International Journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225

ISSN: 2319-4847

COMPARATIVE ANALYSIS OF MACHINE LEARNING METHODS FOR MULTI-LABEL SKIN CANCER CLASSIFICATION

M Neetha [1], Mamidi Sandeep [2], Thoom Siddhartha[3],K Venkata Tirugopala Reddy [4]

[1,2,3] UG Scholar, Department of IT,St. Martin's Engineering College, Secunderabad, Telangana,India – 500100

[4]Assistant Professor, Department of IT,St. Martin's Engineering College, Secunderabad, Telangana,India – 500100

mutha.neethasri@gmail.com

Article Info

Received: 28-03-2025 ***Revised: 05 -04-2025*** ***Accepted: 16-04-2025*** ***Published:27/04/2025***

Abstract:

Skin cancer is one of the most common and dangerous diseases due to a lack of awareness of its signs and methods for prevention. Skin cancer disease can be counted as a fourth burden disease around the world, with the rate of deaths dramatically growing globally. Therefore, early detection at an early stage is necessary to stop the spread of cancer. In this paper, we detect and classify multi-label skin cancer and implement the optimal techniques using machine learning and image processing approaches. However, preprocessing methods assist in removing irrelevant and unnecessary features from the label encoder, and standard features are applied to standardize the range of functionality by scaling the input variance unit. Moreover, various machine learning techniques were applied to check the performance of every classifier on the HAM10000_metadata dataset. The experimental analysis was conducted on the HAM10000_metadata dataset, which consists of seven different types of skin cancer. The results analysis shows that machine learning algorithms such as SVM, DT, and GNB obtained the highest accuracy compared to the other classifiers.

Keywords: Skin cancer, Machine Learning, Classification, SVM, DT, GNB, Multi-class

1.INTRODUCTION

Insufficient The disease of skin cancer is common and effected in many countries, which arises in humans, animals, and in flora, but it is an unusual asset of this disease in the interim. Due to the growing trend of skin cancer, it is having a huge impact on the world and creating a serious burden. Skin cancer is the fourth most commonplace cause of death in the throughout the world. It affects people of all ages, but it primarily affects children and the elderly [1]. The disease can be detected in its early stages and treated with a type of surgery. It is found in several forms, like melanoma, basal, and squamous cells [2]. The unpredictable form of cancer is melanoma. It influences the skin cells, hair follicles, and mucous membranes. Almost every skin cell that gets damaged can lead to skin cancer. It can affect almost anyone and occurs when cells in the skin mutate into cancerous (malignant) cells. Skin cancers are typically called carcinomas, and they can appear on any part of your body. Research has shown that high levels of radiation coming from the sun, such as UV rays, cause skin cancer [3]. Many people in the world suffer from skin cancer, although not everyone knows this. Skin cancer comes from the sun. UV rays are types of radiationpeople in the world suffer from skin cancer, although not everyone knows this. Skin cancer comes from damaging ultraviolet (UV) rays from the sun. UV rays are types of radiation that come from the sun and travel all the way through our atmosphere to get to Earth. It can occur in any area of the skin, and there are several

different types. While it's not always preventable, there are ways to dwindle your uncertainly of unstable skin cancer cells effeteness. Skin cancer is a serious issue and can be very dangerous [4]. It's significant to be conscious of the liability and to take precautions to preserve yourself from cancer waves. In the United States, the disease is very common [5]. According to the skin cancer foundation, more than millions of new cases confronting non-melanoma skin cancers (NMSC) were diagnosed in 2012, while more than 63,000 new cases of melanoma were diagnosed, making it the most dangerous form of skin cancer. Skin cancer occurs when skin cells begin to grow abnormally and multiply rapidly. When this happens at the lower layers of your epidermis, you have non-melanoma skin cancer (NMSC); when skin cells multiply deeper in your dermis, you have a melanoma [6].

2. LITERATURE SURVEY

Over the last two decades, researchers and scientists worked on machine learning techniques to control the multi-class and multi-level skin cancer automated classification. Nazia Hameed et.al, recommended skin cancer sweeping disability global hardship to enhanced approaches numerous diseases through machine learning and deep learning techniques. The performance of algorithms classification has capable to find out the better results enhancing multiple skin lesions [7]. A. Murugan et.al, reviewed implication of skin color purified and segmented to regulate the mean feature extraction anecdotal images segregate the data outcome in human diseases. The combination of SVM + RF algorithms results provide better accuracy as compared to other algorithms [8]. Carolina Magalhaes et.al, proposed on potential infrared thermography observation techniques implement machine learning methods on skin cancer. The exploring diagnosis of skin cancer retrieved the input thermal parameters-based ensemble learning represents better predictive performance of confusion matrix simplifying [9]. Mehwish Dildar et.al, suggested skin cells unrepaired deoxyribonucleic acid transmission continuously in the human parts glitch and alteration melanoma.

The detection of cancer should control at early stages symptoms seriousness medical treatment considering color, shape, symmetry etc. Before many researchers used various methods of machine learning

and deep learning techniques to detect the cancer but still need to improve the results of skin cancer [10].Yuheng Wang et.al, reviewed cancer detection complements polarization deep learning illustration image and malady statistical evaluate pattern ambitious to clarify in human. However, skin cancer categories into two type malignant and benign lesions classifying compared machine learning and deep learning approaches [11]. Rashmi Patil et.al, considered melanoma is an alarming and deadliest cancer detect over machine learning methods classification annoying patient diagnosed. The proposed system applied CNN model to measure loss function and text processing melanoma tumor thickness classification problem [12]. Ravi Manne et.al, deep learning approaches convenience to segregate the skin cancer placed on convolution neural network contribute the superior accuracy of the data by dermatologists. The CNN model reviewed successfully reduce misclassifying images to improve the accuracy through the deep learning techniques vulnerabilities [13].

Cancer Diagnosis Using Deep Learning: A Bibliographic Review.Munir, K., Elahi, H., Ayub, A., Frezza, F., & Rizzi, A In this paper, we first describe the basics of the field of cancer diagnosis, which includes steps of cancer diagnosis followed by the typical classification methods used by doctors, providing a historical idea of cancer classification techniques to the readers. These methods include Asymmetry, Border, Color and Diameter (ABCD) method, seven-point detection method, Menzies method, and pattern analysis. They are used regularly by doctors for cancer diagnosis, although they are not considered very efficient for obtaining better performance. Moreover, considering all types of audience, the basic evaluation criteria are also discussed. The criteria include the receiver operating characteristic curve (ROC curve), Area under the ROC curve (AUC), F1 score, accuracy, specificity, sensitivity, precision, dice-coefficient, average accuracy, and Jaccard index. Previously used methods are considered inefficient, asking for better and smarter methods for cancer diagnosis. Artificial intelligence and cancer diagnosis are gaining attention as a way to define better diagnostic tools. In particular, deep neural networks can be successfully used for intelligent image analysis. The basic framework of how this machine learning works on medical imaging is provided in this study, i.e., pre-processing, image segmentation and post-processing. The second part of this manuscript describes the different deep learning techniques, such as convolutional neural networks (CNNs), generative adversarial models (GANs), deep autoencoders (DANs), restricted Boltzmann's machine (RBM), stacked autoencoders (SAE), convolutional autoencoders (CAE), recurrent neural networks (RNNs), long short-term memory (LTSM), multi-scale convolutional neural network (M-CNN), multi-instance learning convolutional neural network (MIL-CNN). For each technique, we provide Python codes, to allow interested readers to experiment with the cited algorithms on their own diagnostic problems. The third part of this manuscript compiles the successfully applied deep learning models for different types of cancers. Considering the length of the manuscript, we restrict ourselves to the discussion of breast cancer, lung cancer, brain cancer, and skin cancer. The purpose of this bibliographic review is to provide researchers opting to work in implementing deep learning and artificial neural networks for cancer diagnosis a knowledge from scratch of the state-of-the-art achievements.

Introduction to Computer Vision and Real Time Deep Learning-based Object Detection.Shanahan, J., & Dai, L. Computer vision (CV) is a field of artificial intelligence that trains computers to interpret and understand the visual world for a variety of exciting downstream tasks such as self-driving cars, checkout-less shopping, smart cities, cancer detection, and more. The field of CV has been revolutionized by deep learning over the last decade. This tutorial looks under the hood of modern day CV systems, and builds out some of these tech pipelines in a Jupyter Notebook using Python, OpenCV, Keras and Tensorflow. While the primary focus is on digital images from cameras and videos, this tutorial will also introduce 3D point clouds, and classification and segmentation algorithms for processing them.

Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)Collins, G. D., Reitsma, J., Altman, D., & Moons, K. The TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) Statement includes a 22-item checklist, which aims to improve the reporting of studies developing, validating, or updating a prediction model, whether for diagnostic or prognostic purposes. The TRIPOD Statement aims to improve the transparency of the reporting of a prediction model study regardless of the study methods used. This explanation and elaboration document describes the rationale; clarifies the meaning of each item; and discusses why transparent reporting is important, with a view to assessing risk of bias and clinical usefulness of the prediction model. Each checklist item of the TRIPOD Statement is explained in detail and accompanied by published examples of good reporting. The document also provides a valuable reference of issues to consider when designing, conducting, and analyzing prediction model studies. To aid the editorial process and help peer reviewers and, ultimately, readers and systematic reviewers of prediction model studies, it is recommended that authors include a completed checklist in their submission. The TRIPOD checklist can also be downloaded from www.tripod-statement.org.

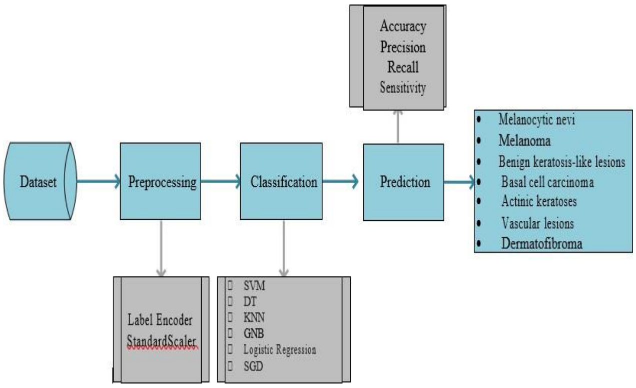
3. PROPOSED METHODOLOGY

International journal of application or innovation in engineering and management
Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

This proposed methodology focused on improving the visibility and quality of images captured under low-light or challenging lighting conditions. The primary goal of the proposed model is to enhance the details and visual appeal of such images, making them clearer and more visually appealing. It employs a deep learning-based approach to enhance low-light images. It utilizes techniques from computer vision, image processing, and deep neural networks to achieve its objectives. Overall, this research is designed to address the challenges posed by low-light images by applying deep learning-based techniques to enhance image quality, improve visibility, and provide visually appealing results. It finds applications in a variety of fields where low-light image enhancement is critical for obtaining meaningful and usable visual data.

Figure 1: Proposed methodology to Skin cancer classification

A. Preprocessing



The data is clean to achieve the maximum performance accuracy and efficiency using different machine learning techniques. Different data preparation techniques are used for the same coefficient of every feature. The label encoder features were used to convert the label data into numeric form so the machine learning technique could easily read the data. Furthermore, a standard scalar component was recycled to institutionalize the range of process of the input data and set by scaling to unit variance [13].

B. Classification

We used different classification techniques (SVM, DT, KNN, GNB, Logistic Regression, and SGD Classifier) for multi-label data.

Multi labels classification SVM

SVM is a supervised method recycling for classification and regression problems. The SVM performs the same principle which is utilized for binary classification problems. The multi- classification problem is divided into smaller sub problems. The One vs All (OVA) approach is a popular method which is used to perform multi-classification on the problem statement. In the One vs All approach, the hyperplane separates the classes and then divides them into two groups, where one group is for one class point and the other group is for all other points. Figure 2 shows that, the Greenline maximize the gap between the green point and all the other points [14].

Multi labels classification DT

The DT classifier is a systematic approach used for multi label classification. It poses different questions based on attributes and features and is visualized in a tree form. Each internal node in the root splits the data into separate records based on different characteristics. The leaves on the tree refer to the class in which the data is split into different categories [15].

Multi labels classification KNN

KNN is a supervised machine learning technique applying for classification. KNN algorithm does not depend on the structure of data. The distance between the two feature vectors can be solved through Euclidean distance [16].

Multi labels classification GNB

NB is a probabilistic ML technique used for classification based on Bye's theorem. Alternative functions are applied to calculate the data transportation, gaussian or normal distribution, for training data. In GNB, we requisite substitute the probability thickness of the normal distribution and also calculate the mean and variance of X [17].

Multi labels classification Logistic Regression

LR is a supervised ML technique used for classification. One-vs-rest allows the LR technique to be used for multi- class classification problems, which includes changing the loss functions and predicting the probability distribution to a multinomial probability [18].

Multi labels classification SGD Classifier

SGD is a straightforward right now very efficient path to proper linear classifiers and regressors held down convex loss functions such as (linear) Support Vector Machines and Logistic Regression. Although SGD has been over on the machine learning association for a long time, it has earned a noticeable amount of attention just newly in the context of large-scale learning [19].

C. Dataset

In this paper, we recycled the HAM10000_metadata data, which is convenient on the Kaggle repository. The data is collected from divergent populations and stored in different modalities. Dataset is machine learning database open source for researchers to observed on big data tackling from ML tasks. The dataset included two type of data training and testing through the machine learning models tracking the real-time prediction of the given features of the dataset [20]. The distribution of the values for different attributes such as gender, age, cell type, and localization are shown in the Figure 3. The behavior of the data shows that performing any ML techniques on such data will not provide

International Journal of application or innovation in engineering and management
Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847
an efficient result if the data is not properly grouped. 80% of the dataset was recycled for training purposes, and 20% of the dataset was applied for testing.

4. EXPERIMENTAL ANALYSIS

This section provides various classification models and their analysis. The six different4 machine learning techniques (SVM, DT, KNN, GNB, Logistic Regression, SGD) were used on the HAM10000_metadata dataset to check the performance of every classifier. First, the data is normalized and standardized before classifiers. The data is trained and tested on different machine learning classifiers in this experiment. 80% of the data was used for training and 20% for testing. The results of different machine learning algorithms is presented in Table 1 to Table 6.

The detail description of label mapping is given below:

Melanocytic Nevi (NV)

Melanoma (MEL)

Benign Keratosis-Like Lesions (BKL)

Basal Cell Carcinoma (BCC)

Intraepithelial Carcinoma / Bowen's Disease (AKIEC)

Vascular Lesions (VASC)

Dermatofibroma (DF).

Figure 1: Home Page

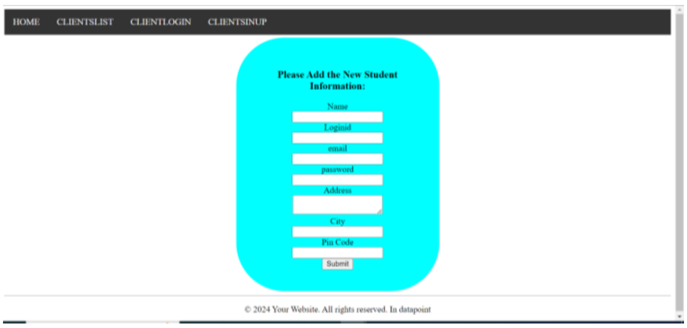


Figure 2: User Registration Form

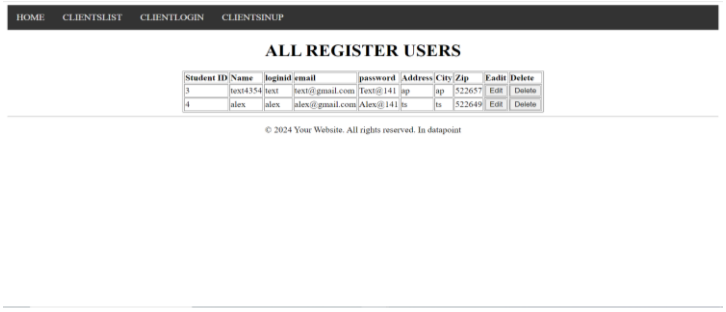


Figure 3: Registered User Form

Figure 4: Admin Login Form

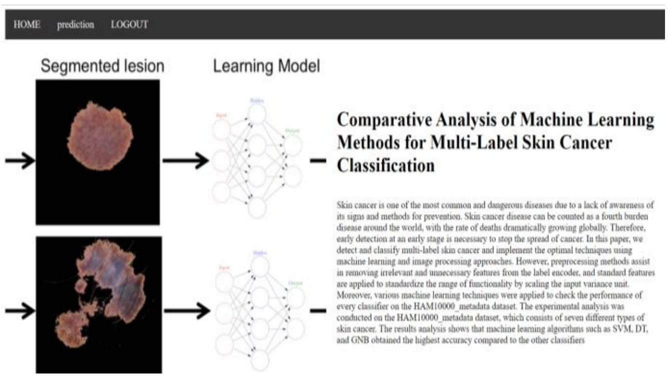


Figure 5: Admin Home page

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225

ISSN: 2319-4847

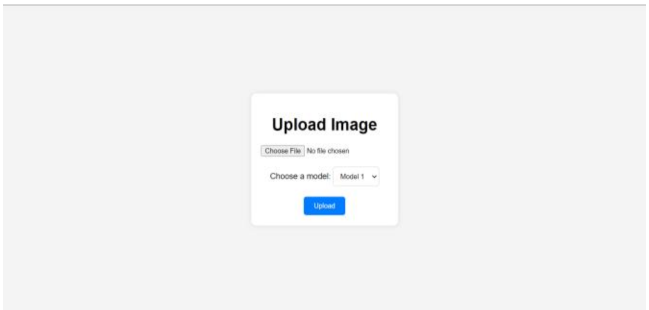


Figure 6: Prediction Form

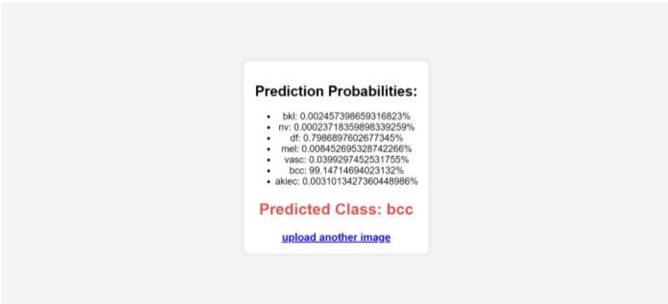


Figure 7: Expected Output

A. Confusion Matrix

Figure 4 represents the confusion matrix of skin cancer classification into seven different classes prediction outcomes. With the help of confusion matrix, we can easily visualize outcomes through machine learning algorithms. Furthermore, most datasets include imbalance data and more irrelevant instances of the data related other classes. The model has the ability to predict all the features to gain the highest accuracy score through confusion matrix approaches.

The confusion matrix shows that the SVM, DT, and GNB achieved better performance within 100% of accuracy, precision, recall, and F1-score during training and testing.

5. CONCLUSION

Skin cancer is a dangerous and widespread disease each year. According to global statistics, 5.4 million new cases of melanoma are reported in the United States each year, and 53.3 percent of melanoma cases are diagnosed. In this study, we utilized the perquisite of disparate machine learning techniques on the skin cancer HAM10000_metadata dataset to review the performance of seven diverse label datasets. The performance of different machine learning techniques shows that SVM, DT, and GNB performance are the maximum (100% accuracy) among other classifiers (KNN95% accuracy, Logistic regression 99%, and SGD achieved 80% accuracy). In the future, we aim to use more classification techniques to check the performance of this dataset.

REFERENCES

[1] A. Salam, F. Ullah, M. Imad, and M. A. Hassan, "Diagnosing of Dermoscopic Images using Machine Learning approaches for Melanoma Detection," in 2020 IEEE 23rd International Multitopic Conference (INMIC), 2020: IEEE, pp. 1-5.

[2] K. Das et al., "Machine Learning and Its Application in Skin Cancer", International Journal of Environmental Research and Public Health, vol. 18, no. 24, p. 13409, 2021. Available: 10.3390/ijerph182413409.

[3] E. Jana, R. Subban and S. Saraswathi, "Research on Skin Cancer Cell Detection Using Image Processing", 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2017. Available: 10.1109/iccic.2017.8524554.

[4] P. Dubal, S. Bhatt, C. Joglekar and S. Patil, "Skin cancer detection and classification", 2017 6th International Conference on Electrical Engineering and Informatics (ICEEI), 2017. Available: 10.1109/iceei.2017.8312419.

[5] I. A. OZKAN and M. KOKLU, “Skin Lesion Classification using Machine Learning Algorithms”, Int J Intell Syst Appl Eng, vol. 5, no. 4, pp. 285–289, Dec. 2017.

[6] D. Wen et al., "Characteristics of publicly available skin cancer image datasets: a systematic review", The Lancet Digital Health, vol. 4, no. 1, pp. e64-e74, 2022. Available: 10.1016/s2589- 7500(21)00252-1.

[7] N. Hameed, A. Shabut, M. Ghosh and M. Hossain, "Multi- class multi-level classification algorithm for skin lesions classification using machine learning techniques", Expert Systems with Applications, vol. 141, p. 112961, 2020. Available: 10.1016/j.eswa.2019.112961.

[8] A. Murugan, S. Nair, A. Preethi and K. Kumar, "Diagnosis of skin cancer using machine learning techniques", Microprocessors and Microsystems, vol. 81, p. 103727, 2021. Available: 10.1016/j.micpro.2020.103727.

[9] C. Magalhaes, J. Tavares, J. Mendes and R. Vardasca, "Comparison of machine learning strategies for infrared thermography of skin cancer", Biomedical Signal Processing and Control, vol. 69, p. 102872, 2021. Available: 10.1016/j.bspc.2021.102872.

[10] M. Dildar et al., "Skin Cancer Detection: A Review Using Deep Learning Techniques", International Journal of Environmental Research and Public Health, vol. 18, no. 10, p. 5479, 2021. Available: 10.3390/ijerph18105479 [Accessed 17August 2022].

[11] Y. Wang et al., "Deep learning enhances polarization speckle for in vivo skin cancer detection".

[12] R. Patil and S. Bellary, "Machine learning approach in melanoma cancer stage detection", Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 6,pp.3285-3293,2022.

[13] S. I. Ullah, A. Salam, W. Ullah, and M. Imad, "COVID-19 Lung Image Classification Based on Logistic Regression and Support Vector Machine," in European, Asian, Middle Eastern, North African Conference on Management & Information Systems, 2021: Springer, pp. 13-23.

[14] M. Imad, N. Khan, F. Ullah, M. A. Hassan, and A. Hussain, "COVID-19 classification based on Chest X-Ray images using machine learning techniques," Journal of Computer Science and Technology Studies, vol. 2, no. 2, pp. 01-11, 2020.

[15] A. Hussain, M. Imad, A. Khan and B. Ullah, "Multi-class Classification for the Identification of COVID-19 in X-Ray Images Using Customized Efficient Neural Network", AI and IoT for Sustainable Development in Emerging Countries, pp. 473-486, 2022. Available: 10.1007/978-3-030-90618-4_23

[16] M. Imad, A. Hussain, M. Hassan, Z. Butt and N. Sahar, "IoT Based Machine Learning and Deep Learning Platform for COVID-19 Prevention and Control: A Systematic Review", AI and IoT for Sustainable Development in Emerging Countries, pp. 523-536, 2022. Available: 10.1007/978-3-030- 90618-4_26

[17] M. Imad, F. Ullah, and M. A. Hassan, "Pakistani Currency Recognition to Assist Blind Person Based on Convolutional Neural Network," Journal of Computer Science and Technology Studies, vol. 2, no. 2, pp. 12-19, 2020.

[18] M. Imad, S. I. Ullah, A. Salam, W. U. Khan, F. Ullah, and M. A. Hassan, "Automatic Detection of Bullet in Human Body Based on X-Ray Images Using Machine Learning Techniques," International Journal of Computer Science and Information Security (IJCSIS), vol. 18, no. 6, 2020.

[19] M. Imad, S. I. Ullah, A. Salam, W. U. Khan, F. Ullah, and M. A. Hassan, "Automatic Detection of Bullet in Human Body Based on X-Ray Images Using Machine Learning Techniques," International Journal of Computer Science and Information Security (IJCSIS), vol. 18, no. 6, 2020.

[20] "Skin Cancer MNIST: HAM10000", Kaggle.com, 2022. [Online].