختبار

يتم التنفيذ و الاختبار علي تجربتين :

- استخدام لغة R programming , و استخدام python مع نفس الخوارزمية و نفس البيانات

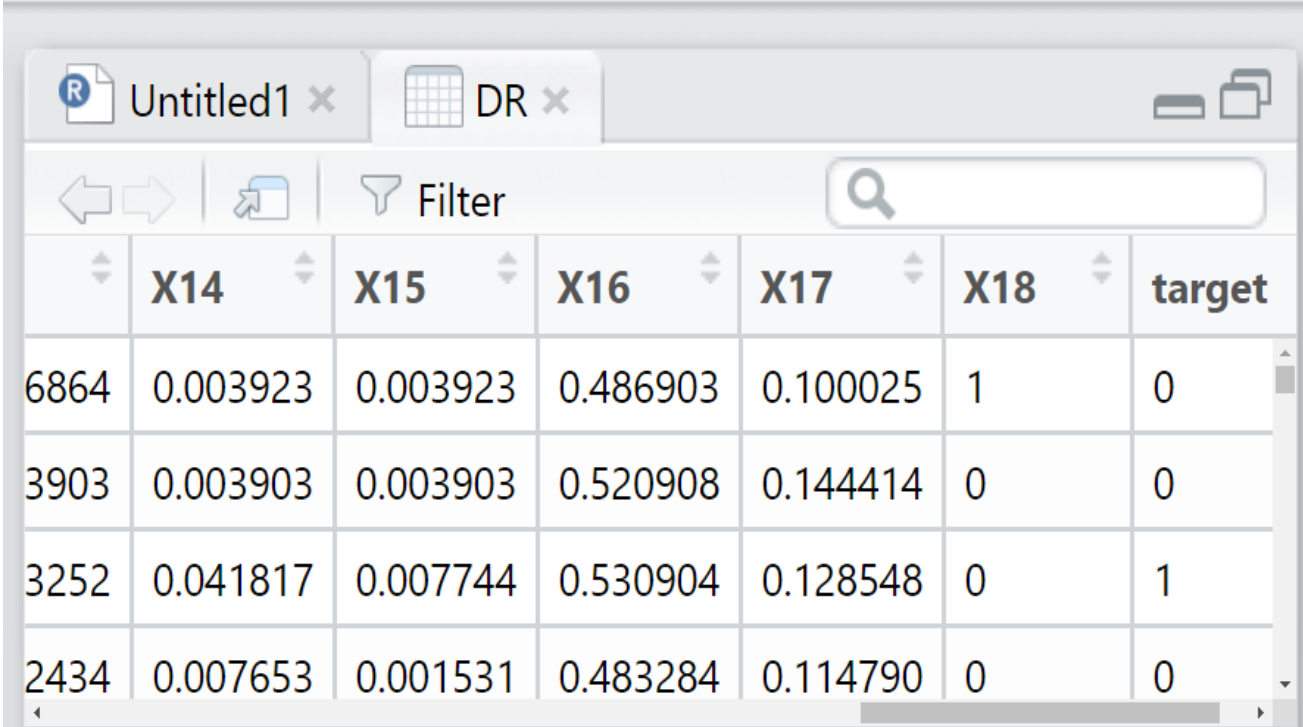
" اولاً استخدام R programming "

1.7 افتح أداة R studio



الشكل (1.7) يوضح الشاشة الرئيسية

2.7 استيراد مجموعة البيانات إلى R studio



The screenshot shows the R Studio interface with a data table displayed. The table has six columns: X14, X15, X16, X17, X18, and target. The rows contain numerical values for each column. The interface includes a toolbar with navigation and search icons, and a search bar.

| | X14 | X15 | X16 | X17 | X18 | target |
|------|----------|----------|----------|----------|-----|--------|
| 6864 | 0.003923 | 0.003923 | 0.486903 | 0.100025 | 1 | 0 |
| 3903 | 0.003903 | 0.003903 | 0.520908 | 0.144414 | 0 | 0 |
| 3252 | 0.041817 | 0.007744 | 0.530904 | 0.128548 | 0 | 1 |
| 2434 | 0.007653 | 0.001531 | 0.483284 | 0.114790 | 0 | 0 |

الشكل (2.7) يوضح الشاشة استيراد البيانات في R studio

3.7 التحقق من هيكل إطار البيانات

```
> str(DR)
'data.frame': 1151 obs. of 21 variables:
 $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ X0      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ X1      : int  1 1 1 1 1 1 0 1 1 1 ...
 $ X2      : int  22 24 62 55 44 44 29 6 22 79 ...
 $ X3      : int  22 24 60 53 44 43 29 6 21 75 ...
 $ X4      : int  22 22 59 53 44 41 29 6 18 73 ...
 $ X5      : int  19 18 54 50 41 41 27 6 15 71 ...
 $ X6      : int  18 16 47 43 39 37 25 2 13 64 ...
 $ X7      : int  14 13 33 31 27 29 16 1 10 47 ...
 $ X8      : num  49.9 57.7 55.8 40.5 18 ...
 $ X9      : num  17.78 23.8 27.99 18.45 8.57 ...
 $ X10     : num  5.27 3.33 12.69 9.12 0.41 ...
 $ X11     : num  0.772 0.234 4.852 3.079 0 ...
 $ X12     : num  0.0186 0.0039 1.3939 0.8403 0 ...
 $ X13     : num  0.00686 0.0039 0.37325 0.27243 0 ...
 $ X14     : num  0.00392 0.0039 0.04182 0.00765 0 ...
 $ X15     : num  0.00392 0.0039 0.00774 0.00153 0 ...
 $ X16     : num  0.487 0.521 0.531 0.483 0.476 ...
 $ X17     : num  0.1 0.144 0.129 0.115 0.124 ...
 $ X18     : int  1 0 0 0 0 0 0 1 0 0 ...
 $ target: int  0 0 1 0 1 1 1 0 1 1 ...
```

الشكل (3.7) يوضح التحقق من هيكل إطار البيانات عن طريق الدالة str (DR)

يوضح لنا الناتج أعلاه أن مجموعة البيانات الخاصة بنا تتكون من 1151 ملاحظة لكل منها

21 خاصية

4.7 تحقق أعلى 6 صفوف من مجموعة البيانات

```
> head(DR)
  id X0 X1 X2 X3 X4 X5 X6 X7      X8
1  1  1  1 22 22 22 19 18 14 49.89576
2  2  1  1 24 24 22 18 16 13 57.70994
3  3  1  1 62 60 59 54 47 33 55.83144
4  4  1  1 55 53 53 50 43 31 40.46723
5  5  1  1 44 44 44 41 39 27 18.02625
6  6  1  1 44 43 41 41 37 29 28.35640
      X9      X10      X11      X12
1 17.775994  5.270920 0.771761 0.018632
2 23.799994  3.325423 0.234185 0.003903
3 27.993933 12.687485 4.852282 1.393889
4 18.445954  9.118901 3.079428 0.840261
5  8.570709  0.410381 0.000000 0.000000
6  6.935636  2.305771 0.323724 0.000000
      X13      X14      X15      X16      X17
1 0.006864 0.003923 0.003923 0.486903 0.100025
2 0.003903 0.003903 0.003903 0.520908 0.144414
3 0.373252 0.041817 0.007744 0.530904 0.128548
4 0.272434 0.007653 0.001531 0.483284 0.114790
5 0.000000 0.000000 0.000000 0.475935 0.123572
6 0.000000 0.000000 0.000000 0.502831 0.126741
  X18 target
1  1      0
2  0      0
3  0      1
4  0      0
5  0      1
6  0      1
```

الشكل (4.7) يوضح أعلى 6 صفوف من مجموعة البيانات

نطاق قيم الخاصية مختلف لكن كل الخصائص تتكون من بيانات رقمية

5.7 تشريح البيانات هو خطوة لتقسيم البيانات إلى مجموعة تدريب واختبار

```
> input <- createDataPartition(y=DR$target , p =0.9 ,list = FALSE)
> training <-DR[input,]
> testing <-DR[-input,]
> dim(training);dim(testing)
[1] 1036  21
[1] 115  21

> |
```

الشكل (5.7) تقسيم البيانات إلى مجموعة تدريب ومجموعة اختبار

الدالة `CreateDataPartition()`:

لتقسيم بياناتنا إلى مجموعة تدريب واختبار و نمرر 3 المعلمات (parameters).

- تأخذ المعلمة "y" قيمة المتغير وفقاً للبيانات التي يجب تقسيمها. في حالتنا ، يكون المتغير المستهدف هو (target) ، لذلك نقوم بتمرير البيانات \$target .

1. `training <- DR [input,]` : هو لوضع البيانات من إطار البيانات إلى بيانات

التدريب

2. `testing <- DR [input,]` : يتم حفظ البيانات المتبقية في إطار بيانات الاختبار

3. `dim (training); dim (testing)` : للتحقق من أبعاد إطار بيانات التدريب لدينا

وإطار بيانات

الاختبار . (استخدام 1036 عينة مجموعة تدريب و 115 مجموعة اختبار

```
> anyNA(DR)
```

```
[1] FALSE
```

الشكل (6.7) يوضح عملية المعالجة

الدالة `anyNA(DR)`:

تحقق مما إذا كانت بياناتنا تحتوي على قيم مفقودة أم لا

7.7 تلخيص التفاصيل لمجموعة البيانات (summary)

```
> summary(DR)
```

| | id | X0 |
|----------|---------|----------------|
| Min. | : 1.0 | Min. :0.0000 |
| 1st Qu.: | 288.5 | 1st Qu.:1.0000 |
| Median : | 576.0 | Median :1.0000 |
| Mean : | 576.0 | Mean :0.9965 |
| 3rd Qu.: | 863.5 | 3rd Qu.:1.0000 |
| Max. | :1151.0 | Max. :1.0000 |

الشكل (7.7) يوضح عملية تلخيص البيانات

الدالة : summary(DR)

التحقق من التفاصيل الملخصة لبياناتنا

Pre-processing:

- centered (21)
- ignored (0)
- scaled (21)

الشكل (8.7) يوضح عملية Standardize للبيانات

الدالة Preprocess (data)

9.7 عملية تحويل المتغير الهدف (Target) الي عامل متغير (factor)

```
training[["target"]] = factor(training[["target"]])
```

الشكل (9.7) يوضح عملية تحويل المتغير الهدف (Target) الي عامل متغير (factor)

الدالة `[[“target”]] = factor (training [[“target”]])`

تستخدم تحويل عمود “target” لإطار بيانات التدريب إلى عامل متغير باستخدام التدريب

10.7 تدريب نموذج الSVM باستخدام الكيرنل الخطي (linear kernel)

```
> svm_Linear <- caret::train(target ~. , data =training ,method="svmLinear",
+                             trainControl=trctrl,           preProcess = c("center","scale"),
+                             tuneLength=10)
>
> svm_Linear
Support Vector Machines with Linear Kernel

1036 samples
 20 predictor
 2 classes: '0', '1'

Pre-processing: centered (20), scaled (20)
Resampling: Bootstrapped (25 reps)
Summary of sample sizes: 1036, 1036, 1036, 1036, 1036, 1036, ...
Resampling results:

Accuracy   Kappa
0.7174028 0.4427454

Tuning parameter 'C' was held constant at
a value of 1
> |
```

الشكل (10.7) يوضح تدريب نموذج الSVM باستخدام الكيرنل الخطي (linear kernel)

الدالة

: train () method

- نمرر المنصف الخطي (svm_Linear) الي الدالة Method

- نمرر المتغير الهدف

- يشير "~." إلى صيغة لاستخدام جميع الخواص في مصنفنا و target كمتغير مستهدف..

يجب تمرير المعلمة "trControl" مع نتائج من طريقة () trianControl لدينا. المعلمة "Preprocess" هي للمعالجة المسبقة لبيانات التدريب لدينا. ينتج نموذج SVM باستخدام svm_Linear، إنه نموذج خطي ، لذلك تم اختياره فقط بالقيمة. $C = 1$

11.7 تتبؤ مجموعة الاختبار (Test Set Prediction)

```
> test_pred <- predict(svm_Linear , newdata =testing)
> test_pred
 [1] 0 0 1 0 1 1 0 0 0 0 0 1 0 0 0 1 0 0 1 0 1
 [22] 0 0 0 1 1 0 0 1 1 0 1 0 0 1 1 0 0 0 1 1 1
 [43] 0 0 1 0 0 1 0 0 0 0 0 1 0 0 0 1 1 1 0 0 1
 [64] 0 0 0 1 0 1 1 0 0 1 0 0 0 0 0 1 1 0 0 0 1
 [85] 1 0 1 0 0 1 1 0 0 0 0 0 1 1 0 0 0 0 0 1 0
 [106] 0 1 0 0 0 1 1 0 0 1
Levels: 0 1
```

الشكل (11.7) يوضح تتبؤ مجموعة الاختبار

الآن ، تم تدريب نموذجنا على قيمة $C \leq 1$. للتنبؤ بالفئات (classes) لمجموعة الاختبار الخاصة بنا. يمكننا استخدام طريقة التنبؤ predict()method

مررنا 2 argument. المعلمة الأولى هي نموذجنا المدرب والمعلمة الثانية "newdata" تحمل إطار بيانات الاختبار الخاص بنا.

الدالة The predict() تقوم بإرجاع قائمة. نحن نحفظه في متغير. test_pred

12.7 مصفوفة التضارب

```
> caret::confusionMatrix(table(test_pred,testing$target))  
Confusion Matrix and Statistics
```

```
test_pred  0  1  
          0 52 21  
          1  8 34
```

Accuracy : 0.7478

95% CI : (0.6583, 0.8242)

No Information Rate : 0.5217

P-Value [Acc > NIR] : 5.457e-07

Kappa : 0.4897

Mcnemar's Test P-Value : 0.02586

Sensitivity : 0.8667

Specificity : 0.6182

Pos Pred Value : 0.7123

Neg Pred Value : 0.8095

Prevalence : 0.5217

Detection Rate : 0.4522

Detection Prevalence : 0.6348

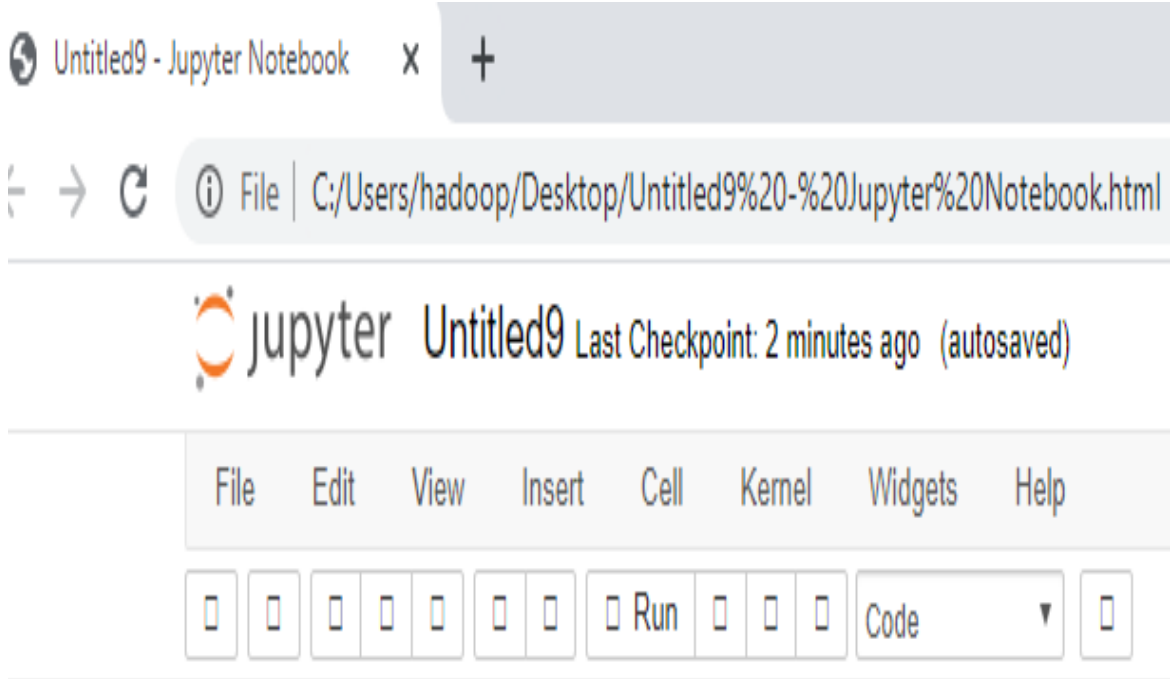
Balanced Accuracy : 0.7424

'Positive' Class : 0 |

الشكل (12.7) يوضح مصفوفة التضارب

“ثانيا استخدام Python”

Jupyter notebook 13.7 فتح محرر



Jupyter notebook شكل (13.7) يوضح الواجهة الرئيسية لمحرر

و مكتبات البايثون

14.7 ادراج البيانات باستخدام مكتبة pandas

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn import metrics
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report

import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib

data = pd.read_csv("C:/Users/hadoop/Desktop/data.csv")

data.head()
```

| | | | | | | |
|----------|----------|------------|----------|----------|-----|--------|
| 0.006864 | 0.003923 | 0.003923.1 | 0.486903 | 0.100025 | 1.2 | target |
|----------|----------|------------|----------|----------|-----|--------|

شكل (14.7) يوضح ادراج البيانات بواسطة مكتبة pandas

15.7 انشاء مجموعة تدريب بواسطة dropping (المتغير المستهدف)

```
train = data.drop('target', axis = 1)
```

```
train.head()
```

| | | | | | | | |
|---|-----|----|------|------|----|----|----|
| 1 | 1.1 | 22 | 22.1 | 22.2 | 19 | 18 | 14 |
| 0 | 2 | 1 | 24 | 24 | 22 | 18 | 16 |
| 1 | 3 | 1 | 62 | 60 | 59 | 54 | 47 |
| 2 | 4 | 1 | 55 | 53 | 53 | 50 | 43 |
| 3 | 5 | 1 | 44 | 44 | 44 | 41 | 39 |
| 4 | 6 | 1 | 44 | 43 | 41 | 41 | 37 |

الشكل (15.7) انشاء مجموعة تدريب بواسطة dropping (المتغير المستهدف)

16.7 انشاء مجموعة للمتغير المستهدف (target set)

```
target = data.target
```

```
target.head()
```

```
0      0
```

```
1      1
```

```
2      0
```

```
3      1
```

```
4      1
```

```
Name: target, dtype: int64
```

شكل (16.7) يوضح انشاء مجموعة للمتغير المستهدف (target set)

17.7 تقسيم مجموعة البيانات لي مجموعة تدريب و مجموعة اختبار

```
X_train,X_test,y_train,y_test= train_test_split(train,target,test_size=0.3,random_state=109)
print("X_train size ==>", X_train.shape)
print("y_train size ==>", y_train.shape)
print("X_test size ==>", X_test.shape)
print("y_test size ==>", y_test.shape)
X_train size ==> (805, 19)
y_train size ==> (805,)
X_test size ==> (345, 19)
y_test size ==> (345,)
```

شكل (17.7) يوضح تقسيم مجموعة البيانات لي مجموعة تدريب و مجموعة اختبار

18.7 انشاء مصنف SVM مع الكيرنل الخطي (linear)

```
clf =svm.SVC(kernel = 'linear')
clf.fit(X_train,y_train)
y_pred =clf.predict(X_test)

print("Accuracy :" ,metrics.accuracy_score(y_test,y_pred))
print("precision :" ,metrics.precision_score(y_test,y_pred))
print("Recall :" ,metrics.recall_score(y_test,y_pred))
print("CM :" ,confusion_matrix(y_test,y_pred))
print("report :" ,accuracy_score(y_test,y_pred))
#print("report:" , classification_report(y_test,y_pred))
```

شكل (18.7) انشاء مصنف SVM مع الكيرنل الخطي (linear)

19.7 تحديد دقة المصنف

```
Accuracy : 0.7844827586206896
precision : 0.833333333333333334
Recall : 0.660377358490566
CM : [[56  7]
      [18 35]]
report : 0.7844827586206896
```

شكل (19.7) يوضح دقة المصنف

الفصل الخامس
النتائج و مناقشتها

8. النتائج

1.8 النتائج التي تم الحصول عليها عن طريق R programming

جدول (1.8) يوضح و صف معلومات التشغيل عن طريق R programming

| | |
|---------------------------|---|
| نموذج | SVM مع نواة خطية (linear SVM) |
| المثيل (instance) | 1151 |
| الخصائص Attribute | 21 |
| مجموعة التدريب و الاختبار | تقسيم 90 % لمجموعة التدريب ، 10 % لمجموعة الاختبار |

يحتوي هذا الجدول على اسم الخوارزمية المستخدمة تسمى SVM مع Kernel الخطي وتتألف مجموعة بيانات اعتلال الشبكية السكري من 1151 instance مع 21 attribute ، ومجموعة البيانات المنقسمة إلى مجموعة التدريب ومجموعة اختبار ، بنسبة 90% مجموعة تدريب بنسبة 10% مجموعة اختبار .

الجدول (2.8) يوضح مصفوفة التضارب (confusion matrix) في R programming

| | | Predicted class | |
|--------------|-----------------|-----------------|-------------|
| | | 0 not having DR | 1 Having DR |
| Actual class | 0 not having DR | 52 TP | 21 FN |
| | 1 Having DR | 8 FP | 34 TN |

= الدقة Accuracy

$$(52+34) / (52+21+8+34) = 74.78\%$$

= معدل الخطأ Error rate

$$(8 + 21) / (52+21+8+34) = 25.21\%$$

= معدل الإيجابي الحقيقي True Positive Rate (TPR)

$$TP / (TP + FN) = 52 / (52 + 21) = 71.23\%$$

= المعدل السلبي الحقيقي True Negative Rate (TPR)

$$TN / (TN + FP) = 34 / (34 + 8) = 80.95\%$$

= معدل الإيجابي الخاطئ False Positive Rate (FPR)

$$FP / (TN + FP) = 8 / (34 + 8) = 19.04\%$$

= False Negative Rate معدل السلبية الكاذبة

$$(FNR) = FN / (TP + FN) = 21 / (52 + 21) = 28\%$$

$$\%TP / (TP + FP) = 52 / (52 + 8) = 86.66\% = \text{precision الدقة}$$

$$TP / (TP + FN) = 52 / (52 + 21) = 71.23\% = \text{recall الاستدعاء}$$

الجدول (3.8) يوضح تلخيص النتيجة النهائية عن طريق R programming

| TPR | FPR | TNR | FNR | Precision | Recall | Error rate | Accuracy |
|--------|--------|--------|-----|-----------|--------|------------|----------|
| 71.23% | 19.04% | 80.95% | 28% | 86.66% | 71.23% | 25.21% | 74.78% |

Precision: النسبة المئوية للسجلات المصنفة حسب المصنف على أنها إيجابية وإيجابية بالفعل

Recall: النسبة المئوية للسجلات الإيجابية المصنفة حسب المصنف على أنها إيجابية

الإيجابي الحقيقي (TP): يصنف المريض على أنه يعاني من اعتلال الشبكية السكري وفي الحقيقة لديه.

إيجابي كاذب (FP): تصنيف المريض بأنه لا يعاني من اعتلال الشبكية السكري وفي الواقع لديه

2.8 النتائج التي تم الحصول عليها عن طريق Python

جدول(4.8): يوضح و صف معلومات التشغيل عن طريق Python

| | |
|---------------------------|---|
| نموذج | SVM مع نواة خطية(linear SVM) |
| المثيل (instance) | 1151 |
| الخصائص Attribute | 21 |
| مجموعة التدريب و الاختبار | تقسيم 90 % لمجموعة التدريب ، 10 % لمجموعة الاختبار |

يحتوي هذا الجدول على اسم الخوارزمية المستخدمة تسمى SVM مع Kernel الخطي وتتألف مجموعة بيانات اعتلال الشبكية السكري من 1151 instance مع 21 attribute ، ومجموعة البيانات المنقسمة إلى مجموعة التدريب ومجموعة اختبار ، بنسبة 90% مجموعة تدريب بنسبة 10% مجموعة اختبار

الجدول (5.8) يوضح مصفوفة التضارب (confusion matrix) في python

| | | Predicted class | |
|--------------|-----------------|-----------------|-------------|
| | | 0 not having DR | 1 Having DR |
| Actual class | 0 not having DR | TP 56 | FN 7 |
| | 1 Having DR | FP 18 | TN 35 |

= الدقة Accuracy

$$(56+35) / (56+7+18+35) = 78.44\%$$

= معدل الخطأ Error rate

$$(18 + 7) / (56+7+18+35) = 21.55\%$$

= معدل الإيجابي الحقيقي True Positive Rate (TPR)

$$TP / (TP + FN) = 56 / (56 + 7) = 88.88\%$$

= المعدل السلبي الحقيقي True Negative Rate (TNR)

$$TN / (TN + FP) = 35 / (35 + 18) = 66.03\%$$

= معدل الإيجابي الخاطئ False Positive Rate (FPR)

$$FP / (TN + FP) = 18 / (35 + 18) = 33.93\%$$

= معدل السلبية الكاذبة False Negative Rate

$$(FNR) = FN / (TP + FN) = 7 / (56 + 7) = 11.11\%$$

83.33% = precision الدقة

= recall الاستدعاء

66.03%

الجدول (6.8) يوضح تلخيص النتيجة النهائية عن طريق Python

| TPR | FPR | TNR | FNR | Precision | Recall | Error rate | Accuracy |
|--------|--------|--------|--------|-----------|--------|------------|----------|
| 88.88% | 33.93% | 66.03% | 11.11% | 83.33% | 66.03% | 21.55% | 78.44% |

الجدول (7.8) يوضح مقارنة النتائج النهائية

| | TPR | FPR | TNR | FNR | Precision | Recall | Error rate | Accuracy |
|------------------------------|--------|--------|--------|--------|-----------|--------|------------|----------|
| R programming SVM(linear) | 71.23% | 19.04% | 80.95% | 28% | 86.66% | 71.23% | 25.21% | 74.78% |
| Python SVM(linear) | 88.88% | 33.93% | 66.03% | 11.11% | 83.33% | 66.03% | 21.55% | 78.44% |

الفصل السادس

الخاتمة و التوصيات و المصادر و المراجع

9. الخاتمة

الحمد لله الذي بنعمته تتم الصالحات و الحمد لله الذي وفقنا لهذا العمل المتواضع و نتمني ان يكون هذا البحث بداية انطلاقا للبحوث في هذا المجال في هذا البحث ، تم استخدام خوارزمية نظام الدعم الالي (SVM) مع الكيرنل الخطي و تمت التجربة علي اداتين مختلفتين و هما البايثون python و R programming و تم الحصول علي نتائج مختلفة حيث حققت خوارزمية SVM التي ادرجت عن طريق R علي دقة (Accuracy) 74.78% و النسبة المئوية للسجلات المصنفة حسب المصنف على أنها إيجابية وإيجابية بالفعل يسمى الضبط (precision) علي 86.66% و النسبة المئوية للسجلات الإيجابية المصنفة حسب المصنف على أنها إيجابية يسمى الاستدعاء (recall) علي 71.23% . عند استخدام خوارزمية نظام الدعم الالي (SVM) التي ادرجت عن طريق python تم الحصول علي دقة (Accuracy) 78.44% و النسبة المئوية للسجلات المصنفة حسب المصنف على أنها إيجابية وإيجابية بالفعل يسمى الضبط (precision) علي 83.33% و النسبة المئوية للسجلات الإيجابية المصنفة حسب المصنف على أنها إيجابية يسمى الاستدعاء (recall) علي 66.03%.

10. التوصيات

1. استخدم المزيد من مجموعة البيانات للحصول على نتائج أفضل
2. التهيئة الصحيحة لمجموعة البيانات للحصول على دقة أفضل ونتائج أفضل
3. بناء عدة نماذج على مجموعة البيانات بدلاً من نموذج واحد

11. المراجع و المصادر

[1] R. Priya¹ and P. Aruna² “ DIAGNOSIS OF DIABETIC MACHINE LEARNING TECHNIQUES” RETINOPATHY USING ICTACT JOURNAL ON SOFT COMPUTING, JULY 2013, VOLUME: 03, ISSUE: 04

الساعة 6:17 م 15/10/2019

[2] Swati Gupta* and Karandikar AM” Diagnosis of Diabetic Retinopathy using Machine Learning” Journal of Research and Development DOI: 10.4172/2311-3278.1000127” Department of Computer Science and Engineering, Shri Ramdeobaba College Of Engineering and Management, Nagpur, India ,2015

الساعة 6:18 م 15/10/2019

[3] Mr. JaykumarLachure¹, M.A. V .Deorankar², Mr. Sagar Lachure³, Miss. Swati Gupta⁴ and M. Romit Jadhav⁵ “Diabetic Retinopathy using Morphological Operations and Machine Learning”, IEEE International Advance Computing Conference (IACC) 978-1-4799-8047-5/15/\$31.00 ©2015 IEEE,2015

الساعة 6:19 م 15/10/2019

[4] Harry Pratt^{a,*}, Frans Coenen^b, Deborah M Broadbent^c, Simon P Harding^{a,c}, Yalin Zheng^{a,c} Convolutional Neural Networks for Diabetic Retinopathy”

“International Conference On Medical Imaging Understanding and Analysis 2016, MIUA 2016, 6–8 July 2016, Loughborough, UK”
2016

الساعة 6:22 م 15/10/2019

[5] Kele Xu¹, Dawei Feng^{2,*} and Haibo Mi² “Deep Convolutional Neural Network–Based Early Automated Detection of Diabetic Retinopathy Using Fundus Image”

الساعة 6:22 م 15/10/2019

[6] Rishab Gargeya,¹ Theodore Leng, MD, MS² “Automated Identification of Diabetic Retinopathy Using Deep Learning “AMERICAN ACADEMY™ OF OPHTHALMOLOGY ,2017

الساعة 6:25 م 15/10/2019

[7] Gwenol'e Qu'el'ec^a, Katia Charri`ere^{b,a}, Yassine Boudi^{b,a}, B'eatrice Cochener^{c,a,d}, Mathieu Lamard^{c,a} ” Deep Image Mining for Diabetic Retinopathy Screening “arXiv:1610.07086v3 [cs.CV] 28 Apr 2017”
France ,2017

الساعة 6:27 م 15/10/2019

[8] Yehui Yang^{1,2}, Tao Li², Wensi Li³, Haishan Wu¹, Wei Fan¹ and Wensheng Zhang² “ proposed “Lesion detection and Grading of

Diabetic Retinopathy via Two-stages Deep Convolutional Neural Networks” arXiv:1705.00771v1 [cs.CV] 2 May 2017, china ,2017

الساعة 6:28 م 15/10/2019

[9] Ms.D.Pavithra¹ , Dr.A.N.Jayanthi² “A STUDY ON MACHINE LEARNING ALGORITHM IN MEDICAL DIAGNOSIS” ,International Journal of Advanced Research in Computer Science Volume 9, No. 4, July – August 2018, Department of Computer Science and Engineering¹, Department of Electronics and Communication Engineering² ,India 2018

الساعة 6:29 م 15/10/2019

”,Molecules journal, School of Information and Communication, National University of Defense Technology, Wuhan 430019, China, 2017

[10] Navoneel Chakrabarty “A Deep Learning Method for the detection of Diabetic Retinopathy” 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) , Computer Science and Engineering Jalpaiguri Government Engineering College Jalpaiguri, India2019

الساعة 6:30 م 15/10/2019

[11] M.Rajeswari¹, R.J.Nithya sri², P.Santhiya³, P.Saranya⁴“Diabetic Retinopathy Detection Using Tensor Flow Based on Machine Learning”

,International Journal of Innovative Research in Science, Engineering
and Technology, Vol. 8, Issue 3, March 2019

الساعة 6:32 م 15/10/2019

[12] JESSE VISLISEL AND THOMAS OETTING, MS, MD”
DIABETIC RETINOPATHY” September 1, 2010

الساعة 6:32 م 15/10/2019

[13] Asa Ben-Hur¹, Jason Weston²” A User's Guide to Support Vector
Machines”, Department of Computer Science Colorado State
University¹, Jason Weston NEC Labs America Princeton, NJ 08540
USA²,

الساعة 6:33 م 15/10/2019

[14] Hadley Wickham, Rice University “ R programming language”

<http://journal.r-project.org>

, <http://www.jstatsoft.org>

, <http://www.jstatsoft.org/v22>

[15] Dataset

<https://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrece+Data+Set>

[16] Python

<https://www.barmaje.com/topics/7>