

# **IMPROVING RACIAL EQUITY IN SKIN CANCER**

## **DETECTION:**

*LEVERAGING ARTIFICIAL INTELLIGENCE DRIVEN SYNTHETIC IMAGE GENERATION, CASCADING CONVOLUTIONAL NEURAL NETWORKS, AND AFFORDABLE DIAGNOSTIC HARDWARE FOR ACCURATE CANCER SCREENING ACROSS ALL SKIN TONES*



**By Kate Choi**

## **PERSONAL STATEMENT:**

On my fifth birthday, my mother gave me a pocket-sized handheld magnifying glass. With a press of a button, the device would flick its lens, allowing me to analyze insect legs, butterfly wings, leaf fragments, and the skin on the back of my hand. A new world opened up to me.

A decade later, my curiosity about the world beyond the human eye continues. After my grandfather's skin cancer was initially missed by his physician a few years ago, I studied online photos of skin lesions and made a shocking discovery. While a disease affecting 5 million Americans each year has extensive data on skin lesions, it is not the type of data that would ever help my family. An abundance of information exists for White patients of European descent, but data for the millions of other darker-skinned victims of skin cancer is jaw-droppingly nonexistent.

I was gripped with unease—and compelled to learn more. I leapt into heaps of literature, discovering how the lack of skin lesion images from ethnic patients renders dermatologists and AI models less accurate in diagnosing cancer in minority groups. The inequality was glaring; the numbers stark. White patients have a 90% malignant melanoma 5-year survival rate. For Black patients? A terrifyingly lower 66%.

So I decided to act—and a 2,000-hour research project began. I brainstormed. I cold-emailed scores of dermatologists on how I might create a tool to improve the diagnostic outcomes of people with darker skin. Their response was unanimous: “without the data, such a task is impossible.”

I was thus challenged to think outside the box of possibility to mitigate the lack of data on minority patients. While scrolling through clips of AI-generated synthetic artwork, I was struck with an idea: could I use AI to generate *synthetic* images of skin lesions from darker skin? Immersing myself in machine learning libraries, I eventually developed a generative AI network that produces realistic images of lesions from darker-skinned patients. Thousands of hours and images later, I used these synthetic images to train the first racially equitable, diagnostic AI convolutional neural network (that still works on humans!).

But one question remained: how could I make this AI-driven software actually accessible? Months later, from confused heaps of sketches, circuitry, and screwdrivers strewn across my bedroom floor, I produced DermScan AI 1.0, a \$30, unassuming, little black box. At the press of a button, my new handheld device flicks its lens, captures an image of a lesion, and within seconds, delivers a malignant or benign diagnosis with an accuracy outperforming dermatologists. Inspired by the goal of helping people like my grandfather, I now could help skin cancer victims from *all* socioeconomic and ethnic backgrounds.

My favorite handheld device these days didn't just further my project—it fueled my personal growth. Inexperienced with coding and artificial intelligence, I taught myself the skills needed to attain

the technical complexity involved in integrating sophisticated algorithms with precise hardware configurations to achieve accurate diagnostic capabilities.

However, beyond the technological pathways my device opened for me, the ability to support communities that resemble me is so meaningful. My project not only highlighted the urgent need for change, but also ignited my personal passion—equity in healthcare. Witnessing firsthand the disparities in medical treatment among different communities, I found my calling. With the support of the National Institute on Minority Health and Health Disparities, my work now extends beyond the world of skin cancer—applying innovative science to understand the intersection of Type 2 Diabetes and socioeconomic disparagement, or healthcare access and red-lining. Inequality is deeply embedded and widespread. However, a future where skin color is not a determining factor in health outcomes is a future worth fighting for.

I can't wait.

### **ABSTRACT:**

Massive underrepresentation of non-White skin in dermatology research contributes to significant disparities in skin cancer outcomes. Since artificial intelligence (AI) algorithms are largely trained on research databases containing a disproportionate number of skin lesion images from White patients, the potential for widening this inequality gulf is being realized: underperformance of AI models to detect skin cancer in patients with Black skin has already been demonstrated.

To rectify racial disparities in AI-based skin cancer detection, this project introduces a novel AI convolutional neural network (CNN) architecture trained on high-quality synthetically generated dark skin lesion images via an AI-driven style transfer network. The CNN also utilizes a unique *Cascading* architectural design which has higher accuracy by allowing the model to focus its training on ambiguous lesions. In addition to its 84% accuracy in diagnosing light skin, the model demonstrates an 83% accuracy in diagnosing skin cancer in patients with dark skin, outperforming both dermatologists and the best published models for skin cancer detection in dark skin. To make this software accessible, a user-friendly device costing under \$30 was developed with the ability to wirelessly connect to the CNN model and capture, upload, and diagnose skin lesion images within seconds. This patent-pending device is designed to serve under-resourced communities worldwide, providing support for individuals of all skin tones who lack immediate access to healthcare. Addressing emerging inequities is imperative to prevent the formation of biases in AI and the medical field. Ongoing advancements demand the fair and responsible application of these new tools.

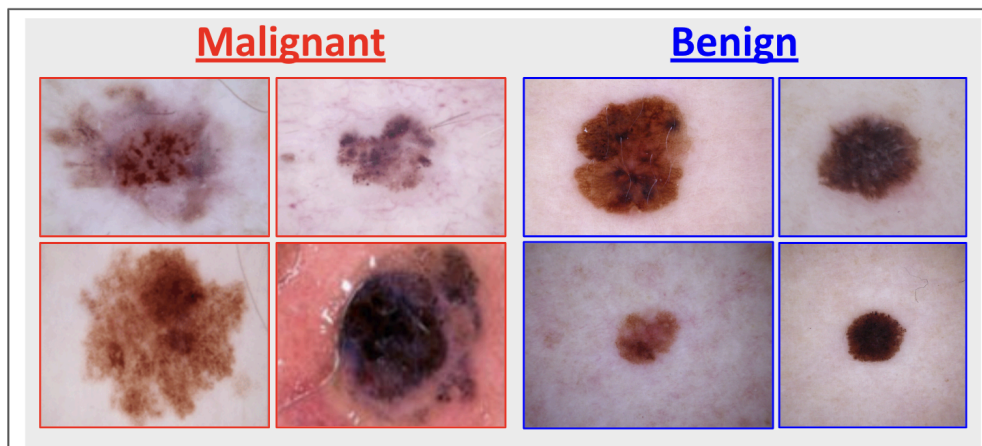
## INTRODUCTION:

*Note: (\* denotes figures created by Kate Choi unless otherwise specified).*

Each year, nearly 5 million Americans are diagnosed with skin cancer (American Cancer Society, 2023), fueling an \$8.1 billion industry dedicated to its treatment and research (Guy et al, 2015). The rise of artificial intelligence (AI) has opened new pathways in this field, enabling researchers to develop AI-powered diagnostic tools that can distinguish between malignant and benign skin lesions with unprecedented precision. If not detected early, skin cancers such as melanoma can become fatal by quickly spreading to other parts of the body. Early-stage melanoma is highly curable through surgery or chemotherapy to remove cancerous cells, but advanced-stage melanomas can be difficult to treat and may have a poor prognosis (Siegel et al, 2024). Thus, finding ways to detect skin cancer in its early stages is crucial for successful treatment and improved outcomes.

Deep convolutional neural networks (CNNs) are a type of artificial neural networks that are particularly effective at processing images and other high-dimensional data into different categories based on specific features (LeCun et al, 2015). CNNs, inspired by the brain's visual cortex, use interconnected layers to detect features in data. Widely applied in fields like computer vision and medical image analysis, CNNs extract image features (e.g., edges and textures) through convolutional filters. Pooling layers reduce spatial resolution, enhancing robustness to small input variations, and fully connected layers complete classification or regression (LeCun et al, 2015).

For example, Soenksen et al created a CNN which takes in different images of skin lesions and uses information based on shape, color, size, and other characteristics to create a tool which can determine



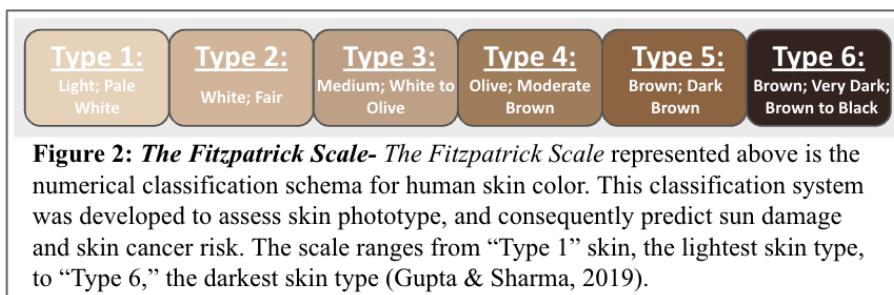
**Figure 1: Malignant (Cancerous) Vs. Benign (Normal) Skin Lesions-** The skin lesion images, drawn from the publicly-available *HAM10000* data set, feature characteristically malignant skin lesions (marked by irregular shapes, colors, sizes, and textures) and characteristically benign skin lesions. *Skin lesion images source: (Tschandl et al, 2018).*

whether a suspicious lesion is malignant or benign (*see Fig. 1\**), or if clinical management should be undertaken (Soenksen et al, 2021). In addition to publicly available images, the researchers

trained this model using a dataset of 20,388 images of skin lesions from 133 patients. Their AI algorithm

was able to accurately identify melanoma in 90.3% of cases, compared to the 86.6% accuracy for dermatologists. AI-assisted tools such as the one developed by these researchers can be used as a complementary tool to improve accuracy, efficiency, and costs in diagnosis (Soensken et al, 2021). This shift has transformed skin cancer detection, making early diagnoses more efficient and accessible for patients, while empowering dermatologists with an advanced tool to improve patient outcomes.

Yet, despite the vast investment and focus, a critical oversight persists: individuals with darker skin tones remain largely neglected in skin cancer research, leaving them vulnerable to delayed diagnoses and poorer outcomes. Melanoma is a type of skin cancer that can affect people of all skin tones, but it is more likely to be diagnosed at a later stage and have a worse prognosis in people with darker skin (i.e.,

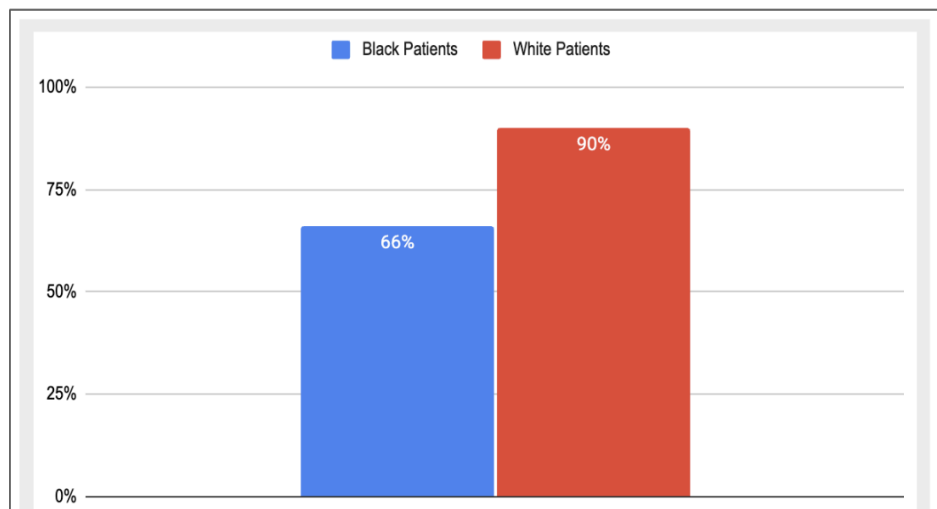


Fitzpatrick V to VI skin type- *see Fig. 2\**). One such reason for this diagnostic delay is because people with darker skin have more melanin which alters the

appearance of melanoma, making it harder to detect (Diao & Adamson, 2022). Consequently, the death rate of melanoma is disproportionately high for African American patients compared to other races. According to a 2019 study, the five-year survival rate in the United States for African American patients with melanoma was 66%, in comparison to the 90% rate for non-Hispanic White patients (Brady et al, 2021) (*see Fig. 3\**).

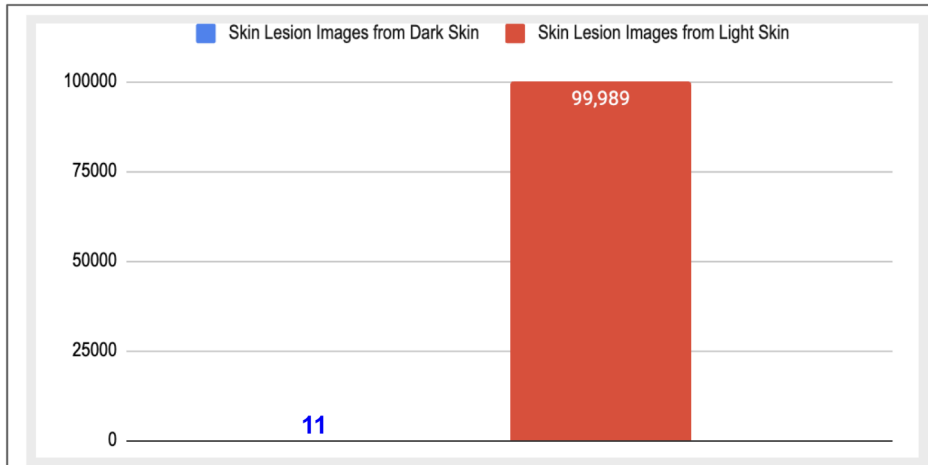
Thus, as AI technologies continue to progress, it is all the more imperative to develop them thoughtfully to avoid reinforcing existing inequities.

Despite AI-assisted tools being revolutionary in improving melanoma detection, the data



**Figure 3: The Malignant Melanoma Five-Year Survival Rate for Black and White Patients-** According to a 2019 study, melanoma has a worse prognosis in Black patients; the five-year survival rate in the United States for African American patients with melanoma was 66%, in comparison to the 90% rate for non-Hispanic White patients (Brady et al, 2021).

available for training these AI networks is lacking in images of darker skin (Wen et al, 2022). AI programs' accuracy is determined by the quality and content of the data sets they are trained on; algorithms trained on scans taken from skin lesions of people with lighter skin will not be as accurate for diagnosing lesions from darker skin. After conducting a study analyzing 21 data sets containing 100,000



**Figure 4: Underrepresentation of Skin Lesion Images from Darker Skinned Individuals-** In a study examining 100,000 skin lesion images, only 11 of them—around 0.011% of the dataset—included dark brown to black skin, illustrating the extreme underrepresentation of darker skin tones (Wen et al, 2022).

images of skin lesions, researchers found that among the data sets where skin color was stated (2,436 pictures), only ten were of brown skin and only one was of dark brown or black skin (Wen et al, 2022) (see Fig. 4\*).

Furthermore, another study found that of 121 data sets that featured

darker skin tones, only 2 contained images of lesions from patients with a VI skin type on the Fitzpatrick Scale (Alipour et al, 2024). Of these two data sets, Fitzpatrick VI skin type accounted for 0.07% (a single image) and 3.97% of all images, respectively (Alipour et al, 2024). The substantial under-representation of darker skin in datasets calls into question the effectiveness of AI in supporting melanoma and skin cancer diagnosis on a global scale (Guo et al., 2021). This lack of representation is deeply rooted in historical inequalities and social stigma, including the misconception that Black people are less susceptible to skin cancer, which has unjustly limited the inclusion of darker skin in dermatological databases (Gupta et al, 2016).

Thus, the resulting social inequality is profound; without inclusive training data, advancements in AI could inadvertently widen the gap in healthcare access and outcomes, leaving marginalized populations at greater risk for misdiagnosis and inadequate treatment.

Hence, this study proposes to develop a CNN that can accurately diagnose malignant lesions in patients across a diverse range of skin tones. By developing a novel CNN architecture based on a publicly available database of benign/malignant skin lesions (i.e., the HAM10000) which have been modified via AI-generated synthesis to improve the sample size of non-White skin, the aim of this study is to achieve a CNN model with high accuracy in diagnosing cancerous lesions in all skin types. Without intentional

efforts to create inclusive AI systems, these tools risk reinforcing systemic healthcare inequalities, ultimately failing to serve all patients equitably.

### **RESEARCH QUESTION:**

The glaring under-representation of darker skin tones in dermatological datasets highlights a critical gap in AI's potential to equitably diagnose melanoma and other skin cancers across diverse populations. By addressing these biases, the conducted work seeks to correct long standing injustices, end historical misconceptions on dark skin, and ensure reliable cancer detection for all. Thus, the central question guiding this research is: *Can one develop an accurate, racially equitable tool via artificial intelligence to diagnose skin cancer?*

### **MATERIALS:**

**Software Components for the Diagnostic AI Model:** The software was designed to address gaps in dermatological data on darker skin and provide medically relevant insights for dermatologists. Programming focused on accuracy and inclusivity, incorporating feedback from dermatologists to optimize the model's diagnostic capabilities and usability. Key components include:

- **Publicly Available HAM10000 Dataset-** The publicly available HAM10000 dataset of skin lesion images was used to train both models (Tschandl et al, 2018). Notably, HAM10000 images are already preprocessed and standardized, eliminating the need for additional preprocessing before running them through the AI models. Out of 10,015 total images, 5,332 were selected (3,762 for training, 941 for validation, and 641 for testing), focusing only on more challenging cases of skin cancer—melanomas, benign keratosis lesions, actinic keratoses and intraepithelial carcinomas, basal cell carcinomas, and melanocytic nevi—to ensure the models would provide *meaningful* assistance to dermatologists. AI support is unnecessary for identifying easily recognizable lesions such as dermatofibromas and vascular lesions. The original HAM10000 images (primarily lighter skin tones) were used to train the CNN for diagnosing light skin, while modified “synthetic” versions of these images (adjusted to mimic darker skin tones) were used to train the CNN for diagnosing dark skin, thereby addressing data gaps for darker skin tones.
- **Publicly Available Diverse Dermatology Images (DDI) Dataset-** Despite limited quality and quantity (making the dataset only usable for testing), the publicly available DDI skin lesion images (Daneshjou et al, 2022) were used for testing and validating the synthetically-trained model's performance on *real* images of darker skin lesions.

- **Publicly Available Journal of the American Academy of Dermatology Images-** Images from the publicly available Journal of the American Academy of Dermatology (Elston et al, 2024) were used as style references for creating the synthetic images.
- **Google Colaboratory (Jupyter Notebook)-** Google Colaboratory served as the programming platform for (1) developing the *Cascading* diagnostic CNNs (*via Python coding*) for both light and dark skin, (2) building the VGG-hybrid network (*via Python coding*) for synthetic image generation, and (3) creating the application to integrate the diagnostic AI model with the physical diagnostic device (*via C++ coding*) (see below).

### **Hardware Components for Applying the Software to Hardware (Total Material Cost ~ \$30):**

To make this software accessible to the public, a physical, Internet of Things (IoT) device was developed to apply the AI diagnostic model in real-world settings. Key components include:

- **ESP-32 Wrover Cam Controller:** Serves as the primary processing unit for capturing and analyzing images. This controller connects to WiFi, enabling access to the diagnostic CNN model in order to generate skin lesion diagnoses.
- **Bread Board:** Serves as a base for wiring and connections.
- **Blue, Red, Yellow LED Bulbs:** Used to indicate a benign, malignant, and further test categorization, respectively.
- **24 and 26 Gauge Wires:** Connect the various parts to the central controller.
- **2 MP Camera:** Captures detailed images of skin lesions for diagnostic analysis.
- **USB-C Port (5V Power Supply):** When plugged into an outlet, the USB port powers the device.
- **Push Button SW for Image Capture:** Initiates the image capture of skin lesions.
- **6x4x2 Inch Plastic Box:** Provides protective housing for the device's components.
- **Screwdriver, Screws, Drill:** For assembling and securing the hardware elements.
- **Sticker Labels:** For organizing and identifying device components.

### **METHODOLOGY:**

The methodology is broken into **4 sections:** (1) CNN Architecture Design; (2) Model Training (via Synthetic Image Development); (3) Statistical Evaluation; (4) Software to Hardware Application.

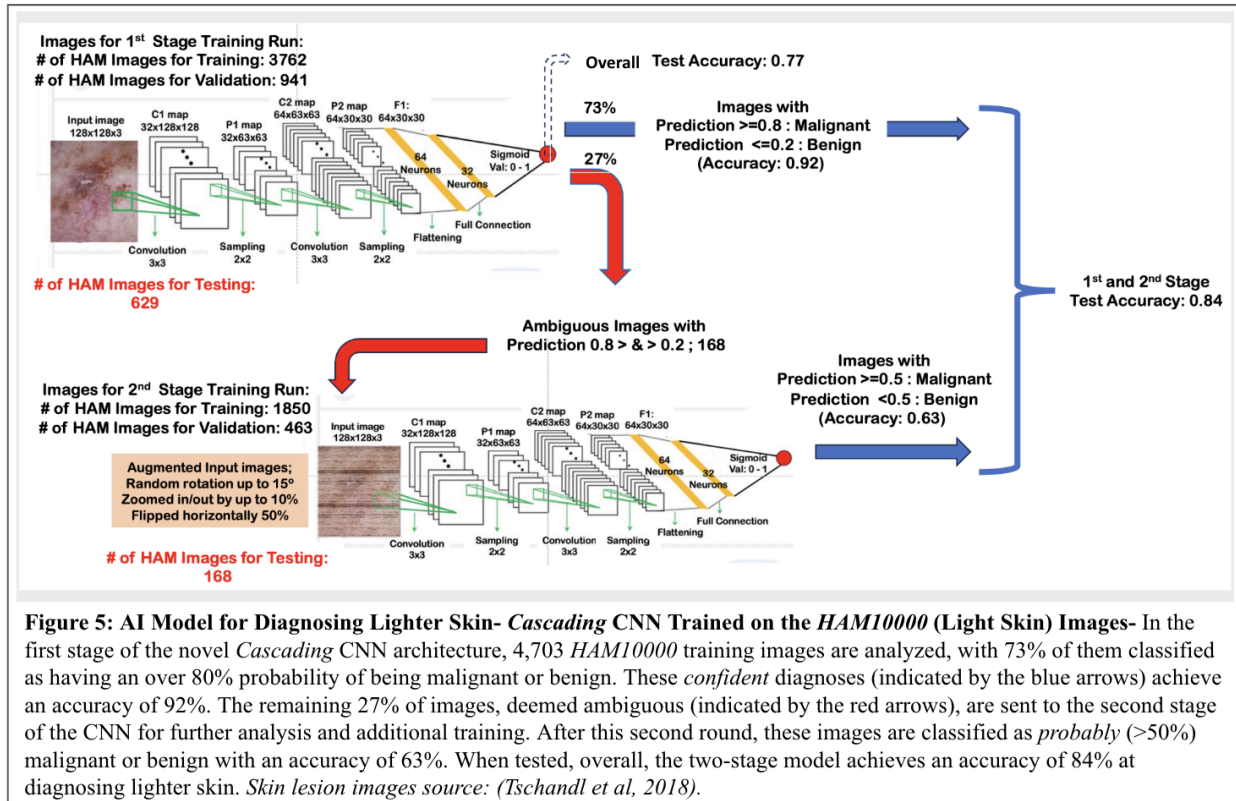
**Section 1- Designing the Novel CNN *Cascading* Architecture:** The CNN model architecture was developed through rigorous iteration. The proposed CNN is developed to detect key image differences of malignant skin lesions (e.g., the abnormal color, size, texture, and shape of malignant skin lesions) by analyzing pixel variations in each distinct characteristic of the skin lesion, such as edges, color gradients, and texture patterns, captured in the input image. By developing commonalities between pixel

variations and the signs of malignancy with which they are correlated, CNNs are able to make generalizations about which pixel patterns connote to the image signature of cancer. The extracted characteristics are then passed through coded pooling layers which condense and refine the information, eventually leading to the diagnoses.

The developed CNN consists of two layers of CNN models, each containing a sequence of convolutional, sampling, convolutional, sampling, flattening, and fully connected layers, ending with a sigmoid activation. In this process, the convolutional layers capture fine details, the sampling (or pooling) layers reduce spatial dimensions to emphasize key features, and the flattening layer organizes the data for classification. The fully connected neurons, combined with the sigmoid function, then produce a probability of malignancy, enabling the model to accurately classify the skin lesion as malignant or benign.

However, in order to focus training and develop a model that is relevant and truly applicable to dermatological settings, the developed model uniquely focuses on learning to diagnose more ambiguous cases of skin cancer. Dermatologists do not need assistance in learning how to diagnose obviously benign, or obviously malignant skin lesions— they need assistance in diagnosing ambiguous cases. Thus, in order to design a model that can assist dermatologists with these more difficult cases and consequently be more medically useful, the model breaks away from traditional design and utilizes a novel *Cascading* architecture:

In the first stage of its novel *Cascading* architecture, the model is programmed to run through the images and based upon its database, denote a “1” if it believes an image has a >80% chance of being malignant, a “0” for a <20% of being malignant (meaning it is benign), and a “T” for ambiguity (chance of malignancy is between 20% and 80%). The “T” images are then augmented—meaning they are modified in various ways, such as by adjusting brightness, rotation, or scale—to create slightly different versions of the same image. This process of augmentation prevents overfitting, which occurs when a model becomes too specialized to the training data, thus performing poorly on new data. The augmented images are then run through a second-stage CNN, allowing the model to better recognize malignant skin cancer in more ambiguous cases. In this second stage, the model denotes a “1” if it decides an image has a  $\geq 50\%$  of being malignant, and a “0” for a <50% chance of malignancy. This process is summarized in Figure 5, with the blue arrows indicating the selection of the benign vs. malignant lesions using the first-stage CNN, while the red arrows indicate the additional analysis of the ambiguous lesions through the *Cascading*, second-stage CNN (see Fig. 5\*).



**Section 2- Model Training (via Synthetic Image Development):** In the case of a CNN designed to detect malignancies from images of skin lesions, the training process involves showing the CNN thousands of labeled images of skin lesions, some of which are malignant and others that are not. As the CNN processes each image, it learns to recognize patterns and features that are characteristic of malignant skin cancer lesions, and uses this information to make accurate predictions about the label of new, previously unseen (testing) skin lesion images. The better the training process, the more accurate and reliable the model's predictions will be on unseen data.

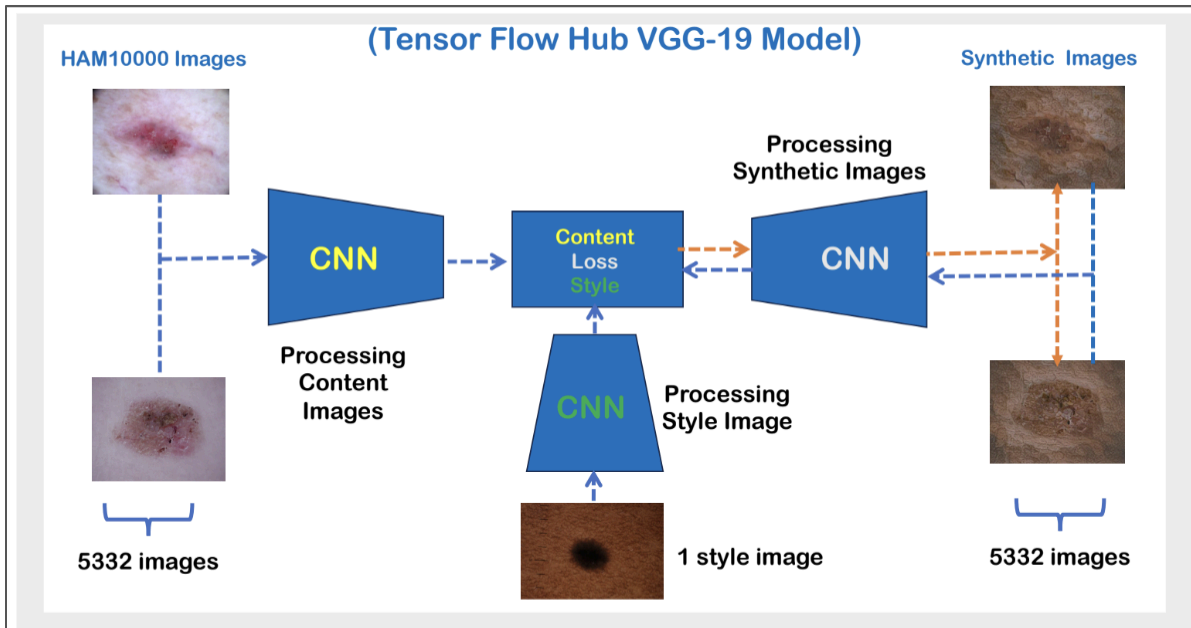
Thus, the publicly-available HAM10000 dataset was the basis for these images. 4,703 skin lesion images from the database (selected as described above) were for training and validating the CNN model for diagnosing light skin (with the remaining 629 images being saved for statistical evaluation). However, in order to develop a racially equitable model, the model must be trained on high-quality skin lesion images from a diverse range of skin tones. Given that there is no publicly available data set containing high-quality images of skin lesions from darker skin, AI-generative image methods were utilized to mitigate this racial gap in skin cancer data and transform the light skin images from the HAM10000 dataset to mimic skin lesions from darker skin.

The HAM10000 images were transformed to darker skin using an AI-driven style transfer network. Contrary to conventional approaches which simply permute the baseline image, the AI style

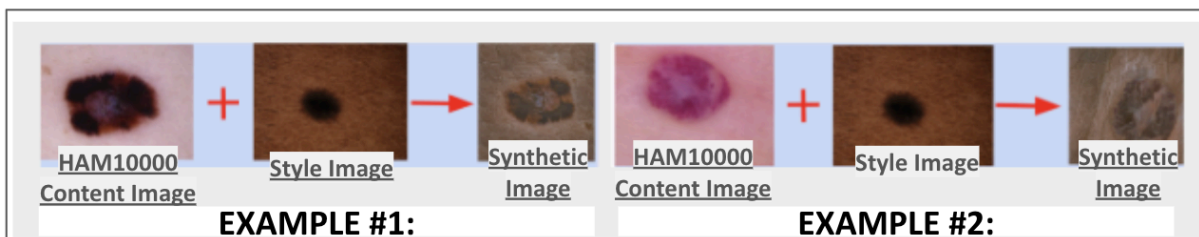
transfer algorithm implemented is generative in nature. In other words, instead of merely applying a darker color to a skin lesion image, the algorithm synthesizes new image details. This technique enables a more nuanced integration of the “style” skin tone with the lesion’s underlying characteristics, producing a realistic representation that more accurately reflects how skin cancer manifests across different skin tones.

To do so, the generative multi-layer VGG-hybrid network takes in two inputs: a vectorized “content” image, which represents the baseline skin lesion, and a “style” image, which provides the skin color to be applied onto the baseline image. By computing the similarity between the spatial statistics of the style image (via VGG) and the generative output, the style transfer network in effect “grafts” the nuances of ethnic skin (beyond mere “pixel colorings”) onto the HAM10000 baseline image. For instance, the earlier layers of the model will integrate features such as color, texture, and edges—by identifying and incorporating variations in pixel intensity and placement to create the fine lines, pores, and subtle surface irregularities that make skin look realistic—from the ethnic skin image into the HAM10000 baseline image. In the later layers, the model will further blend the deeper nuances of the style image into the content.

Fidelity to the original content image is maintained via a simultaneous Euclidean Mean Square Error (MSE) loss function between the generative and initial content image. In effect, the network infuses the intricacies of ethnic skin lesions (the style images from the Journal of the American Academy of Dermatology) onto the HAM10000 dataset (the content images); the Cascading CNN model is then re-trained to diagnose cancer on the high-quality synthetically-generated dermatoscopic dark skin (*see Figs. 6 and 7\**).



**Figure 6: Generating Synthetic Images-** The *HAM10000* images (on the left) are transformed to represent darker skin tones using a generative multi-layer VGG-hybrid network. This AI-driven model takes two inputs: a vectorized “content” image (the *HAM10000* images) and a “style” image (skin lesion images from darker skin sourced from the public *Journal of the American Academy of Dermatology*). By analyzing the similarity between the spatial statistics of the “style” image (using VGG) and the generative output, the style transfer network effectively grafts the unique characteristics of ethnic skin from the “style” image onto the baseline “content” image, resulting in the synthetic skin lesion images (on the right). *Skin lesion images sources: (Tschandl et al, 2018); (Elston et al, 2024).*



**Figure 7: AI-Driven “Dark Skin” Synthetic Image Generation-** As shown in this more condensed view of the synthetic image generation process, the *HAM10000* content image of a skin lesion from lighter skin is combined with a style image of a skin lesion from darker skin. This process results in a synthetic yet realistic representation of skin lesions from darker skin. *Skin lesion images sources: (Tschandl et al, 2018); (Elston et al, 2024).*

These AI-generated images were then used to train the AI model for dark skin, which follows the same *Cascading* architecture as the model trained on unaltered *HAM10000* images for diagnosing lighter skin (see Fig. 8\*), thereby creating a model specifically designed to be racially-inclusive. Importantly, the CNN for diagnosing light skin was trained on the same images as the CNN for dark skin, with the latter using synthetically generated surrounding dark skin. Consequently, both models are expected to exhibit



benign diagnoses out of the total, determined by comparing the AI model's diagnoses with their published pathological results.

The **second statistical evaluation** involved assessing the efficacy of using the novel Cascading architecture to increase diagnostic accuracy. In order to do so, the testing images were run through both the Cascading CNNs for light skin and dark skin, as well as standard one-stage CNN models. The accuracies produced by each CNN were then compared to determine the extent to which the Cascading architecture improved diagnostic accuracy for both the light-skin and dark-skin models.

The **third statistical evaluation** involved assessing the efficacy of using AI-generated images to train a model that can diagnose skin cancer in real darker-skinned patients. In order to do so, both the HAM10000 trained model for diagnosing lighter skin, and the synthetically trained model for diagnosing darker skin were tested on the DDI images containing real skin lesion images from darker skin. The two models' accuracies were then compared to assess how training on synthetic images of darker skin improves the accuracy in diagnosing real darker skin. (*NOTE: Validation with DDI is flawed. DDI images contain noise and irrelevant background details- see Fig. 9; therefore, only a limited number of images could be used from the DDI for validation. Further clinical validation using standardized image acquisition, i.e., by dermatoscope, will be conducted during anticipated future clinical trials.*)

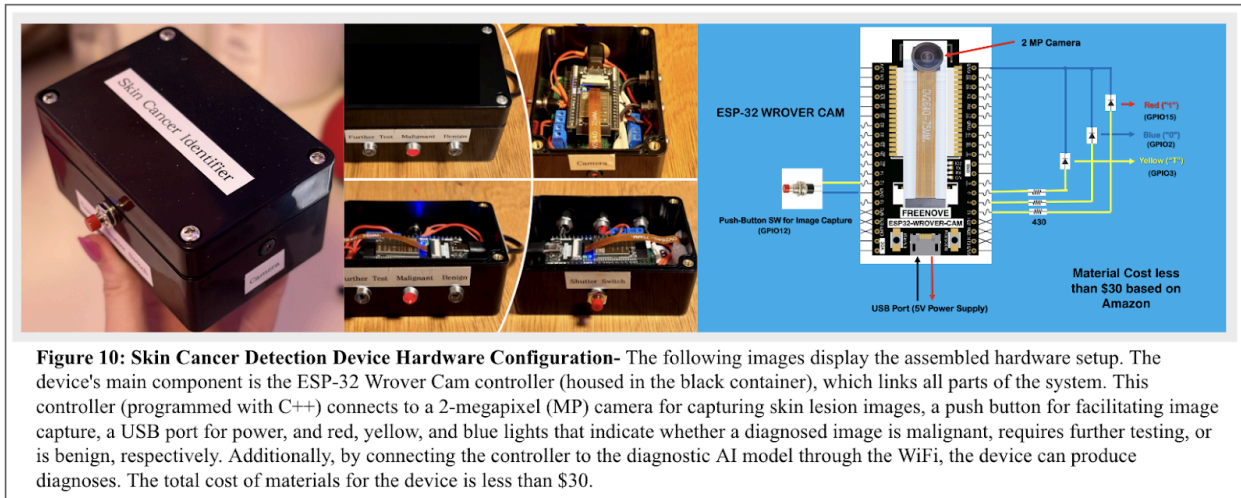


**Figure 9: Poor Image Quality of the *Diverse Dermatology Images (DDI)* Data Set-** The images above showcase samples from the public *Diverse Dermatology Images (DDI)* dataset, featuring real skin lesion images from patients with darker skin that were used to validate the effectiveness of the *synthetic* darker-skinned images. As shown, the quality of the dataset is quite poor, as it includes irrelevant background footage (such as bed linens, floors, or tabletops) and lacks focus on the skin lesions themselves. Therefore, testing with these compromised DDI images will result in lower performance scores; further testing using higher-quality images is consequently necessary for more accurate validation and (most likely) higher performance scores. *Skin lesion images source: (Daneshjou et al, 2022).*

The **fourth statistical evaluation** was a comparison of the developed model to other existing models as well as dermatologists. The models in this study for diagnosing DDI images were compared to published results from widely-used AI models—ModelDerm, DeepDerm, and the official HAM10000

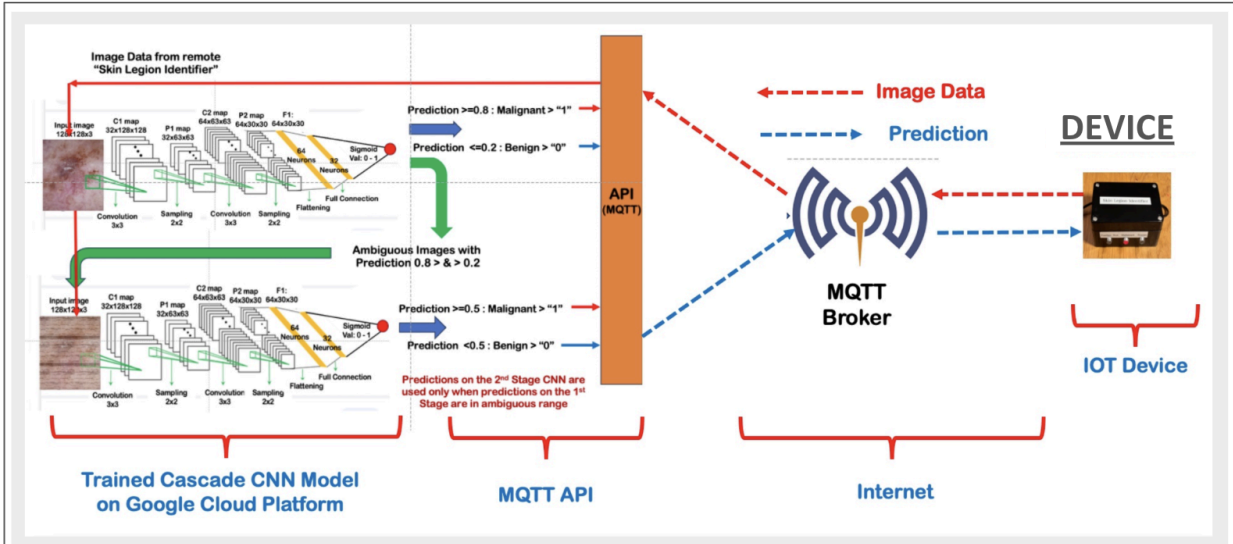
model—as well as dermatologists, in diagnosing both real lighter skin and darker skin images from the same DDI dataset presented by Daneshjou et al (2022). The performance of the model will be gauged via sensitivity score, i.e., the rate of correct malignant diagnoses. Misdiagnosis is a critical issue in skin cancer diagnosis that must be addressed. Therefore, sensitivity score is the best metric for evaluating the models' performance, as it emphasizes its ability to avoid misdiagnosis.

**Section 4- Software to Hardware Application:** To ensure the AI model is accessible for practical use, a user-friendly IoT hardware component was developed. The current device is compact, functional, and the material costs equate to under \$30. It features an ESP-32 Wrover Cam controller, which integrates all system components. The controller, programmed in C++, is connected to a 2-megapixel (MP) camera for capturing skin lesion images, a push button for image capture, a USB port for power, and indicator LED lights (red, yellow, and blue) signaling whether the diagnosis is malignant, requires further testing, or is benign, respectively (*see Fig. 10\**).



**Figure 10: Skin Cancer Detection Device Hardware Configuration-** The following images display the assembled hardware setup. The device's main component is the ESP-32 Wrover Cam controller (housed in the black container), which links all parts of the system. This controller (programmed with C++) connects to a 2-megapixel (MP) camera for capturing skin lesion images, a push button for facilitating image capture, a USB port for power, and red, yellow, and blue lights that indicate whether a diagnosed image is malignant, requires further testing, or is benign, respectively. Additionally, by connecting the controller to the diagnostic AI model through the WiFi, the device can produce diagnoses. The total cost of materials for the device is less than \$30.

The CNN connects to the hardware through an internet-based interface, using an API and MQTT requests and responses that connect to the ESP-32 controller (*see Fig. 11\**). This allows for seamless communication between the hardware and the CNN, enabling real-time diagnosis directly on the device.



**Figure 11: System Summary of Diagnosis Sequence from Cascading CNN to Device Output-** The CNN model (software) is connected to the hardware implementation via an internet interface- Application Programming Interface (API). Through the MQTT Broker, the device is able to send captured images to the CNN uploaded on Google Cloud. The CNN then sends a signal back to the device indicating a “malignant,” “benign,” or “further test” prediction. *Skin lesion images source: (Tschandl et al, 2018).*

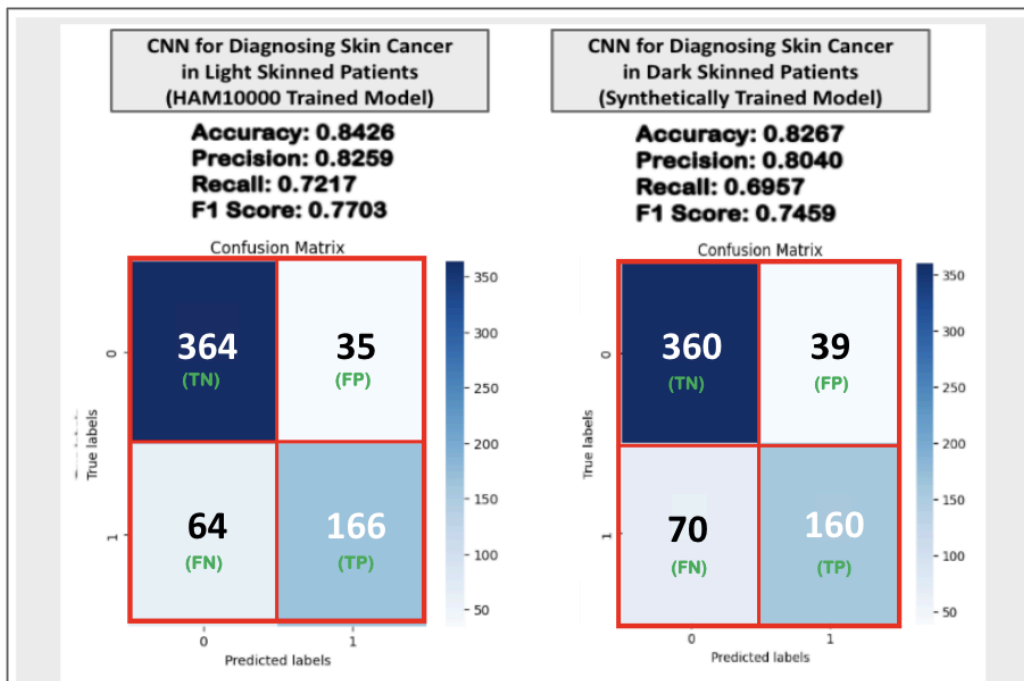
Upon the development of the hardware component, statistical evaluation for the device was also conducted to assess its diagnostic reliability and responsiveness. This involved comparing the device's diagnostic results to confirmed pathological outcomes to determine overall accuracy in detecting malignant and benign lesions.

Moreover, the software connected to the device is designed to automatically delete each image after processing. This ensures that no personal data is retained, prioritizing privacy and data security.

## RESULTS:

The results of the undertaken study can be broken into **5 parts**: **(1)** An evaluation of the total diagnosis system; **(2)** An evaluation of the novel *Cascading* CNN architecture; **(3)** An evaluation of the efficacy of the synthetic training images; **(4)** A comparison of the developed model to existing models and dermatologists; **(5)** Validation of the diagnostic device.

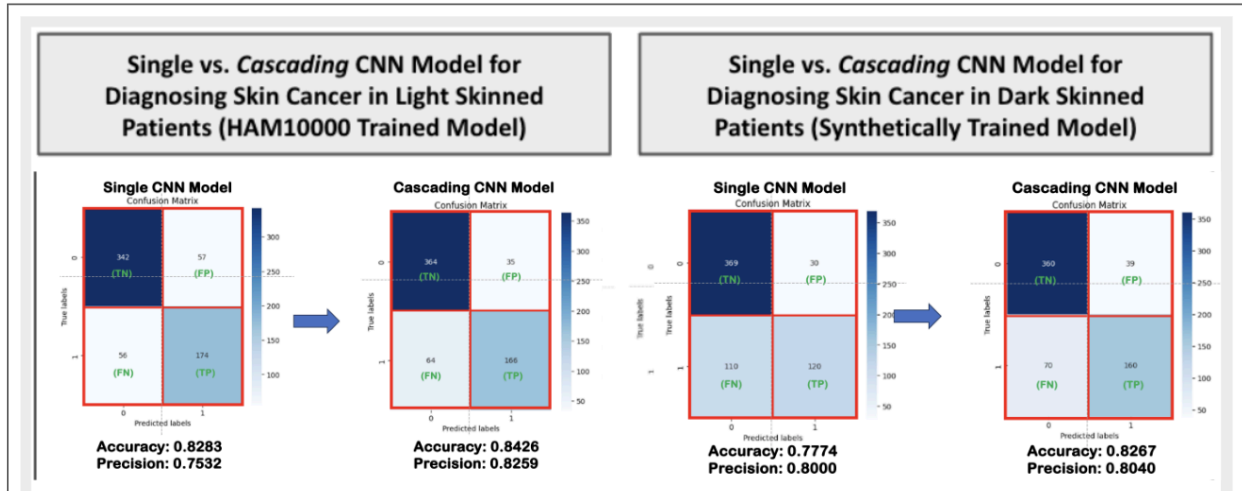
**Part 1- An Evaluation of the Total Diagnosis System:** Overall, when tested on 629 skin lesion images, the two developed CNNs for diagnosing skin cancer in patients with light and dark skin achieved an accuracy of 84% and 83%, respectively. The CNN trained to diagnose skin cancer in lighter skinned patients diagnosed 364 true negatives (TN), 166 true positives (TP), 64 false negatives (FN), and 35 false positives (FP). The CNN trained to diagnose skin cancer in darker skinned patients diagnosed 360 true negatives, 160 true positives, 70 false negatives, and 39 false positives (*see Fig. 12\**).



**Figure 12: Confusion Matrix and Performance Metrics for the Total AI Diagnosis System-** Overall, after testing the two CNNs on 1,258 skin lesion images (629 of lighter and 629 of darker skin), the developed models for light and dark skin achieved accuracies of 84% and 83% (in addition to precision scores of 83% and 80%, recall scores of 72% and 70%, and F1 scores of 77% and 75%) respectively. Accuracy reflects the overall correctness of diagnoses across all images, precision indicates the percentage of true positive malignant diagnoses out of all malignant predictions, recall measures the model's ability to identify actual malignant cases, and F1 score represents a balance between precision and recall, offering a single metric to assess both false positives and false negatives.

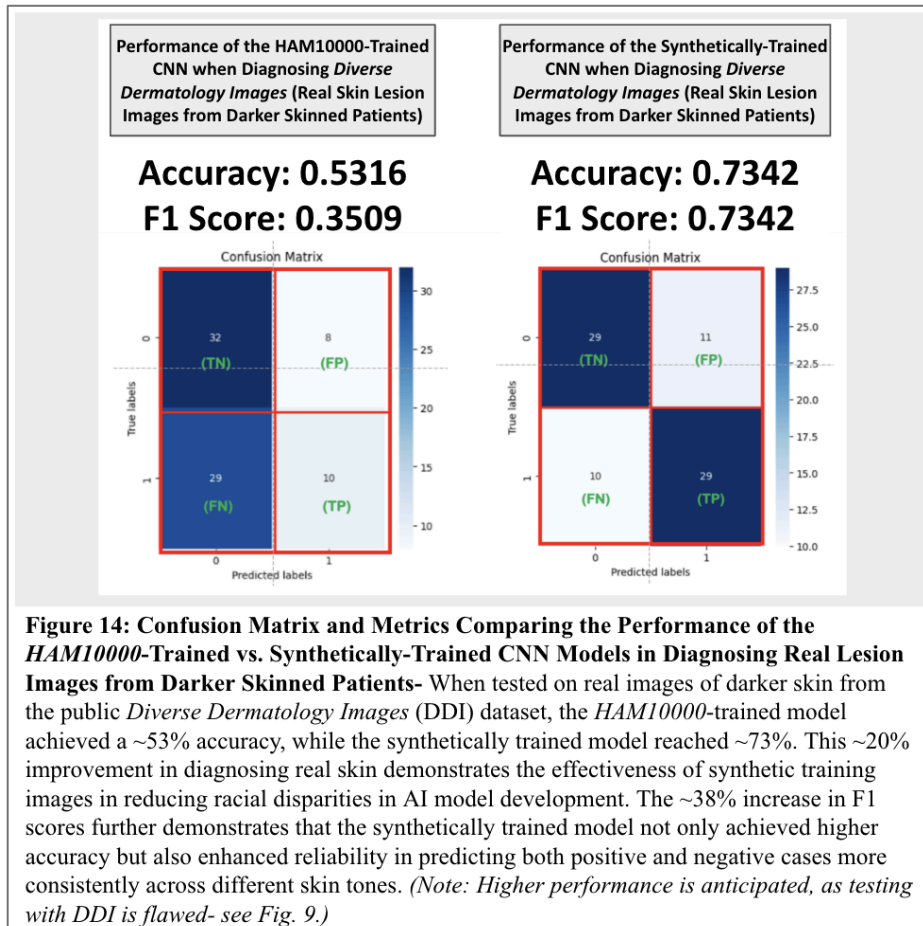
**Part 2- An Evaluation of the Novel *Cascading* CNN Architecture:** In both cases (in models diagnosing light and dark skin), the novel *Cascading* models resulted in higher diagnostic accuracies in comparison to the standard (single) CNN models. The *Cascading* CNN Model for diagnosing skin cancer

in lighter skin achieved a 1.43% increase in accuracy compared to the standard (single) CNN model. Similarly, the *Cascading* CNN Model for diagnosing skin cancer in darker skin achieved a 4.89% increase in accuracy compared to the standard (single) CNN model (see Fig. 13\*).



**Figure 13: Confusion Matrix and Performance Metrics Comparing the Novel *Cascading* CNN Architecture with Standard CNN Architecture-** Compared to standard CNN models, the novel *Cascading* CNN architecture improved accuracy by 1% for diagnosing lighter skin (increasing from 83% to 84%), and 5% for diagnosing darker skin (increasing from 78% to 83%). Precision scores also increased by 8% for diagnosing lighter skin (increasing from 75% to 83%), and 0.4% for diagnosing darker skin (increasing from 80.0% to 80.4%). This increase in performance validates the usage of the novel *Cascading* architecture to increase diagnostic capabilities in dermatological settings.

**Part 3- An Evaluation of the Efficacy of the Synthetic Training Images:** When evaluated on real skin lesion images from darker skinned patients (from the DDI dataset), the HAM10000-trained CNN (which was only trained on images of light skin) achieved a 53% accuracy. In comparison, the model trained on synthetic dark skin lesion images achieved a 73% accuracy. Thus, the synthetic images improve accuracy in diagnosing skin cancer in dark skin by 20%. Additionally, the synthetic images increased the models' F1 score by 38%, improving from 35% to 73% (see Fig. 14\*). (NOTE: Due to the quality of the DDI images, validation using DDI is flawed. With the racially diverse and dermatoscopic acquired images expected from prospective clinical trials, one would expect to see accuracy levels exceeding 73%).

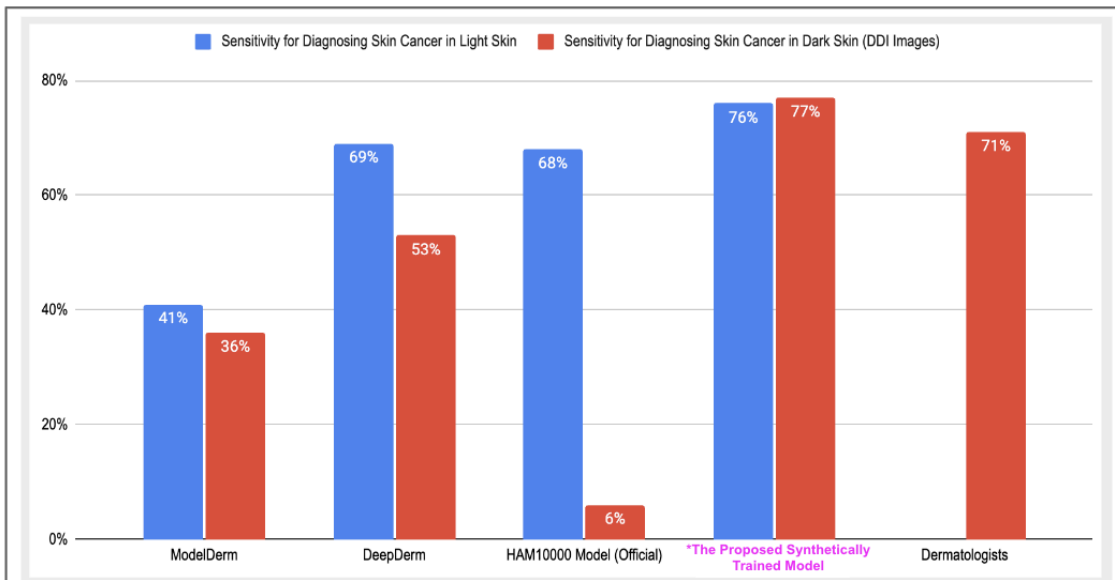


These three portions of the results can be summarized in the following table (see Tbl. 1):

	Overall System Architecture Accuracy	Improvement in Accuracy when Utilizing a <i>Cascading</i> Architecture	Accuracy in Diagnosing Patients with <i>Real</i> Dark Skin
Model for Light Skin (Trained via <i>HAM10000</i> Images)	<b>84%</b>	<b>1%</b>	<b>53%</b>
Model for Dark Skin (Trained via Synthetic Images)	<b>83%</b>	<b>5%</b>	<b>73%</b>

**Table 1: Summary of Results-** The following table summarizes the three former parts of results. First, overall, the AI systems achieved an 84% accuracy in diagnosing cancer in lighter skin and 83% in darker skin. Second, using the novel *Cascading* CNN architecture instead of standard CNN models, diagnostic accuracy improved by 1% for lighter skin and 5% for darker skin. Third, when comparing a CNN trained solely on lighter skin data with one trained on *synthetically generated* darker skin, accuracy for diagnosing *real* darker skin increased by 20%, from 53% to 73%. The percent accuracies demonstrate the superiority of the *Cascading* architecture, the effectiveness of synthetic images in developing AI models for diagnosing real darker skin despite limited data, and the models' racial equity.

**Part 4- Comparison to Existing Research:** In comparison to other models, the model presented here showed stable sensitivity after training on synthetic images, thereby demonstrating the validity of training AI on synthetic data for skin cancer when actual patient data availability may be limited. Other models like DeepDerm, ModelDerm, and the HAM10000 Model show a decrease in sensitivity when tested on skin lesion images—collected from the same DDI dataset—from darker skin (Daneshjou et al., 2022). When tested on the same images, the developed model, with a sensitivity of 77%, also outperforms dermatologists with a sensitivity of 71% and uniquely showed an increase in sensitivity in darker skinned images from the DDI dataset (*see Fig. 15\**).



**Figure 15: Evaluating the Proposed Model to Current Diagnostics, Including Both AI Models and Dermatologists Performance-** The proposed model maintains a stable sensitivity score for diagnosing skin cancer in “real”patients with darker skin, achieving 76% sensitivity for lighter-skinned and 77% for darker-skinned patients. In contrast, when evaluated on the same dataset, the published AI models *ModelDerm*, *DeepDerm*, and the official *HAM10000* show declines in sensitivity for light versus dark skin, dropping from 41% to 36%, 69% to 53%, and 68% to 6%, respectively. The proposed model also outperforms dermatologists (with a 71% sensitivity- the % of correct malignant diagnoses) by 6% in diagnosing these same darker-skinned patient images, validating the use of synthetic images for training reliable, racially-equitable diagnostic AI models. Thus, in addition to outperforming dermatologists, the proposed model is the only racially-equitable model.

**Part 5- Validation of the Diagnostic Device:** To evaluate the physical device’s capability to diagnose real skin lesions in patients with dark skin, 44 – 22 malignant and 22 benign – randomly selected DDI images (representing dark skin lesions) were tested. Comparing the device’s predictions with the published pathological confirmations, the device captured (via its attached camera) and within seconds, accurately identified 19 malignant and 19 benign lesions, achieving an overall accuracy and sensitivity of 86.4% (*see Tbl. 2 and Fig. 16\**).



## **FURTHER STEPS:**

**Clinical Trials:** Given the promising results, a faculty investigator and dermatologist has agreed to conduct a clinical trial of the developed algorithm and device. Patients, anticipated to be two-thirds non-White based on the clinic's typical demographics, will be provided consent to participate during their routine follow-up appointments. This trial will provide critical validation of the model and device's effectiveness before it is implemented in clinical care.

## **DISCUSSION:**

The results presented affirm the hypothesis, showing that the developed Cascading and synthetically-trained CNN model, together with the physical device, effectively enhances the accuracy of skin cancer diagnosis for darker-skinned patients. This objective was systematically achieved by:

1. The development of a novel *Cascading* CNN model that achieved a 5% and 1% superior accuracy compared to single CNN models for dark and light skin, respectively. This increase in performance indicates that the novel *Cascading* architecture is an effective design for improving diagnostic accuracy.
2. The creation of synthetic images of skin lesions from darker skin by employing a generative multi-layer VGG-hybrid AI model that would alternatively have been unavailable given the lack of existing data. The research faced challenges due to the lack of quality images of darker skin. However, this issue was addressed by generating AI-created data, demonstrating that AI image generation can effectively mitigate data shortages for marginalized populations. This method improved diagnostic accuracy for *real* dark skin by 20%, and F1 score by 38%. As such, this improvement indicates that synthetic images are an effective way to train an accurate and racially equitable model. The increase in F1 score also indicates improved reliability, showing the model's enhanced accuracy in correctly identifying true positives while reducing errors, making it a more consistent and fair diagnostic tool across diverse skin tones.
3. Training the *Cascading* CNN model on the original HAM10000 images as well as the synthetically created images of skin lesions from darker skin. Overall, the diagnostic accuracy was comparable with an accuracy rate of 84% for light skin and 83% for dark skin. The relatively similar diagnostic accuracies between the model for light skin and the model for dark skin indicate the development of a racially equitable tool.
4. The validation of the synthetically trained model on real skin lesions from darker skin via the DDI dataset. The developed model demonstrated a superior sensitivity (77%) in classifying dark skin compared to other AI models (with sensitivities ranging from 6% to 53%) and even dermatologists (with a sensitivity of 71%). Consequently, the results support the hypothesis,

demonstrating the efficacy of the developed *Cascading* and synthetically-trained CNN model in improving the accuracy of skin cancer diagnosis among darker skinned patients. The developed model showed stable sensitivity after training on synthetic images, thereby demonstrating the validity of training AI on synthetic data for skin cancer when actual patient data availability may be limited. It also outperforms dermatologists and other published models. Thus, the accomplished CNN rectifies racial inequities in skin cancer diagnosis by providing a usable tool for patients of all skin tones to diagnose their skin lesions.

5. The proof-of-concept creation of a low-cost IoT device which sends images remotely through MQTT protocol to the *Cascading* CNN model residing on a cloud platform to provide a diagnosis on the remote device. The device can also serve as a tool to support impoverished communities across the globe who do not have access to immediate healthcare. By pressing a button, a user can capture an image of their skin lesion and receive an immediate diagnosis. Thus, one device can serve an entire community of people. This affordable, user-friendly device enables secure, private, widespread access to accurate skin cancer diagnostics, particularly for underserved communities.

Without intentional efforts to create inclusive AI systems, these tools risk reinforcing systemic healthcare inequalities, ultimately failing to serve all patients equitably. These results indicate that this method and device have the capacity to capture skin lesion images and accurately diagnose them with a sensitivity outperforming dermatologists and existing AI models across different skin tones.

## **CONCLUSION:**

Thus, this research represents a profound step toward social justice in healthcare by addressing critical gaps in representation and diagnostic accuracy. The diagnostic tool, designed to perform reliably across all skin tones, offers a path to equitable medical resources for communities that have historically been underserved in dermatology. For individuals with darker skin—often at a higher risk of late-stage skin cancer diagnoses due to insufficient data representation—this advancement means earlier detection, improved treatment outcomes, and, ultimately, a greater chance of survival. While additional validation is expected and necessary, the development of synthetic data for darker skin and a low-cost IoT ensures that essential, life-saving technology reaches those who need it most, irrespective of socioeconomic barriers.

As AI continues to develop, it is imperative that the technology propelling society forward does not perpetuate the same inequities that have historically held society back. In an increasingly AI-driven future, scientists must guide its development as a force for progress that empowers and uplifts, ensuring that the benefits of innovation are distributed equitably and meaningfully across all people.

**BIBLIOGRAPHY:** *Note: (\* denotes figures created by Kate Choi unless otherwise specified).*

Alipour N, Burke T, Courtney J (2024). Skin type diversity in skin lesion datasets: a review. *Curr Dermatol Rep*, 13(3):198–210. doi: 10.1007/s13671-024-00440-0.

American Cancer Society (2023). *Cancer Facts & Figures 2023*. Atlanta: American Cancer Society.

Brady J, Kashlan R, Ruterbusch J (2021). Racial disparities in patients with melanoma: a multivariate survival analysis. *Clin Cosmet Investig Dermatol*, 14:547-550. doi: 10.2147/CCID.S311694.

Daneshjou J, Vodrahalli K, Novoa RA, et al (2022). Disparities in dermatology AI performance on a diverse, curated clinical image set. *Sci Adv*, 8(32):eabq6147. doi: 10.1126/sciadv.abq6147.

Diao JA, Adamson AS (2022). Representation and misdiagnosis of dark skin in a large-scale visual diagnostic challenge. *J Am Acad Dermatol*, 86(4):950-951. doi:10.1016/j.jaad.2021.03.088.

Elston DM, Sloan B, Kantor J (2024). JAAD Skin of Color Image Atlas. *J Am Acad Dermatol*. <https://www.jaad.org/collection-skin-of-color-images>.

Gupta AK, Bharadwaj M, Mehrotra R (2016). Skin cancer concerns in people of color: risk factors and prevention. *Asian Pac J Cancer Prev*, 17(12):5257–5264. doi: 10.22034/APJCP.2016.17.12.5257.

Guy GP, Jr., Machlin SR, Ekwueme DU, Yabroff KR (2015). Prevalence and costs of skin cancer treatment in the U.S., 2002-2006 and 2007-2011. *Am J Prev Med*, 48(2):183-187. doi:10.1016/j.amepre.2014.08.036.

LeCun Y, Bengio Y, Hinton G (2015). Deep learning. *Nature*, 521(7553):436-44. doi: 10.1038/nature14539.

Siegel RL, Miller KD, Wagle NS, Jemal A (2023). Cancer statistics, 2023. *CA Cancer J Clin*, 73(1):17-48. doi: 10.3322/caac.21763.

Soenksen LR, Kassis T, Conover ST, et al (2021). Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. *Science Translational Medicine*, 13(581). doi: 10.1126/scitranslmed.abb3652.

Tschandl P, Rosendahl C, Kittler H (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data*, 5: 180161. doi: 10.1038/sdata.2018.161.

Wen D, Khan SM, Xu AJ et al (2022). Characteristics of publicly available skin cancer image datasets: a systematic review. *The Lancet Digital Health*, 4(1): E63-E74. doi: 10.1016/S2589-7500(21)00252-1.