# Automatic Skin Cancer Detection with Convolutional Neural Networks

*@justgrossi*
*https://github.com/justgrossi/Portfolio.git*

***Abstract*—A Convolutional Neural Network-based model was developed for skin cancer detection using dermoscopic images. Trained on 10,000 images representing various skin lesions, the model underwent resizing and normalization. The finalized architecture included 3 convolutional and 3 max-pooling layers, a flattening layer, a fully connected layer, and an output layer. Tuning of epochs, batch size, and learning rate, coupled with data augmentation, achieved a peak accuracy of 74.5% with 20 epochs, a batch size of 32, and a learning rate of 0.0001.**

***Index Terms*—Convolutional Neural Network, Skin Cancer, Classification, Computer Vision.**

## I. Introduction

In 2020 the World Health Organization estimated 1.5 million new global cases of skin cancer. Over 120,000 of those caused premature deaths [1]. Genetic, environmental and risk factors such as a weakened immune system, the consumption of tobacco and the frequent use of sunbeds all increase the probability of developing malignant skin lesions. Continued exposure to excessive amounts of ultraviolet radiation, however, is the primary trigger. Ultraviolet radiation can neither be seen nor felt while it reaches skin's deeper layers damaging cells' deoxyribonucleic acid. Skin cancer is the most common type of cancer in Ireland. In 2020 over 13,000 people were diagnosed and one in four died prematurely [2]. Melanoma is the most common and deadliest form. In the United States alone and by the end of 2022 around 100,000 new melanomas will be discovered, and approximately 7,650 people will perish. Melanoma rates have been rapidly increasing over the past few decades [3]. As often happens with most diseases, skin cancer is easier to fight if diagnosed at an early stage. Features such as colour, size, shape and uniformity have been usually visually evaluated for decades to discriminate between malignant and benign skin lesions. Traditional detection methodologies are expensive and require time and professionally trained practitioners. In recent years artificial intelligence and image processing techniques have been progressively adopted to facilitate and optimize the process [4] [5] [6]. Among these computer vision and Convolutional Neural Networks (i.e. CNN) specifically play a key role. Early diagnosis dramatically improves survival probabilities. With new technologies constantly on the rise, self-checking through dedicated applications could become the new normal helping detect early signs of illness. That would allow the possibility to consult a specialist and begin treatment when the likelihood of success is still high. In recent years multiple and highly variable image-processing tasks have been approached using CNNs. Esteva et al. suggest they have the potential to achieve an accuracy matching dermatologists [7]. It is vital to emphasize though that computer-assisted diagnostics isn't meant to replace specialists. Dermatologists' role remains essential for both diagnosis and treatment.

## II. Related Works

Dildar et al. reviewed 1,483 papers on early skin cancer detection systems based on deep learning methodologies. Convolutional Neural Networks, Kohonen' self-organizing neural networks, and generative adversarial neural networks were the most promising techniques, and CNNs recorded the best results. Authors also reviewed the most common challenges. Hardware able to sustain the intensive training required, and the difficulty to correctly categorize smaller than 1mm or 2mm skin lesions being the most significant ones. The lack of diverse data sets encompassing samples of dark-skinned populations and/or of older people was highlighted, too. Current models, hence, appear to be relevant only for European, Australian, and North Americans younger than 65. Lastly, imbalanced data sets not including large-enough samples of less common skin cancers make it difficult to obtain conclusive categorizations [8]. Similar results and challenges were found by Takiddin et al. who reviewed 906 papers considering *shallow* and *deep* methodologies. The former use simpler architectures (e.g. Support Vector Machine, Näive Bayes, Logistic Regression, Random Forest and K-Nearest Neighbour) while the latter multy-layered ones. CNNs once again exhibited the best results. Given the diverse range of measures of performance, though, a direct comparison was not possible. Imbalanced data sets including fewer samples of less frequent cancer types once again hindered classification accuracy. The most accurate models were those trained on smaller data sets. Models' generalisability was, thus, dubious [9].

In [10] an accuracy of 89.5% was achieved by processing a sample of 23,907 dermoscopic images extracted from "*The International Skin Imaging Collaboration*" [11] archive. Due to computational limitations, the images had to be resized. Additionally, to alleviate the computational load during network training, grayscale images were derived from the original ones. Beyond 80 neurons per hidden layer, reductions in accuracy were observed. Comparable accuracies were reported in [12] and [13]. In [12] an ensemble comprising *Inceptionv3* and *DenseNet-201* was formed, resulting in an

enhanced test set accuracy of 88.5%. Various CNNs—*VGG16, AlexNet, DenseNet201*, and *Inceptionv3*—were compared against standard models. The ensemble model of *Inceptionv3* and *DenseNet-201* emerged as the most effective approach. Employing the whale optimization algorithm during network training yielded an accuracy of 88.5%. In [14] authors pre-processed images by employing LAB transformation to emphasize lesion regions. They employed a pre-trained *ResNet-50* model consisting of 175 layers, including convolutional, *ReLU*, pooling, batch normalization, sum, and fully connected layers. Subsequently, they applied sigmoid activation on probe layers and executed a *ResNet101* pre-trained model comprising 347 layers, featuring convolutional, *ReLu*, pooling, batch normalization, addition, fully connected, and classification layers, followed by a FC layer returning an output. Feature selection using kurtosis controlled PCA (KcPCA) was then conducted, and the chosen features were input to an SVM classifier, exploring multiple kernel functions (SVM linear, quadratic, cubic, FGSVM, MGSVM, CGSVM, RBF), with RBF delivering the best outcomes. Their model achieved an 89.71% accuracy.

In [15] Google's *MobileNet* is adapted to skin lesion classification preserving the initial layers: convolutional layers, 13 depthwise convolution layers, and pointwise convolution layers. Batch normalization (BN) and *ReLU* activation were applied after each convolution layer, followed by a global average pooling layer to reduce feature map size. The last five layers were replaced by a dropout layer and Fully Connected layer (employing *Softmax*) in their adapted version. Additionally, data up-sampling was implemented to balance the dataset. Their achieved results were an accuracy of 83.23%, specificity of 87%, sensitivity of 85%, and F1 score of 82%. In [16] researchers delved into the relationship between dataset size, the number of diagnostic classes, and model performance metrics. Their findings questioned the reliability of small and deep models, particularly when trained and tested on datasets containing only a few diagnostic classes.

In [17] a CNN-based model's performance is compared with that of 145 medical professionals in classifying dermoscopic images. The CNN demonstrated performance on par with or better than the medical professionals. Uncertainties however remain regarding the influence of sample resolution and the predominance of fair-skinned samples on outcomes suggesting potential variations compared to darker skin cases. Addressing the common issue of relatively small sample sizes in image classification, [18] proposed a novel ensemble CNN approach to build a model based on a small dataset of 2000 images. The authors utilized a combination of *GoogLeNet, AlexNet, ResNet*, and *VGGNet* architectures commonly used in similar problems. This approach offers a promising avenue for further exploration in selecting the most suitable architecture for the specific problem under consideration. In [19] researchers employed CNN with transfer learning from dermoscopic melanoma images, achieving an F1 score of 83.61% using *MobileNet* within a total training time of 11 seconds. [20] introduced ODIN, a method to detect out-of-distribution data samples in neural networks without significant retraining, substantially improving the baseline method. Titus J. utilized CNNs in [21] on 804 biopsy dermoscopic images of melanoma and nevi. They found that AI algorithms effectively aided dermatologists in melanoma detection but advised implementing these algorithms after the clinical diagnosis to avoid dermatologist bias. In [22] deficiencies and safety issues in AI diagnostic systems for skin cancer were highlighted. This study was the first to identify lesions where expert readers outperformed automated approaches. Comparisons between KNN, Bag-of-Features, HOG with SVM classifiers in [23] revealed SVM's superiority in skin cancer detection, achieving 96% accuracy compared to KNN's 91.3%. They suggested improving accuracy through a larger training dataset and better pre-processing techniques. A systematic review in [24] evaluated the use of AI and machine learning algorithms for early skin cancer detection in primary care settings. The paper identified 11,296 studies using these algorithms and critiqued the poor model descriptions and unspecified dataset appropriateness. It emphasized the early stage of most studies and the limited clinical support, suggesting a combined approach involving algorithms and clinical expertise for greater success. The paper by [25] introduces a different approach by building a Full Resolution Convolutional Network (FrCN) model without reshaping or resizing images but using higher resolution images for training. This FrCN model operates at full image resolution, learning from each individual pixel, and enhances learning by enlarging images in the training dataset, thus incorporating each pixel as a training sample. The method's emphasis on detailed pixel-level learning contrasts with other reviewed papers, showing that increased granularity and a larger quantity of images lead to improved accuracy in models.

In contrast, [26] evaluates a deep learning convolutional model against fifty-eight dermatologists. This study diverges from typical methodologies by directly comparing the performance of the model against a group of dermatologists, presenting a unique assessment perspective. The paper acknowledges numerous individual tests on CNNs and medical trials for melanoma detection but highlights a scarcity of simultaneous comparative trials. In this study, a relatively small sample of 300 images was chosen for testing, acknowledging potential inaccuracies due to the sample size. The CNN model was trained on 100 images, and the remaining 200 were used for testing. Dermatologists, considering the time feasibility, assessed 100 images each. The comparison revealed the CNN model's superior accuracy, albeit under unequal testing conditions, as stated above. The study emphasizes that dermatologists employ additional techniques beyond visual examination for skin lesion analysis, noting that integrating clinical information significantly enhances their accuracy. In [27] the focus is on integrating AI-based support systems into clinical care. Given challenges in patient access to clinical experts, especially during the COVID pandemic, the study

proposes leveraging AI-based image detection accuracy alongside human opinions for optimal patient care. The study trains a CNN model from a group of images using methodologies akin to those described in other related papers. In this study, human participants comprised 169 (56.0%) board-certified dermatologists, 77 (25.5%) dermatology residents, and 38 (12.6%) general practitioners. Raters initially diagnosed a batch of images without AI-based support and subsequently, with support from CNN results. The study observed a notable improvement in human raters' accuracy, rising from 63.6% to 77.0%, with decision support from AI-based multiclass probabilities. This aligns with the conclusion from [26] and [27], suggesting that collaboration between clinical practitioners and AI would benefit individuals seeking treatment. Additionally, in the initial section of [28], it is highlighted that similar types of skin lesions often lead to inaccurate diagnoses, potentially resulting in undetected cancerous lesions. The paper emphasizes the potential of Computer-Aided Diagnosis (CAD) in aiding dermatologists with early diagnosis, potentially saving lives based on accuracy.

## III. Methodology

The Cross Industry Standard Process for Data Mining (CRISP-DM) [29] was chosen as the paper's guiding methodology. The research question was approached and facilitated by framework's six stages.

### A. Business Understanding

The aim is to construct a model capable of discerning diverse types of skin lesions, especially those with the potential for skin cancer development, with the ultimate goal of reducing mortality rates through early detection. These models hold promise in aiding medical professionals by early identification of possibly cancerous lesions, guiding patients to specialists for further diagnostic tests to confirm the presence of cancer. Furthermore, as technology advances, the prospect of developing apps enabling regular self-examinations by patients becomes feasible, facilitating the detection of even minor skin changes at an early stage. However, given that no technology is infallible, the intention isn't for individuals to solely rely on diagnostic support. If noticeable and significant changes are noticed, seeking consultation with a medical specialist is crucial, even if the technology didn't detect cancer. Constructing models capable of categorizing skin lesions solely from images provides a cost-effective means of initial screening before referral to a dermatologist. This model, as proposed, streamlines the process by allowing prioritization of appointments based on image classifications, avoiding the need for multiple preliminary appointments.

### B. Data Understanding

The HAM10000 dataset [30] consists of 10,015 skin lesion images meticulously gathered by researchers to address the scarcity of dermatoscopic image datasets. This comprehensive collection encompasses Melanocytic nevi, Melanoma, Benign keratosis-like lesions, Basal cell carcinoma, Actinic keratoses, Vascular lesions, and Dermatofibroma lesions. These images were obtained from the Harvard Database, and an accompanying metadata file categorized each image based on lesion type, patient gender and age, lesion location, and the diagnostic test conducted. Figure 1 exhibits a selection of the initial 25 images, showcasing the distinct cancer types each image represents through accompanying labels.

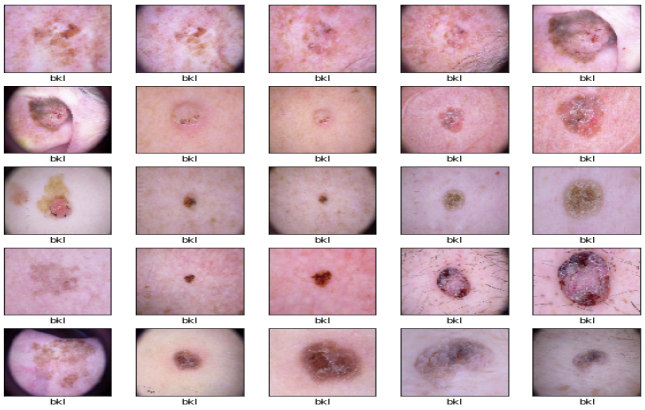

Fig. 1. Samples of dermoscopic images

The dermoscopic images in the dataset pertained to "Males" for the 54%, "Females" for the 45.5% and to "Unknown" for the 0.5%. Melanocytic nevi significantly outnumber the other six types of lesions (Table I). This type of lesion is a frequently occurring benign condition characterized by the localized proliferation of pigment cells. Despite its benign nature, Melanocytic nevi can occasionally be discovered alongside Melanoma, the second most prevalent lesion within the dataset.

TABLE I
Lesion Type Distribution

| Lesion Type | Count |
|---|---|
| Melanocytic nevi | 6705 |
| Melanoma | 1113 |
| Benign keratosis-like lesions | 1099 |
| Basal cell carcinoma | 514 |
| Actinic keratoses | 327 |
| Vascular lesions | 142 |
| Dermatofibroma | 115 |

Histology, the study of tissues under a microscope, stands as the most prevalent method for diagnosing lesions (Table II). The second most frequent diagnosis method is follow-up. This aligns logically with the most common benign lesion type, as individuals visiting a doctor may be requested to return for further observation of the lesion.

TABLE II
DIAGNOSIS TYPE DISTRIBUTION

| Diagnosis Type | Count |
|---|---|
| Histologically | 5340 |
| Follow Up | 1113 |
| Consensus | 1099 |
| Confocal | 514 |

Examining the patients' ages from whom the samples were collected (Fig. 2), the distribution shows a slight left skew. The majority of patients fall between 35 and 75 years old. Several reasons could contribute to this skewed representation. For instance, younger individuals might have fewer regular check-ups and could potentially be less susceptible to developing cancerous lesions.

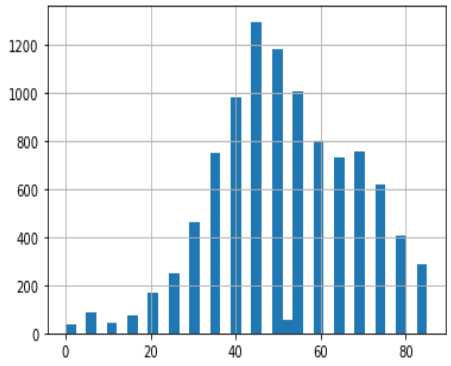


Fig. 2. Age of Patients

Regarding the lesion locations, the most frequent areas are the back and lower extremities, followed by the trunk and upper extremities (as depicted in Figure 3). Conversely, the least common sites are the toes, fingers, and genitals, potentially attributed to their smaller size compared to other body parts.

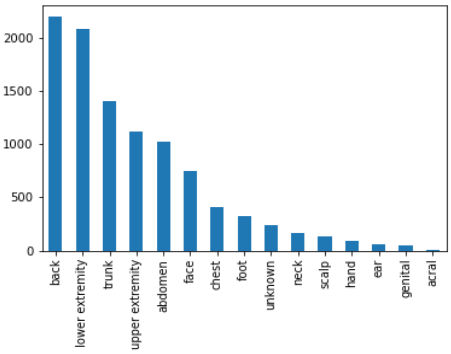


Fig. 3. Localisation of Lesions

## C. Data Preparation

The images initially had a resolution of 450x600 with 3 RGB channels. However, due to equipment limitations, they were resized to an 80x80 format. The square format was chosen to facilitate potential transfer learning applications. Following this, the values underwent normalization, which aids faster model convergence and simplifies the training of the cost function. Since the RGB channels contained values ranging from 0 to 255, the array values were divided by 255 to fall within the 0 to 1 range. Finally, the dataset was divided into training and validation sets using an 80/20 split. The training set comprised 8012 images, while the test set contained 2003 images.

## D. Modelling

Initially, a base model was constructed to observe the behavior of the data as it passed through several convolutional and pooling layers. The model began with a convolutional layer comprising 32 channels and utilized a ReLU activation function. Its input shape was set to (80, 80, 3). Subsequently, a max-pooling layer was added with a pool size and stride of 2. Following this, another convolutional layer with 64 channels and ReLU activation function, along with another max-pooling layer, were incorporated. Then, a subsequent combination of a convolutional layer and a max-pooling layer was introduced. Since the output of the last convolutional layer formed a 3D tensor, it had to be flattened to be compatible with the subsequent Dense layer, as Dense layers accept only 1D vectors as input. Consequently, a fully connected layer with 128 channels was applied. Finally, a Dense layer with 7 outputs corresponding to the 7 available classes in the dataset was implemented. The architecture of the model, along with the layers and the number of parameters at each layer, is depicted in Figure 4.

```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_18 (Conv2D)           (None, 78, 78, 32)        896

max_pooling2d_18 (MaxPoolin  (None, 39, 39, 32)        0
g2D)

conv2d_19 (Conv2D)           (None, 37, 37, 64)        18496

max_pooling2d_19 (MaxPoolin  (None, 18, 18, 64)        0
g2D)

conv2d_20 (Conv2D)           (None, 16, 16, 64)        36928

max_pooling2d_20 (MaxPoolin  (None, 8, 8, 64)          0
g2D)

flatten_6 (Flatten)          (None, 4096)              0

dense_13 (Dense)             (None, 128)               524416

dense_14 (Dense)             (None, 7)                 903

=================================================================
Total params: 581,639
Trainable params: 581,639
Non-trainable params: 0
_________________________________________________________________
```

Fig. 4. Sequential Model

Adam was the selected optimizer and parameter adjustments were made after obtaining the base model. Varying combinations of batch sizes, learning rates, and number of epochs were tested over multiple iterations. Given that 10,000 images might not provide sufficient information for the model to learn, data augmentation techniques were applied to enhance data variability. Despite that strategy, however, limitations persisted regarding the quantity of images being fed into the model, and to achieve high accuracy scores on both the training and validation sets, a considerable amount of input is essential.

## IV. Evaluation/Results

The model was trained multiple times testing different combinations of batch size, learning rate and number of epochs. An epoch refers to one iteration over the entire dataset, hence, the number of epochs is a hyper-parameter that determines how many times the learning algorithm will work through the entire dataset. Weights are updated with every iteration until the optimal state is achieved. Choosing the right number of epochs is important because having too few epochs may result in an under fitted model while having too many epochs can cause an overfitted model. The approach was to start with a small number and increase it gradually until the model's performance started to plateau. By comparing the training and validation accuracy it was possible to get a sense of whether the model was overfitting or underfitting. If the training accuracy was high, but the validation accuracy was low, this could have indicated the model was overfitting the training data and not generalising well to new data. On the other hand, if the training accuracy was low, but the validation accuracy was high, this could have indicated the model was underfitting and was not learning the patterns in the training data well. A model's training loss is the measure of how much the model improves after each iteration of optimisation.

The expected outcome is that the training and validation loss decrease and the training and test accuracy increase with every iteration. In deep learning, batch size is an essential hyper-parameter. Selected batch sizes may lead to various testing and training accuracies and runtimes. As suggested in the cited papers, small batches usually perform better, therefore, only batch sizes between 8 and 64 were tested. With regard to the learning rate, models usually converge too fast if it is too large. On the other hand, a very small learning rate can cause models to require enormous amount of time to train. Smaller learning rates require a larger number of epochs. As shown in the table below, a number of 20 epochs and a learning rate of 0.0001 provided to be the best combination.

| Model | Epoch | Batch Size | Training Size | Learning Rate | Tr_Accuracy | Tr_Loss | Val_acc | Val_Loss |
|---|---|---|---|---|---|---|---|---|
| Model 1 | 10 | Default(32) | 8012 | Default(0.001) | 0.791 | 0.564 | 0.742 | 0.713 |
| Model 2 | 10 | 8 | 8012 | 0.0001 | 0.767 | 0.641 | 0.719 | 0.753 |
| Model 3 | 10 | 32 | 8012 | 0.0001 | 0.738 | 0.712 | 0.72 | 0.759 |
| Model 4 | 20 | 32 | 8012 | 0.0001 | 0.774 | 0.61 | 0.745 | 0.689 |
| Model 5 | 10 | 64 | 8012 | 0.0001 | 0.714 | 0.805 | 0.678 | 0.931 |
| Model 6 | 10 | 128 | 8012 | 0.0001 | 0.708 | 0.822 | 0.69 | 0.852 |
| Model 7 | 10 | 8 | 8012 | 0.001 | 0.842 | 0.413 | 0.717 | 0.896 |
| Model 8 | 10 | 32 | 8012 | 0.001 | 0.786 | 0.573 | 0.72 | 0.78 |
| Model 9 | 20 | Data augmentation | | | 0.81 | 0.501 | 0.744 | 0.717 |

Fig. 5. Hyperparameters vs Performance

Figure 6 recaps the performance of the best-performing architecture which was run using 20 epochs, a batch size of 32, and a learning rate of 0.0001. Training accuracy was 74% with a loss of 61% while validation accuracy was 75% with a loss of 69%.

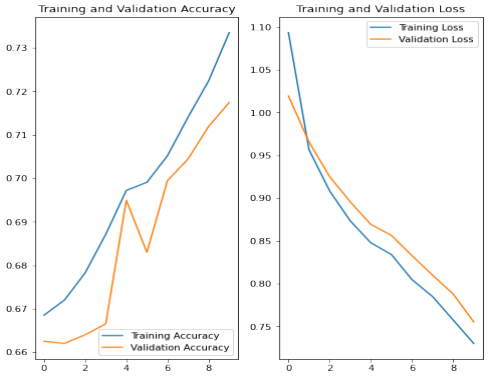


Fig. 6. Training and Validation Accuracy vs Loss

Data augmentation techniques were then applied as 10,000 images are not a massive number of samples the network can avail to learn from. This was done with the intent of improving dataset's variability. The training accuracy of the resulting architecture was 81% with a loss of 50%, and a validation accuracy of 74% and a loss of 71% (Fig.7).

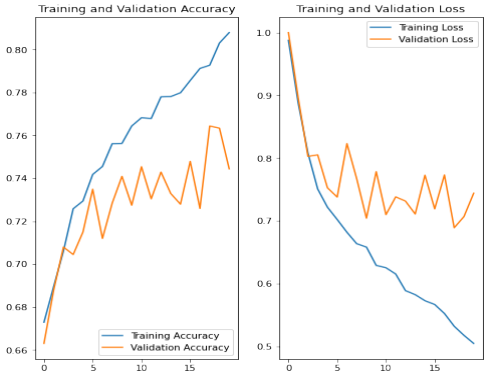


Fig. 7. Training and Validation Accuracy vs Loss

## V. Conclusion and Future Work

A Convolutional Neural Network type of architecture was adopted to detect skin cancer through dermoscopic images. This non-invasive and cost-effective technique can support dermatologists' decision-making process and reduce the gap between diagnosis and treatment. Early detection significantly improves the odds of a successful recovery underscoring deep learning's potential to enhance survival rates. Self-detection of early signs of illness could become reality thanks for instance to internet-of-things devices, or dedicated mobile apps enabling real-time feedback. Considering results were not as good as some of the reviewed related works, future research should aim at bridging the gap by expanding the dataset with both more diverse records and features. Enhancements in modeling and training duration, coupled with increased computational capacity, are anticipated to benefit performance. Exploring various combinations of learning rates and epochs is crucial too, considering their interdependence. Optimal batch

sizes were inferred from literature and experimental outcomes [31] [32]. Addressing class imbalances within the dataset is imperative too, to tackle significant disparities among different skin lesion categories. Additionally, considering transfer learning could capitalize on pre-existing tested predictive frameworks, potentially advancing the model's capabilities.

## REFERENCES


[1] The World Health Organization, "Ultraviolet radiation". Available at: https://www.who.int/news-room/fact-sheets/detail/ultraviolet-radiation

[2] Irish Cancer Society, "Skin cancer". Available at: https://www.cancer.ie/cancer-information-and-support/cancer-types/skin-cancer

[3] American Cancer Society, "Melanoma skin cancer statistics". Available at: https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html

[4] N. Razmjooy, B.,S. Mousavi, F. Soleymani, "A hybrid neural network Imperialist Competitive Algorithm for skin color segmentation", 2013, Mathematical and Computer Modelling, Volume 57, Issues 3–4, Pages 848-856, ISSN 0895-7177, https://doi.org/10.1016/j.mcm.2012.09.013.

[5] N. Heller, E. Bussmann, A. Shah, J. Dean, and N. Papanikolopoulos, "Computer aided diagnosis of skin lesions from morphological features", 2018, retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/216030.

[6] A. Kulkarnu, M. Debajyoti, "SVM classifier based melanoma image classification", 2017, Research Journal of Pharmacy and Tech, 10(12): 4391-4392. doi: 10.5958/0974-360X.2017.00808.3

[7] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun, "Dermatologist-level classification of skin cancer with Deep Neural Networks", 2017, Nature 542, 115–118. https://doi.org/10.1038/nature21056

[8] M. Dildar, S. Akram, M. Irfan, H. U. Khan, M. Ramzan, A. R. Mahmood, S. A. Alsaiari, A. H. M. Saeed, M. O. Alraddadi, M. H. Mahnashi, "Skin cancer detection: A review using Deep Learning Techniques", 2021, International Journal Environment Research Public Health. doi: 10.3390/ijerph18105479

[9] A. Takiddin, J. Schneider, Y. Yang, A. Abd-Alrazaq, M. Househ, "Artificial Intelligence for Skin Cancer Detection: Scoping Review", 2021, Journal Medical Internet Research, 23(11):e22934. doi: 10.2196/22934

[10] G. Sharma, R. Chadha, "Skin Cancer and Oral Cancer Detection using Deep Learning Technique", 2022, IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), pp. 1-5, doi: 10.1109/ICDCECE53908.2022.9792688

[11] The International Skin Imaging Collaboration, "Isic Archive". Available at: https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main

[12] A. Ray, A. Gupta, A. Al, "Skin lesion classification with deep convolutional neural network: Process development and validation", 2020, JMIR Dermatology, doi: 10.2196/18438

[13] Ni Zhang, Yi-Xin Cai, Yong-Yong Wang, Yi-Tao Tian, Xiao-Li Wang, and Benjamin Badami, "Skin cancer diagnosis based on optimized convolutional neural network," Artificial Intelligence in Medicine, 08-Nov-2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365719301460.

[14] M. A. Khan, M. Y. Javed, M. Sharif, T. Saba and A. Rehman, "Multi-Model Deep Neural Network based Features Extraction and Optimal Selection Approach for Skin Lesion Classification," 2019 International Conference on Computer and Information Sciences (ICCIS), 2019, pp. 1-7, doi: 10.1109/ICCISci.2019.8716400.

[15] W. Sae-Lim, W. Wettayaprasit and P. Aiyarak, "Convolutional Neural Networks Using MobileNet for Skin Lesion Classification," 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2019, pp. 242-247, doi: 10.1109/JCSSE.2019.8864155.

[16] K. Wilson, C. M Umemneku Chikere, S. Graziadio, L. Vale, and J. Allen , "Diagnostic Test Evaluation Methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - an update," PloS one. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31603953/.

[17] T. J. Brinker, A. Hekler, A. H. Enk, J. Klode, A. Hauschild, C. Berking, B. Schilling, S. Haferkamp, D. Schadendorf, S. Fröhling, J. S. Utikal, and C. von Kalle, "A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task," European Journal of Cancer, 08-Mar-2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959804919301443.

[18] H. Balazs, "Skin lesion classification with ensembles of deep convolutional Neural Networks," Journal of Biomedical Informatics, 10-Aug-2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046418301618.

[19] J. Millenia, M. Farid Naful, and J. Siswantoro, "Melanoma Detection using Convolutional Neural Network with Transfer Learning on Dermoscopic and Macroscopic Images". Available: http://repository.ubaya.ac.id/42809/1/

[20] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," arXiv.org, 30-Aug-2020. [Online]. Available: https://arxiv.org/abs/1706.02690.

[21] T. J. Brinker, A. Hekler, A. H. Enk, C. Berking, S. Haferkamp, A. Hauschild, M. Weichenthal, J. Klode, D. Schadendorf, T. Holland-Letz, C. von Kalle, S. Fröhling, B. Schilling, and J. S. Utikal, "Deep neural networks are superior to dermatologists in melanoma image classification," European Journal of Cancer, 08-Aug-2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959804919303491.

[22] M. Combalia, N. Codella, V. Rotemberg, C. Carrera, S. Dusza, D. Gutman, B. Helba, H. Kittler, N. R. Kurtansky, K. Liopyris, M. A. Marchetti, S. Podlipnik, S. Puig, C. Rinner, P. Tschandl, J. Weber, A. Halpern, and J. Malvehy, "Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: The 2019 International Skin Imaging Collaboration Grand Challenge," The Lancet Digital Health, 01-May-2022. [Online]. Available: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(22)00021-8/fulltext.

[23] K. Renganathan, V. Bhaskar, and J. Vishnuvardhan, "Melanoma Skin Cancer Detection Using Knn And Svm Classifier," Ilkogretim Online - Elementary Education Online, vol. 19, no. 2, pp. 2076–2085, 2020.

[24] O T Jones, R N Matin, M van der Schaar, K Prathivadi Bhayankaram, C K I Ranmuthu, M S Islam, D Behiyat, R Boscott, N Calanzani, J Emery, H C Williams, F M Walter, "Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review", The Lancet Digital Health, Volume 4, Issue 6, 2022, Pages e466-e476.

[25] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," Computer Methods and Programs in Biomedicine, 19-May-2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260718304267.

[26] H.A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists", Annals of Oncology, Vol. 29, Issue 8, 2018, Pages 1836-1842,ISSN 0923-7534, https://doi.org/10.1093/annonc/mdy166. Available: https://www.sciencedirect.com/science/article/pii/S0923753419341055

[27] P. Tschandl, C. Rinner, Z. Apalla, G. Argenziano, N. Codella, A. Halpern, M. Janda, A. Lallas, C. Longo, J. Malvehy, J. Paoli, S. Puig, C. Rosendahl, H. P. Soyer, I. Zalaudek, and H. Kittler, "Human–computer collaboration for Skin cancer recognition," Nature News, 22-Jun-2020. [Online]. Available: https://www.nature.com/articles/s41591-020-0942-0.

[28] M. A. Kassem, K. M. Hosny and M. M. Fouad, "Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning," in IEEE Access, vol. 8, pp. 114822-114832, 2020, doi: 10.1109/ACCESS.2020.3003890.

[29] N. Hotz, "What is CRISP DM?," Data Science Process Alliance, 05-Dec-2022. [Online]. Available: https://www.datascience-pm.com/crisp-dm-2/.

[30] Medical University of Vienna. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Available: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

[31] R. Lin, "Analysis on the Selection of the Appropriate Batch Size in CNN Neural Network," 2022 International Conference on Machine

Learning and Knowledge Engineering (MLKE), 2022, pp. 106-109, doi: 10.1109/MLKE55170.2022.00026.
[32] I. Goodfellow, "Deep Learning (Adaptive Computation and Machine Learning series)", 2016, pp. 429