ISSN: 2249-7196
IJMRR/Sep. 2024/ Volume 14/Issue 7/32-37
Sibghatullah / International Journal of Management Research & Review

ENHANCED DERMATOSCOPIC SKIN LESION CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

Sibghatullah[1], Mustafa Ali[2], Shaik Obaid Rahman[3], Khutaija Abid
[1,2,3] B.E. Student, Department of IT, Lords Institute of Engineering and Technology,Hyderabad
[4]Assistant Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
khutaija@lords.ac.in

Abstract: *Malignant melanoma poses significant impact on public health. Extensive research has focused on distinguishing between benign and malignant skin lesions through dermatoscopic image analysis. Our study emphasizes the classification aspect, using the MNIST HAM 10000 dataset. Initially, we addressed the challenge of imbalanced data by applying Synthetic Minority Oversampling Technique (SMOTE) to balance the dataset, significantly improving accuracy across various machine learning algorithms. Among these algorithms—Decision Tree (using Gini index and Entropy), Naïve Bayes, XGBoost, Random Forest, Logistic Regression, and Support Vector Machine (specifically with Polynomial kernel)—we found that Support Vector Machine with Polynomial kernel achieved the highest accuracy of 96.825%. While XGBoost, being a Gradient Boosting algorithm, showed varying results, we verified its accuracy using k-fold cross validation (k=10), achieving 95.984%. Ultimately, our findings highlight Support Vector Machine with Polynomial kernel as the most effective for this classification task[1]*

I. Introduction

In the realm of medical diagnostics, the classification of malignant skin lesions holds critical importance. Skin lesions often differ significantly from the surrounding skin area and can range from harmless to potentially dangerous. Identifying these lesions amidst environmental factors like noise, hair, and skin color poses a formidable challenge. The classification process involves multiple stages from imaging to final categorization.

Our study utilizes the MNIST HAM 10000 dataset, comprising 10,015 images representing various skin lesion diseases across seven types. We approach classification using a range of Machine Learning and Deep Learning algorithms, leveraging pixel-based image data stored in CSV format. Previous research (referenced in [12] and [13]) has explored different techniques such as Multiple Instance Learning, K-Nearest Neighbor, Decision Trees, Logistic Regression, Support Vector Machine, and Deep Learning for automated melanoma detection.

In our proposed system, we conducted two main experiments. Firstly, we enhanced classification accuracy by upsampling the dataset, which notably improved performance. Secondly, we compared the effectiveness of various algorithms, including Support Vector Machine (with Polynomial kernel), Naïve Bayes, Random Forest, Decision

Tree (using Gini index and Entropy), Logistic Regression, and XGBoost. Our findings indicated that Support Vector Machine with Polynomial kernel achieved the highest accuracy. We also evaluated a Baseline Convolutional Neural Network (CNN) model, which yielded an accuracy of 82.424%, but opted to focus on Machine Learning models due to our CSV data format.

The paper is structured into five sections: Introduction, Literature Survey on skin lesion classification (Section II), our Proposed System (Section III), Implementation Details and Results (Section IV), and Evaluation by comparing machine learning model accuracies (Section V)[2-4]

II. Literature Survey

1). Skin lesion classification of dermoscopic images using machine learning and convolutional neural network

Author:Bhuvaneshwari Shetty, Roshan Fernandes, Anisha P. Rodrigues

Detecting dangerous illnesses connected to the skin organ, particularly malignancy, requires the identification of pigmented skin lesions. Image detection techniques and computer classification capabilities can boost skin cancer detection accuracy. The dataset used for this research work is based on the HAM10000 dataset which consists of 10015 images. The proposed work has chosen a subset of the dataset and performed augmentation. A model with data augmentation tends to learn more distinguishing characteristics and features rather than a model without data augmentation. Involving data augmentation can improve the accuracy of the model. But that model cannot give significant results with the testing data until it is robust. The k-fold cross-validation technique makes the model robust which has been implemented in the proposed work. We have analyzed the classification accuracy of the Machine Learning algorithms and Convolutional Neural Network models. We have concluded that Convolutional Neural Network provides better accuracy compared to other machine learning algorithms implemented in the proposed work. In the proposed system, as the highest, we obtained an accuracy of 95.18% with the CNN model. The proposed work helps early identification of seven classes of skin disease and can be validated and treated appropriately by medical practitioners.[5]

2) Skin lesion classification from dermoscopic images using deep learning techniques

Author: Adria Romero Lopez; Xavier Giro-i-Nieto; Jack Burdick; Oge Marques

The recent emergence of deep learning methods for medical image analysis has enabled the development of intelligent medical imaging-based diagnosis systems that can assist the human expert in making better decisions about a patients health. In this paper we focus on the problem of skin lesion classification, particularly early melanoma detection, and present a deep-learning based approach to solve the problem of classifying a dermoscopic image containing a skin lesion as malignant or benign. The proposed solution is built around the VGGNet convolutional neural network architecture and uses the transfer learning paradigm. Experimental results are encouraging: on the ISIC Archive dataset, the proposed method achieves a sensitivity value of 78.66%, which is significantly higher than the current state of the art on that dataset.[6]

3)Skin Lesion Classification and Detection Using Machine Learning Techniques: A Systematic Review
Author:Taye GirmaDebele

Skin lesions are essential for the early detection and management of a number of dermatological disorders. Learning-based methods for skin lesion analysis have drawn much attention lately because of improvements in computer vision and machine learning techniques. A review of the most-recent methods for skin lesion classification, segmentation, and detection is presented in this survey paper. The significance of skin lesion analysis in healthcare and the difficulties of physical inspection are discussed in this survey paper. The review of state-of-the-art papers targeting skin lesion classification is then covered in depth with the goal of correctly identifying the type of skin lesion from dermoscopic, macroscopic, and other lesion image formats. The contribution and limitations of various techniques used in the selected study papers, including deep learning architectures and conventional machine learning methods, are examined. The survey then looks into study papers focused on skin lesion segmentation and detection techniques that aimed to identify the precise borders of skin lesions and classify them accordingly. These techniques make it easier to conduct subsequent analyses and allow for precise measurement and quantitveevaluations. The survey paper discusses well-known segmentation algorithms, including deep-learning-based, graph-based, and region-based ones. The difficulties, datasets, and evaluation metrics particular to skin lesion segmentation are also discussed. Throughout the survey, notable datasets, benchmark challenges, and evaluation metrics relevant to skin lesion analysis are highlighted, providing a comprehensive overview of the field. The paper concludes with a summary of the major trends, challenges, and potential future directions in skin lesion classification, segmentation, and detection, aiming to inspire further advancements in this critical domain of dermatological research.[7]

III.System Analysis

The incidence of skin cancer around the world is increasing year by year. However, early diagnosis and treatment can greatly improve the survival rate of patients. Skin lesion boundary segmentation is essential to accurately locate lesion areas in dermatoscopic images. It is true that accurate segmentation of skin lesions is still challenging dues to problems such as blurred borders, which requires an accurate and automatic skin lesion segmentation method. In existing system using KNN algorithm but it provides less accuracy. Disadvantage of exisitin system:

1. low efficiency.
2. knn algorithm is not sutable for skin lesion segmentation.

Algorithm:KNN algorithm[8]

Proposed system: In this paper, our focus is on preprocessing and classification of skin lesion images. Previous studies often used imbalanced datasets, leading to suboptimal accuracy. Many relied on limited or non-standardized datasets. Our approach uses the standardized HAM10015 dataset, consisting of 10,015 skin lesion images categorized into 7 types. The flow diagram of our proposed methodology is illustrated in Figure 1.

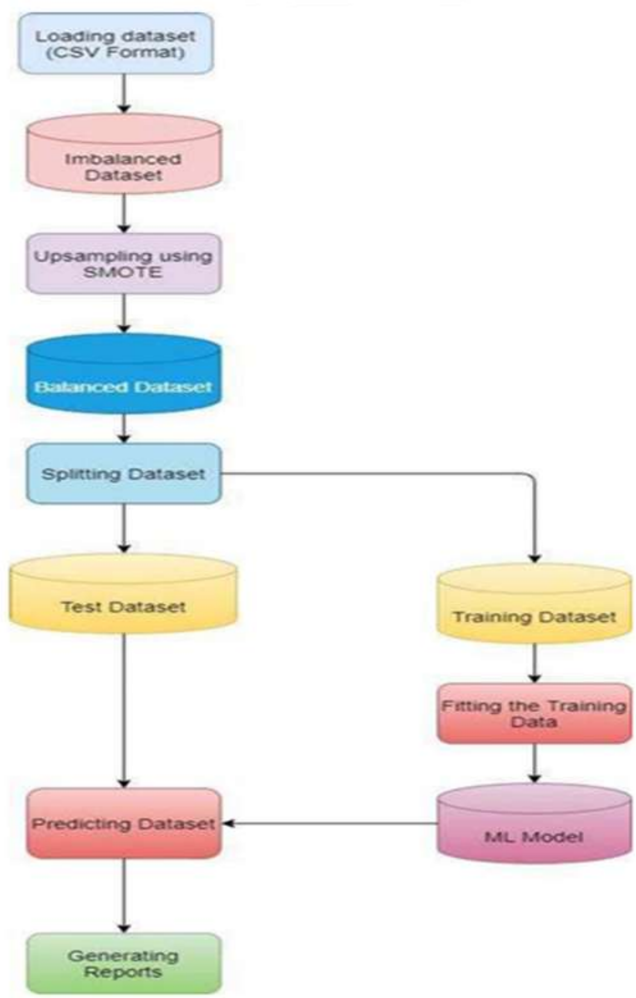
A. Preprocessing: To address the imbalance in the dataset, we employed Synthetic Minority Oversampling Technique (SMOTE) for up-sampling. This step resulted in each type of skin lesion having 6,705 samples.

B. Classification: Once the dataset was balanced, the next stage involved training machine learning models to classify the samples. We evaluated various algorithms to determine which provided the best accuracy. The algorithms considered include Naïve Bayes, Logistic Regression, Decision Tree classifier using Gini index and Entropy, Support Vector Machine with all kernel types, and XGBoost.

C. Algorithm: We utilized default parameters for each algorithm. The workflow of our proposed system is outlined as follows:

1. Load the dataset.
2. Balance the dataset using SMOTE.
3. Split the dataset into training and testing data.
4. Develop machine learning models (Naïve Bayes, Logistic Regression, Decision Tree with Gini index and Entropy, Support Vector Machine with all kernels, and XGBoost) using the training dataset.
5. Evaluate the models using the testing dataset.[9]

Figure 1 depicts the flow diagram of our proposed system.



IV.System Study

In recent studies, several approaches have been proposed to improve the classification and segmentation of skin lesions. For instance, bi-directional dermatoscopic feature learning (biDFL) [1] integrates dermatoscopic features with multi-scale consistent decision fusion (mCDF) to enhance classification accuracy. The system achieved 57.8% and 70.3% accuracy on the ISBI 2016 and 2017 databases, respectively.

In another study [3], a computer-aided diagnosis (CAD) system compared ABCD and three point checklist methods using algorithms like J48 Decision Tree, Random Forest, and Logistic Regression, finding better accuracy with the three point checklist method. They used the PH2 dataset with 200 images.

A system proposed in [4] used MobileNet V1 and Inception V3 pre-trained CNN models to classify skin lesions from the MNIST HAM 10000 dataset, achieving 72% accuracy with Inception V3 and 58% with MobileNet V1.

To address challenges like noise and intra-class discrepancy in melanoma classification, [5] proposed a Multiple Convolutional Neural Network approach on the ISIC 2016 dataset, achieving 97.78% AUC for training and 85.22% for testing.

For Atopic Dermatitis classification in [6], a multiclass SVM classifier trained on color, texture, and redness features achieved 86% accuracy with 10-fold cross-validation.

Skin lesion segmentation was addressed in [8] using a fully convolutional neural network (CNN) with a focus on minimizing negative multi-label Dice F1 scores, achieving an AUC value of 0.895%.

In [9], a novel approach using Machine Learning and Information Theory, specifically Harris Corner Detector and Havrda Entropy, achieved 92.45% sensitivity for skin lesion detection.

In [10], skin lesion classification based on ABCDE symptoms using Xception pre-trained neural network achieved promising results on the ISIC dataset.

Lastly, segmentation of dark skin lesions was targeted in [14] using the SDI+ unsupervised algorithm, which includes pre-processing, segmentation, and post-processing steps, resulting in improved accuracy for dark skin lesion segmentation."

This paraphrased version captures the essence of each study's methodology and findings while presenting the information concisely.[10-11]

V. Conclusion

Based on our findings, balancing the dataset significantly improves accuracy. In our study, we achieved this by up-sampling using SMOTE. We concluded that the Support Vector Machine algorithm with Polynomial kernel outperforms other machine learning algorithms in accuracy for dermatoscopic image classification. Our current research focuses exclusively on preprocessing and classification of these images.Looking ahead, our future work will concentrate on the segmentation and feature extraction processes of dermatoscopic images using deep learning models that take images directly as input data[12]

VI. References

[1] X.Wang,X. Jian,H.Ding, and J. Liu, "Bi-directional dermoscopic feature learning and multi-scale consistent decision fusion for skin lesionsegmentation,"IEEETransactionsonImageProcessing,vol.29, pp.3039–3051,2020.

[2]Kaushik Viknesh, P.Normal Kumar, and R.Seetharaman, "Computer aided diagnostic system for the classification of skin cancer using dermoscopic images," Proc. 2019 Third International Conference on, Inventive Systems and Control (ICISC),2019,pp.342–345.

[3] S.Krishna and A. S.Remya Ajai, "Analysis of three point check list,and,ABCD methods,for the, feature extraction of dermoscopic images to detect melanoma, "Proc. Ninth International Symposium on Embedded Computing and System Design(ISED),2019,.

[4] K. E. Purnama, A. K. Hernanda, A. A. Putri Ratnat, I. Nurtanio, A. N. Hidayati,M. H. Purnomo, S.M. Susiki Nugroho,and,R.F.Rachmadi,"Disease,classificationbasedondermoscopic skin images using convolutional neural network in tele dermatology system,"Proc. InternationalConferenceofComputerEngineering,Network,and,Intel,ligent,Multimedia(CENIM),2019,pp.1–5.

[5] Y. Guo, A. S. Ashour, L. Si, andP Mandalaywala, "Multiple convolutional neural network for skin dermoscopic image classification," Proc. 2018 IEEE International Symposium on Signal Processing and InformationTechnology (ISSPIT),2018,pp.365–369.

[6] R. Suhendra, F.Arnia, R. Idroes,N. Earlia, E. Suhartono, "A novel approach to multi-classa Topic Dermatitis Diseaseseverity, Scoringusing ,multi,classSVM,"Proc.,IEEEInternationalConferenceonCybernet,icsandComputational,Intelligence(IEEECYBERNETICSCOM), 2019, pp.35–39.

[7] N.M.Kinyanju, T.Odonga, C. Cintas,N. C. F. Codella, R. Panda, P.Sattigeri, and K.R.Varshney, "Estimating skin tone, effects on classification perform, an cein dermatology, datasets,"

[8]:J.Kawahara and G.Hamarneh,"Fully convolutional neural networks to detect clinical dermo scopic features,"IEEE Journal of Biomedical and Health Informatics, vol.23,no.2019,pp.578–585,2019.

[9]H.Fabelo,V.Melián,B.Martínez,P.Beltrán,S.Ortega,M.Marrero, G.M.Callicó,R.Sarmiento, I.Castaño,G.Carretero,P.Almeida,A. García,J.A.Hernández,andF.Godtliebsen,"Dermatologic hypers pectral imaging system fo rskin cancer diagnosis assistance, "Proc.of 2019 XXXIV Conferenceon Design of Circuits and Integrated Systems (DCIS), pp.1–6,2019.

[10] D. Gavrilov, L. Lazarenko, andE. Zakirov, "AI recognition in skin pathologies detection"Proc.2019International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI),2019,pp.54–542.

[11] Y. Tang, Z. Fang, S. Yuan, C. Zhan, Y. Xing, J. T. Zhou, and F. Yang," iMSC Gnet: Iterative multi-scalec on text-guided segmentation of skin lesion in Dermoscopic images "Published.

[12]E.Vocaturo,D.Perna,and E.Zumpano, "Machine learning techniques for automated melanoma detection," Proc. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM),2019,pp.2310 2317.