Journal of Big Data and Artificial Intelligence

WWW.JBDAI.ORG

ISSN: 2692-7977

JBDAI Vol. 2, No. 1, 2024 DOI: 10.54116/jbdai.v2i1.32

MACHINE LEARNING STUDY: IDENTIFICATION OF SKIN DISEASES FOR VARIOUS SKIN TYPES BY USING IMAGE CLASSIFICATION

Gulhan Bizel
Saint Peter's University
gbizel@saintpeters.edu

Albert Einstein
Saint Peter's University
aeinstein@saintpeters.edu

Amey G. Jaunjare
Saint Peter's University
ajaunjare@saintpeters.edu

Sharath Kumar Jagannathan
Saint Peter's University
sjagannathan@saintpeters.edu

ABSTRACT

Increased machine learning methods have helped improvise human interaction with digital devices, which helps in skin disease identification, prediction, and classification by using algorithms. Image classification for skin disease application algorithms can detect Caucasian skin tones but poorly performs when analyzing other skin colors. In this research, a deep learning algorithm was used to address the problem that other applications perform poorly with the classification of skin disease types.

Convolutional neural network, a machine-learning algorithm was used to classify images and add the predicted images within the data set. The images in the data set covered a lot of patient factors, such as age, sex, disease site (e.g., hand, feet, head, nail), skin color (white, yellow, brown, black), and different periods of lesions (early, middle, or late). Multiple private applications can detect skin diseases during the analysis. For the darker color skin population, the performance was poor, and skin cancer detection was not possible even with the help of image recognition. This research aims to conduct an analysis of visual searches within skin-related health searches to identify opportunities to provide digital health consumers with visual search results that are more representative of America's diverse populations.

Keywords *convolutional neural network (CNN), neural network, deep learning, machine learning, image classification.*

1. Introduction

A widely used dermatology diagnostic technique is the examination of diseased skin against healthy skin. A field that focuses on skin is now recognizing the importance of skin color (Enderling 2019). Images are the critical resource for diagnoses in dermatology. The lack of images of darker skin tones creates a barrier for proper treatment. Skin conditions that present unusual color patterns, for example, redness in light skin, can be harder to see in dark skin. Physicians who lack diagnostic experience with such image patterns may struggle with diagnosing people of color. It is unknown whether "Dreamscape Immersive," a company that focuses on creating immersive virtual reality (VR) experiences, will improve the diagnostic accuracy for all types of pigmented skin lesions or only for those that are melanocytic (Errichetti 2020). The issue is becoming more serious for dermatologists and may influence different outcomes for a different color of skin. The lack of images is one reason why dermatologists may misdiagnose skin illnesses for patients who are darker skinned, which may result in serious health-threatening medical issues.

A skin lesion is a nonspecific term that refers to any change in the skin surface. Skin lesions may have color (pigment); be raised, flat, large, small, or fluid filled; or exhibit other characteristics (Skin Lesions 2022). With current technology, smartphones are showing promise as diagnostic tools. During dermatology appointments, patients can send their physicians pictures of their skin lesions before their visit. Patients who are comfortable with using this technology seem to have adopted this new tool for their health care. Smartphone photographs have provided valuable relevant information for a physician's diagnosis and treatment decisions, which is usually based only on the medical history reported by patients (Hubiche 2016).

Google's artificial intelligence initiatives, often referred to as Google AI, (with its primary location in Mountain View, California and offices and research centers around the world) encompasses a wide range of tools, research, and applications aimed at pushing the boundaries of what AI can do. Google's mission with AI is to make it universally accessible and useful, ensuring that the benefits of AI are spread broadly across all sectors of society. Google AI is used to detect common skin cancer problems, which has improved the clinical trials to treat their patients effectively. There are, according to a Google search history, billions of people searching for answers and information related to skin, hair, and nails. Approximately 2 billion people worldwide experience dermatologic issues, and there is a shortage of dermatologists to diagnose and treat them. Using a search bar makes it difficult to understand what type of skin diseases they have because many of the terms are scientifically complex and difficult for a patient to understand.

Google has developed a web-based application tool in which the users can take their pictures by using their personal devices and can then upload them in that tool. The model analyzes the images and has been programmed to identify 288 common skin disease problems. It then provides information about the disease, which helps the patient research his or her skin problems and suggests common questions asked to a dermatologist. This tool is not for a proper diagnosis, but it assists users in making their decision on what further steps are required (Bui 2021). Although social media is a powerful tool for users to share their information with regard to their health, there is a risk of misleading and inaccurate information when dermatologists are encouraged to present their answers in multiple social media and other applications, which counteract other misleading information (Powell 2019).

In recent years, machine learning has been pivotal in advancing skin cancer diagnosis and understanding. Das et al. (2021) explored the utility of machine learning in skin cancer identification, which has become a significant focus given the rising incidence of skin cancer globally. Their work, published in the *International Journal of Environmental Research and Public Health*, has shed light on the profound implications of machine learning techniques in early skin cancer detection and treatment planning. On a similar note, Li et al. (2021) delved into the application of deep learning for skin disease diagnosis in their review published in *Neurocomputing*. Their comprehensive review underscores the remarkable progress and the challenges encountered in deploying deep learning algorithms for skin disease diagnosis, which contributes to a broader understanding of the potential and the limitations of these technologies.

The quest for efficient and accurate skin disease image classification has led to the development of innovative algorithms and frameworks. Chen et al. (2021) introduced the "Interactive Attention Sampling Network for Clinical Skin Disease Image Classification," presented at the 4th Chinese Conference on Pattern Recognition and Computer Vision. Their work delineates the efficacy of the Interactive Attention Sampling Network in categorizing clinical skin disease images, thus offering a promising avenue for enhancing diagnostic accuracy. Furthermore, the challenge of class imbalance in skin disease recognition was addressed by Yang et al. (2020) through their proposed self-paced balance learning algorithm. Published in the "IEEE Transactions on Neural Networks and Learning Systems," their research presents a robust methodology to mitigate the class imbalance issue, which is prevalent in many classification tasks, especially in clinical skin disease recognition. Through these diversified studies, it is evident that the fusion of machine learning and dermatologic expertise holds immense promise for the evolution of skin disease diagnosis and management.

A skin disease, also known as dermatosis, refers to any condition that affects the skin. These conditions can range from temporary to chronic, mild to severe, and can be caused by various factors, including genetics, environmental

factors, allergens, and pathogens. Skin diseases are very often experienced during one's lifetime. Skin diseases prevade all cultures and affect between thirty to seventy percent of individuals (Hay et al. 2014). People can be affected by skin disease anytime during their lifetime. Skin disease is twofold: skin infection and skin neoplasm with thousands of specific skin conditions (Hurt 2012). Skin disease may impact the quality of a person's life, mostly related to his or her psychosocial situation. However, only a small portion of people can recognize these diseases without seeing a physician or dermatologist.

There are several over-the-counter medications for treating the regularly occurring skin diseases in daily life. So, people who decide to use these treatments need to choose the correct medicines without the benefit of seeing a physician. Should people want to take this route, a visual analyzing system would be useful, even if it is not fully accurate. For example, if a skin disease presents itself, a person can submit a photo of the skin condition to his or her physician for an initial diagnosis. Interestingly, there are already several applications that use computer visual techniques to recognize many common skin diseases based on simple photo images.

Although there are several related tools, skin disease recognition has not yet been a hundred percent accurate by image recognition. Skin disease images have no consistently distinct pattern, like a fingerprint. For instance, it is still very difficult to find an accurate description of a simple allergic skin disease such as eczema. In addition, there are many variations, such as contact issues between lesions and surrounding skin or the coloring inside the lesion, which makes it complicated for skin disease recognition. A recent report indicates that performing clinical skin disease recognition by image analysis is of major importance because skin disease is one of the most common diseases that appear in medicine (Esteva et al. 2017).

There are some related developments in skin disease recognition such as disease classification (Barata et al. 2021; Yu et al. 2017) and detection and localization (Marchetti et al. 2018). Examining clinical skin disease images is economical and getting the digital image from a portable electronic device is convenient for patients who can easily run a self-diagnosis. One drawback to self-diagnosis is that there are few open data sources that are needed to develop deep learning technology in this field. The challenge that researchers face is that clinical imaging is easily affected by light intensity, camera angle, uncertain background, and other light-related factors and interferences (Yang et al. 2018). Moreover, most current research addresses binary skin disease recognition problems.

Smartphones can be considered a new potential source of medical data. It has been observed that an increasing number of patients present pictures of their skin problems to their physicians. To benefit from their full utility, the application determined the proportion of patients whose smartphone photos provided information relevant to their diagnosis. Patients are now practicing this type of new technology as part of their own health care. Relevant information by using mobile phone photos can be provided for diagnostic and treatment decisions. A result is determined based on the patient's medical history. Because these reports may influence the patient's diagnosis and treatment, it is crucial to keep the photos in the patient's medical records. With these digital applications, patients can better partner with their physicians and become more self-aware of their health status.

The paper is outlined as follows: the methodology section gives the overview of the convolutional neural network (CNN) model approach used in skin cancer detection of multiple skin types. The results section shows the graphic representation of the number of trainings performed and researched about the model findings. The final section includes the best results found in the model and the best batch size when considering the best accurate score.

2. Material and Methods

2.1 Dataset

In this model, the image data sets used are from skin diseases, ranging from various types of eczema and acne to diverse cancerous conditions. The dimensions of the pixels vary from 640×480 pixels to 1640×1181 pixels. The images collected were publicly available and have been obtained legally. The objective is to extract the class names from the images and train the model on different skin colors, such as white, black, brown, and the male and female genders. The key task was to import the dataset of images into the dataset and divide the dataset into two parts, training (80% of images) and validation (20% of images) (Skin Disease Images 2023). The different types of classes and datasets that are being used in this research are shown in Figure 1.

2.2 Research Model

The programming language used in this research is Python, developed by Guido van Rossum and first released in 1991, which is one of the most frequently used high-level programming languages. Its high-level, built-in data structure, combined with its dynamic typing and binding, make it very attractive for rapid application development

Figure 1: Classes of images within the dataset.

as well as scripting or glue language to connect existing components (https://www.python.org/). The application used is Jupyter Notebook, which is an open-source web-based application that allows the creation and sharing of documents that integrate live code, equations, visualizations, and other multimedia resources. The Jupyter Notebook evolved from the IPython project in 2014. Project Jupyter is a nonprofit initiative primarily based in the U.S. and operates under NumFOCUS, a 501(c)(3) charitable organization in Austin, Texas. Tensor Flow is used to create large-scale neural networks with many layers (Introduction to Convolutional Neural Networks 2022; Abadi et al. 2016) It is mainly used for deep learning or machine learning problems, such as classification, perception, understanding, discovering, prediction, and creation (Kiran 2020). The flow of the model that has been applied in this research is shown in Figure 2, which covers the steps in the following order: training image, validation data, data augmentation, and CNN. Eventually, the applied model led to the output of the study.

2.2.1 Training image and validation data

After extracting class names and dividing the dataset, the images were resized into 128 × 128 pixels, so all the pixels were set to a common size for all images, which made the model train better. A batch size of sixty four was used to help train the models to minimize the loss function (Chang 2021). Mathematically, it is calculated and defined as in Equation (1). The theta shows the number of parameters used in the model, n represents the number of training samples used for training, i is the single element of the training dataset, and xi is the loss function of the single element of each training sample. Using batch with the option of other hyperparameters will help improve training performance and increase the accuracy of the model. The data were then shuffled by using the default true

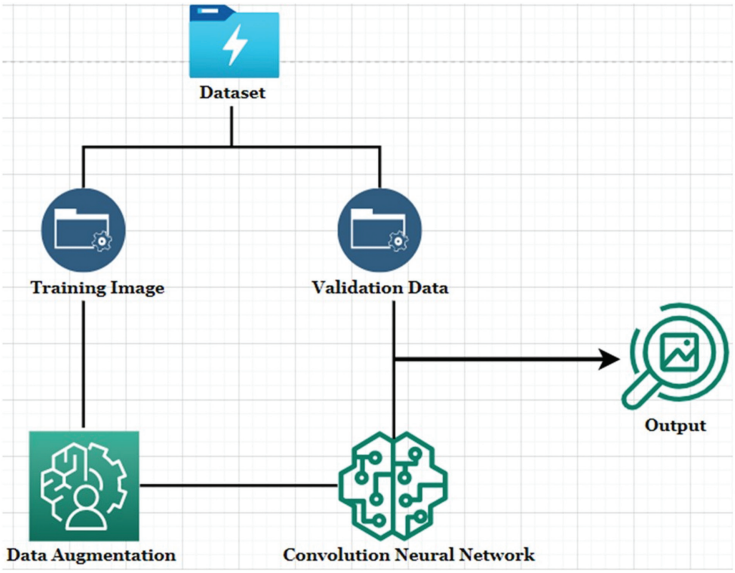


Figure 2: Flow chart of the research model.

function, which determined if the data needed to be shuffled, and arranged the data in alphanumeric order. On creating data training and validation cache function,

$$z(\theta) = 1 \Big/ n \sum_{(i=1)}^{n.x_{i(\theta)}} \tag{1}$$

2.2.2 Data augmentation

Overfitting is the main concern while training the dataset. When the dataset trains too long, it starts to learn noise in the dataset, which matches the irrelevant data once the model is trained. To overcome this problem, data augmentation was used to make the data more stable. Data augmentation was not only used for overfitting but also to increase the accuracy performance of the model. Data augmentation was used for audio, text, images, and other types of data. In this research, data augmentation was used for images when images were rotated vertically and horizontally so that the cancer names and augmented images were trained together to reduce overfitting and increase the model's performance. Data augmentation is useful for improving the performance and outcomes of machine learning models by forming new examples for training of the datasets. If the dataset in a machine learning model is rich and sufficient, then the model performs better and more accurately (Takimoglu 2021). The image shown in Figure 3 is an example of data augmentation performed in the dataset before applying to the model training. It shows the augmentation image of basal cell carcinoma disease from various angles.

2.2.3 Deep learning and CNN classification

A neural network works like the human brain where each neuron is programmed to solve complex problems where neurons learn from other neurons to provide the required outcome. Neurons are the connection between the hidden layers and output layers, which split into parts of the image that are true or false (1 or 0). Neurons have many hidden layers that help in the segmentation of images into kernels to detect the type of image. The output layer is the output of all the segmented images performed by the hidden layers (Shinozaki 2021). In this paper, the CNN classification is being used, which is a part of deep learning for the prediction of images with regard to skin cancers.

Deep learning is a subset of a machine learning model used to perform complex tasks. This means that algorithms are trained like the human brain, and can perform speech recognition, image identification, and prediction (What is Deep Learning 2022). Deep Learning uses layers of neurons to predict the patterns in the input datasets to present the outcome. One of the deep learning models used in this research was the CNN. The CNN is described as the heart of deep learning, which is primarily used for the classification of images, clustering images, and object detection. The process of detection of images is processed through tensors, which are known to be an array of numbers with additional dimensions, is shown in Figure 4. CNN receives the input of images and converts them into RGB

Figure 3: Augmentation image of basal cell carcinoma disease from various angles.

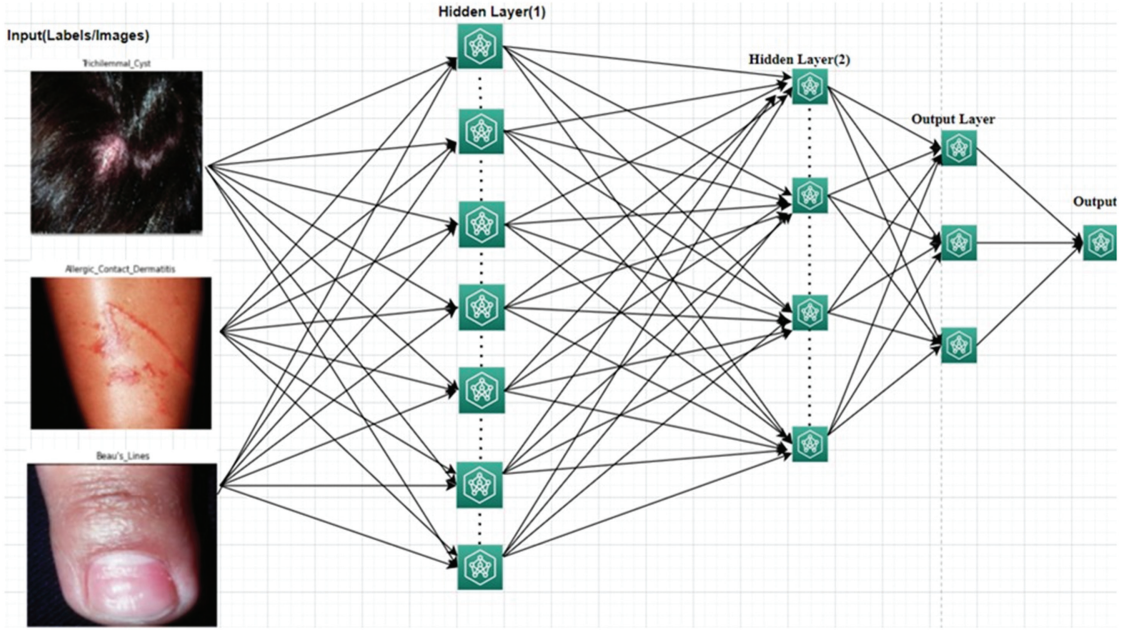


Figure 4: Example of tensors work in the convolutional neural network (CNN).

(red, blue, and green) stacked one above the other, where they are measured by the number of pixels converting them into matrices of multiple dimensions (Liu et al. 2021).

CNN tensors convert the image into an array of matrices which in turn are divided into RGB colors stacked on top of one another. The filter then weights the increases in the strength of the connection as shown in Figure 4. Bias is known to be constantly represented by the number bias 1, which is marked in the right bottom corner in red. According to CNN, bias vector is used to extract the features of the images and determine what classes need to be assigned for that image. CNN takes the images from the raw pixels and trains the images to extract their meaningful features for better prediction and classify the whole image.

CNN is supervised machine learning in which the labels of the classes are trained together for the outcome respective of the class to which the object belongs. CNN trains through the inputted images and works through the class labels and compresses them together. CNN is a multilayer neural network, which has several hidden layers stacked up one after the other, which allows the neural network to learn complex features. The hidden layers in the CNN comprise convolutional layers that have activation layers (ReLU), max pool layer, fully connected layer used for learning and predicting the output (Introduction to Convolutional Neural Networks 2022).

Convolution layers as the primary parameter for a CNN model are represented in Figure 5. These layers are composed of multiple filters that are defined by the height, width, and depth of the input image, which converts the image into weighted matrices known as kernels. Filters range in between $3 \times 3''$ to $11 \times 11''$ which determines the dimension's size of the image. Filters will perform element-wise multiplications, which result in values that will determine the edge and batch of the color in an image. It is computed as in Equation (2) (McDermott 2022),

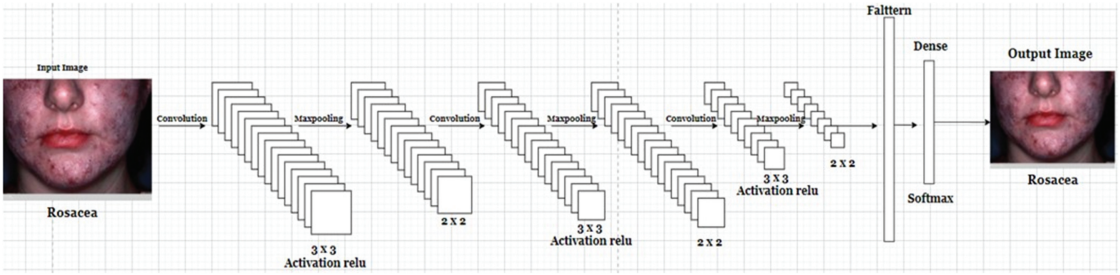
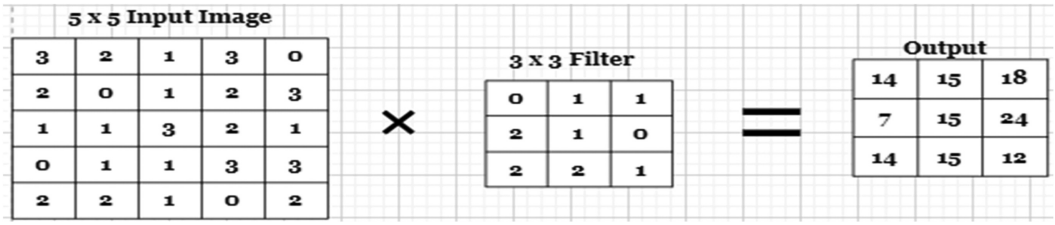


Figure 5: Convolutional neural network (CNN) model flow used in this research with 13 layers.



(3x0) + (2x1) + (1x1) + (2x2) + (0x1) + (1x0) + (2x1) + (3x1) = 14

Figure 6: Example of input image filtered by using con2d with a 3 × 3 kernel.

in which the input image is denoted by *i*, the filter of the image is known as *f*, and *m* and *n* are the rows and columns of the kernels given, respectively. Features of the images are extracted by using kernels in which convolution is a two-dimensional array, which contains correlations of the image with respect to the number of filters applied.

$$z[\mathrm{m}, \mathrm{n}] = (\mathrm{i} * \mathrm{f})[\mathrm{m}, \mathrm{n}] \tag{2}$$

The input image, which is n × n matrix, and the filtered image with m × m matrix, which results in the output size as (n – m + 1) × (n – m + 1), is shown in Figure 6, and shows the calculation of the values slide of one column to the right each time and slides down one row at a time. This is called striding, which is represented as stride = k and the following result is written as ([n – m]/k + 1) × ([n – m]/k + 1) (DeepAI 2020).

From Figure 6. Strides with the 3 × 3 matrix reduce the output size so to retain the same input image size, which is the 5 × 5 matrix. Padding is used to obtain the original input size by adding zeros at the edges of the image array, as shown in Figure 7.

As shown in Figure 7, the padding is added in the first layer because the image size remains the same in which the equation is represented as input of the image size, which is n × n matrix, and filter size, which is m × m matrix, when following the output size as (n + 2p – m + 1) × (n + 2p – m + 1). Convolutions with the striding and

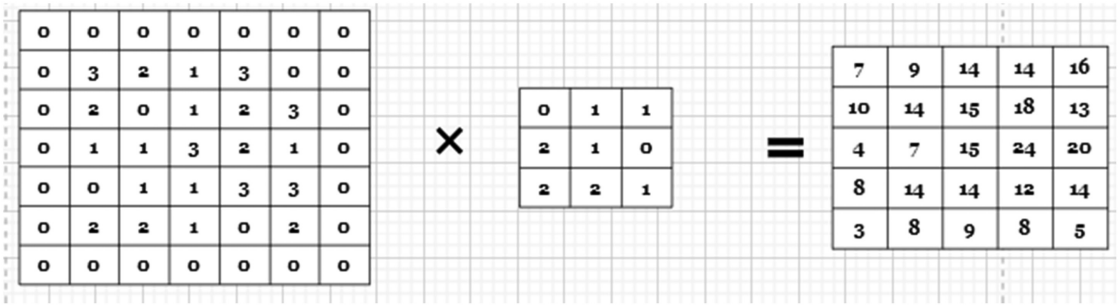


Figure 7: Example of padding of the filter to retain the original input size.

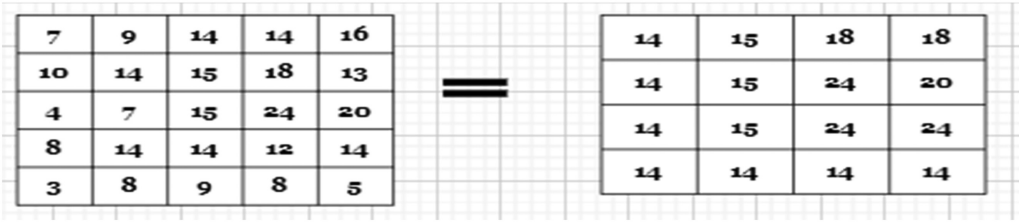


Figure 8: Example of max pooling filter by using 2 × 2 kernel.

padding need an activation function, which is relu, a multi-layer neural network, which is represented as Equation (3) (DeepAI 2020).

$$f(x) = \max(0, x) \tag{3}$$

According to the equation, ReLU is a non-linear activation function that represents a sigmoid where, if the value is negative, then it is referred to as zero and, if the value is positive, then the value remains the same. The following, Equation (4), can be written

$$f(x) = 0, \text{ if } x \leq 0 \tag{4}$$
$$X, \text{ if } x \geq 0$$

From Equation (4), if the value is f(–1) is < 0, then it is restricted to the value 0; conversely if the value is f(2), it remains the same as the input as the value of $x > 0$. The Relu activation function is used to accelerate the training speed of the neural networks, which saves additional computation time than traditional activations. Pooling layers are used to reduce the size of feature maps, which helps in compressing the dimensions of the features in the image. The subsampling takes the minimum, maximum, or average of the image array cells to proceed further with the output, as shown in Figure 8.

In Figure 8, the max pooling operation with a 2 × 2 window is illustrated. For every 2 × 2 region in the feature map, the maximum value is extracted. Flattening is the step in which the important parts of the images are captured before applying the final filter. The final step is to create a vector where the classification can work on that part of the values provided, which converts a multidimensional array n × n matrix to n × 1 matrix.

In Figure 9, it is seen that the layer that is processed after flattening is the dense layer, which uses softmax as the activation layer. It is well known as the extension for logistic regression where the predicted output is the probability between 0 and 1, it is determined by Equation (5) (Saluja 2022).

$$p(y = j|x^{(i)}) = e^{x^{(i)}} \Big/ \sum_{j=0}^{k} e^{x_k^{(i)}} \tag{5}$$

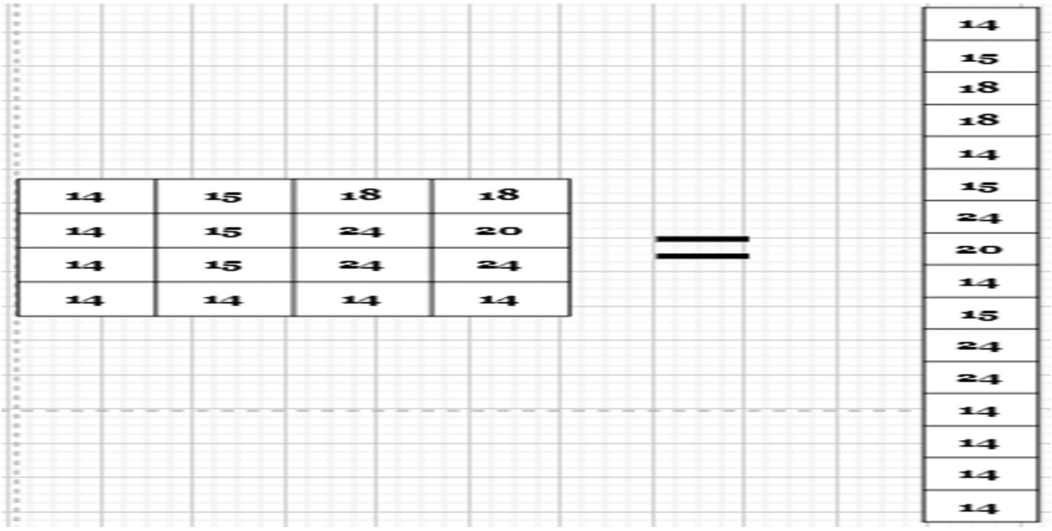


Figure 9: Example of flatten filter n × 1 kernel.

2.2.4 Cost function

In this prediction model after softmax, the loss of the function must be calculated, which is cross-entropy or is known as a log loss function, which calculates the distance between the probabilities created by the softmax layer. The use of the loss function is to determine if the prediction has predicted the correct observation. In this study, sparse cross-categorical entropy is used, which is like cross-entropy to determine the labels that belong to two or more classes in which the labels are transformed into integers by using one-hot representation. It is represented as in Equation (6) (Anis 2021).

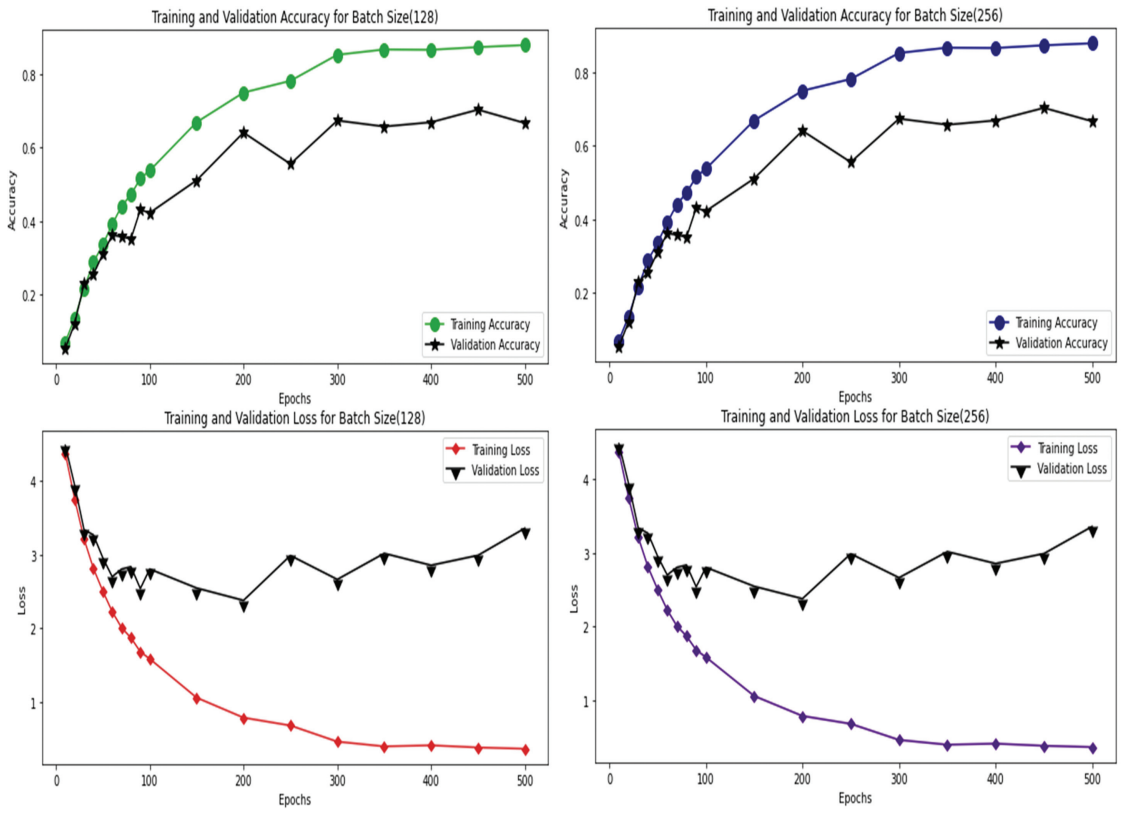
3. Results

In this section, the research results will be presented in detail in different aspects, starting from finding the right batch size to follow up accuracy and loss trends. The next step will be checking the training and validation accuracy comparison to verify the selected batch size. Eventually the selected results of the image classification will be introduced with further analysis of other random image data, which will be explained in greater detail.

3.1 Accuracy and Loss Trend

When the model is trained with several layers of accuracy, a loss check is required to notice which batch size can be used for skin cancer recognition. The trend for accuracy between 0 and 1 while the trend for loss is from 0 to 100 percent is represented in Figure 10. Both accuracy and loss metrics are tracked over a span of 0 to 500 epochs for training and validation sets. This tracking helps determine the presence of any errors or issues with the model's learning process. It helps for understanding model solidity.

The validation loss is a metric, and it is calculated after training each epoch. Validation loss is encountered when under- and overfitting the dataset. If the training loss is close to zero and the validation loss is greater than the



(a) 128 batch size **(b)** 256 batch size

Figure 10: Accuracy and loss trend for different batch sizes (a) 128 batch size, (b) 256 batch size.

training loss, then it is considered as overfitting as the predicted model. The model then starts to accidentally predict some unseen data as true. If the training loss is greater than the validation loss, then that means that the model is inaccurate, which causes large errors.

A batch size for thirty two understanding accuracy and loss on the training and validation are demonstrated in Figure 10(a). The accuracy is greater than seventy percent, and the validation accuracy is nearly sixty percent. As per the loss, the training line is less than 1.0 percent and validation losses are too high (approximately 2.9 percent). Although the accuracy is good, the loss represents high impact for the model.

In the research, a batch size of sixty four is used. The graph shown in Figure 10(b) indicates the distance between the training and validation accuracy. When considering the loss function after training the model, it shows that the sixty-four–batch size was better in terms of the distance between the training and the validation loss. It indicates that the model is good for predicting the training accuracy of 86.34 percent, validation accuracy of 64.22 percent. the validation loss of 2.4 percent, and training loss of 0.4 percent.

$$j(W) = -1/n \sum_{i=1}^{n} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \tag{6}$$

3.2 Training and Validation Accuracy Comparison for Different Batch Sizes

After considering the accuracy results and losses on the four different batch sizes, it is important to compare them to finalize the best batch size and move on with the model application. The training and validation accuracy percentage comparison is explained in Figure 11, and it shows the accuracy for all the batch sizes used to determine which model to use for this research.

The thirty-two–batch size shows a training accuracy of seventy-nine percent and sixty-five percent on validation, in which the images taken for the training in this batch were 206 items. For the 64-batch size, the images taken for training were 103 items, which is a reasonable amount and shows an accuracy of 86.34 percent and validation accuracy of 64 percent. Checking batch size of 128, the images used for training were fifty-two items, which makes the model skip some patterns in skin cancer detection, even though the accuracy is above ninety percent and validation is sixty-seven percent. The loss function is too high for the detection of skin cancer. As for a batch size of 256, the images used for training are thirty-two items, which has fewer chances for the model to understand the labels for the skin cancer and the pattern of the images. When considering all the evidence retrieved from the models of batch sizes in this research, a batch size of sixty four is optimal.

3.3 Image Prediction

Once the batch size has been settled, the next step is testing the model with the image dataset created for the research. The image predictions performed by the model for the validation data, which are twenty percent of the image dataset and eighty percent for training dataset when uploading unseen data for the final prediction, are shown in Figure 12. The objective was to see if the actual label of the skin cancer and the predicted skin cancer label were

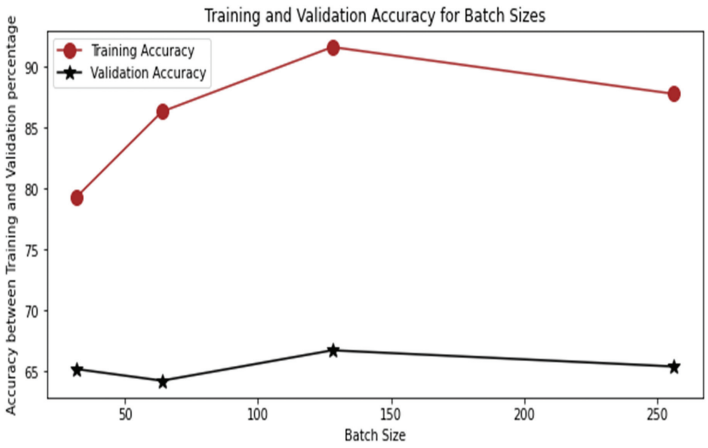


Figure 11: Training and validation accuracy percentage check for batch sizes 32, 64, 128, and 256.

Figure 12: Prediction results for the dataset.

the same. For example, Hailey disease shows the actual label and the predicted label are the same, with the confidence of 99.79 percent. Also, the kerion label shows the predicted label as alopecia. Although the image looks like alopecia, the labels did not match due to overfitting. Also, the Blue Nevus disease class label was predicted incorrectly, with a confidence of 90.63 percent, which helps to conclude that the model is encountering overfitting issues for the same diseases that have very similar skin patterns. To overcome this problem, a data augmentation library can be used in the model to get the required accuracy with the image class labels. The lowest confidence level observed was 70.1 percent for fibroma molle disease, which still shows a strong pattern for medical image classification.

It was decided to test the model image dataset library that has been built for this research. Random disease images could give the indication whether the model was responding to only the dataset library or whether it was coming out with the results also for unintroduced images. Several images were gathered randomly from the internet by using the Google search engine to check if the model can predict the unseen data as the class labels are added to the twenty cancer images shown six of them in Figure 13. The expectation was the model will recognize the class names for some diseases, such as actinic solar damage yet predict acute eczema. However, the prediction results show that the highest confidence level of prediction is 99.22 percent and the lowest prediction confidence is 76.19 percent, with an outlier of a confidence level of 47.32 percent. However, the model needs to be tested by adding more data to understand if the labels are predicted correctly and if there needs to be any other additional layers implemented in the future.

Figure 13. Prediction results for sample images.

4. Conclusion

According to the training and validation accuracy of the model, it was found that the model built for this research performed better predictions with the batch size of 64 compared with batches with 32, 128, and 256 sizes. Although the validation loss is favorable compared to other batch sizes, there are still other models worth testing, including ReLU, VGG-16, and Unet for object detection. When considering that the CNN model has set an important baseline for image classification, masking skin cancers would assist in the prediction with less loss and high accuracy.

Analysis of the research has shown that, with an image library built by the dataset, the results were satisfying in terms of the confidence level starting from 70.1 percent. Despite the same image classifications that may be misread, the confidence levels are within acceptance levels not only for the dataset but also for the random images. The second check point of the model was to test it with some random images retrieved from the Google search engine by simply searching for images under the "common skin diseases images" category. The confidence level for this study began at 76.19 percent, a commendably acceptable level. This suggests that the model is accurate and can effectively handle datasets it has not been previously trained on. The dataset used for the research contains 6,584 images of 198 fine-grained skin disease categories. Medical image banks are hard to compile, so it is a challenging task to label them.

This study also raises a challenging problem for automatic visual classification of clinical skin disease images. An obvious fact is that traditional methods are not sufficient in terms of effectiveness. The challenging work was to build a clinical skin disease images dataset, including samples of real-world images from twenty categories. Each sample in the benchmark is well labeled. The research intended to release the labeled dataset to the community to promote related research and facilitate its expansion. Hence, to increase efficiency of the model, more data are required to train it. Increasing the data in the CNN model would increase the accuracy of the model and minimize the loss function in this research model that is made for the image recognition of skin diseases.

This research helped establish some proven record analysis based on clinical disease images. However, all these works are built on smaller datasets, which only contain very few species and are not publicly available. The absence of benchmark datasets is a barrier to compare the outcomes of the completed research. Consequently, in this research, one of the important milestones would be considered as the introduction of a new publicly available dataset for real-world skin disease image recognition. It would also be a strong baseline for different skin colors, which was the starting point of this research. The challenge is to improve image recognition for the dark colors between brown and black because it makes it more complicated to get the correct disease identification.

References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. 2016. "Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems." Preprint, submitted March 16. https://doi.org/10.48550/arXiv.1603.04467

Anis, A. 2021. "How to Build Custom Loss Functions in Keras for Any Use Case." Cnrg.io blob. Accessed February 11, 2022. https://cnvrg.io/keras-custom-loss-functions/

Barata, C., M. E. Celebi, and J. S. Marques. 2021. "Explainable Skin Lesion Diagnosis Using Taxonomies." *Pattern Recognition* **110**: 107413. doi: 10.1016/j.patcog.2020.107413

Bui, P. 2021. "Using AI to Help Find Answers To Common Skin Conditions." Google AI. Accessed February 4, 2022. https://blog.google/technology/health/ai-dermatology-preview-io-2021/

Chang, D. 2021. "Effect of Batch Size On Neural Net Training—Deep Learning Experiments." Accessed February 9, 2022. https://medium.com/deep-learning-experiments/effect-of-batch-size-on-neural-net-training-c5ae8516e57

Chen, X., D. Li, Y. Zhang, and M. Jian. 2021. "Interactive Attention Sampling Network for Clinical Skin Disease Image Classification." *Proceedings of the Pattern Recognition and Computer Vision, 4th Chinese Conference, PRCV 2021*, Beijing, China, October 22.

Das, K., C. J. Cockerell, A. Patil, P. Pietkiewicz, M. Giulini, S. Grabbe, and M. Goldust. 2021. "Machine Learning and Its Application in Skin Cancer." *International Journal of Environmental Research and Public Health* **18**, no. 24: 13409. doi: 10.3390/ijerph182413409

DeepAI. 2020. "ReLu." Accessed February 11, 2022. https://deepai.org/machine-learning-glossary-and-terms/relu

Enderling, H. 2019. "Towards a Quantitative Personalised Oncology." *Research Outreach* **107**, no. 107: 130–133. doi: 10.32907/ro-107-130133

Errichetti, E. 2020. "Dermoscopy in Monitoring and Predicting Therapeutic Response in General Dermatology (Non-Tumoral Dermatoses): An Up-to-Date Overview." *Dermatology and Therapy* **10**, no. 6: 1199–1214. doi: 10.1007/s13555-020-00455-y

Esteva, A., B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. 2017. "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks." *Nature* **542**, no. 7639: 115–118. doi: 10.1038/nature21056

Hay, R. J., N. E. Johns, H. C. Williams, I. W. Bolliger, R. P. Dellavalle, D. J. Margolis, R. Marks, L. Naldi, M. A. Weinstock, S. K. Wulf, C. Michaud, C. J. L. Murray, and M. Naghavi. 2014. "The Global Burden of Skin Disease in 2010: An Analysis of the Prevalence and Impact of Skin Conditions." *Journal of Investigative Dermatology* **134**, no. 6: 1527–1534. doi: 10.1038/jid.2013.446

Hubiche, T., L. Valério, F. Boralevi, E. Mahe, C. B. Skandalis, A. Phan, P. del Giudice. 2016. "Visualization of Patients' Skin Lesions on Their Smartphones." *JAMA Dermatology* **152**, no. 1: 95. https://jamanetwork.com/journals/jamadermatology/full-article/2453323. doi: 10.1001/jamadermatol.2015.2977

Hurt, M. A. 2012. "Weedon D. Weedon's Skin Pathology." *Dermatology Practical & Conceptual* **2**, no. 1 3rd ed. doi: 10.5826/dpc.0201a15

Introduction to Convolutional Neural Networks. 2022. "IBM Developer." Accessed February 11, 2022. https://developer.ibm.-com/articles/introduction-to-convolutional-neural-networks/

Kiran, T. T. J. 2020. "Computer Vision Accuracy Analysis with Deep Learning Model Using Tensorflow." *International Journal of Innovative Research in Computer Science & Technology* **8**, no. 4: 319–325. doi: 10.21276/ijircst.2020.8.4.13

Li, H., Y. Pan, J. Zhao, and L. Zhang. 2021. "Skin Disease Diagnosis with Deep Learning: A Review." *Neurocomputing* **464**: 364–393. ACM Digital Library. doi: 10.1016/j.neucom.2021.08.096

Liu, J., F. Chao, C. M. Lin, C. Zhou, and C. Shang. 2021. "DK-CNNs: Dynamic Kernel Convolutional Neural Networks." *Neurocomputing* **422**: 95–108. doi: 10.1016/j.neucom.2020.09.005

Marchetti, M. A., N. C. F. Codella, S. W. Dusza, D. A. Gutman, B. Helba, A. Kalloo, N. Mishra, C. Carrera, M. E. Celebi, J. L. DeFazio, N. Jaimes, A. A. Marghoob, E. Quigley, A. Scope, O. Yélamos, and A. C. Halpern. 2018. "Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging Challenge." *Journal of the American Academy of Dermatology* **78**, no. 2: 270–277.e1. doi: 10.1016/j.jaad.2017.08.016

McDermott, J. 2022. "Convolutional Neural Networks: Image Classification with Keras–A Comprehensive Guide." Accessed February 11. https://www.learndatasci. com/tutorials/convolutional-neural-networks-image-classification/

Powell, K. 2019. "Searching by Grant Number: Comparison of Funding Acknowledgments in NIH RePORTER, PubMed, and Web of Science." *Journal of the Medical Library Association* **107**, no. 2: 172–178.

Saluja, A. 2022. "Softmax Function and Layers Using Tensorflow." OpenGenus IQ: Computing Expertise & Legacy. Accessed February 11, 2022. https://iq.opengenus.org/softmax-tf/

Shinozaki, T. 2021. "Biologically Motivated Learning Method for Deep Neural Networks Using Hierarchical Competitive Learning." *Neural Networks* **144**: 271–278. doi: 10.1016/j.neunet.2021.08.027

Skin Disease Images. 2023. "The Image Dataset With the Python Code Is Available in the GitHub Repository." Accessed October 15, 2023. https://github.com/sharathkumarphd/Skin-diseases.git

Skin Lesions. 2022. "Newofmc." Accessed February 22, 2022. https://www.ocalafamilymedicalcenter.com/skin-lesions

Takimoglu, A. 2021. "What is Data Augmentation? Techniques, Benefit & Examples." AI Multiple. Accessed February 9, 2022. https://research.aimultiple.com/data-augmentation/

What is Deep Learning? 2022. "SAS." Accessed February 4, 2022. https://www.sas.com/en_us/insights/analytics/deep-learning.html

Yang, K., Z. Sun, A. Wang, R. Liu, Q. Sun, and Y. Wang. 2018. "Deep Hashing Network for Material Defect Image Classification." *IET Computer Vision* **12**, no. 8: 1112–1120. doi: 10.1049/iet-cvi.2018.5286

Yang, J., X. Wu, J. Liang, X. Sun, M.-M. Cheng, P. L. Rosin, and L. Wang. 2020. "Self-Paced Balance Learning for Clinical Skin Disease Recognition." *IEEE Transactions on Neural Networks and Learning Systems* **31**, no. 8: 2832–2846. doi: 10.1109/TNNLS.2019.2917524

Yu, L., H. Chen, Q. Dou, J. Qin, and P. A. Heng. 2017. "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks." *IEEE Transactions on Medical Imaging* **36**, no. 4: 994–1004. doi: 10.1109/TMI.2016.2642839