وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم البرمجيات

تصنيف أمراض الجلد عند الأطفال باستخدام التعلم التجميعي

رسالة مقدمة
إلى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة ماجستير علوم في
البرمجيات

من قبل

فاطمه معن عبداللطيف حسين

بإشراف

أ. م. اسيل وليد علي حميد

١٤٤٧ هـ ٢٠٢٥

الملخص

تعد الإصابة بالأمراض الجلدية ولاسيما لدى الأطفال من الإشكالات الصحية بالغة الأهمية وتصنف كأحد أكثر الأمراض انتشاراً بين الأطفال في مراكز رعاية الطفل والمدارس، على الرغم من أن الاضطرابات المصاحبة للأمراض لا يتطلب الدخول الفوري إلى المشفى إلا أنها تمثل النسبة الكبيرة من المراجعات في استشارات طب الأطفال، تنعكس هذه الأمراض بشكل كبير على الاطفال وتؤثر على سلامتهم النفسية والجسدية. فقد تؤدي الى اضطرابات نفسية تنعكس على تفاعلاتهم اليومية وحياتهم بشكل عام.

تهدف هذه الرسالة إلى استخدام تقنيات المتقدمة لمعالجة هذه المشكلات عن طريق استخدام اساليب التعلم التجميعي Ensemble Learning(EL) حيث ساهمت في رصد هذه الأمراض في مراحلها الاولية وبدقة عالية للحد من تفاقم الإصابة وانتشارها بين الأطفال، كما يمكن إن تعد وسيلة فعالة لترشيد الوقت والموارد البشرية والمالية.

سَلطت هذه الرسالة الضوء على استخدام خمس فئات تعد من الأمراض الاكثر انتشاراً بين الاطفال بواقع 4141 صورة للأمراض (جدري الماء Chickenpox، الاكزيما Eczema، القوباء Impetigo، الشرى Urticaria، الثآليل Warts) وجمعت البيانات الخاصة بهذه الامراض يدوياً من مصادر متعددة من مكتبات Kaggle، ومن موقع DermNet ، ومن موقع atlas dermatologico. وبعد قراءة البيانات عولجت بتحويلها الي صيغة لونية موحدة وتغيير حجمها الى(224,224) وقُسمت إلى 80% البيانات المستخدمة في التدريب و20% البيانات المستخدمة في عملية الاختبار. تم اقتراح أربعة نماذج من اساليب EL لتصنيف أمراض جلد الأطفال وتمت تجزئتها إلى جزأين يتناول الجزء الأول نماذج التعلم التجميعي التقليدية والمتمثل (الغابة العشوائية Random Forest, XGBoost, Adaboost) والتي استخدمت في تصنيف الأمراض عن طريق تدريب (اشجار القرار Decision Tree) كمتعلمين أساسيين بعد إجراء المعالجة عليها واستخراج الميزات عبر استخدام تقنية HOG وتقليل الميزات المستخرجة باستخدام تقنية PCA والجزء الثاني تناول التعلم التجميعي العميق والمتمثل باستخدام نموذج التكديس Stacking الهجين RESLR والذي يجمع بين خوارزميتي Resnet50 والانحدار اللوجستي Logistic Regression كمتعلمين أساسيين والانحدار اللوجستي كمتعلم فوقي

لتصنيف البيانات بعد معالجتها واستخراج ميزاتها باستخدام خوارزمية ResNet50. تم استخدام مقاييس تقييم الأداء (Accuracy, Precision, Recall, F1-Score) لقياس أداء النماذج.

حصلت النماذج المقترحة على نتائج دقة متفاوتة حيث حصل نموذج التكديس Stacking الهجين (RESLR) على أعلى نسبة بلغت 97.33% ونموذج XGBoost على 93.71% ونموذج الغابة العشوائية Random Forest على 93.35% وبينما حصل نموذج AdaBoost على أقل دقة بلغت 92.87% ويلاحظ تفوق التعلم التجميعي العميق والذي يتمثل بنموذج التكديس Stacking على النماذج الأخرى وذلك بسبب الاستفادة من الدمج بين القوة والتعقيد لخوارزمية ResNet50 وسرعة ومرونة خوارزمية الانحدار اللوجستي، حيث ساعد ResNet50 على استخراج وتدريب ميزات عميقة ومفيدة ومعقدة من البيانات بينما استغل الانحدار اللوجستي هذه الميزات بطريقة فعالة وبسيطة مما جعل النموذج يتميز بالأداء العالي والكفاءة.

Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Software

Classification of Skin Diseases in Children using Ensemble Learning

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Master of Science
in
Software**

By

Fatima Maan Abdulatif Hussein

Supervised by

A. P. Aseel Waleed Ali Hamid

2025 A.D. **1447 A.H**

Abstract

Skin diseases, especially in children, is a very important health problem and are believed to be among the most common infectious diseases among children in schools and childcare centers. Although skin diseases do not require immediate admission to a hospital, they constitute the largest percentage of pediatric consultations. These diseases affect children's bodily and mental health and may cause mental problems that impact their general lives.

This thesis aims to use advanced technologies to address these problems by using Ensemble Learning methods, which contributed to monitoring these diseases in there initial stages with high accuracy to reduce the aggravation of the infection and its spread among children. It can also be considered an effective means of rationalizing time and human, and financial resources.

This thesis highlighted the use of five categories of the most common diseases among children (chickenpox, Eczema, Impetigo, Urticaria and Warts). Data on these diseases were collected manually from multiple sources from Kaggle libraries, DermNet, and the atlas dermatologico. After reading the data, it was processed by converting it to a unified color format and changing its size to (224,224) pixel and divided into 80% of the dataset used in training and 20% of the dataset used in testing process. Four ensemble learning models were proposed to classification skin diseases in children and were divided into two parts. The first part is traditional ensemble learning, represented by models (Adaboost, XGBoost, and Random Forest) used in disease classification, by training decision trees as base learners after preprocessing and extracting features using the HOG technique and reducing the extracted features using

PCA technique. The second part deals with deep ensemble learning, represented by using the hybrid stacking model (RESLR), which combines the ResNet50 and Logistic Regression algorithms as base learners and Logistic Regression as an meta learner to classification data after processing and extracting features using the ResNet50 algorithm. Using performance evaluation metrics (Accuracy, Precision, Recall, F1-Score) to measuring Models Performance.

The proposed models achieved varying accuracy results, as the hybrid Stacking model (RESLR) achieving the highest accuracy rate of 97.33%, the XGBoost model achieving 93.71%, and the Random Forest model achieving 93.35%. The AdaBoost model achieved the lowest accuracy rate of 92.87%. It is obvious that the deep ensemble learning model represented by the Stacking model outperformed the other models due to the combination of the power and complexity of the ResNet50 algorithm with the speed and flexibility of the logistic regression algorithm. ResNet50 helped extract and train deep, useful, and complex features from the data, while logistic regression exploited these features in an effective and simple way, making the model distinguished by its high performance and efficiency.