Skin Cancer Diagnosis Using Deep Convolutional Neural Network - Project Report

Md. Farhadul Islam
Department of Computer Science and Engineering
School of Data and Sciences
BRAC University
Dhaka, Bangladesh
md.farhadul.islam@g.bracu.ac.bd

Fardin Bin Rahman
Department of Computer Science and Engineering
School of Data and Sciences
BRAC University
Dhaka, Bangladesh
fardin.bin.rahman@g.bracu.ac.bd

Sarah Zabeen
Department of Computer Science and Engineering
School of Data and Sciences
BRAC University
Dhaka, Bangladesh
sarah.zabeen@g.bracu.ac.bd

Md. Azharul Islam
Department of Computer Science and Engineering
School of Data and Sciences
BRAC University
Dhaka, Bangladesh
md.azharul.islam@g.bracu.ac.bd

Abstract

A third of the proportion of the population that is diagnosed with cancer is accounted for by Skin Cancer worldwide, and the numbers are only rising. While the use of dermoscopy has greatly aided in the diagnostic capability of skin cancer, it still remains a challenge to appropriately detect and distinguish between different types of skin cancer, which may appear similar in appearance. To tackle this issue, researchers have conducted several experiments and made great progress in developing automated tools to assist dermatologists in decision making. Our proposed approach involves an automated computer-aided diagnosis system for multi-class skin (MCS) cancer classification with incredible accuracy. Our model exceeds both expert dermatologists and contemporary deep learning methods for MCS cancer classification in many evaluation metrics. We conduct our study with seven classes of HAM10000 dataset and our proposed model achieves 98.4% accuracy.

Preprint. Under review.

1 Introduction

Skin cancer is an invasive disease caused by the abnormal growth of melanocyte cells in the body, which tend to replicate and spread through lymph nodes to destroy surrounding tissues [18]. The damaged cells manifest themselves in the form of a mole on the outer layer of skin. They may or may not be malignant. On the other hand, Melanoma is classified as cancer since it is comparatively more threatening. Across the world, skin cancer accounted for over 1 million deaths every month, and 300,000 more monthly cases in 2018 [2]. Melanoma is listed as the 19th most common disease with the highest death rate [2]. Although it has been a tough journey coping with the high mortality rate, the recent developments in artificial intelligence and image processing has led us to believe that the survival rate may very well increase over time. More importantly, computer-aided diagnostic (CAD) tools are more time and labor efficient than the currently applied clinical approaches.

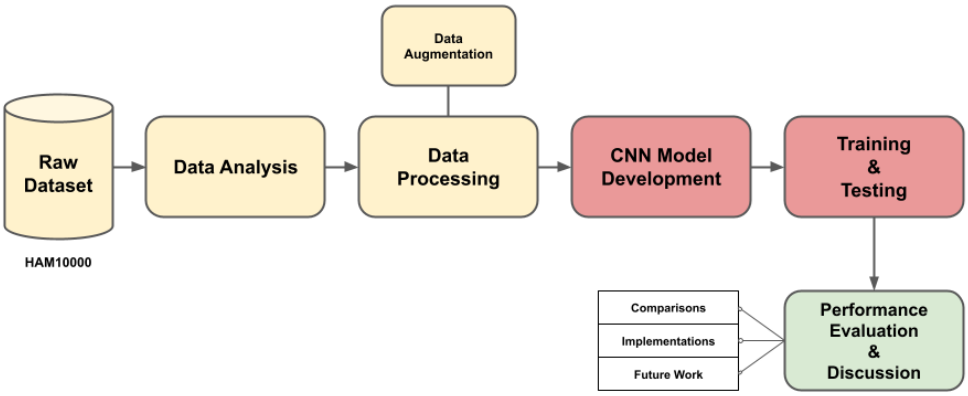


Figure 1: Workflow of the Study

A skilled dermatologist must execute a step by step approach for diagnosis [13] which is rather expensive both in terms of time consumed and labor exerted. Yet, the results of the diagnosis may vary depending upon the expertise of the dermatologist, and reportedly the accuracy for properly diagnosing skin lesions is below 80% [17]. The statistics are further lowered by the finite number of expert dermatologists available worldwide.

Skin cancer is mainly divided into three categories: basal, squamous, and melanocyte [3]. Basal cell carcinoma is the most common sub-category of skin cancer. Its growth is rather slow and it does not disseminate to other parts of the body, but it has a tendency to recur. Thus, taking care of it as swiftly as possible is crucial. Squamous cell carcinoma has a comparatively higher chance of spreading and reaches far deeper below the surface of the skin. Lastly, melanocytes are responsible for causing skin to darken in color when exposed to sunlight. This is due to the melanin produced by the melanocytes, which protects the skin from sunlight. However, if melanin amasses under the skin, it causes malignant moles to develop. This is known as melanoma cancer. Unlike the first two types of cancer cells, melanoma cancer causes significant damage to surrounding tissues. It is classified as dangerous and can be life-threatening.

In this work, we propose a Deep CNN model which achieves high accuracy, keeping the model weight low. Our main contributions are:

- Using CNN to classify seven distinct categories diseases with high accuracy.
- Creating a lightweight, user friendly model for applications.

2 Related Work

In early 1990, computer-aided diagnosis systems were developed to help dermatologists overcome the difficulties in skin cancer categorization [23]. The first studies using dermoscopy images were limited to classifying benign and melanoma skin cancer lesions [22]. Since then, several approaches for tackling this difficult issue have been published. Several studies [12, 4, 22] follow the commonly used manual evaluation methods based on the ABCD rules proposed by Nachbar et al. [21]. Moreover, traditional Multimedia Applications and Tools In the search for a more accurate and trustworthy technique, machine learning classifiers such as Super Vector Machines [7], K-Nearest Neighbors [5], Decision Trees [1], and Artificial Neural Networks [15] were also considered for skin cancer classification.

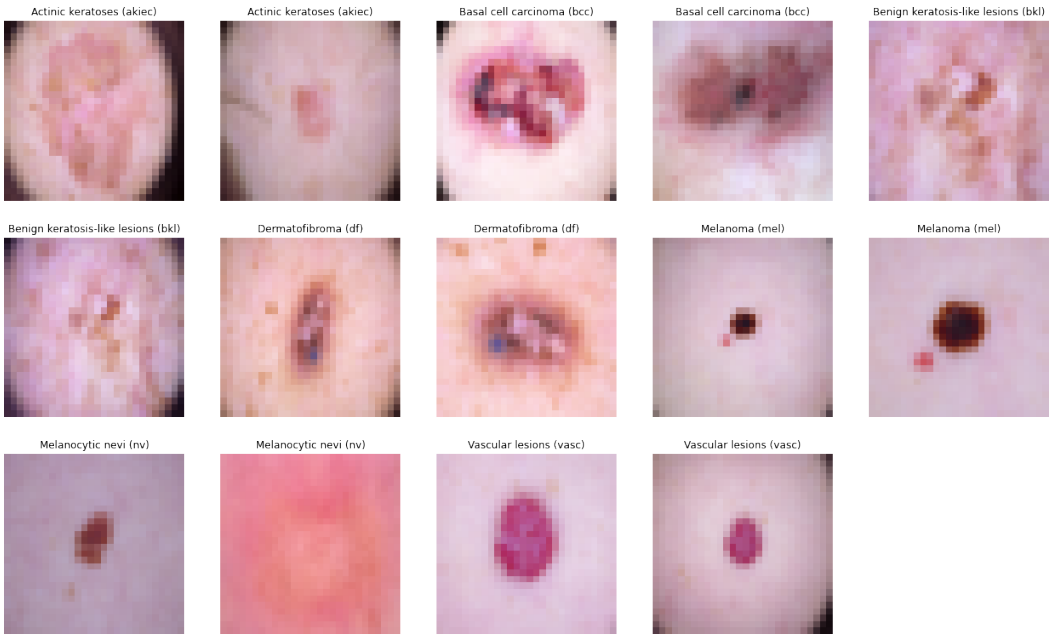


Figure 2: Image Samples from the Dataset.

Convolutional neural networks have made great strides in existing problems and quickly became the preferred choice for skin cancer classification [11]. Since CNNs require large datasets to become familiar with the problem [19], current literature [13, 16] primarily uses transfer learning to solve large dataset problems. This is a technique where a model trained for a source task is partially recycled for a new target task.

Previous studies of computational classification by dermoscopy are not only lacking in general capabilities [19, 6], but also unable to improve the accuracy of MCS skin cancer classification [13, 14]. Unfortunately, most of the research so far has not used large datasets. This is the basis for a deep learning model to work well. Unlike these studies, our proposed model performs much better on seven classses of skin cancer.

3 Research Methodology

3.1 Dataset

For our research, the HAM10000 dataset [10] is used, which consists of a grand selection of thousands of multi-source dermoscopy images of common pigmented skin lesions. Among these are 115 images showing Dermatofibroma,142 images of Vascular Lesions, 327 pictures showing Actinic keratosis, 514 images of Basal cell carcinoma, 1099 pictures of Benign keratosis, 1113 images of Melanoma and 6705 images depicting Melanocytic nevi. This amounts to 10,015 dermoscopy images displaying seven different types of skin cancer. Some sample images of skin cancer types from HAM10000 are represented in 2.

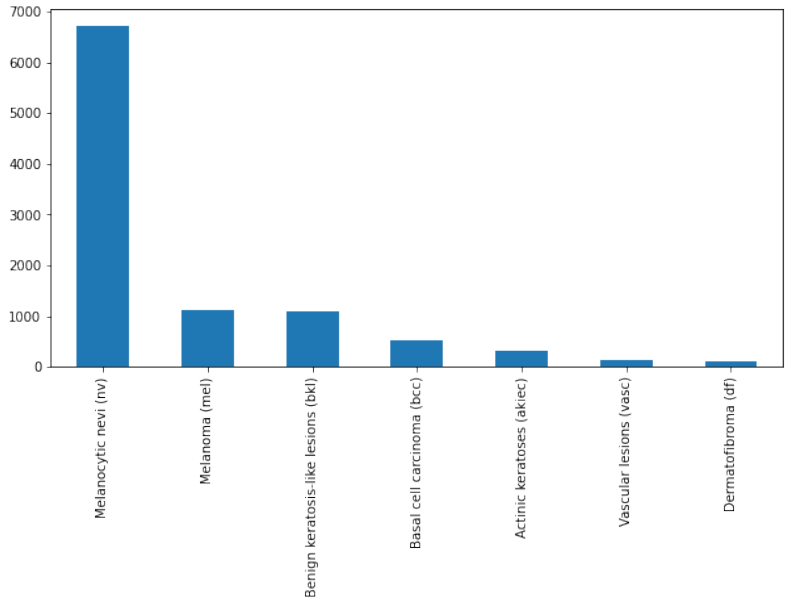


Figure 3: Frequency of images from each class

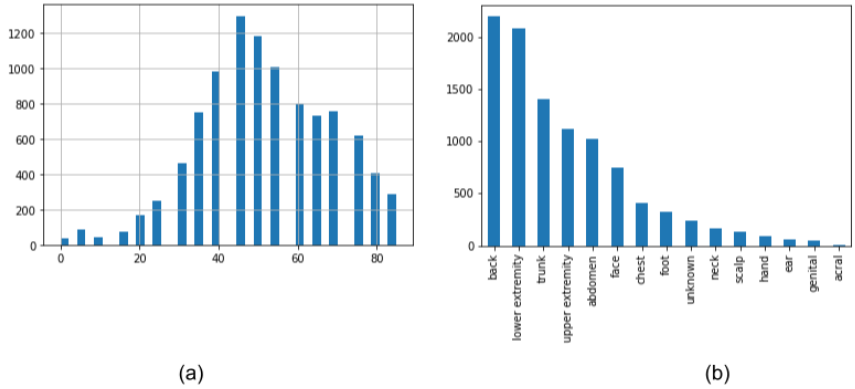


Figure 4: Histogram of ages (a), Histogram of affected are positions (b)

3.2 Data Analysis and Processing

Firstly we check the missing values in our dataset. After finding the missing values we replace them with Mean values of that column or feature. We then resize the image to 28 x 28 pixel with 3 channel (RGB). We Augment our data by using factors such as rotation, width shift, height shift, shear, horizontal and vertical flip. We augment the data so that there are no imbalance in the dataset, so that we can get better accuracy. The dataset is divided into 60:20:20 training, validation, and test datasets. The imbalance in the original dataset can be seen in 3. Other exploratory data analysis can be seen in 4.

3.3 CNN Architecture

Our CNN model starts with a three-channel picture with a resolution of 28×28 pixels from the dataset. We use five Conv2D layers with (3,3) kernel size. Moreover, we utilize MaxPooling layers with a pool size of (2,2) to minimize the computational cost of the layers. Additionally, we utilize default strides of (1,1) for all Conv2D layers. We choose ReLU as the activation function since its gradient is not saturated, which significantly speeds the evolution of stochastic gradient descent

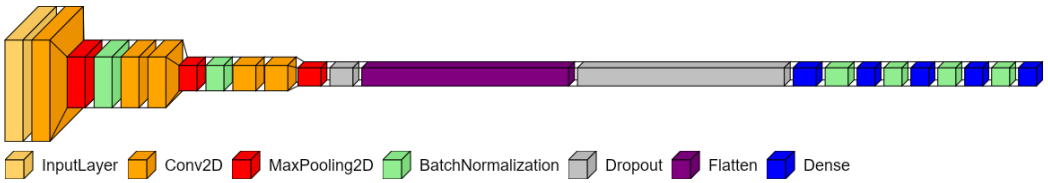


Figure 5: Visual Representation of Proposed Deep CNN Architecture

Layer	Output Shape	Parameters
Input Layer	(None, 28, 28, 3)	0
2D Convolutional Layer	(None, 28, 28, 64)	1792
2D Max Pooling	(None, 14, 14, 64)	0
Batch Normalization	(None, 14, 14, 64)	256
2D Convolutional Layer	(None, 14, 14, 128)	73856
2D Convolutional Layer	(None, 14, 14, 128)	147584
2D Max Pooling	(None, 7, 7, 128)	0
Batch Normalization	(None, 7, 7, 128)	512
2D Convolutional Layer	(None, 7, 7, 256)	295168
2D Convolutional Layer	(None, 7, 7, 256)	590080
2D Max Pooling	(None, 3, 3, 256)	0
Dropout	(None, 3, 3, 256)	0
Flatten	(None, 2304)	0
Dropout	(None, 2304)	0
Dense Layer	(None, 256)	590080
Batch Normalization	(None, 256)	1024
Dense Layer	(None, 128)	32896
Batch Normalization	(None, 128)	512
Dense Layer	(None, 64)	8256
Batch Normalization	(None, 64)	256
Dropout	(None, 32)	2080
Batch Normalization	(None, 32)	128
Dense Layer	(None, 7)	231
Total Parameters: 1,744,711		

Table 1: Output Shape and Parameter Size of Each Layer of The Proposed Model

(SGD) as compared to other activation functions such as the Tanh / Sigmoid function. Here Max Pooling layer and Batch Normalization Layers comes after each convolutional layer as a block. Eventually, we transform the values to a one-dimensional array and begin adding fully connected (FC) layers to the CNN, first with 2304 nodes, then 256, 128, 64 nodes, and finally 32 nodes. To avoid overfitting in the model, we utilize a 15% dropout after the final convolutional block and after the flatten layer. The linear activation function is used as a network classifier at the output layer. It is a simple straight-line activation function in which our function is proportionate to the weighted sum of neurons or input. Linear activation functions provide a wider range of activations, and a positive sloped line may enhance the activation level as the input rate increases. Our proposed model is represented in 5. Finally comes the output Dense layer with seven nodes, each being a class or label.

Table 2 shows the optimum hyperparameters applied in different datasets for the proposed Deep CNN model. The total number of parameters in this model is 1,744,711. We showed the model architecture in details in 1.

Hyperparameter					
Image Input Size	Epoch	Batch Size	Learning Rate	Dropout Rate	Parameters
28 x 28	30	8	0.0001	15%	1,744,711

Table 2: Hyperparameter of Our Proposed Model

After experimenting with various tweaks in the model, we come up with our Deep CNN architecture. This architecture provides the best performance while minimizing the number of parameters as much as possible and therefore it uses less computing resources.

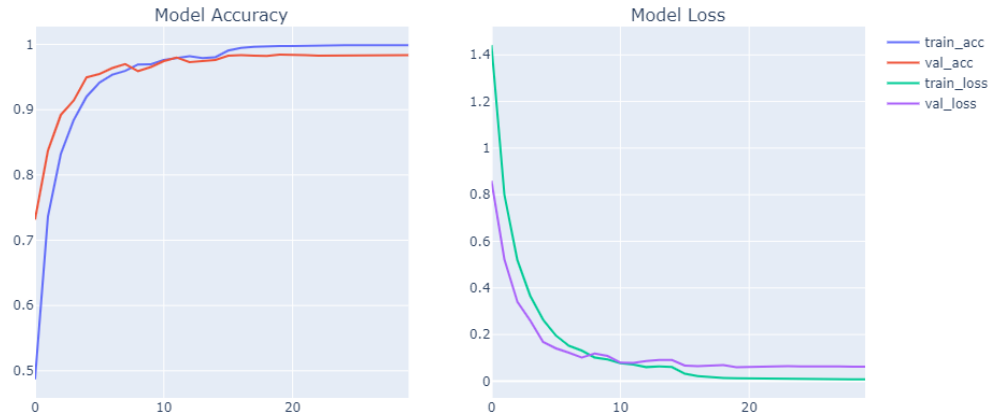


Figure 6: Accuracy and Loss Graph

4 Experimental Evaluation

Following the development of our suggested architecture, we test our model using commonly used metrics for regression-based tasks. Our primary objective is to create a model that is gives the best possible result regardless parameters or other hyperparameters.

4.1 Experimental Setup

Tensorflow, Keras, Pillow, and OpenCV Python libraries are used to make the training and testing protocols for this Deep CNN model, as well as for all other cutting-edge models. The models are trained and evaluated with an NVIDIA RTX 3080TI GPU which has 34.1 TeraFLOPs of performance.

4.2 Evaluation Metrics

Following any image classification task, evaluation of the quality of the model depends on performance evaluation metrics. For quantitative evaluation, renowned metrics of performance evaluation, such as Accuracy1, Precision2, Precision3, and F1-Score4 are employed to determine the performance of the proposed approach.

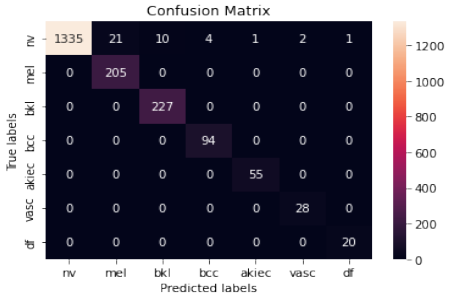
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

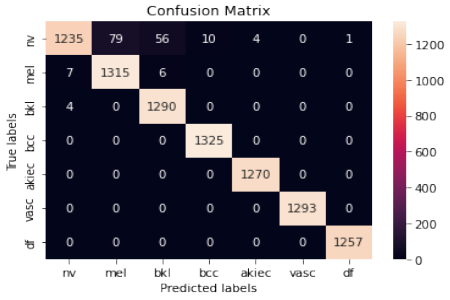
$$F1\,Score = \frac{2 \times Sensitivity \times Precision}{Sensitivity + Precision} \tag{4}$$

The accuracy is defined as the ratio of the model's correct predictions to the total number of predictions made. The Precision is used to determine a proportion of correct identifications. Precision is calculated



.5

Figure 7: Confusion Matrix of Training Dataset



.5

Figure 8: Confusion Matrix of Test Dataset

Figure 9: Confusion Matrix

Original Dataset				
Test Accuracy: 98.053%				
Class	Precision	Recall	f1-score	Support
Melanocytic Nevi	1.00	0.97	0.99	1374
Melanoma	0.91	1.00	0.95	205
Benign keratosis-like lesions	0.96	1.00	0.98	227
Basal cell carcinoma	0.96	1.00	0.98	94
Actinic keratoses	0.98	1.00	0.99	55
Vascular lesions	0.93	1.00	0.97	28
Dermatofibroma	0.95	1.00	0.98	20
Accuracy			0.98	2003
Macro Average	0.96	1.00	0.98	2003
Weighted Average	0.98	0.98	0.98	2003

Table 3: Classification Report on Original Dataset

by dividing the number of true positive outcomes (TP) by the number of predicted positive outcomes (TP + FP). The recalls are used to figure out what percentage of true positives is appropriately detected. Recall is calculated by dividing the number of true positives (TP) by the total amount of data (TP + FN) is the important factor. Similarly, the F1 Score is a popular metric for assessing the effectiveness of machine learning algorithms. It is calculated as the arithmetic mean of accuracy and recall. It has a value between 0 and 1. The number of occurrences properly identified by the learning models is reflected in the F1 scores.

4.3 Experimental Findings

After a rigorous research, we get the accuracy of 98.4%, 98.5% precision and 98.4% recall. Which outperforms all popular methods with high accuracy such as, [8, 9, 20]. We get 98.053% accuracy using the original dataset. Which also outperformes the methods mentioned in 5. Detailed classification report of the training and testing dataset can be seen in 3 and 4 respectively. 9 shows the confusion

Augmented Dataset				
Test Accuracy: 98.405%				
Class	Precision	Recall	f1-score	Support
Melanocytic Nevi	0.99	0.90	0.95	1385
Melanoma	0.96	0.99	0.98	1328
Benign keratosis-like lesions	0.96	1.00	0.98	1294
Basal cell carcinoma	0.99	1.00	0.99	1325
Actinic keratoses	1.00	1.00	1.00	1270
Vascular lesions	1.00	1.00	1.00	1293
Dermatofibroma	1.00	1.00	1.00	1257
Accuracy			0.98	9152
Macro Average	0.98	0.98	0.98	9152
Weighted Average	0.98	0.98	0.98	9152

Table 4: Classification Report on Augmented Dataset

Ref.	**Method**	**Accuracy**	**Precision**	**Recall**
[8]	MobileNet	83.1	89.0	83.0
[20]	InceptionResnetV2	70.0	N/A	N/A
	PNASNet-5-Large	76.0	-	-
	SENet154	74.0	-	-
	InceptionV4	67.0	-	-
[9]	InceptionV3	91.56	89.0	89.0
	ResNeXt101	93.20	88.0	88.0
	InceptionResnetV2	93.20	87.0	88.0
	Xception	91.47	89.0	88.0
	NASNetLarge	91.11	86.0	86.0
Ours	**Proposed Deep CNN**	**98.4**	**98.5**	**98.4**

Table 5: Comparison among other existing models

matrix of both training and testing dataset. The matrix show how well the model performs in almost all of the classes of the dataset.

5 Discussion

From our study we find that, our model performs well if we design the CNN model with 64 kernel size, instead of a smaller value, model reaches comparatively higher accuracy. The fully connected layers is built using a series of dense layers where the number of nodes is reduced by half. This increases our accuracy by 1 to 2%. Moreover, our model is not as heavy as models mentioned in 5. Our model only has 1,744,711 parameters, which can be considered as a lightweight CNN. This also helps in app or web implementations as lightweight model requires less space and are faster to work with. Overall, we can conclude that our study contributes not only in performance but also in implementations.

6 Conclusion and Future Work

The number of people diagnosed with skin cancer is increasing day by day. Addressing this global issue to the public has become urgent. Recent progress in the field of deep learning, has enabled medical professionals to automate the diagnosis with high accuracy. Our work has outperformed all the popular CNN models, making our model ready-to-use for computer aided diagnosis. The future work may deal with interpretable diagnosis which will help medical professionals visually as well. The highlited regions will tell what our proposed model has predicted and why it has predicted such outcomes. This issue will ensure lesser risk in medical practises. Moreover, uncertainty quantification can ensure more reliability while diagnosing patients.

References

[1] Automatic Detection of Blue-White Veil and Related Structures in Dermoscopy Images. 32. ISSN 0895-6111. doi: 10.1016/j.compmedimag.2008.08.003. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3160648/.

[2] Skin cancer statistics | World Cancer Research Fund International, . URL https://www.wcrf.org/cancer-trends/skin-cancer-statistics/.

[3] Skin Cancer | Skin Cancer Facts | Common Skin Cancer Types, . URL https://www.cancer.org/cancer/skin-cancer.html.

[4] Qaisar Abbas, M. Emre Celebi, Irene Fondón, and Waqar Ahmad. Melanoma recognition framework based on expert definition of abcd for dermoscopic images. *Skin research and technology : official journal of International Society for Bioengineering and the Skin (ISBS) [and] International Society for Digital Imaging of Skin (ISDIS) [and] International Society for Skin Imaging (ISSI)*, 19, 06 2012. doi: 10.1111/j.1600-0846.2012.00614.x.

[5] Lucia Ballerini, Robert Fisher, Ben Aldridge, and Jonathan Rees. *A Color and Texture Based Hierarchical K-NN Approach to the Classification of Non-melanoma Skin Lesions*, volume 6, pages 63–86. 01 2013. ISBN 978-94-007-5388-4. doi: 10.1007/978-94-007-5389-1_4.

[6] Marco Burroni, Rosamaria Corona, Giordana Dell'Eva, Francesco Sera, Riccardo Bono, Pietro Puddu, Roberto Perotti, Franco Nobile, Lucio Andreassi, and Pietro Rubegni. Melanoma computer-aided diagnosis reliability and feasibility study. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 10:1881–6, 04 2004. doi: 10.1158/1078-0432.CCR-03-0039.

[7] M. Emre Celebi, Hassan A. Kingravi, Bakhtiyar Uddin, Hitoshi Iyatomi, Y. Alp Aslandogan, William V. Stoecker, and Randy H. Moss. A methodological approach to the classification of dermoscopy images. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 31:362–373, September 2007. ISSN 0895-6111. doi: 10.1016/j.compmedimag.2007.01.003. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3192405/.

[8] Saket S. Chaturvedi, Kajol Gupta, and Prakash S. Prasad. Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using mobilenet. *ArXiv*, abs/1907.03220, 2019.

[9] Saket S. Chaturvedi, Jitendra V. Tembhurne, and Tausif Diwan. A multi-class skin cancer classification using deep convolutional neural networks. *Multimedia Tools and Applications*, pages 1 – 22, 2020.

[10] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, and Allan Halpern. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic), 2019. URL https://arxiv.org/abs/1902.03368.

[11] Noel C. F. Codella, Junjie Cai, Mani Abedini, Rahil Garnavi, Alan Halpern, and John R. Smith. Deep learning, sparse coding, and svm for melanoma recognition in dermoscopy images. In *MLMI*, 2015.

[12] Nayara Holanda de Moura, Rodrigo M. S. Veras, Kelson Rômulo Teixeira Aires, Vinicius Ponte Machado, Romuere R. V. Silva, Flávio H. D. Araújo, and Maíla de Lima Claro. Abcd rule and pre-trained cnns for melanoma diagnosis. *Multimedia Tools and Applications*, 78:6869–6888, 2018.

[13] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, February 2017. ISSN 1476-4687. doi: 10.1038/nature21056. URL https://www.nature.com/articles/nature21056.

[14] Balázs Harangi, Ágnes Baran, and András Hajdu. Classification of skin lesions using an ensemble of deep neural networks. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2575–2578, 2018.

[15] Hitoshi Iyatomi, Hiroshi Oka, Masataka Saito, Ayako Miyake, Masayuki Kimoto, Jun Yamagami, Seiichiro Kobayashi, Akiko Tanikawa, Masafumi Hagiwara, Koichi Ogawa, Giuseppe Argenziano, H. Peter Soyer, and Masaru Tanaka. Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system. *Melanoma Research*, 16(2):183–190, April 2006. ISSN 0960-8931. doi: 10.1097/01.cmr.0000215041.76553.58.

[16] Jeremy Kawahara and Ghassan Hamarneh. Multi-resolution-tract cnn with hybrid pretrained and skin-lesion trained layers. In Li Wang, Ehsan Adeli, Qian Wang, Yinghuan Shi, and Heung-Il Suk, editors, *Machine Learning in Medical Imaging*, pages 164–171, Cham, 2016. Springer International Publishing. ISBN 978-3-319-47157-0.

[17] R. Marks. Epidemiology of melanoma. *Clinical and Experimental Dermatology*, 25(6):459–463, September 2000. ISSN 0307-6938. doi: 10.1046/j.1365-2230.2000.00693.x.

[18] Rebecca S. Mason and Jörg Reichrath. Sunlight vitamin D and skin cancer. *Anti-Cancer Agents in Medicinal Chemistry*, 13(1):83–97, January 2013. ISSN 1875-5992.

[19] Ammara Masood and Adel Al-Jumaily. Computer aided diagnostic support system for skin cancer: A review of techniques and algorithms. *International Journal of Biomedical Imaging*, 2013, 2013.

[20] Md Ashraful Alam Milton. Automated skin lesion classification using ensemble of deep neural networks in isic 2018: Skin lesion analysis towards melanoma detection challenge. *ArXiv*, abs/1901.10802, 2019.

[21] Franz Nachbar, Wilhelm Stolz, Tanja Merkle, Armand B. Cognetta, Thomas Vogt, Michael Landthaler, Peter Bilek, Otto Braun-Falco, and Gerd Plewig. The abcd rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology*, 30(4):551–559, 1994. ISSN 0190-9622. doi: https://doi.org/10.1016/S0190-9622(94)70061-3. URL `https://www.sciencedirect.com/science/article/pii/S0190962294700613`.

[22] Nilkamal S. Ramteke and Shweta Jain. Abcd rule based automatic computer-aided skin cancer detection using matlab ®. 2013.

[23] R White, Darrell S. Rigel, and Robert J. Friedman. Computer applications in the diagnosis and prognosis of malignant melanoma. *Dermatologic clinics*, 9 4:695–702, 1991.