Melanoma Detection with Domain Specific Artifact Debiasing and Augmentation

Varun Venkat Rao, Vibhuti Ravi, Devishi Kambiranda, Sudharsan Babu Srikanthan
varu@umich.edu, vibhutir@umich.edu, devishis@umich.edu, sudhsrik@umich.edu

Abstract

The problem of early and accurate detection of different kinds of melanoma ranging from the most common to the rarest is of prime importance. Currently, expert dermatologists use the widely accepted "Asymmetry, Border, Color and Diameter" metric to diagnose the form of cancer. Deep learning techniques show promising performance in their ability to replicate melanoma diagnosis. In this project, we address the two major problems of dataset imbalance and improved AUC scores through the proposed augmentation technique and incorporation of hand crafted features. We achieve a 56% improvement using unlearning methods over using augmentation techniques and upto 24% increase in model performance compared to baseline on clinical and dermoscopic data.

1 Introduction

Melanoma is one of the most aggressive forms of skin cancer. It is diagnosed in more than 132,000 people worldwide each year, according to the World Health Organization. Hence, it is essential to detect melanoma early before it spreads to other organs in the body and becomes more difficult to treat. While visual inspection of suspicious skin lesions by a dermatologist is normally the first step in melanoma diagnosis, it is generally followed by dermoscopy imaging for further analysis. Dermoscopy is a noninvasive imaging procedure that acquires a magnified image of a region of the skin at a very high resolution to clearly identify the spots on the skin, and helps identify deeper levels of skin, providing more details of the lesions. Moreover, dermoscopy provides detailed visual context of regions of the skin and has proven to enhance the diagnostic accuracy of a naked eye examination, but it is costly, error prone, and achieves only average sensitivity in detecting melanoma[14]. This has triggered the need for developing more precise computer-aided diagnostics systems that would assist in early detection of melanoma from dermoscopy and clinical images. Further more the recent success of Deep learning models to accurately diagnose skin lesion images at the level of a dermatologist[15] has proved it to be a valuable tool to assist with task of skin cancer classification.

However, despite significant strides in skin lesion recognition, melanoma detection remain a challenging task to date due to various reasons. One of the factors contributing to this is the highly imbalanced nature of most datasets being used to train such deep learning models. For a deep learning model to work well, it requires a large amount of balanced data with sufficient examples of each class for the model to learn from. However in the case of most skin lesion datasets there is a huge imbalance. Common skin cancers, such as nevus, seborrheic keratosis, solar lentigo, etc., contain thousands of images, but rarer forms, such as melanoma, contain very few images which makes it difficult to generalize from only visual features.

Additionally, the presence of domain specific artifacts such as medical markings used by dermatologists to highlight regions of interest in an image, microscope scales, and body hair behave as distractions making it harder for the deep learning model to accurately classify the skin lesion sample. Prediction irregularities due to these biases induced by artifacts makes the problem harder (Figure 1).

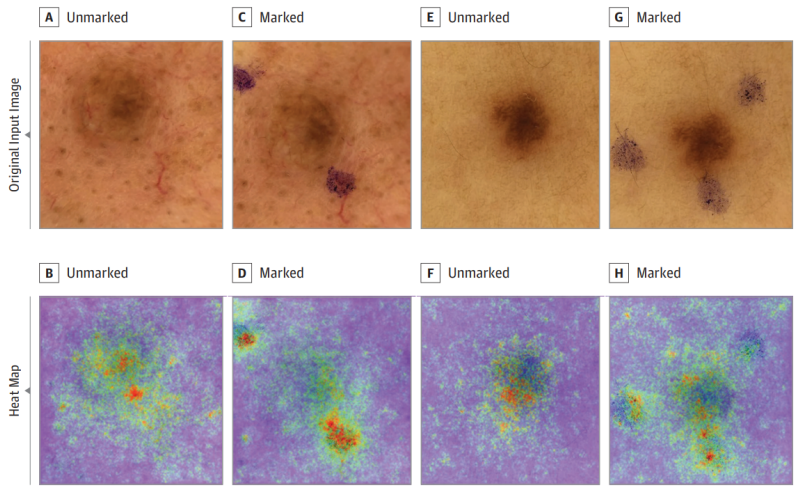

Figure 1: **Effect of medical markers on skin lesion images.** Heat maps reveal relevant pixels for convolutional neural network's (CNN's) prediction of the skin lesion. Heat maps reveal that medical markings on the skin are of high relevance for CNN's prediction of the skin lesion, while the nevus itself is ignored.[7]

As such, in order to successfully tackle the problem of melanoma detection it is imperative for us to address the imbalanced nature of the data as well as As a result, by tackling the problem of class imbalance in the data along minimizing the impact of domain specific artifacts on classification results, we can develop a robust model that is capable of accurately identifying melanoma sample.

2 Related Work

Automating skin cancer has been a long standing challenge due to a variety of reasons. The earliest forms of skin cancer detection involved capturing the ABCD aspects of skin cancer. Computer Aided Design [1] is used in preprocessing, feature extraction, feature fusion, and classification. This method is computationally very heavy and is not feasible for general applications. In automated diagnosis systems, texture and colour features are considered as two fundamental visual characteristics, which are vital for the detection of melanoma. Employing divergence-based colour features was presented in 2015 by Møllersen et al. [2]. Several studies have already been done employing texture-based features like generalized co-occurrence matrices [3], gradient histograms [4], and RSurf features [5]. Codella et al. [6] present a combination of deep learning, sparse coding, and support vector machine (SVM) learning algorithms for melanoma classification. Their method achieved a classification accuracy of 93.1 % on a dataset containing 2624 clinical cases of melanoma, atypical nevi, and benign lesions. The problem of markings affecting model performance was conducted in detail in [7]. Several different techniques were introduced to help remove this, one such effective method was proposed in [8]. We draw inspiration from this method of debiasing and incorporate it in our work.

A complete switch to deep learning techniques involved the exclusive use of deep convolutional neural networks, and detecting over 750 classes [9]. Since these models required a lot of hardware for both training and deployment, lighter models such as MobileNetv3 etc. began to become more popular [10] at the expense of increased output uncertainity. Ensemble techniques were the solution to this and so the winner [11] of the 2019 ISIC Challenge used an ensemble of supervised and unsupervised techniques to achieve high AUC scores. However it proved to be computationally very inefficient during training. Around the same time, incorporation of patient metadata while performing classification started gaining widespread popularity due to the surprising efficacy [12]. As seen in the previous literature mentioned above, either work has exclusively focused on overcoming the dataset imbalance through the use of multiple supervised and unsupervised techniques or the focus has been increasing model ability to recognise the rarest forms of cancer using different methods such as use of patient metadata.

Our method differs significantly from the previous work and SOTA methods due to the following approach: (1) Addressing dataset imbalance (3% positive samples) through incorporating carefully selected hand crafted features which provide a dermatologist intuition to the model and aid in correct classification. (2) Teaching the model to ignore artifact noise such as the presence of medical markers, microscope scale, body hair, etc. [shown in Figure 2] by utilizing domain specific augmentation and deep unlearning methods. (3) Performing a comparative analysis between the domain specific augmentation and deep unlearning methods to determine the optimal technique for artifact debiasing. (4) Validating model robustness by evaluating model performance on independent testing datasets, which contain challenging skin lesion images from a different distribution.

3 Proposed Method

3.1 Dataset

Dermoscopic skin lesion datasets with diagnosis labels and metadata from International Skin Imaging Collaboration (ISIC) challenge is used. A combination of data from 2017 to 2020 from the ISIC Challenge (35,574 images) is used, since in some of the years, a higher representation of artifacts is present as compared to other competition years. Images are pre-processed (center cropped and resized to an image size of 256×256) before training and the same techniques are applied during testing to ensure uniformity. 33%,(3326 images) of the 2018 challenge data used as the validation set for hyperparameter tuning.

To test model generalization ability, it is tested on different datasets and the results are presented in the Results section. These datasets include the MClass public human benchmark used as a test dataset for assessing domain generalisation, also providing a human benchmark. The dermoscopic MClass data is made up of images from the ISIC archive, some of which were also present in the ISIC training data, so these were removed from the training data to prevent data leakage. Two additional test sets, the Interactive Atlas of Dermoscopy dataset, and the Asan test dataset, are used to further test domain generalisation. The Atlas dataset comprises 1,011 lesions across 7 classes, with one dermoscopic and one clinical image per lesion. The Asan test dataset comprises 852 clinical images across 7 classes of lesions. Whilst the ISIC training data is mostly white Western patients, the Atlas and Asan datasets seem to have representation from a broad variety of ethnic groups, which provides a good test of a model's ability to deal with domain shift.

3.2 Artifact Debiasing

The first approach we experiment with to overcome the problems to algorithmic bias introduced by certain artifacts present in skin lesion images and domain generalisation in melanoma classification i.e. to find an instrument-invariant feature representation without compromising performance, is to remove bias and spurious variation from an automated melanoma classification pipeline using leading bias 'unlearning' techniques.

We add a debiasing architecture called 'Turning a Blind Eye' (TABE) to our baseline model. We use a joint loss over primary and secondary data to learn a feature representation that simultaneously

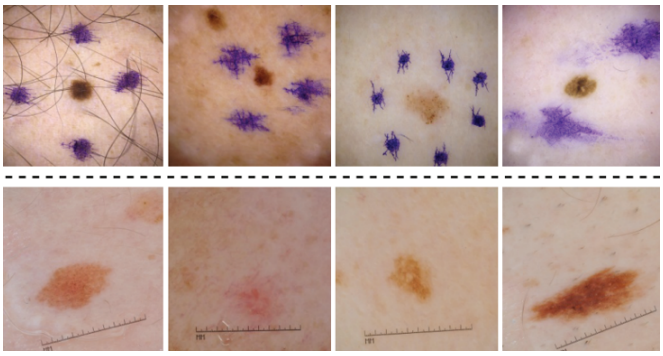

Figure 2: **Artifacts seen in ISIC 2020 data** : Medical Markers [top], Rulers [bottom]

learns to classify the primary task but becomes invariant to secondary tasks, the spurious variations. The spurious variation classification loss and confusion loss act in opposition to learn the classifier on the feature embedding and change the feature embedding to confuse the classifier, respectively. The complete architecture is shown in Figure 3.

We aim to remove unwanted bias using the auxiliary classifier, θ_m, where m is the m-th unwanted bias. One approach is to calculate the cross entropy loss for the artifact classification, multiply it with a negative scalar and add it to the primary classification loss. This results in a min-max game wherein we try to minimize the primary loss while maximising the auxillary loss at the same time resulting in the model having the worst performance on the artifact classification. Additionally, we add an auxillary confusion loss defined as the cross entropy loss between the softmax output of the artifact classes and a uniform distribution. Since a uniform classification output is equivalent to a random guess, training will produce the worst possible model performance for artifact classification. We also propose to add a fully connected layer to integrate the hand crafted features with the CNN features.

The auxiliary classifier minimises an auxiliary classification loss, $\mathcal{L}_s$, used to identify bias in the feature representation, $\theta repr$, as well as an auxiliary confusion loss, $\mathcal{L}_{conf}$, used to make θ_{repr} invariant to the unwanted bias. Since these losses stand in opposition to one another, they are minimised in separate steps: first Ls alone, and then the primary classification loss, $\mathcal{L}_p$, together with $\mathcal{L}_{conf}$. The confusion loss is defined as follows:

$$\mathcal{L}_{conf,m}(x_m, y_m, \theta_m; \theta_{repr}) = -\sum_{n_m} \frac{1}{n_m} \log p_{n_m} \tag{1}$$

where x_m is the input, y_m is the bias label, p_{n_m} is the softmax of the auxiliary classifier output and n_m is the number of auxiliary classes. This confusion loss works towards finding a representation in which the auxiliary classification head performs poorly by finding the cross entropy between the output predicted bias and a uniform distribution. The complete joint loss function being minimised is:

$$\mathcal{L}(x_p, y_p, x_s, y_s \theta_p, \theta_s, \theta repr) = \mathcal{L}_p(x_p, y_p; \theta repr, \theta p) + \mathcal{L}_s + \alpha \mathcal{L}_{conf} \tag{2}$$

where α is a hyperparameter which determines how strongly the confusion loss impacts the overall loss.

The results obtained using this method is seen in Table 1 and Table 2 under the architecture of ResNeXt 101 + TABE.

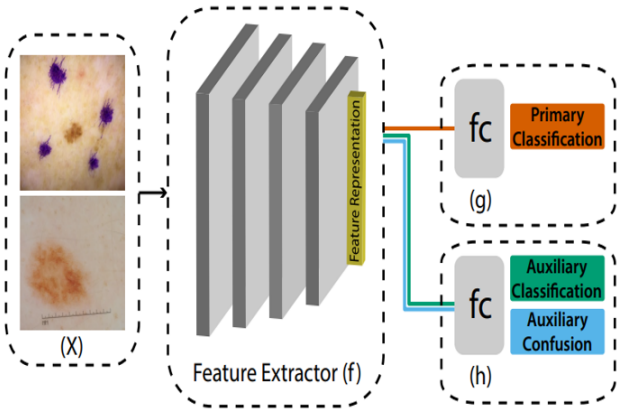

Figure 3: **'Turning a Blind Eye' generic architecture.** Feature extractor, f, is implemented as a convolutional architecture such as ResNeXt or EfficientNet in this work. 'fc' denotes a fully connected layer.

3.3 Data Augmentation

The second approach to solving the problems of artifact bias and imbalanced data is to perform data augmentation through the generation of synthetic images with custom designed artifacts on sampled

images. This has the effect of introducing randomness, forcing model to look past the noise while also alleviating the problem of imbalanced data. This method adds new data points to the input space and preserves semantic and temporal information. Some of the techniques used to do this are: blackHat filtering, morphological sampling, random artifact stamping, hair incorporation or random removal etc. Artifacts mimic real world noise and teach the model to generalize across images thus preventing overfitting and making it region unspecific.

First, functions that generate hair strands, medical markers, rulers etc are developed. Since the 2018 ISIC challenge is the only dataset which contain Ground truth masks used for segmentation, we train a UNet model to generate masks for other cancer images in our dataset. During the pre-processing step, a random decision is made to use either one of the artifacts described above or none at all. When this decision is made, the function generating this artifact is called which creates a stamped mask that is used to introduce noisy artifacts to the original image [shown in Figure 4]. Once these images are processed and prepared, they are fed as input to the baseline model of ResNeXt-101. The results obtained using this method is seen in Table 1 and Table 2 under the architecture of Baseline + Domain Augmentation.

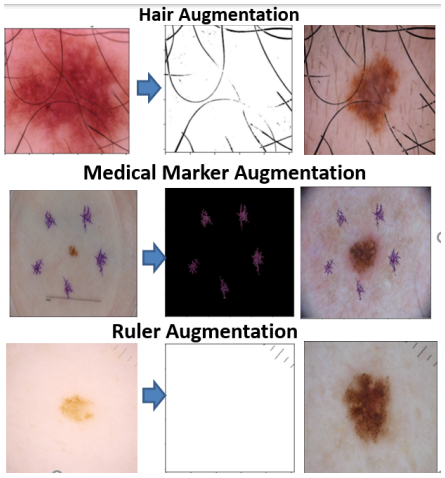

Figure 4: **Domain specific artifact augmentation.** Augmented images with synthetic artifacts (right) were created using artifact masks (middle) extracted from training samples (left)

3.4 Hand-crafted features

There are several studies that focus on classifying melanoma based on the "Asymmetry, Border, Color, and Diameter" (ABCD) metric to classify whether a lesion is benign or malignant. Scientists have been trying to find different ways to somehow use mathematical models to describe the ABCD metric. Previous work involves using SVMs, GMMs and kNNs for categorising the different features in the appropriate feature space but the disadvantage of using such methods is the processing time, computational inefficiency and the general difficulty in finding an appropriate feature space with linear separability. Others manipulated the dataset (either a very small number of images or images with visible differences) which made modelling easier. Clearly this dilutes the very purpose of automating and generalising the task of skin lesion classification. Hence our proposed usage of the following described hand crafted features captures the essence of the ABCD metric specified by the American Association of Dermatology (AAD):
1. **Asymmetry:** Typically symmetric lesions are benign while asymmetric lesions are irregularly shaped.

2. **Border:** Melanoma cases at most have uneven, asymmetric boundaries and appear as random Alia borders while benign cases have smooth edges.

3. **Colour:** Lesions appear in a variety of colours and it is important to take into consideration the effect of dyes and lighting techniques during segmentation tasks.

4. **Diameter:** Generally melanoma has higher diameter than benign lesions.

To incorporate this intuition based on the certified classification metrics, we feed the model carefully engineered hand-crafted features that capture the configuration, texture, and appearance of the lesion. Amongst others, Hu, Zernike Moments, Haralick features capture different aspects of the lesion's configuration. They are translation invariant, good shape descriptors and help capture and describe the **Asymmetry and Border** criteria. Local Binary Patterns (LBP) provide depth perception which is often absent in most computational models. The LBP in combination with shape descriptors help describe the **Diameter** of the lesions. Color Histograms assist in differentiating the lesions based on their appearance and perfectly describe the **Colour** metric.These hand crafted features were chosen after extensive experimentation and are added to our Baseline model with the extracted features from the CNN in the penultimate linear layer. The results obtained using this method is seen in Table 1 and Table 2 under the architecture of Baseline + Hand Crafted.

3.5 Selecting our final model

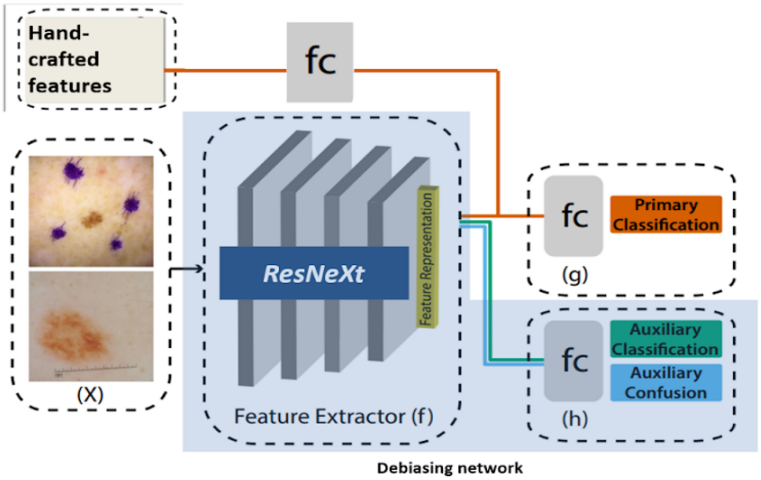

Figure 5: **Proposed model: ResNeXt-101 with Artifact Debiasing & hand-crafted features**

We have conducted experiments with three different methods: artifact Debiasing [Section 3.2], Data Augmentation [Section 3.3], and adding Hand-crafted features [Section 3.4]. We summarize the results [Table 1 & 2] of each of these and conduct a comparative analysis to pick the best performing approach.

Our experimental results provide evidence that the effects of each of the aforementioned biases are notably reduced, with the debiasing technique. Debiasing also shows generalisation benefits of 'unlearning' spurious variation relating to the imaging instrument used to capture lesion images. Clearly, debiasing provides better results than in the case of data augmentation.

We also note that Baseline + Hand-crafted shows an improvement over the baseline model results. We therefore take the route of artifact debiasing with the addition of hand-crafted features to build our final model.

3.5.1 Artifact Debiasing with hand-crafted features

The proposed model uses the ResNext-101 architecture for image feature extraction. It consists of a basic building block made up of a set of convolutional transformations in a homogenous, multibranch topology. Repetitions of this block constitute the ResNext-101 model [13]. Early experimentation showed ResNeXt-101 to be the overall best performing architecture and it is therefore used as the feature extractor in the domain generalisation experiments. The ResNext feature extractor is a 2-stage processing pipeline that sequentially performs local and global feature extraction which helps capture size, color and shape variations of lesions. A fully connected layer uses the feature representation generated by the ResNext model to perform skin cancer classification using a softmax loss function. This is known as the primary classification head and makes up the first stage of the classification pipeline. The second stage of the classification pipeline consists of an auxillary classification head

which classifies the artifacts present in the image (i.e. marker, ruler, hair, etc.). The objective is to teach the model to ignore the presence of artifacts while classifying skin lesions.

To address the imbalance in data, we concatenate hand crafted features with the extracted CNN features. To do this, we propose to add another fully connected layer to process the hand engineered features. This combined output is further processed algebraically to produce the final prediction score. The final architecture is shown in Figure 5. The concatenation has the effect of positively reinforcing pre-prediction decisions to confidently produce output scores.

4 Experimental Results

Following a grid search, the learning rate (searched between 0.03 and 0.00001) and momentum (searched between 0 and 0.9) are selected as 0.0003 and 0.9 respectively. The learning rate of the TABE heads is boosted by a factor of 10 (to 0.003). The best performing values of the hyperparameters α in Equation 2 is also empirically chosen to be $\alpha = 0.03$.

A weighted loss function is implemented for all model configurations to tackle class imbalance, with each weighting coefficient, $\mathcal{W}_n$, being the inverse of the corresponding class frequency, c. Since the proportion of benign and malignant lesions is highly imbalanced in the test sets, accuracy proved not to be a descriptive metric to use. Instead, AUC and F1 scores were used as metrics to measure model performance. Further more given the imbalanced nature of the dataset, the F1 metric was used to determine the best model as it represents the models ability to correctly identify the melanoma samples.

The results of our domain augmentation, debiasing and hand-crafted featured in different configurations with the baseline are shown in Table 1 (AUC scores) and Table 2 (F1 scores).

4.1 Domain specific augmentation vs. Deep unlearning

An Ablation study was performed on the two debiasing techniques by first evaluating the baseline (ResNext-101) performance in isolation before incorporating the debiasing techniques. Based on the results in Table 2 we see that the deep unlearning technique performance supersedes the data augmentation method on 3/5 test datasets. In certain cases the unlearning method shows up to a **56%** improvement compared to the augmentation method. As such we the model configuration utilizing deep unlearning methods (specifically with ruler debiasing) was proposed as the desired technique to handle artifact bias.

Experiment	**Architecture**	**Atlas**		**ASAN**	**MClass**	
		Dermoscopic	Clinical	Clinical	Dermoscopic	Clinical
Dermatologist	–	–	–	–	0.671	0.769
Baseline	ResNeXt-101	0.791	0.589	0.736	0.895	0.697
Baseline + Domain Aug	ResNeXt-101	0.828	0.682	0.874	0.852	0.789
Marker Debias	ResNeXt-101 + TABE	0.825	0.642	0.781	0.909	0.822
Ruler Debias	ResNeXt-101 + TABE	0.821	0.641	0.755	0.854	0.926
Baseline + Hand Crafted	ResNeXt-101	0.814	0.57	0.724	0.9225	0.765
Baseline + Domain Aug + Hand Crafted	ResNeXt-101 + DNN	0.803	0.594	0.698	0.743	0.797
Marker Debias + Hand Crafted	ResNeXt-101 + TABE + DNN	0.611	0.27	0.216	0.667	0.552
(Proposed) Ruler Debias + Hand Crafted	ResNeXt-101 + TABE + DNN	0.823	0.632	0.773	0.8825	0.866

Table 1: **AUC scores for the Model with different configurations.** The best performance for between the two debiasing techniques are highlighted in blue. Additionally, the best performance for the model configurations with handcrafted features is highlighted in red. All configurations were trained to converge to a minimum loss of 0.05

Experiment	Architecture	Atlas		ASAN	MClass	
		Dermoscopic	Clinical	Clinical	Dermoscopic	Clinical
Baseline	ResNeXt-101	0.46	0.133	0.028	0.703	0.3125
Baseline + Domain Aug	ResNeXt-101	0.43	0.21	0.26	0.62	0.42
Marker Debias	ResNeXt-101 + TABE	0.351	0.052	0.026	0.516	0.182
Ruler Debias	ResNeXt-101 + TABE	0.503	0.328	0.225	0.632	0.333
Baseline + Hand Crafted	ResNeXt-101	0.523	0.26	0.128	0.842	0.2
Baseline + Domain Aug + Hand Crafted	ResNeXt-101 + DNN	0.567	0.274	0.281	0.661	0.547
Marker Debias + Hand Crafted	ResNeXt-101 + TABE + DNN	0.611	0.27	0.216	0.667	0.552
(Proposed) Ruler Debias + Hand Crafted	ResNeXt-101 + TABE + DNN	0.618	0.333	0.307	0.619	0.533

Table 2: **F1 scores for the Model with different configurations.** The best performance for between the two debiasing techniques are highlighted in blue. Additionally, the best performance for the model configurations with handcrafted features is highlighted in red. All configurations were trained to converge to a minimum loss of 0.05

4.2 Ablative analysis with handcrafted features

Next an Ablation study was performed to study the impact of handcrafted features by evaluating the four model configurations (baseline, domain specific augmentation, marker unlearning, and ruler unlearning) in isolation before incorporating the handcrafted features to each of the configurations. From Table 2 we can see that incorporating handcrafted features almost always resulted in an improvement of model performance. Utilizing the handcrafted features in certain cases (Marker Debiasing) results in an improvement of F1 scores from 0.351 to 0.611, which is a **74%** increase. In more extreme cases the F1 score increases from 0.026 to 0.216 (which is a 10x improvement).

4.3 Comparing proposed model with baseline

Next, we compare our proposed model which consists of artifact debiasing + hand crafted features with the baseline ResNeXt model. From Table 1, we see a consistent increase in AUC scores which goes up to a **24%** increase in performance. From Table 2, we see a jump in F1 scores of up from 0.028 to 0.307.

5 Future Work

Currently, the debiasing method described in the previous section is able to unlearn single artifacts. Our future goal would be to extend the model to unlearn multiple noisy artifacts simultaneously. Given the computational efficiency of the proposed method, we believe it can further be made much more efficient. This would help us develop mobile friendly versions of the proposed architecture which would help in portability, remote access and democratisation. We also plan on exploring the effect of incorporation of patient metadata.

6 Conclusion

To summarize, we designed a developed a 2-stage pipeline where images are subjected to spatial operations that determine the relevant from irrelevant features and remove unwanted noise in the form of artifacts. Additionally, we incorporate hand crafted features to capture variations in appearance, texture and shape, to mimic the traditional weighted metric used in the determination of malignant lesions. Our approach introduces a novel method of data augmentation using custom designed masks for each artifact to generate synthetic data and introduce domain generalization. Given the highly imbalanced composition of the dataset, we synthesized auxiliary artifacts in addition to medical markers to successfully teach the deep unlearning model to ignore irrelevant features. We also integrated custom hand-crafted features into the model to increase prediction confidence which is in line with the medical community which attempts to identify the salient features of the images for the presence of a melanoma. Future work consists of improving sensitization of the deep unlearning model to multiple artifacts parallelly and to extend our work to predict the subtype of melanoma as well.

References

[1] Jose-Agustin Almaraz-Damian, Volodymyr Ponomaryov, Sergiy Sadovnychiy and Heydy Castillejos-Fernandez Melanoma and Nevus Skin Lesion Classification Using Handcraft and Deep Learning Feature Fusion via Mutual Information Measures

[2] K. Møllersen, J. Y. Hardeberg, and F. Godtliebsen, "Divergencebased colour features for melanoma detection," in Colour and Visual Computing Symposium. IEEE, 2015, pp. 1–6.

[3] L. Ballerini, R. B. Fisher, B. Aldridge, and J. Rees, "A color and texture based hierarchical k-NN approach to the classification of non-melanoma skin lesions," in Color Medical Image Analysis, pp. 63–86. Springer, 2013.

[4] C. Barata, M. Ruela, M. Francisco, T. Mendonc¸a, and J. S. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," IEEE Systems Journal, vol. 8, no. 3, pp. 965–979, 2014.

[5] T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "Efficient Melanoma Detection Using Texture-Based RSurf Features," in International Conference Image Analysis and Recognition. Springer, 2016, pp. 30–37.

[6] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," in International Workshop on Machine Learning in Medical Imaging. Springer, 2015, pp. 118–126.

[7] Winkler JK, Fink C, Toberer F, Enk A, Deinlein T, Hofmann-Wellenhof R, Thomas L, Lallas A, Blum A, Stolz W, Haenssle HA. Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition. JAMA Dermatol. 2019 Oct 1;155(10):1135-1141. doi: 10.1001/jamadermatol.2019.1735. PMID: 31411641; PMCID: PMC6694463.

[8] Peter J. Bevan and Amir Atapour Abarghouei,Skin Deep Unlearning: artifact and Instrument Debiasing in the Context of Melanoma Classification, CoRR, volume abs/2109.09818, 2021.

[9] Andre Esteva1, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun, Dermatologist-level classification of skin cancer with deep neural networks

[10] Adi Wibowo, Satriawan Rasyid Purnama, Panji Wisnu Wirawana, Hanif Rasyidib, Lightweight encoder-decoder model for automatic skin lesion segmentation

[11] Qishen Ha, Bo Liu, Fuxu Liu, Identifying Melanoma Images using EfficientNet Ensemble: Winning Solution to the SIIM-ISIC Melanoma Classification Challenge

[12] Shih-Cheng Huang, Anuj Pareek, Saeed Seyyedi, Imon Banerjee & Matthew P. Lungren, Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770–778, Las Vegas, NV, USA, June 2016. IEEE. 3, 4, 5, 6

[14] Zunair, Hasib, and A. Ben Hamza. "Melanoma detection using adversarial training and deep transfer learning." Physics in Medicine & Biology 65.13 (2020): 135005.

[15] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," IEEE Transactions on Medical Imaging, vol. 36, no. 4, pp. 994–1004, 2016.

[16] A. A. A. Al-abayechi, X. Guo, W. H. Tan and H. A. Jalab, "Automatic skin lesion segmentation with optimal colour channel from dermoscopic images," ScienceAsia, vol. 40S, pp. 1–7, 2014.

[17] D. N. H. Thanh, N. N. Hien, V. B. S. Prasath, L. T. Thanh and N. H. Hai, "Automatic Initial Boundary Generation Methods Based on Edge Detectors for the Level Set Function of the Chan-Vese Segmentation Model and Applications in Biomedical Image Processing," in The 7th International Conference on Frontiers of Intelligent Computing: Theory and Application (FICTA-2018), Danang, 2018.

[18] Al-Masni MA, Al-Antari MA, Choi MT, Han SM, Kim TS. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. Computer Methods and Programs in Biomedicine. 2018;162:221–231. doi: 10.1016/j.cmpb.2018.05.027.

[19] Burdick J, Marques O, Weinthal J, Furht B. Rethinking Skin Lesion Segmentation in a Convolutional Classifier. Journal of Digital Imaging. 2018;31(4):435–440. doi: 10.1007/s10278-017-0026-y.