EDITH COWAN UNIVERSITY

CSG2341: INTELLIGENT SYSTEMS

LITERATURE REVIEW

Machine learning in dermoscopic diagnostics of malignant cutaneous neoplasms

Authors

Mark JAMSEK (10502496)

Joshua SOTO-KITCHER (10471169)

Jason KNOX (10467971)

Supervisor

Dr. Jumana Mahmoud ABU-KHALAF

Group Set 2 Team 10

https://cvs.jamsek.dev/csg2341-project

Due: September 20, 2020

Submitted: 20 September 2020

Abstract

Skin cancer is a fatal disease that impacts one million Australians every year. Due to the inherent difficulty of accurately diagnosing skin lesions, and the critical importance of early treatment, the purpose of this report is to review the literature on machine learning enhanced diagnostics of cutaneous neoplasms. Initial research found that convolutional neural networks produce the most accurate image classification models, which led to further investigation into the binary classification models developed with the AlexNet, ResNet, and VGGNet architectures that were employed in various skin cancer detection projects. Further, with the problem of distinguishing not only benign from malignant naevi but also melanocytic and non-melanocytic skin cancers, additional research into the multiclass classification model implemented with EfficientNet was also performed. Subsequent findings reveal that SoftMax and Support Vector Machine (SVM) image classification functions were favoured, and consistently produced models testing above 90% accuracy. Interestingly, these results were irrespective of the dataset size. In addition, an on-device inference application was reviewed, which highlighted both the challenges and prospects of bringing this technology to mobile devices. As a final point, we propose a detailed plan to design and develop a VGGNet convolutional neural network to classify images of skin lesions from the PH^2 and/or HAM10000 datasets using a SoftMax and/or SVM classifier; with the secondary objective of researching the viability of adapting this technology to mobile devices.

Contents

Introduction

According to the Australian Institute of Health and Welfare (2016), Australia has the highest rate of skin cancer in the world, with approximately one million cases diagnosed each year. Consequently, as the most common form of cancer in Australia, two-thirds of the population will receive a positive diagnosis by the age of seventy. However, although malignant melanoma is the most aggressive, there are several more common types such as squamous and basal cell carcinoma that are much more responsive to treatment if identified early (Australian Institute of Health and Welfare, 2016). For this reason, this report will present findings from a comprehensive review of the machine learning literature insofar as it pertains to the image classification of cutaneous neoplasms and its diagnostic application.

Initially, skin cancers are primarily identified via a visual diagnosis before being confirmed with a histological examination; however, this method is not trivial and is susceptible to irreproducible results (Adegun & Viriri, 2020, p. 7160; Yu et al., 2017, 238–239). Because of this, there has been extensive research in developing intelligent systems to enhance the dermoscopic diagnosis of skin lesions, which allows for myriad improvements such as, inter alia, increased screening rates (Yap et al., 2018, p. 1261), and reduced healthcare costs (Fisher et al., 2020, p. 87). Further, more recently, given the ubiquity of smartphones that can provide high definition photographic images, research into developing automated classification using images that can be taken by anyone, anywhere has yielded promising results (Dai et al., 2019). Subsequently, given the nature of this problem, which in sum and substance involves discerning a category of skin lesion from a given image, the large body of research is invariably comprised of the machine learning type known as supervised learning; in particular, using convolutional neural networks—considered most optimal for the given task (Naeem et al., 2020, p. 110575). Accordingly, the focus of this report is primarily on this approach, and will explicate various binary and multiclass classification techniques that fall under this category.

First, the approaches reviewed in this report will be classified to provide a taxonomical overview of the subset of machine learning that is pertinent to the problem of image classification. Second, through a comprehensive albeit concise review of multiple works, a description of different convolutional neural networks used to enhance dermoscopic skin cancer diagnostics by classifying images of skin lesions will be presented. Next, a comparison of the performance and results of these approaches will be discussed and viewed through the lens of challenges and corresponding implications with an eye toward future research. Lastly, before considering inherent limitations and potential improvements, the denouement will summarise the findings of this review and highlight future prospects.

Classification

Machine learning is a type of artificial intelligence, typically further divided into three main types (Table 1): supervised learning; unsupervised learning; and reinforcement learning. In the supervised approach, the system receives a set of paired values $\{(x_i, y_i)\}$—typically called the training set—and learns a mapping from input x to output y (Negnevitsky, 2011, p. 427). The result of this mapping function enables the model to generalise to previously unseen examples, thereby producing a categorical response, known as classification or pattern recognition (Murphy, 2012, p. 2).

Table 1

Machine Learning Types

Type	Description
Supervised	Supervised learning algorithms experience a dataset where each data point is associated with a label; by observing these associations the model learns to predict y from x by estimating $p(y \mid x)$.
Unsupervised	Unsupervised learning algorithms experience a dataset that contains myriad features, and identifies interesting properties of its structure by observing multiple data points of a random vector x and learning the probability distribution $p(x)$.
Reinforcement	Reinforcement learning algorithms interact with a dataset, creating a feedback loop between the system and its experiences, and discovering actions that yield the best reward through experimentation.

Note. Adapted from *Deep Learning* (pp. 102–103), by I. Goodfellow, Y. Bengio, and A Courville, 2016, The MIT Press. Copyright 2013 by Massachusetts Institute of Technology.

For this reason, we aver that supervised learning is most applicable to the problem of diagnosing cutaneous neoplasms. By training a model on a dataset of skin lesions, the intelligent system will estimate a function that is capable of outputting predictions on novel images; that is, the model will learn how to generalise beyond the training set to classify new images of skin lesions. This position is supported by Kaymak et al. (2018, p. 2), who state that, "[c]onvolution neural networks are among the strongest methods in deep learning which are used for detection, classification and segmentation of melanoma." However, there are various types of classification, which can be categorised as follows: binary classification, which has a single target variable $t \in \{0, 1\}$ such that $t = 1$ denotes class C_1 and $t = 0$ denotes class C_2; multiclass classification where each input maps to one of K mutually exclusive classes where target variables $t_k \in \{0,\ 1\}$ have a 1-of-K encoding designating the class; and multilabel classification where each input instance x is mapped to vector y, such that $y_i \in \{0, 1\}$, to assign multiple classifications per output (Bishop, 2006). Accordingly, given the fundamental problem of establishing whether a particular skin lesion is either benign or malignant, this review will focus on several binary classification techniques. However, due to there being several types of skin cancer—melanoma, and squamous and basal cell carcinoma—the multiclass classification model will also be reviewed. In addition, with the secondary objective of ascertaining the viability of using mobile devices to assist in skin cancer detection, an explication of an on-device inference application (app) will conclude the reviewed approaches, which have been classified and presented in Table 2 along with several more common binary and multiclass classification algorithms.

Table 2

Classification Types

Binary	Multiclass
AlexNet (ECOC SVM)	EfficientNet (Bayesian)
ResNet (SVM)	On-device (SoftMax)
VGGNet (SoftMax)	K-Nearest Neighbours
Decision Forest	Naive Bayes

Note. Reviewed approaches displayed with architecture and classifier denoted as *A (B)*.

Therefore, in the following section, a description of the CNN is provided, together with reviews of binary classification models developed with the AlexNet, ResNet, and VGGNet architectures that were used in various skin cancer detection projects.

Further, to provide contrast, multiclass classification models developed with the EfficientNet architecture and an on-device inference app have also been reviewed.

Approaches

Various network architectures have been used to implement a supervised learning, intelligent system that classifies images of skin lesions as a malignant neoplasm diagnostician with accuracy rates comparable to that of the human eye (Table 3).

Table 3

Comparison of skin cancer detection projects

Study	Dataset	Layers	Architecture	Classifier	Ac	Se	Sp
Attia et al., 2017	ISBI 2016 1,275 (900/375)	14	CNN/RNN Hybrid	-	98	95	94
Dorj et al., 2018	3,753 (2985/768)	12	AlexNet	ECOC SVM	94	98	91
Ge et al., 2017	MoleMap 26,584	41 & 22	VGG-16 & GoogleNet	Novel - SoleNet, ShareNet & TripleNet	97	-	-
Yu et al., 2017	ISBI 2016 1,279 (900/379)	50	ResNet-50	SVM (Chi-squared kernel)	87	43	98
Yap et al., 2018	2,917	41×2	$2 \times$ ResNet-50 + fusion	SoftMax	72	-	-
Alom et al., 2019	Kaggle 1,279 (900/379)	34	RU-Net & R2U-Net	SoftMax	95	92	95
Kumar et al., 2019	PH^2 100	9	SWT	SVM	93	92	90
Yan et al., 2019	ISIC 2016 & 2017 379	41	VGG-16	SoftMax	-	-	-
Swain et al., 2020	1,630 (1400/230)	4	CNN	SoftMax	93	-	-

Note. Where available, the *Dataset* column lists collection name with training and testing subsets denoted as (x/y), otherwise only total images experienced are shown. Ac = accuracy; Se = sensitivity; and Sp = specificity, with values in %.

Interestingly, the literature indicates no predominant network architecture, while some skin cancer detection projects use an ensemble. Conversely, in recent years, Support Vector Machine (SVM) and SoftMax classification algorithms have been favoured (Attia et al., 2017; Ge et al., 2017; Yap et al., 2018). The varied results of these approaches can be attributed to different architectures, layer depths, classification algorithms, and datasets. A direct comparison between AlexNet, VGG-16 and ResNet-50 network architectures using sensitivity (Sen), specificity (Spec), mean average precision (MAP), accuracy (Acc), and area under receive operation curve (AUC) metrics have been documented in Table 4.

Table 4

Impact of network architectures on skin lesion classification results

Network	Parameter	Layer	Sen	Spec	mAP	Acc	AUC	Time
AlexNet	61 M	Conv5	40.00	95.72	61.37	84.70	82.08	0.94 s
VGG-16	138 M	Conv5_3	45.33	94.08	57.66	84.43	81.18	2.72 s
ResNet-50	25.6 M	Conv5_9	45.33	96.71	65.08	86.54	81.49	1.33 s

Note. All values in % unless otherwise noted. Adapted from 'Aggregating Deep Convolutional Features for Melanoma Recognition in Dermoscopy Images' by Z. Yu, X. Jiang, T. Wang, & B. Lei in Q. Wang, Y. Shi, H.I. Suk, & K. Suzuki (Eds), *Machine Learning in Medical Imaging: Lecture Notes in Computer Science* (p. 244), 2017, Springer. Copyright 2017 by Springer International Publishing AG.

Convolutional Neural Network

According to Patterson and Gibson (2017, pp. 125–126), convolutional neural networks (CNNs) are remarkably successful in computer vision. This makes CNN models ideal in the nascent field of image classification for medical diagnosis. For this reason, the CNN approach will be elucidated, with different architectures and classifiers described, to present a clear and concise analysis of this highly effective artificial intelligence technique.

Dorj et al. (2018, p. 9911) explain that the CNN is comprised of three distinct layers: convolutional; pooling; and fully connected. The convolutional layer is the primary component, although the three work in concert to produce a synergistic effect; that is, the first two perform feature extraction, while the fully connected are classification layers (Patterson & Gibson, 2017, pp. 128–129). In addition, the convolutional layer has three configurable hyperparameters—depth, stride, and padding—that optimise performance by controlling its output volume with the formula $\frac{W-K+2P}{S}+1$

where W is the size of the input neuron, K is the size of the filter, P is the padding size, and S is the stride length (Dorj et al., 2018, p. 9913). However, in the case of diagnosing skin lesions where the input is a multidimensional array of image data, convolution involves a linear operation of input matrices (M_A, N_A) such that $C(i,j) = \sum_{m=0}^{(M_A-1)} \sum_{n=0}^{(N_A-1)} A(m,n)B(i-m, j-n)$ where $0 \leq i < M_A + M_B - 1$ and $0 \leq j < N_A + N_B - 1$ (Dorj et al., 2018, p. 9913). More specifically, the filter is applied to the input to extract pertinent features. First, the depth value sets how many filters are to be applied, which invariably corresponds to the number of channels in the input image; for example, an RGB colour image will necessitate three filters—one for each channel (Patterson & Gibson, 2017, pp. 128, 138; Dorj et al., 2018, p. 9912). Next, Patterson and Gibson (2017, p. 139) explain, the stride parameter determines how many pixels the filter will traverse before performing another calculation on the input matrix, with the stride and output size inversely related. Lastly, padding controls the spatial size of the image by effectively encompassing the input with additional layers to both preserve information of corner and edge pixels and allow more convolutions thereby enabling deeper networks (Patterson & Gibson, 2017, p. 299). The output of the convolution layer—$C(i,j)$—is the feature map that is then passed into the pooling layer, which effectively reduces spatial dimensionality to lessen computational demands and mitigate the risk of overfitting (Patterson & Gibson, 2017, p. 140). Most commonly, average- and max-pooling reduction methods are employed to define mean and peak values, respectively, in its data-reducing calculations; the results of which are input into the fully connected layer, which then classifies the image (Patterson & Gibson, 2017, pp. 128–129).

In the following sections, the binary classification models AlexNet, ResNet and VGGNet will be reviewed before multiclass classification models EfficientNet, and a novel on-device CNN inference app, are explored in more detail to provide a contrast of these popular convolutional neural networks.

Binary Classification

AlexNet

AlexNet is a large, deep convolutional neural network that won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) in 2012 (Khan et al., 2018). The network is comprised of eight parameter layers including five convolutional layers and three fully connected layers and implements dropout techniques to reduce overfitting (Figure 1). According to Krizhevsky et al. (2017, p. 88), AlexNet models learned best with a minimal weight decay of 0.0005, momentum of 0.9, and an adaptive learning rate initialised to 0.01 that would decrease gradually over three adjustments, using the error updating formula:

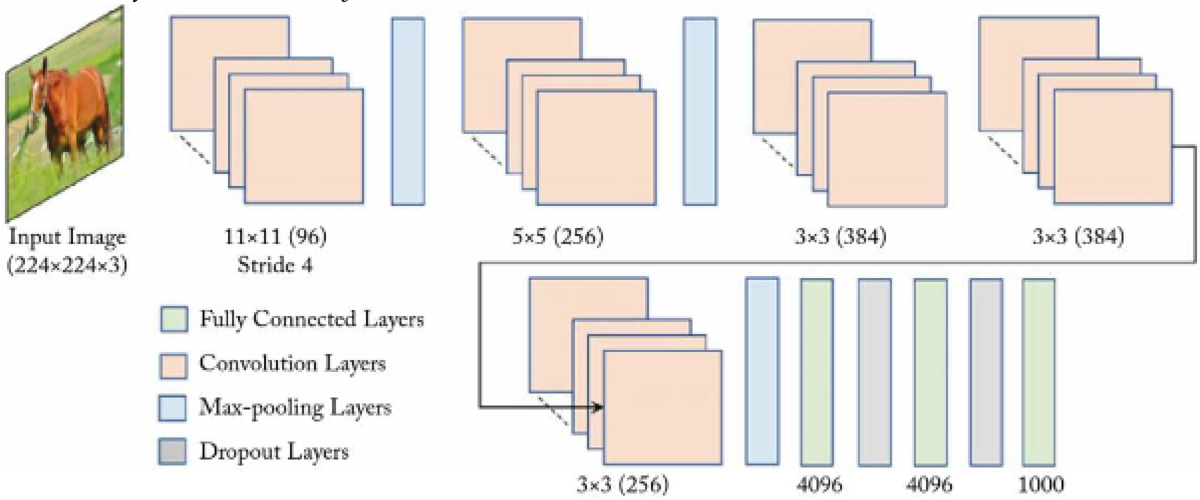
$$
\begin{aligned}
v_{i+1} &:= 0.9 \cdot v_i - 0.0005 \cdot \epsilon \cdot w_i - \epsilon \cdot \left\langle \frac{\partial L}{\partial w} |_{w_i} \right\rangle_{D_i}, \\
w_{i+1} &:= w_i + v_{i+1},
\end{aligned}
$$

where i is the current iteration, v is momentum, and ϵ the learning rate, which was consistently applied across all layers.

AlexNet has been used on various skin cancer detection projects as well as being used in an ensemble of network architectures with comparable accuracy rates ranging from 87% to 94% (Dorj et al., 2018; Harangi et al., 2018; Yu et al., 2017).

Figure 1

Rendition of convolutional layers



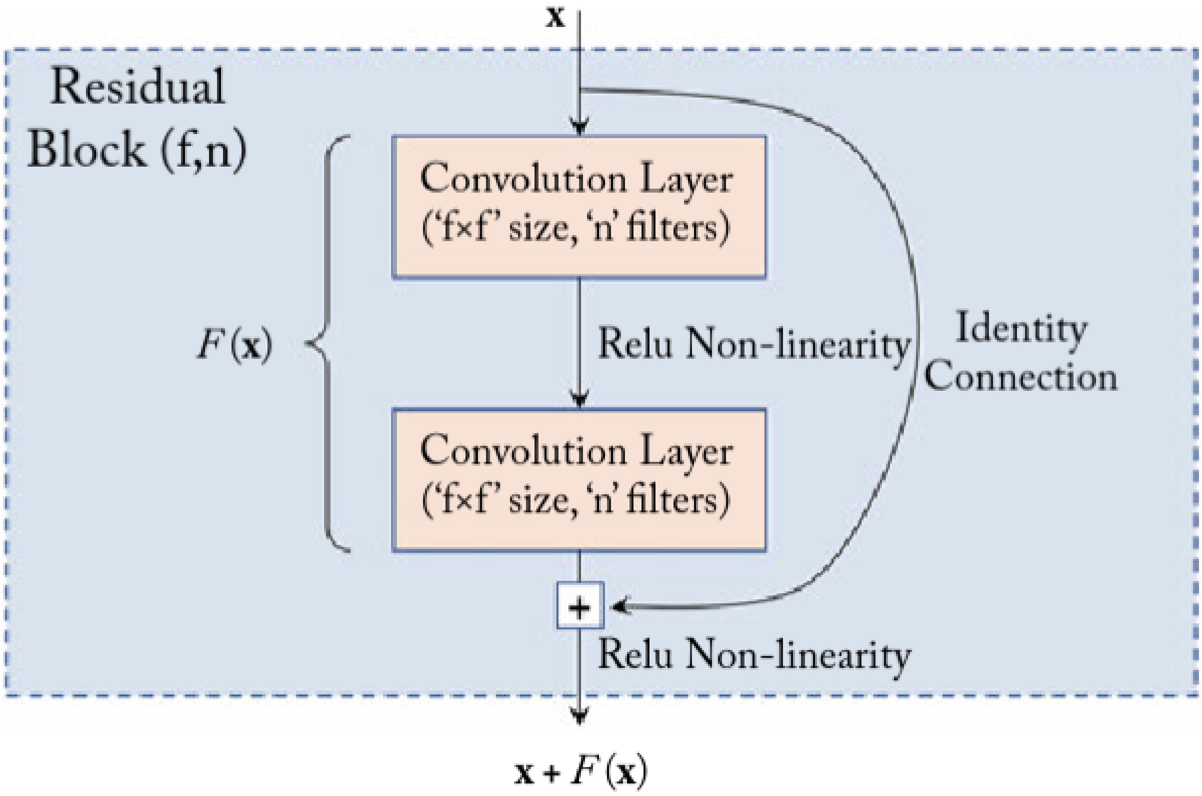
Note. From "*A Guide to Convolutional Neural Networks for Computer Vision*" by S. Khan, H. Rahmani, S. Shah, & M. Bennamoun, 2018, p. 103 (https://doi.org/10.2200/S00822ED1V01Y201712COV015). Copyright 2018 by Morgan & Claypool Publishers.

ResNet

A Residual Neural Network (ResNet), comprised of stacked residual blocks (Figure 2) with similarities to an inception model used by GoogleNet, was developed by Microsoft Research and won the ILSVRC in 2015 with a top-5 error rate of 3.57% after experiencing an ImageNet dataset using a depth of up to 152 layers (Khan et al., 2018). Meanwhile, He et al. (2019) report that the network enables easier training of deep CNN architectures due to an identity mapping feature that skips identity connections in the residual blocks. Similar to AlexNet, ResNet has been used in various skin cancer detection projects (Alom et al., 2019; Yap et al., 2018; Yu et al., 2017).

Figure 2

The Residual Block



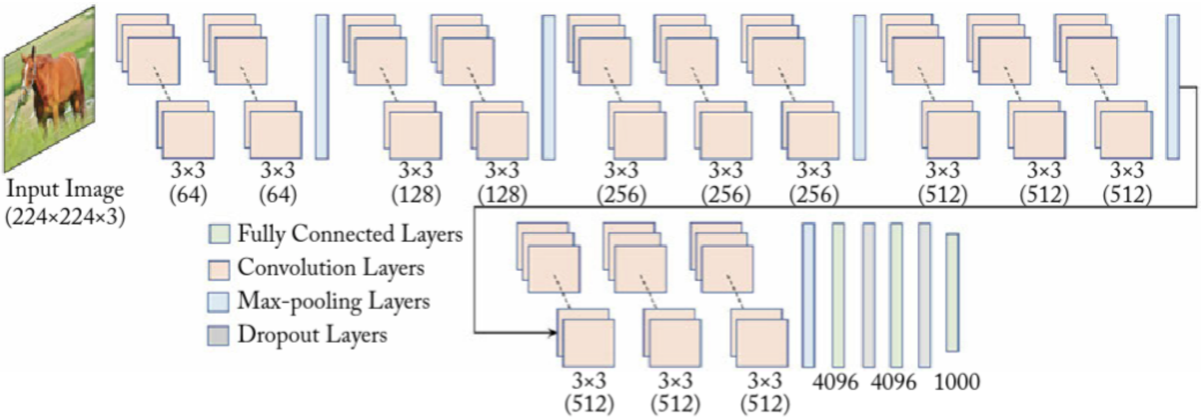
Note. From "*A Guide to Convolutional Neural Networks for Computer Vision*" by S. Khan, H. Rahmani, S. Shah, & M. Bennamoun, 2018, p. 109 (https://doi.org/10.2200/S00822ED1V01Y201712COV015). Copyright 2018 by Morgan & Claypool Publishers.

VGGNet

According to Khan et al. (2018, p. 104), an important feature of the VGGNet architecture is its smaller filters. This facilitates more layers, which enables deeper networks, and results in improved performance on vision tasks—an important advantage in image classification. Similarly, the use of smaller kernels can result in a reduced number of parameters, which improves efficiency in training and testing (Khan et al., 2018). In addition, VGGNet uses activation dropouts in the fully connected layers to reduce over-fitting (Khan et al., 2018). Topologically, the network is comprised of 3x3 convolution kernels with each layer followed by a ReLU layer, max-pooling layers, and three fully connected layers shown in Figure 3 (Khan et al., 2018). These characteristics combined with its model simplicity make VGGNet a popular choice for image classification.

Figure 3

The VGGNet-16 architecture



Note. From "*A Guide to Convolutional Neural Networks for Computer Vision*" by S. Khan, H. Rahmani, S. Shah, & M. Bennamoun, 2018, p. 104 (https://doi.org/10.2200/S00822ED1V01Y201712COV015). Copyright 2018 by Morgan & Claypool Publishers.

Multiclass Classification

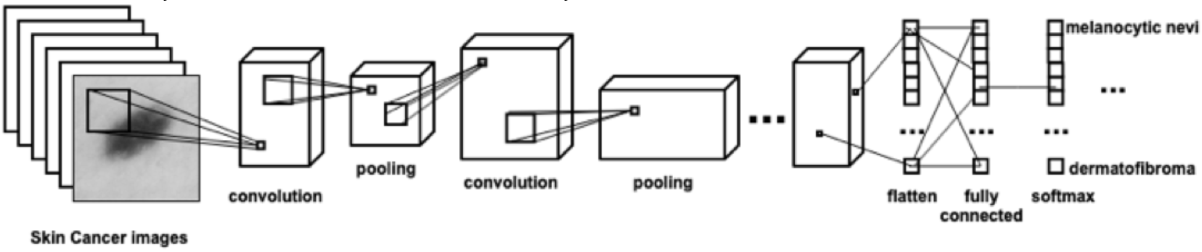
On-Device Inference

Traditionally, machine learning applications are extremely computationally intensive; hence, if a machine learning model was built to be used on a mobile device (e.g., smartphone, tablet) it would be executed in the cloud. In contrast, Dai et al. (2019) discuss an approach of using a CNN to be executed locally on a mobile device. The objective of this approach is to create a classification system which can execute almost instantly with limited computational power.

First, as shown in Figure 4, the architecture of the machine learning model has a series of convolution and pooling layers that perform feature extraction. After this, the outer layers flatten the data before the probability is calculated via a function in the

Figure 4

Architecture of Convolutional Neural Networks for Skin Cancer Detection



Note. From 'Machine Learning on Mobile: An On-device Inference App for Skin Cancer Detection' by X. Dai, I. Spasić, B. Meyer, S. Chapman, & F. Andres, 2019, *Fourth International Conference on Fog and Mobile Edge Computing*, p. 302. (https://doi.org/10.1109/FMEC.2019.8795362). Copyright 2019 by IEEE.

connected layer. Dai et al. (2019, p. 303) report that the algorithms are implemented with Python, TensorFlow, Scikit-learn, and Keras. The convolutional layers are used for the machine learning algorithm to recognise and activate upon certain patterns within the image. Whereas pooling, Dai et al. (2019, p. 303) explain, is used to reduce the parameters and remove non-essential features to mitigate overfitting. In this case, max-pooling is implemented using the MaxPool2D Python module (Dai et al., 2019, p. 303).

The outer layer, which flattens the data, receives a multidimensional array from the previous convolutional layer and converts it to a two-dimensional array. The connected layer uses the SoftMax Function (Figure 5) to calculate the probabilities of each type of malignant neoplasm and presents the most likely classification to the end-user.

Figure 5

The SoftMax Function

$$\sigma(Z)_j = \frac{e^{Z_j}}{\sum_{k=1}^{K} e^{Z_k}} \text{ for } j = 1, \ldots, K$$

Note. Z = input vector; j = output units. From 'Machine Learning on Mobile: An On-device Inference App for Skin Cancer Detection' by X. Dai, I. Spasić, B. Meyer, S. Chapman, & F. Andres, 2019, *Fourth International Conference on Fog and Mobile Edge Computing*, p. 303. (https://doi.org/10.1109/FMEC.2019.8795362). Copyright 2019 by IEEE.

Dai et al. (2019, p. 303) argue that to successfully implement the above model with limited computational power, the input data had to be augmented to minimise the number of input parameters. This was achieved by randomly performing cropping, rotating, shifting, or zooming with the image focal point remaining on the skin lesion.

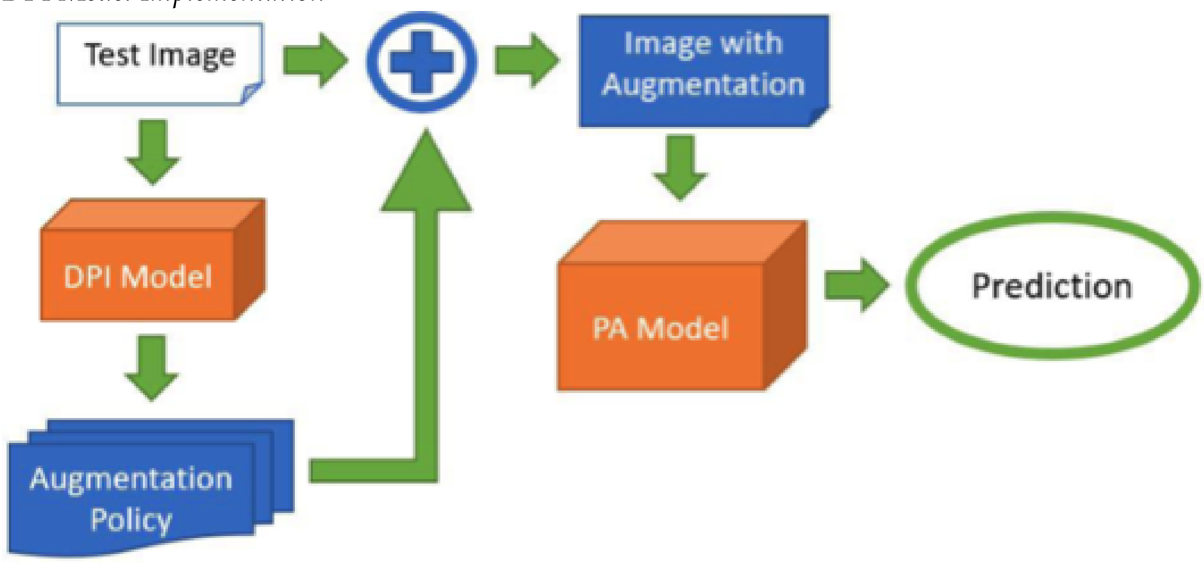
After the model has been trained on a dataset on a powerful computer, the model is stored on the mobile device. As a result, the user can take a picture of their skin privately, and the algorithm is computed locally on the phone. However, if the model is stored on an iOS app, there are additional steps that must be performed to be integrated by Apple's *Core ML* framework, which allows for more effective use of the hardware (Dai et al., 2019, p. 304). Conversely, if on an Android device, the TensorFlow model can be converted via TensorFlow Lite, which skips the conversion process entirely.

EfficientNet

Putra et al. (2020, p. 40544) propose the idea of a Dynamics Pre-processing Inference (DPI), which is a Bayesian optimisation method used to select an augmentation policy for a machine learning model. It allows separate images in a dataset to dynamically have an augmentation policy rather than having one statically assigned for each image in each dataset.

Figure 6

DPI Model Implementation



Note. From 'Enhanced Skin Condition Prediction Through Machine Learning Using Dynamic Training and Testing Augmentation' by T. A. Putra, S. I. Rufaida, & J-S Leu, 2020, *IEEE Access*, p. 40541. (https://doi.org/10.1109/ACCESS.2020.2976045). Copyright 2020 by IEEE.

As shown in Figure 6, the image from the dataset goes through the DPI model to find the augmentation policy for that specific image. Once the policy is found, the test image is modified based on the augmentation policy, and is used as an input to the EfficientNet CNN (Putra et al., 2020, p. 40537). This model is used to find a probability of a type of skin cancer based on an image of a skin lesion. Importantly, Putra et al. (2020) note that this process of dynamically assigning an augmentation policy can work with any CNN.

To achieve dynamic augmentation for inference, the following is defined as a convolutional neural network function: $P_\vartheta : \chi_i - > \Psi$. This function maps χ_i to the augmentation space Ψ that contains all augmentations, which enables defining the training objective shown in Figure 7.

Figure 7

Augmentation Function

$$\min_{\vartheta} \sum_i \sum_j z_{i,j} \times \log(P_{\vartheta,j}(x_i)),$$

Note. z_i = top-k augmentation of the i-th dataset. From 'Enhanced Skin Condition Prediction Through Machine Learning Using Dynamic Training and Testing Augmentation' by T. A. Putra, S. I. Rufaida, & J-S Leu, 2020, *IEEE Access*, p. 40540. (https://doi.org/10.1109/ACCESS.2020.2976045). Copyright 2020 by IEEE.

Putra et al. (2020, p. 40540) explain that the index j (on $z_{i,j}$) is a boolean value identifying when the j-th augmentation is in the top-k, such that:

$$z_i = top_k(-\ell(f_\theta(\Psi(\chi_i), y_i)))$$

where ℓ is the loss function of the main machine learning model. Subsequently, the task of finding the top-k is now a multilabel classification problem. The inference augmentation model is then trained using x_i and z_i, which then finds which augmentation works best for the selected input image. Once the network is trained with the multilabel data on inference, the maximum logit value can be selected to get the best augmentation for a given image.

Challenges

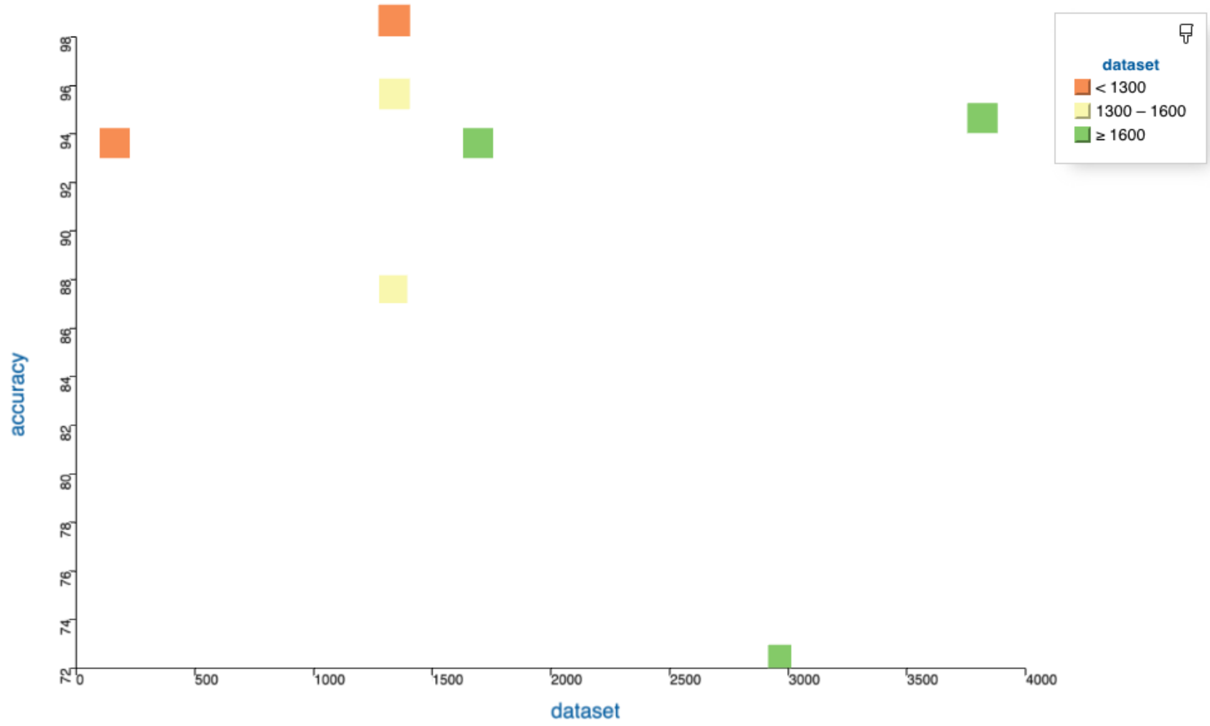
Despite the various intricacies of neural network models used for image classification, the reviewed approaches exhibit many similarities and display comparable performance (Table 3). Most notably, multiple models produced similar accuracy rates irrespective of the size of the dataset experienced. For example, Ge et al.

(2017, p. 257) presented a VGG-16 network trained on a dataset of 26,584 images using a Novel classification algorithm that realised a 97% accuracy rate. On the other hand, a hybrid recurrent-convolutional neural network running a SoftMax algorithm achieved 98% with just 1,275 data points (Attia et al., 2017, p. 294). Consequently, the findings of this meta-analysis somewhat counterintuitively indicate a weak correlation between dataset size and predictive ability (Figure 8). This raises some interesting questions regarding the apparent dominance of convolutional neural networks, and suggests that further investigation into hybrid models is warranted. Similarly, the observation that the SoftMax algorithm, which is commonly implemented in convolutional networks, returned the most disparate results—72% to 98% accuracy—suggests that the performance of these models are multifactorial, with a nexus of variables contributing to the results. As such, careful consideration should be given to the design and implementation of neural networks. For instance, rather than adjusting multiple variables at once, single, deliberate tuning of one hyperparameter or topological component at a time enables more reliable inferences to be drawn. To demonstrate, Ozkan and Koklu (2017, p. 287) trained multiple models on the PH^2 dataset while holding the hyperparameters shown in Table 5 constant, and manipulating just the number of neurons in the hidden layer to identify which number produced the best result. Accordingly, it was established that 18 hidden layer neurons were optimal, with an accuracy rate of 92.5%.

Adegun and Viriri (2020, p. 7161) explain that, due to their visual complexity (e.g., irregular demarcations, inhomogeneous features), the accurate diagnosis of skin lesions is an inherently challenging task. This is further compounded by the similarity between benign and malignant naevi, which confounds both clinician and intelligent system alike (Ge et al., 2017, p. 250). Given the present era of Big Data, however, and increasingly inexpensive computational power, future research is promising. Nonetheless, as greater dependency on machine learning to solve complex problems continues to grow, Yap et al. (2018, p. 1266) argue that efforts to mitigate biases should feature prominently in future work. This is particularly important in problem domains that impact the health sector with both personal and public ramifications.

Figure 8

Dataset vs Accuracy



Note. Scatter plot of seven cited studies illustrates a weak correlation between dataset size and performance.

Table 5

Backpropogation Training Parameters

Parameter	Value
Learning Rate	0.01
Momentum Constant	0.9
Epochs	1000–10000
Performance Function	Cross-entropy
Minimum Performance Gradient	$1/(e^{-10})$

Note. From 'Skin Lesion Classification using Machine Learning Algorithms' by I. A. Ozkan and M. Koklu, 2017, *International Journal of Intelligent Systems and Applications in Engineering, 5*(4), 287. (https://doi.org/10.18201/ijisae.2017534420). Copyright 2017 by Advanced Technology & Science.

Conclusion

In this work, we acknowledge the severity of skin cancer among the Australian demographic, and propose convolutional neural networks as a supervised machine learning solution for enhanced dermoscopic diagnosis of cutaneous neoplasms.

Specifically, trained models offer more robust image classification that enable more accurate diagnosis, and the potential for higher rates of more accessible screening. As a result, several popular convolutional neural networks that performed binary and multiclass classification were compared and several important observations were made. As noted, multiple architectures of various topologies, dataset size notwithstanding, obtained comparable results. This provides an explanation for why no dominant architecture exists. On the other hand, our findings indicate a preference for SoftMax and Support Vector Machine algorithms in both binary and multiclass classification, which were prevalent across all architectures with accuracy rates consistently above 90%. Closer analysis of the AlexNet architecture revealed that the best results were achieved with a minimal weight decay of 0.0005, a relatively high momentum of 0.9, and an adaptive learning rate initialised to 0.01, which recorded accuracy rates as high as 94%. Another important observation is that max-pooling appears to be preferred, as it was invariant in all five approaches studied. Similarly, smaller filters were also favoured, to which the resultant deeper networks and improved image classification were attributed. Meanwhile, our review of an on-device inference app highlighted the potential of mobile devices in delivering more accessible skin cancer diagnostic tools to a larger demographic. The prospects of which are of great import: almost instantaneous diagnosis delivered to users in the privacy of their own home; widespread screening; and private and public healthcare savings. Despite the numerous benefits, several key limitations exist. First, the visual complexity of skin lesions foment a reliance on high quality dermoscopic images. Second, the computation constraints of embedded devices pose engineering challenges that must be met. Nonetheless, possible improvements also exist; in particular, hybrid convolutional-recurrent neural networks, despite experiencing a relatively small dataset, obtained the highest accuracy, which shows considerable promise for future research. Further, the potential influx of large-scale image aggregation through the development of skin lesion diagnostic apps for mobile devices would yield great benefit. In summary, machine learning enhanced skin cancer diagnostics will improve the accuracy and rate of diagnosis and early detection; the implications of which will not only reduce healthcare costs, but, more importantly—reduce the mortality rate.

Research Plan

The working title for our project implementation report is *Skin Lesion Classification using VGG Convolutional Neural Network*.

Aims and Objectives

The overarching goal of the project is to design and develop a convolutional neural network (CNN) to perform binary and/or multiclass image classification of skin lesions. In addition, we aim to determine what factors optimise network performance, and, vis-à-vis hyperparameters, identify values that are most conducive to low error and high accuracy rates. Accordingly, research questions include:

1. Does the size of the dataset influence accuracy?
   - optimal training and testing subsets
2. What values for certain configurable hyperparameters deliver the best performance?
   - learning rate
   - momentum constant
   - epochs
   - stride
   - depth
   - padding
3. What topology delivers the best performance?
   - number of layers
   - size of each layer
4. Which pooling type delivers the best performance?
   - average pooling
   - max pooling

Further, an important sub-objective of this project is to establish the viability of bringing machine learning enhanced skin cancer detection technology to mobile devices.

Methodology

The problem to be solved requires using machine learning to detect malignant cutaneous neoplasms. As such, a CNN will be implemented and trained to learn to detect skin cancer from skin lesion images. Successful completion of this project will require several steps. First, data will need to be collected and allocated to both training and testing subsets. Second, further research of existing skin cancer detection models using CNNs will be continued to: identify potential architectures; establish appropriate topologies, hyperparameters, feature extraction protocols, and classification functions; and determine suitable training and testing percentages for the given dataset(s). Next, certain performance metrics such as accuracy, sensitivity, specificity, and dice similarity coefficient will be defined to evaluate the efficiency of the system (Table 6). Lastly, given these particulars, the network will be designed and developed; after which training and testing will commence. All results will be aggregated, and CSV files created to facilitate analysis.

Table 6

Performance Metrics

Metric	Description	Formula
Accuracy	Measure the proportion of correctly classified positive and negative results among total test cases.	$\frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity	Measure the proportion of correctly classified positive results among total true positive test cases.	$\frac{TP}{TP+FN}$
Specificity	Measure the proportion of correctly classified negative results among total true negative test cases.	$\frac{TN}{TN+FP}$
Dice coefficient	Measure the similarity between the prediction and ground truth by measuring correctly classified positive results and penalising for false positives.	$\frac{2TP}{FP+2TP+FN}$

Note. From 'Deep Learning-Based System for Automatic Melanoma Detection' by A. A. Adegun and S. Viriri, 2020, *IEEE Access, 8*, 7168. (https://doi.org/10.1109/ACCESS.2019.2962812). Copyright 2020 by IEEE Access.

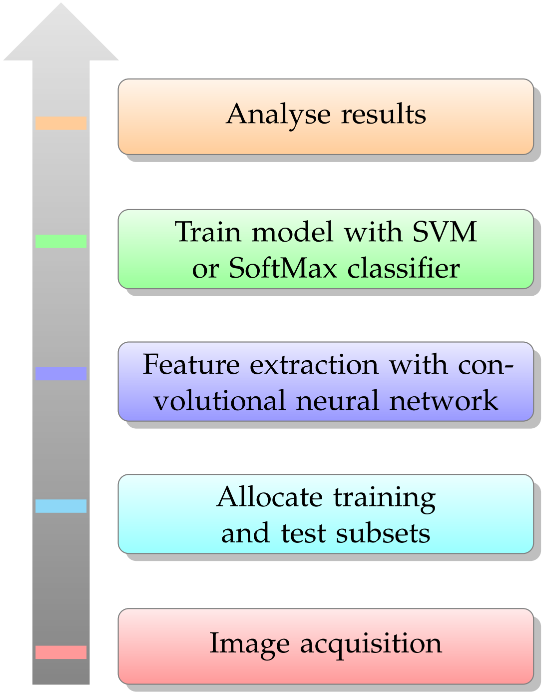
Implementation

The general procedure to develop this model is illustrated in Figure 9, and requires utilising Sci-kit Learn and Keras to implement a VGG convolutional neural network using a SVM and/or SoftMax classifier activation function to experience either one, or both, of the PH^2 or HAM10000 datasets[1].

[1] PH^2 is obtained from https://www.fc.up.pt/addi/ph2%20database.html. HAM10000 is obtained from https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T.

Figure 9

Skin lesion classification procedure



The former is a collection of 200 dermoscopic skin lesion images comprised of 40 melanomas, and 80 common and atypical naevi that were obtained from Hospital Pedro Hispano in Portugal (ADDI Project, 2020). The latter is a much larger collection of over 10,000 dermoscopic images of common pigmented skin lesions obtained from various sources (Tschandl, 2018). The Sci-kit Learn module will be imported (Listing 0.1) to implement the Support Vector Machine (adaptive boost) algorithm detailed with pseudocode in Listing 0.2.

```python
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets, svm
```

Listing 0.1: Python Sci-kit module with SVM implementation.

```
set stdev ← σ, minimum stdev ← σ_min, and step stdev ← σ_step   ❶
for i ∈ {1,2,...,n}, do   ❷
    set data weight w ← 1/n   ❷a
    while (σ ≥ σ_min), do:   ❸
        use radial basis function SVM to train w   ❸a
        set error Ωm ← ∑_{i=1}^{n} w^T y_i   ❸b
        if Ωm > 0.5, then:   ❸c
            set σ ← σ − σ_step
            break
        set classification weight δ_m ← (1/2) log((1−Ω_i)/Ω_i)   ❸d
        set weight vector {w_i^{m+1}} ← w_i^m exp{−δ_m y_i H_m(x_i)} / B_i   ❸e
set H(x) ← signum(∑_{m=1}^{M} Ω_m · H_m(x))   ❹
```

Listing 0.2: SVM Adaptive Boost algorithm: M = combined classifier; B_i = normalisation constant. From 'Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms' by J. Premaladha & K. S. Ravichandran, 2016, *Journal of Medical Systems, 40*(4), p.95. (https://doi.org/10.1007/s10916-016-0460-2). Copyright 2019 by Springer.

References

ADDI Project. (2020). *PH^2 database*. Retrieved September 18, 2020, from https://www.fc.up.pt/addi/ph2%20database.html

Adegun, A. A. & Viriri, S. (2020). Deep learning-based system for automatic melanoma detection. *IEEE Access*, *8*, 7160–7172. https://doi.org/10.1109/ACCESS.2019.2962812

Alom, M. Z., Yakopcic, C., Hasan, M., Taha, T. M. & Asari, V. K. (2019). Recurrent residual U-Net for medical image segmentation. *Journal of Medical Imaging*, *6*(1), 1–16. https://doi.org/10.1117/1.JMI.6.1.014006

Attia, M., Hossny, M., Nahavandi, S. & Yazdabadi, A. (2017). Skin melanoma segmentation using recurrent and convolutional neural networks. *Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging*, 292–296. https://doi.org/10.1109/ISBI.2017.7950522

Australian Institute of Health and Welfare. (2016). *Skin cancer in australia* (Cat. no. CAN 96). Canberra, AIHW. https://www.aihw.gov.au/reports/cancer/skin-cancer-in-australia

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Dai, X., Spasić, I., Meyer, B., Chapman, S. & Andres, F. (2019). Machine learning on mobile: An on-device inference app for skin cancer detection. *Fourth International Conference on Fog and Mobile Edge Computing*, 301–305. https://doi.org/10.1109/FMEC.2019.8795362

Dorj, U.-O., Lee, K.-K., Choi, J.-Y. & Lee, M. (2018). The skin cancer classification using deep convolutional neural network. *Multimedia Tools and Applications*, *77*(8), 9909–9924. https://doi.org/10.1007/s11042-018-5714-1

Fisher, R. B., Rees, J. & Bertrand, A. (2020). Classification of ten skin lesion classes: Hierarchical knn versus deep net. In Y. Zheng, B. M. Williams & K. Chen (Eds.), *Medical image understanding and analysis* (pp. 86–98). Springer International Publishing.

Ge, Z., Demyanov, S., Chakravorty, R., Bowling, A. & Garnavi, R. (2017). Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images. In M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins & S. Duchesne (Eds.), *Medical image computing and* *computer assisted intervention - miccai 2017* (pp. 250–258). Springer International Publishing.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. The MIT Press.

Harangi, B., Baran, A. & Hajdu, A. (2018). 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)*, 2575–2578. https://doi.org/10.1109/EMBC.2018.8512800

He, X., Wang, S., Shi, S., Tang, Z., Wang, Y., Zhao, Z., Dai, J., Ni, R., Zhang, X., Liu, X., Wu, Z., Yu, W. & Chu, X. (2019). Computer-aided clinical skin disease diagnosis using cnn and object detection models. *2019 IEEE International Conference on Big Data (Big Data)*, 4839–4844. https://doi.org/10.1109/BigData47090.2019.9006528

Kaymak, S., Esmaili, P. & Serener, A. (2018). Deep learning for two-step classification of malignant pigmented skin lesions. *2018 14th Symposium on Neural Networks and Applications (NEUREL)*, 1–6. https://doi.org/10.1109/NEUREL.2018.8587019

Khan, S., Rahmani, H., Shah, S. & Bennamoun, M. (2018). *A guide to convolutional neural networks for computer vision* (G. Medioni & S. Dickinson, Eds.). Morgan & Claypool Publishers. https://doi.org/10.2200/S00822ED1V01Y201712COV015

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Murphy, K. P. (2012). *Machine Learning A Probabilistic Perspective*. The MIT Press.

Naeem, A., Farooq, M. S., Khelifi, A. & Abid, A. (2020). Malignant melanoma classification using deep learning: Datasets, performance measurements, challenges and opportunities. *IEEE Access*, *8*, 110575–110597. https://doi.org/10.1109/ACCESS.2020.3001507

Negnevitsky, M. (2011). *Artificial Intelligence A Guide to Intelligent Systems* (3rd. ed.). Pearson Education Limited.

Ozkan, I. A. & Koklu, M. (2017). Skin lesion classification using machine learning algorithms. *International Journal of Intelligent Systems and Applications in Engineering*, *5*(4), 285–289. https://doi.org/10.18201/ijisae.2017534420

Patterson, J. & Gibson, A. (2017). *Deep Learning A Practitioner's Approach*. O'Reilly Media, Inc.

Premaladha, J. & Ravichandran, K. S. (2016). Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms. *Journal of Medical Systems*, *40*(4), 96. https://doi.org/10.1007/s10916-016-0460-2

Putra, T. A., Rufaida, S. I. & Leu, J. (2020). Enhanced skin condition prediction through machine learning using dynamic training and testing augmentation. *IEEE Access*, *8*, 40536–40546. https://doi.org/10.1109/ACCESS.2020.2976045

S. Mohan Kumar, K. G., J. Ram Kumar. (2019). Skin cancer diagnostic using machine learning techniques-stationary wavelet transform and random forest classifier. *International Journal of Innovative Technology and Exploring Engineering*, *9*(2), 4705–4708. https://doi.org/10.35940/ijitee.B9016.129219

Swain, D., Bijawe, S., Akolkar, P., Mahajani, M., Shinde, A. & Maladhari, P. (2020). Virtual dermoscopy using deep learning approach. In P. K. Mallick, P. K. Pattnaik, A. R. Panda & V. E. Balas (Eds.), *Cognitive computing in human cognition: Perspectives and applications* (pp. 61–71). Springer International Publishing. https://doi.org/10.1007/978-3-030-48118-6_6

Tschandl, P. (2018). *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions*. https://doi.org/10.7910/DVN/DBW86T

Yan, Y., Kawahara, J. & Hamarneh, G. (2019). Melanoma recognition via visual attention. In A. C. S. Chung, J. C. Gee, P. A. Yushkevich & S. Bao (Eds.), *Information processing in medical imaging* (pp. 793–804). Springer International Publishing.

Yap, J., Yolland, W. & Tschandl, P. (2018). Multimodal skin lesion classification using deep learning. *Experimental Dermatology*, *27*(11), 1261–1267. https://doi.org/10.1111/exd.13777

Yu, Z., Jiang, X., Wang, T. & Lei, B. (2017). Aggregating deep convolutional features for melanoma recognition in dermoscopy images. In Q. Wang, Y. Shi, H.-I. Suk & K. Suzuki (Eds.), *Machine learning in medical imaging* (pp. 238–246). Springer International Publishing.