International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

Skin Cancer Detection Using Deep Learning

Nutan Chaudhari [1,*] , Sakshi More [2] , Virendra Gaikwad [3] Amruta Saraf[4]

[1]Department of Computer Engineering, Matoshri College of Engineering Nashik, Savitribai phule pune University

[2]Department of Computer Engineering, Matoshri College of Engineering Nashik, Savitribai phule pune University

[3]Department of Computer Engineering, Matoshri College of Engineering Nashik, Savitribai phule pune University

[4]Department of Computer Engineering, Matoshri College of Engineering Nashik, Savitribai phule pune University

Abstract: In cancer, there are over 200 different forms. Out of 200, melanoma is the deadliestform of skin cancer. The diagnostic procedure for melanoma starts with clinical screening, followed by dermoscopic analysis and histopathological examination. Melanoma skin cancer is highly curable if it gets identified at the early stages. Thefirst step of Melanoma skin cancer diagnosis is to conduct a visual examination of the skin's affected area. Dermatologists take the dermatoscopic images of the skinlesions by the high-speed camera, which have an accuracy of 65-80% in the melanoma diagnosis without any additional technical support. With further visual examination by cancer treatment specialists and dermatoscopic images, the overallprediction rate of melanoma diagnosis raised to 75-84% accuracy. The project aimsto build an automated classification system based on image processing techniques to classify skin cancer using skin lesions images.

Keywords: deep learning; machine learning; convolutional neural network; HAM10000; skin le- sion; ESRGAN

Introduction

1. Among all the skin cancer type, melanoma is the least common skin cancer, but itis responsible for **75%** of death SIIM-ISIC Melanoma Classification, 2020. Being a lesscommon skin cancer type but is spread very quickly to other body parts if not diagnosed early. The **International Skin Imaging Collaboration (ISIC)** is facilitatingskin images to reduce melanoma mortality. Melanoma can be cured if diagnosed and treated in the early stages. Digital skin lesion images can be used to make a teledermatology automated diagnosis system that can support clinical decision.

2. Currently, deep learning has revolutionised the future as it can solve complex problems. The motivation is to develop a solution that can help dermatologists better support their diagnostic accuracy by ensembling contextual images and patient-level information, reducing the variance of predictions from the model.tr

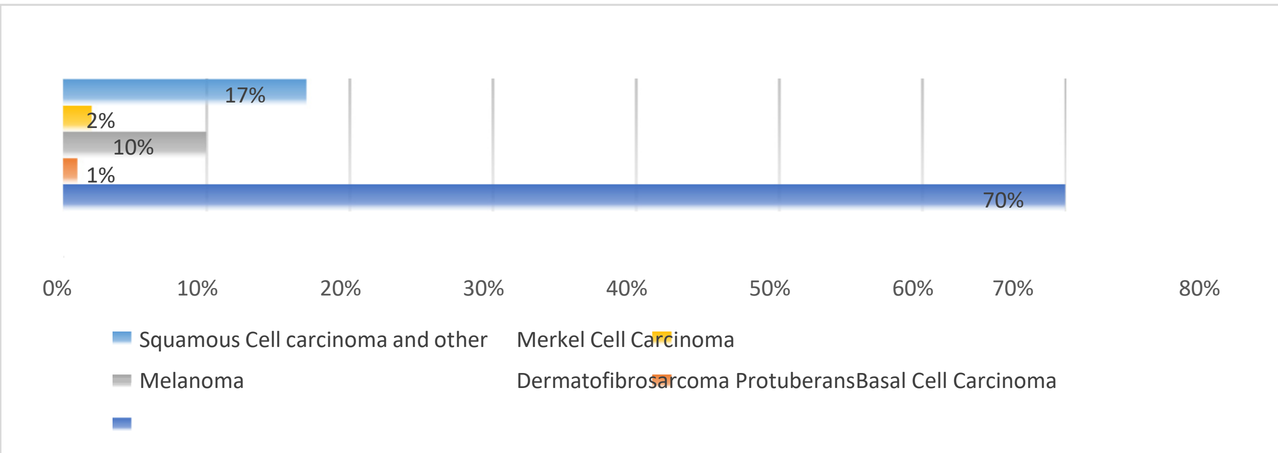
Motivation

The overarching goal is to support the efforts to reduce the death caused by skin cancer. The primary motivation that drives the project is to use the advanced image classification technology for the well-being of the people. Computer vision has made good progress in machine learning and deep learningthat are scalable across domains. With the help of this project, we want to reduce the gap between diagnosing and treatment. Successful completion of the project with higher precision on the dataset could better support the dermatological clinic work. The improved accuracy and efficiency of the modelcan aid to detect melanoma in the early stages and can help to reduce unnecessary biopsies

Application

We aim to make it accessible for everyone and leverage the existing model and improve the current system. To make it accessible to the public, we buildan easy-to-use website. The user or dermatologist can upload the patient demographic information with the skin lesion image. With the image and patient demographic as input, the model will analyse the data and return theresults within a split second. Keeping the broader demographic of people in the vision, we have also tried to develop the basic infographic page, which provides a generalised overview about melanoma and steps to use the online tool to get the results.

Figure 1. According to Reference [4], a number of different types of skin cancer are widespread.



A magnifying lens and light in dermatology are used to see medical patterns better such as hues, veils, pigmented nets, globs, and ramifications [7,8].. Visually impaired people can see the morphological structures that are otherwise hidden. These include the ABCD (Asymmetrical form, Border anomaly, Color discrepancy, Diameter, and Evolution) [9], 7-point checklist [10], and pattern analysis [11]. Non-professional dermoscopic images have a predictive value of 75% to 80% for Melanoma, but the interpretation takes time and is highly subjective, depending on the experience of the dermatologist [12]. Computer-Aided Diagnosis (CAD) approaches have made it easier to overcome these difficulties [8,12]. CAD of malignancies made a giant leap forward thanks to Deep Learning (DL)-based Artificial Intelligence (AI) [13,14]. In rural areas, dermatologists and labs are in poor supply; therefore, using DL approaches to classify skin lesions could help automate skin cancer screening and early detection [15,16]. To classify images in the past, dermoscopic images strongly depended on the extraction of handcrafted characteristics [17,18]. Throughout these promising scientific advances, the actual deployment of DCNN-based dermoscopic pictures has yielded amazing results. Still, future development of diagnosis accuracy is hampered by various obstacles, such as inadequate training data and imbalanced datasets, especially for rare and comparable lesion types. Regardless of the restrictions of the dataset, it is vital to maximize the performance of DCNNs for the correct Classification of skin lesions [14,19].
Models such as CNN, and modified Resnet50, are used in this research. We found that the invented CNN model beats existing DCNNs in classification accuracy while testing their performance on the HAM10000 dataset. To select the best network for diverse medical imaging datasets, it may be necessary to conduct multiple experiments. Accordingly, the paper's primary contributions can be summarized in this way:
We used an enhanced generative adversarial network with super-high resolution (ESRGAN) with 10,000 training photos to produce high-quality images for the Human Against Machine dataset (HAM10000 dataset [20]) to enhance the visibility of the images. ERSGAN improves the accuracy of Classification.
Segmentation is performed for each image in the dataset to specify ROI to facilitate the learning process.
We used Augmentation to ensure that the HAM10000 dataset had an even distribution of data.
The feasibility of the suggested system is determined by a thorough comparative eval- uation using numerous assessment measurements, such as accuracy, recall, precision, confusion matrix, top 1 accuracy, top 2 accuracy, and the F-score.
Pre-trained networks' weights are fine-tuned with the help of the HAM10000 dataset and a modified version of Resnet-50.
The recommended technique's overall effectiveness has been enhanced due to this change. Overfitting is prevented by using an alternative training process supported by applying various training strategies (e.g., batch size, learning rate, validation patience, and data augmentation).
This study provides an optimization strategy incorporating a CNN model a transfer learning model for detecting multiple skin lesions. Additionally, we utilized a revised form of Resnet-50 to train the weights of each Model before using it. Comparing the models' output using images of skin lesions from the HAM10000 dataset is necessary. The dataset has a class imbalance, necessitating an oversampling approach. The paper will proceed in accordance with this arrangement. Section 2 describes the relevant research work; after that, Section 3 illustrates the dataset and the proposed approach. The following Section 4 provides and analyzes the outcomes of the suggested technique described in Section 3; this study concludes with Section 5.

Related Work

The development of a CAD procedure for skin cancer has been the basis of several investigations [21,22]. CAD systems have followed the standard medical image analysis pipeline using classical machine learning approaches for skin lesion image processing [21]. In this pipeline, image preparation, fragmentation, extraction of features, and classifications have all been tried numerous times with little success. In skin cancer research, image processing, machine learning, CNN, and DL have all been used [23] in the past. Traditional image identification algorithms necessitate feature estimate and extraction, whereas deep learning can automatically exploit the images' deep nonlinear relationships [24,25]. CNN was the first DL model employed for skin lesion image processing. Some of the most recent deep learning studies are summarized in the following lines. For instance, Haenssle et al. [26] analyzed a Google Inception V4 deep learning model against 58 dermatologists' diagnoses. The data collection includes one hundred patients' images (dermoscopic and digitalized) and medical records. Additional research presented by Albahar [24] generated an improved DL model for detecting malignant Melanoma. Model results were compared to dermatologist diagnoses from 12 German hospitals, where 145 dermatologists used the Model to arrive at their conclusions. Li et al. [27] reviewed CNN deep learning models with 99.5 percent of the time; residual learning and separable convolution are the greatest methods for constructing the most accurate Model. This level of precision, however, was only possible since the problem was binary in nature.

For automated Diagnosis, Pacheco et al. [25] developed a smartphone app that used images of skin lesions and clinical data to identify them. The study looked at the skin lesions of 1641 persons with six types of cancer. An experimental three-layer convolutional neural network, GoogleNet, ResNet, VGGNet, and MobileNet, was compared by researchers. Initially, images of lesions taken with smartphones were used as teaching aids, but later, images of both sorts of lesions were included (clinical descriptions and images of skin lesions). The original Model's accuracy was 0.69 percent, but clinical data increased that to 0.764 percent. To improve upon Pacheco's findings, a new study was proposed. Based on dermal cell images, a model-driven framework for melanoma diagnosis was created by Kadampur and Riyaee [27]. With the help of the HAM10000 dataset, several deep-learning models attained an area under the curve (AUC) of 0.99. To categorize malignant and benign skin lesions, two CNN models were employed by Jinnai et al. [28]. Results of the Model were compared to dermatologists' diagnoses and found to have a superior classification accuracy than dermatologists, according to the results.

Furthermore, Prassanna et al. [29] proposed a deep learning-based system for high- level skin lesion segmentation and malignancy detection by building a neural network. It accurately recognizes the edge of a significant lesion and designs a mobile phone model using deep neural network transfer learning and fine-tuning to improve prediction accuracy. Another approach presented by Panja et al. [30] classifies skin cancer as melanoma or benign; feature extraction was used to retrieve damaged skin cell features using a CNN model after segmenting skin images. In [31], researchers classify ISIC 2019 Dataset photos into eight classes. ResNet-50 was used to train the Model by evaluating initial parameter values and altering them using transfer training. Images outside these eight classifications are classified as unknown.

Skin cancer detection relied heavily on the transfer learning idea. According to Kassem et al. [32], a study utilizing the GoogleNet pre-trained model for eight categories of skin cancer lesions produced an accuracy of 0.949. This time, a dermatoscope, a medical device used to examine skin lesions, was used to test the proposed YOLOv2-SqueezeNet's segmentation and drawback performance. Using the equipment considerably improved the capacity to make an early diagnosis. Table 1 shows that several deep-learning models have been implemented to categorize skin cancer in current history.

Dataset

The project dataset is openly available on Kaggle (SIIM-ISIC Melanoma Classification, 2020). It consists of around forty-four thousand images from thesame patient sampled over different weeks and stages. The dataset consists of images in various file format. The raw images are in **DICOM (Digital Imagingand COmmunications in Medicine)**, containing patient metadata and skin lesion images. DICOM is a commonly used file format in medical imaging. Additionally, the dataset also includes images in **TFRECORDS (TensorFlow Records)** and JPEG format.

Furthermore, thirty-three thousand are in training set among the forty-four thousand images and around eleven thousand in the test set. However, our quick analysis found a significant class imbalance in the training dataset. Thirty-two thousand are labelled as **benign (Not Cancerous)** and only five hundred marked as **malignant (Cancerous)**. That is, the training set contains only ±1.76% of malignant images (Figure 1). Along with the patient's images, the dataset also has a CSV file containing a detail about patient-level contextual information, which includes patient id, gender, patient age, location of benign/malignant site, and indicator of malignancy for the imagedlesion.

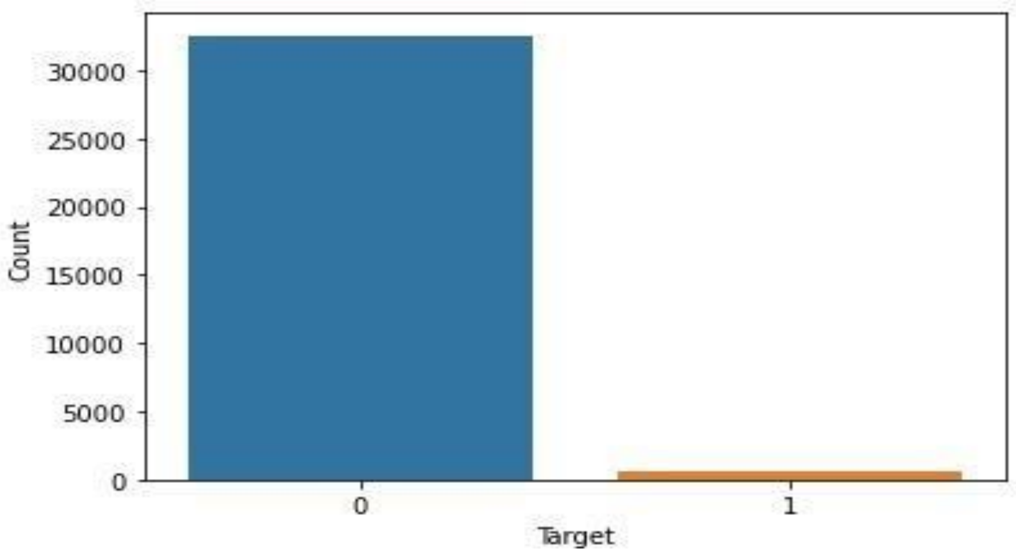


Figure 1: Class imbalance

 Overview of the Architecture

The project contains two flow diagrams.

he web UI contains five pages, of which four of them are used to explain the project andhow to use the proposed CAD system (Figure 3). The inference page named **"Our Solution"** is where the inference is made on the skin lesion images. All the validation is performed on the client-side to reduce the server overload. If the inserted information is not correct, then an error notification popup is shown; any user can easily understand that.Validated data is passed onto the server, where inference is performed by Onnx network, and response is return in the JSON format.

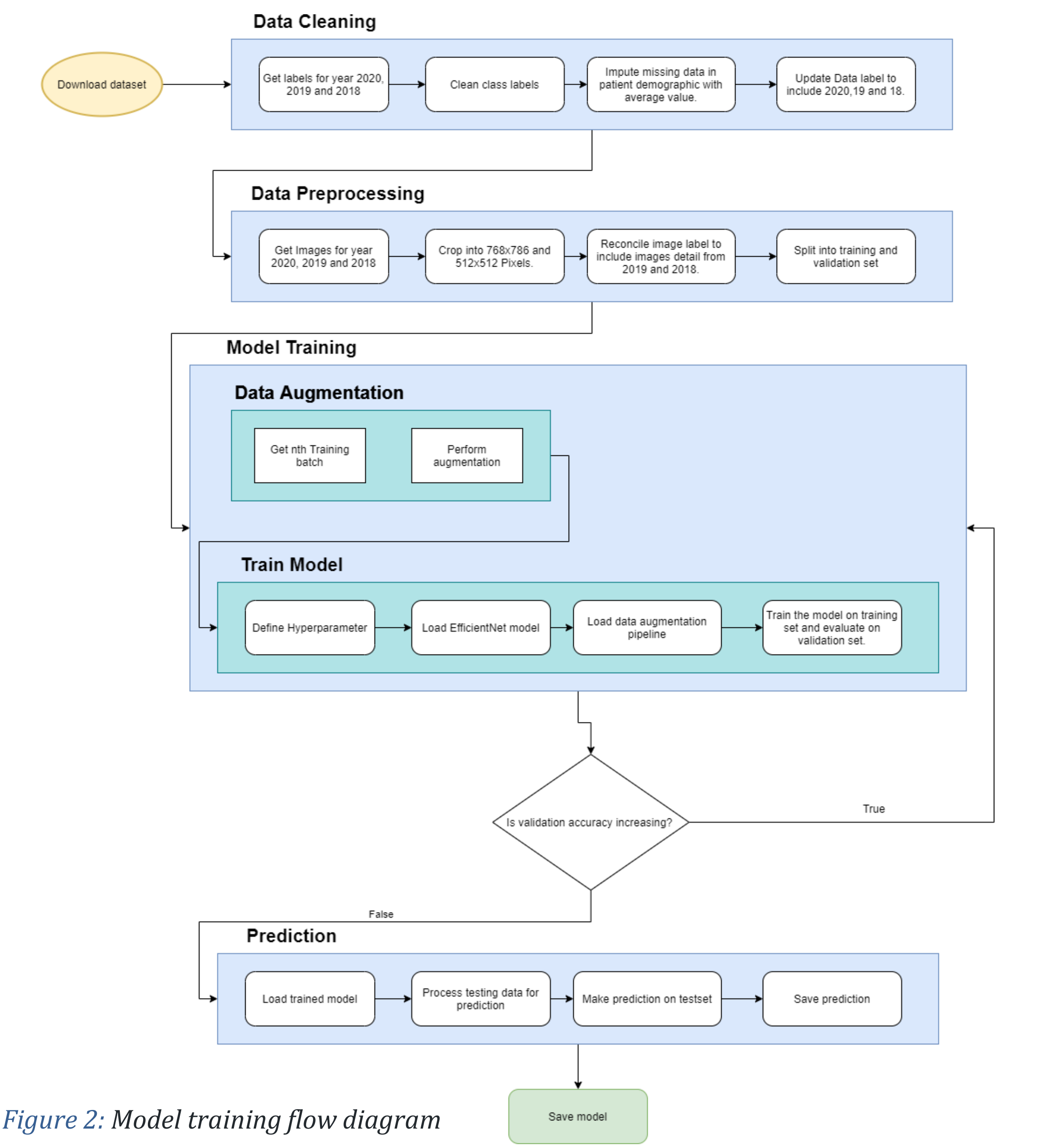


Figure 2: Model training flow diagram

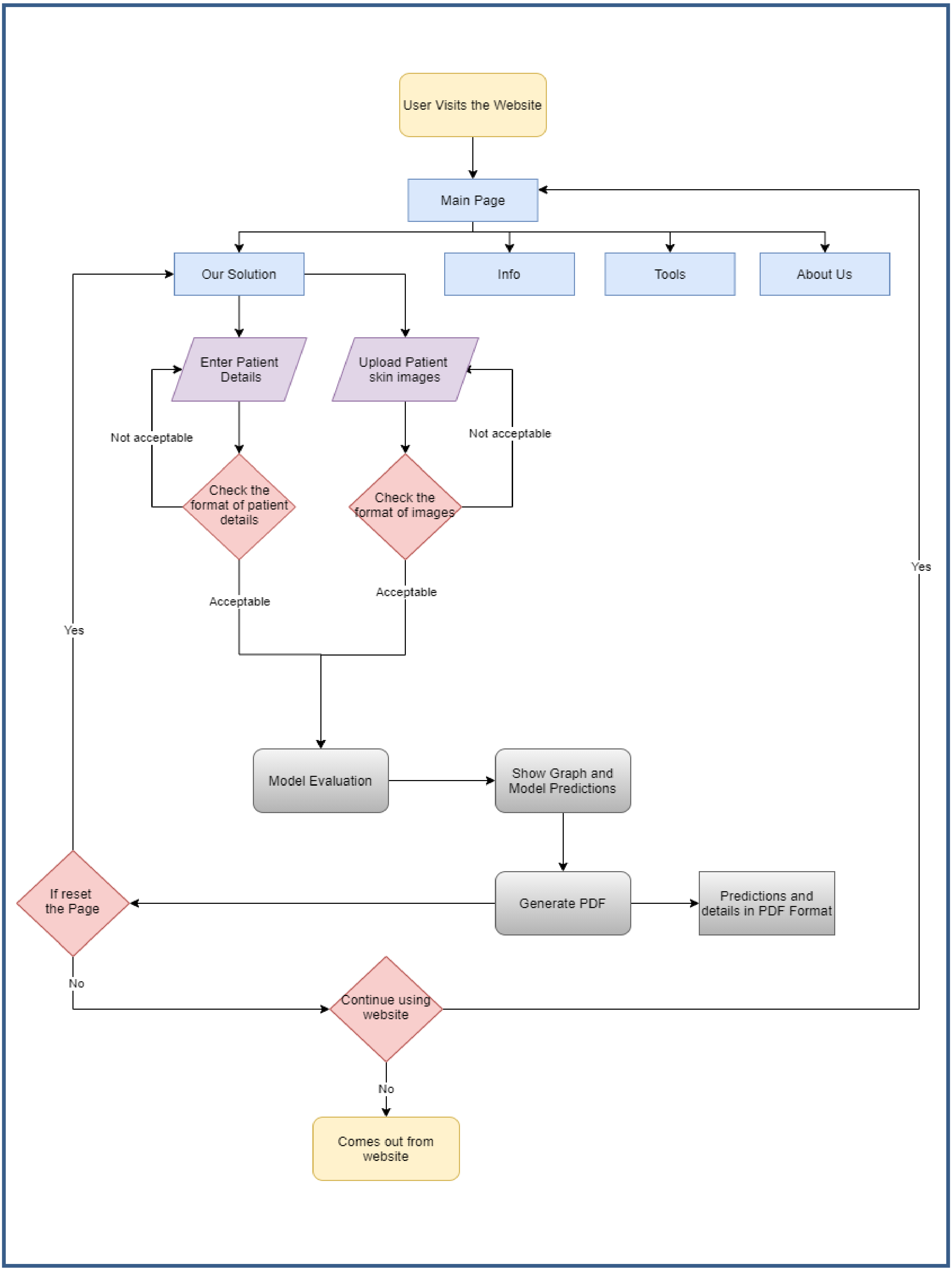


Figure 3: Web UI flow diagram

Module Of The Project

- Dataset
- Pre-processing
- Feature Extraction
- Classification

Description each model

Dataset –

Provide Dataset (This Means That The Data Collected Should Be Made Uniform And Understandable For A Machine That Doesn't See Data The Same Way As Humans Do.)example, HAM10000

Pre-processing –

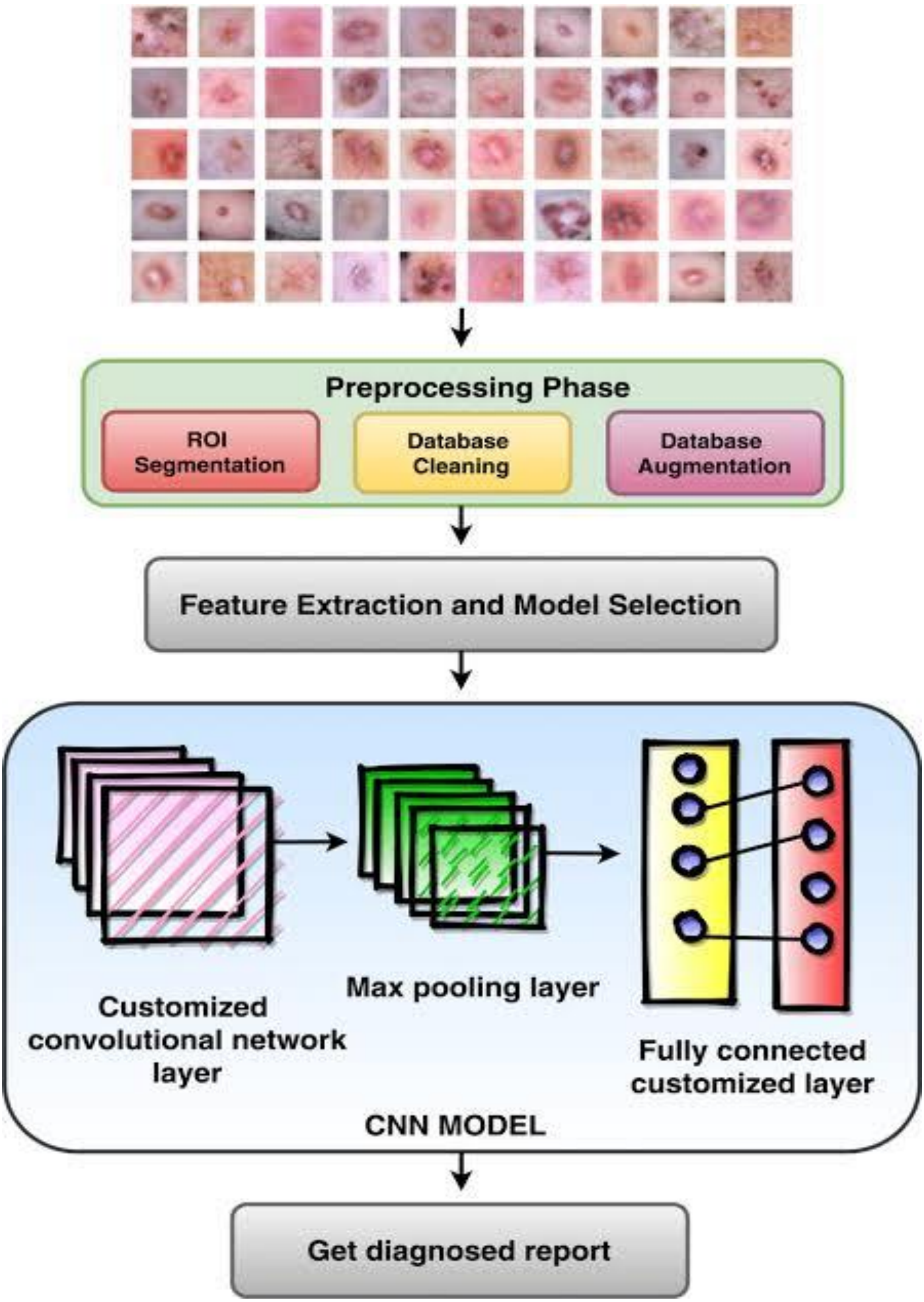
A Real-world Data Generally Contains Noises, Missing Values, And Maybe In An Unusable Format Which Cannot Be Directly Used For Machine Learning Models. Data Pre-processing IsRequired Tasks For Cleaning The Data And Making It Suitable For A Machine Learning ModelWhich Also Increases The Accuracy And Efficiency Of A Machine Learning Model

Feature Extraction –

Feature Extraction Aims To Reduce The Number Of Features In A Dataset By Creating NewFeatures From The Existing Ones (And Then Discarding The Original Features). These New Reduced Set Of Features Should Then Be Able To Summarize Most Of The Information Contained In The Original Set Of Features.

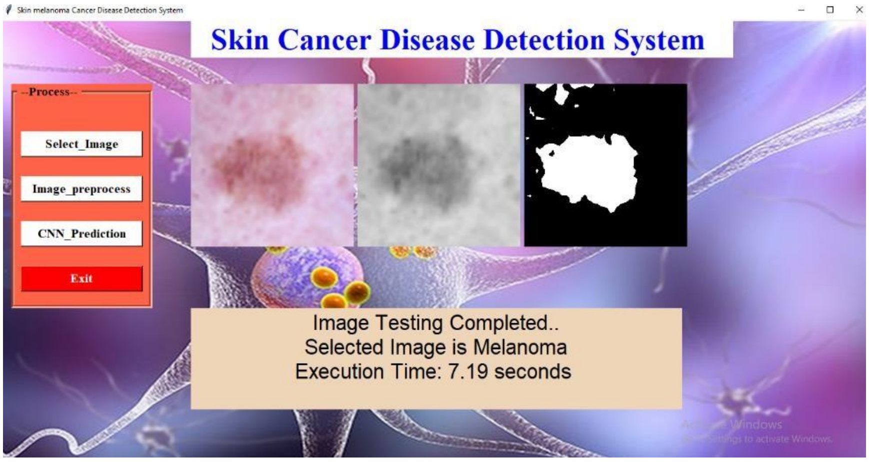
Classification –

The Classification Algorithm Is A Supervised Learning Technique That Is Used To Identify TheCategory Of New Observations On The Basis Of Training Data. In Classification, A Program Learns From The Given Dataset Or Observations And Then Classifies New Observation Into ANumber Of Classes Or Groups


Preprocessing Phase
ROI Segmentation
Database Cleaning
Database Augmentation
Feature Extraction and Model Selection
Customized convolutional network layer
Max pooling layer
Fully connected customized layer
CNN MODEL
Get diagnosed report

6.1 Conclusion

In this project, different phases of image processing were applied on skin Nodules. From these different image processing techniques, the fuzzy filter will provide the efficient de noising. Segmentation done by marker based watershed algorithm, gives various region of image. GLCM is used to extract thedifferent features of image and which takes less time for generating the result. This results are passed through CNN Classifier, which classifies the nodules as benign or malignant. CNN classifier provides 92.5 percentage accuracy.



6.2 Future Scope

System can be further modified into web application. The modules used inthis system can be used in Cognitive Computing model.

.

References

1. Saeed, J.; Zeebaree, S. Skin Lesion Classification Based on Deep Convolutional Neural Networks Architectures. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 41–51. [CrossRef]
2. Albahar, M.A. Skin Lesion Classification Using Convolutional Neural Network with Novel Regularizer. *IEEE Access* **2019**, *7*, 38306–38313. [CrossRef]
3. Khan, I.U.; Aslam, N.; Anwar, T.; Aljameel, S.S.; Ullah, M.; Khan, R.; Rehman, A.; Akhtar, N. Remote Diagnosis and Tri-aging Model for Skin Cancer Using EfficientNet and Extreme Gradient Boosting. *Complexity* **2021**, *2021*, 5591614. [CrossRef]
4. Nikitkina, A.I.; Bikmulina, P.Y.; Gafarova, E.R.; Kosheleva, N.V.; Efremov, Y.M.; Bezrukov, E.A.; Butnaru, D.V.; Dolganova, I.N.; Chernomyrdin, N.V.; Cherkasova, O.P.; et al. Terahertz radiation and the skin: A review. *J. Biomed. Opt.* **2021**, *26*, 043005. [CrossRef]
5. Xu, H.; Lu, C.; Berendt, R.; Jha, N.; Mandal, M. Automated analysis and classification of melanocytic tumor on skin whole slide images. *Comput. Med. Imaging Graph.* **2018**, *66*, 124–134. [CrossRef]
6. Namozov, A.; Im Cho, Y. Convolutional neural network algorithm with parameterized activation function for mela-noma classification. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 17–19 October 2018. Ozkan, I.A.; KOKLU, M. Skin lesion classification using machine learning algorithms. *Int. J. Intelli-Gent Syst. Appl. Eng.* **2017**, *5*, 285–289. [CrossRef]
7. Thamizhamuthu, R.; Manjula, D. Skin Melanoma Classification System Using Deep Learning. *Comput. Mater. Contin.* **2021**, *68*, 1147–1160. [CrossRef]
8. Stolz, W. ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma. *Eur. J. Der-Matol.* **1994**, *4*, 521–527.
9. Argenziano, G.; Fabbrocini, G.; Carli, P.; De Giorgi, V.; Sammarco, E.; Delfino, M. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Arch. Dermatol.* **1998**, *134*, 1563–1570. [CrossRef]
10. Pehamberger, H.; Steiner, A.; Wolff, K. In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions. *J. Am. Acad. Dermatol.* **1987**, *17*, 571–583. [CrossRef]
11. Reshma, G.; Al-Atroshi, C.; Nassa, V.K.; Geetha, B.; Sunitha, G.; Galety, M.G.; Neelakandan, S. Deep Learning-Based Skin Lesion Diagnosis Model Using Dermoscopic Images. *Intell. Autom. Soft Comput.* **2022**, *31*, 621–634. [CrossRef]
12. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
13. Yao, P.; Shen, S.; Xu, M.; Liu, P.; Zhang, F.; Xing, J.; Shao, P.; Kaffenberger, B.; Xu, R.X. Single model deep learning on im-balanced small datasets for skin lesion classification. *arXiv* **2021**, arXiv:2102.01284.
14. Adegun, A.; Viriri, S. Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art. *Artif. Intell. Rev.* **2020**, *54*, 811–841. [CrossRef]
15. Yang, J.; Sun, X.; Liang, J.; Rosin, P.L. Clinical skin lesion diagnosis using representations inspired by dermatologist cri-teria. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018, Salt Lake City, UT, USA, 18–23 June 2018.
16. Satheesha, T.Y.; Satyanarayana, D.; Prasad, M.N.G.; Dhruve, K.D. Melanoma Is Skin Deep: A 3D Reconstruction Technique for Computerized Dermoscopic Skin Lesion Classification. *IEEE J. Transl. Eng. Health Med.* **2017**, *5*, 1–17. [CrossRef] [PubMed]

17. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 180161. [CrossRef] [PubMed]

18. Smith, L.; MacNeil, S. State of the art in non-invasive imaging of cutaneous melanoma. *Ski. Res. Technol.* **2011**, *17*, 257–269. [CrossRef]

19. Masood, A.; Ali Al-Jumaily, A. Computer aided diagnostic support system for skin cancer: A review of techniques and algorithms. *Int. J. Biomed. Imaging* **2013**, *2013*, 323268. [CrossRef]

20. Efimenko, M.; Ignatev, A.; Koshechkin, K. Review of medical image recognition technologies to detect melanomas us-ing neural networks. *BMC Bioinform.* **2020**, *21*, 270. [CrossRef]

21. Brinker, T.J.; Hekler, A.; Enk, A.H.; Klode, J.; Hauschild, A.; Berking, C.; Schilling, B.; Haferkamp, S.; Schadendorf, D.; Fröhling, S.; et al. A convolutional neural network trained with dermoscopic images performed on par with 145 derma-tologists in a clinical melanoma image classification task. *Eur. J. Cancer* **2019**, *111*, 148–154. [CrossRef]

22. Pacheco, A.G.; Krohling, R.A. The impact of patient clinical information on automated skin cancer detection. *Comput. Biol. Med.* **2020**, *116*, 103545. [CrossRef] [PubMed]

23. Haenssle, H.A.; Fink, C.; Schneiderbauer, R.; Toberer, F.; Buhl, T.; Blum, A.; Kalloo, A.; Hassen, A.B.H.; Thomas, L.; Enk, A.; et al. Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2018**, *29*, 1836–1842. [CrossRef] [PubMed]

24. Li, L.F.; Wang, X.; Hu, W.J.; Xiong, N.N.; Du, Y.X.; Li, B.S. Deep learning in skin disease image recognition: A review. *IEEE Access* **2020**, *8*, 208264–208280. [CrossRef]

25. Kadampur, M.A.; Al Riyaee, S. Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. *Inform. Med. Unlocked* **2020**, *18*, 100282. [CrossRef]

26. Jinnai, S.; Yamazaki, N.; Hirano, Y.; Sugawara, Y.; Ohe, Y.; Hamamoto, R. The Development of a Skin Cancer Classification System