Ministry of Higher Education and
Scientific Research
University of Diyala
College of Science
Department of Computer Science

Skin Disease Classification Approach Based on Metaheuristic Optimization

A Thesis
Submitted to the Department of Computer Science\ College of Science\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science

By

Saja Salim Mohammed

Supervised By

Asst. Prof. Dr. Jamal Mustafa Al-Tuwaijari

2021 A.D. 1442 A.H.

بسم الله الرحمن الرحيم

﴿ خلق السموات والأرض بالحق وصوركم فأحسن صوركم واليه المصير ﴾

صدق الله العظيم

سورة التغابن

الآية 3

Dedication

I would like to dedicate this work to:

To who taught me that the champions will never be defeated, but they convert it to victory.

Our Prophet Muhammad

Peace be Upon Him (PBUH)

My Father, My Mother, My Husband, My Sister, My Brothers, and My Children.

Acknowledgment

*Firstly, all my prayers go to **Allah**, the Almighty, who helped me to accomplish this work until its end.*

*I wish to express my thanks to my college (college of science), my supervisor, **Ass- Prof. Dr. Jamal Mustafa Al-Tuwaijari** for supervising this research, and for the generosity, patience, and continuous guidance throughout the work. It has been my good fortune to have advice and guidance from him. My thanks to the academic and administrative staff at the Department of the computer sciences.*

I would like to thank my family who has supported me throughout and encouraging me to further my studies and help to complete this project, and I wish to express my thanks to my friends.

Saja Salim

ABSTRACT

Specialist doctors face great difficulties in diagnosing certain types of skin diseases due to the high similarity among these diseases such as seborrheic dermatitis, psoriasis, lichen planus, chronic dermatitis, pityriasis rosea, and pityriasis rubra. Predicting skin diseases became increasingly important in the general health of society, so automated, effective and accurate diagnosis systems of skin diseases are very important for biomedical analysis. Automated diagnostic technology proposes Feature Selection (FS) and classification methods to improve the efficiency of current diagnostic systems where metaheuristic algorithms were used in some of these systems.

This thesis presents Skin Disease Classification Approach based on a new FS technique by utilizing three metaheuristic algorithms: Sine Cosine Algorithm (SCA), Particle Swarm Optimization (PSO), and Antlion Optimization Algorithm (ALO). Thereafter, the original algorithms domain were converted to a binary version towards applying the classification to determine the best subset of features based on the wrapper model. Then, the mutation operator was applied as an internal function to preserve diversity and strengthen the SCA's exploration capabilities, and the new outcome of this mechanism is called Enhanced Sine Cosine Algorithm (ESCA). At this point, the system produces four results from the above four approaches.

The output of the used approaches is: ALO achieved rating accuracy of (94%) with the ratio of selected features (0.86) while PSO achieved rating accuracy of (89%) with the ratio of selected features (0.81) as for SCA achieved accuracy (95%) with the ratio of selected features (0.70) and finally ESCA achieved excellent diagnostic accuracy of (98%) with the ratio of selected features (0.62).

List of Tables

List of algorithms

List of Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
BALO	Binary Antlion Optimization Algorithm
BPSO	Binary Particle Swarm Optimization
BSCA	Binary Sine Cosine Algorithm
DM	Data Mining
DT	Decision Tree
ESCA	Enhanced Sine Cosine Algorithm
FS	Feature Selection
GA	Genetic Algorithm
IT	Information Technology
ML	Machine Learning
pm	Probability of Mutation
SVM	Support Vector Machines

Chapter One

Introduction

CHAPTER ONE

INTRODUCTION

1.1 Introduction

The skin is the outer covering of the body, as it protects the body from external influences and preserves the body's internal organs [1]. Skin diseases are a complex health problem that causes inconvenience and anxiety to humans due to their association with the aesthetic aspect, especially if it appears on the face, hands, or hair. Skin disorders come in a variety of shapes and types [2]. Skin diseases constitute a large percentage compared to other diseases due to the large area of human skin, as every part or area of the skin is exposed to different types of injuries [3]. Skin cancer has been increasing in rates in recent years [4][5]. Skin cancer is known as abnormal growth in skin cells [6]. The technical innovations that are emerging in the fields of health, medicine, and remote diagnosis have the potential to make a quantum leap in many countries to help provide medical dermatological advice and reduce financial costs[7].

Artificial Intelligence (AI) is complex mathematical algorithms that implement tasks similar to a human brain, means it can learn and understand the world as human do. Machine Learning (ML) is a main part of AI that uses data to build models that contribute to making decisions [8]. Recently, their applications have been widely used in medicine [9].

Diagnostic classification algorithms for different diseases have been developed to provide high accuracy for disease prediction [10][11][12][13]. AI is increasingly being used in the diagnosis of skin diseases and has achieved

promising results [14]. From diagnosis to personalized care, ML can change dermatologists' practices [15].

There are several applications for ML, the most significant of which is Data Mining (DM) [16]. DM can optimize decision-making in healthcare organizations by extracting new patterns and information from large amounts of data [17]. Healthcare organizations store a big amount of information in the daily records. Information Technology (IT) allows for the automation of DM, which aids in the discovery of hidden patterns that would be interested which eliminate manual works management and speed-up access to information by a direct method through electronic records systems that could minimize the cost of services [18]. Early diagnosis of diseases leads to better treatment for the patient and saves lives [19].

1.2 Related Works

Many authors have contributed to designing classification systems for the diagnosis of dermatology diseases.

In 2014, Menai and Altayash [20], proposed three models for the differential diagnosis of skin diseases which are the unpruned decision tree, pruned decision tree, and an ensemble of unpruned decision trees. The experimental results show that boosting decision trees can achieve a significant increase in performance in comparison with a single decision tree, especially when unpruned decision tree is used as the base learner. The best accuracy obtained was 96.72%.

In 2015, Rambhajan et al. [21], have employed the Bayesian technique for classifying the six skin diseases along with the Best First Search as a Feature Selection (FS) technique. The performance of the model is calculated using some

measures like accuracy, sensitivity, and specificity. After removing 20 features from the database, the obtained taxonomy rate is 99.31%.

In 2015, Liu et al. [22], suggested a new diagnostic procedure for the six skin disease diagnoses. The main novelty lies in the integration of Minimum Relevance Minimum Redundancy (MRMR) and Extreme Learning Machine (ELM), which allows classifying diseases in an efficient and fast way. Actual trials of the disease database demonstrated the accuracy of the proposed approach. The system can be employed to aid in the diagnostic decision-making of the six dermatological diseases as it has achieved promising results with an accuracy of 98.55%.

In 2016, Maghooli et al. [23], proposed to use the Classification and Regression Tree (CART) to predict the differential diagnosis of the six skin diseases. The proposed method is the Cross-Industry Standard Process for DM (CRISP-DM). The accuracy recorded by the proposed method reached 94.84%.

In 2016, Lahijanian et al. [24], developed a diagnostic approach based on multiple classifier systems for the diagnosis of the six diseases of the skin. The proposed approach includes two parts. Firstly, rough-set-based FS is applied to choose the best attributes from the original attributes set to shorten the response time and improve the accuracy. Secondly, building A group of three classifiers including MLP, KNN, SVM to make their decision on specific attributes; Finally, a polling technique is used to select the best among the classifiers to stabilize the critical decision. The empirical results showed an excellent classification accuracy of 97.78%.

In 2017, Tuba et al. [25], proposed a method for optimizing the Support Vector Machines (SVM) through a modern metaheuristic algorithm to achieve a promising diagnostic result of the six skin diseases. Parameters of the SVM

were adjusted by the Elephant Herding Optimization (EHO) algorithm. This method has achieved the highest accuracy with a training/test ratio of 80/20% is 99.07%.

In 2017, Maryam et al. [26], presented an approach by using a multiclass SVM classifier. The hybrid FS method, named ChiGA (Chi-Square and Genetic Algorithm) was used to optimal diagnose skin diseases. Chi-square is used as a filter method to remove redundant features and GA as a wrapper method to select the ideal feature subset with SVM used as a classifier. The resulting diagnostic accuracy was 99.18%.

In 2018, Idoko et al. [27], proposed a promising manner based on a fuzzy concept. The integrated fuzzy neural network (FNN) utilized the Takagi-Sugeno-Kang (TSK)-type rule as a taxonomy tool. The system merges between the ML abilities of the neural network and the fuzzy concept. The clustering technique was used for feature extraction. The total accuracy for the FNN classifier was 98.37%.

In 2019, Verma et al. [28], proposed different DM techniques for classifying skin disease. Five ML techniques SVM, CART, Decision Tree (DT), Random Forest (RF), and GBDT are used to classify the six skin diseases. The top accuracy obtained among these various techniques is 95.90% from GBDT. After that, scaled the dataset and again utilized these techniques and get the highest accuracy 97.27% in the case of ScaledRF. A multi-model ensemble approach is then applied combining these five DM techniques and obtained the highest accuracy of 98.64%. ML-based multi-modal collection method provides more information by using the first-stage prediction as a feature rather than a separate training and reduces generation errors. Also, by using ML techniques,

the complex relationships between classifiers are automatically learned, allow the collection method for better predictions.

In 2019, Bisen et al. [29], build a detection system using the Best First search(BFS), Sequential Floating Backward Search(SFBS), and Bayesian net. The proposed approach consists of three stages. In the first stage, the dimensions of the six skin diseases are reduced to 19 features using the BFS. In the second stage, SFBS is being exploited as a pre-processing step before the major classifier. Then, in the third stage, the Bayesian net is used as a classifier. This method has been named (BFSSFBS). The obtained classification accuracy of the proposed approach was 99.31%.

In 2020, Verma et al. [30], a hybrid approach were proposed by using three FS mechanisms Chi-Square, Information Gain, and Principle Component Analysis (PCA), and then combining them to pick the best subset of the data set. Six standard classifiers were used to evaluate model performance, Gaussian Naïve Bayesian, KNN, SVM, DT, RF, and Multilayer Perceptron (MLP). Boosting, Bagging, and Stacking ensemble techniques are utilized on-base classifiers to enhance the results of the proposed approach. The accuracy obtained by Bagging, Boosting, and staking are 95.94%, 97.70%, and 99.67%.

1.3 Problem Statement

The main problem with this work is that the classification of six skin diseases is difficult to distinguish for many dermatologists. The specialist doctors encounter confusion when making a clinical diagnosis of a patient with one of the six skin diseases due to the similarity of the external characteristics. To confirm the medical decision in such cases, a skin biopsy must be approved to obtain more information about the type of disease. ML techniques have shown the ability to detect skin diseases with high accuracy. Despite such high

accuracy, the use of all attributes in the dataset limits their performance in two main manners. First, the higher dimensionality of the dataset imposes the need for more complex DM models, which increases the computations required to achieve the required task. Hence, existing methods suffer from the need for a long execution time to predict the class of each input. The other limitation is imposed by the negative influence, i.e., confusion, that some of the features may impose on the DM model, when these features contain irrelevant data. To handle such a limitation, existing methods use different types of feature-selection techniques. However, in recent years, metaheuristic algorithms have shown significant improvement in the performance of DM in several applications. The dataset used is taken from the UCI repository site http://archiv.ics.uci.edu/ml/dataset/ dermatology [31].

1.4 The Aim of Thesis

Thesis aims to design an efficient and accurate approach for classifying skin diseases by using metaheuristic algorithms to assist specialists in diagnosing dermatological diseases in an accurate and fast manner based on computer technologies.

1.5 Thesis Overview

The remaining chapters are classified as follows:

Chapter2: Theoretical Background

This chapter presents an explanation of machine learning and its relationship to medical sciences and illustrates the dermatological diseases that have been used in this thesis, in addition, it gives a picture of metaheuristic algorithms and feature selection methods. Moreover, this chapter explains the

algorithms used in the dermatology classification approach, and then a summary of the evaluation criteria was performed.

Chapter3: The Proposed Approach

This chapter describes the proposed classification approach, its design, and implementation.

Chapter4: Experimental Results

This chapter shows the implementation results of the proposed approach steps and evaluates these results.

Chapter5: Conclusions And Future Works

It is the last chapter in which a set of conclusions are clarified and some future improvements are suggested.