

Original Paper

# Deep Learning Multi-Modal Melanoma Detection: Algorithm Development and Validation

Nithika Vivek<sup>1</sup>; Karthik Ramesh<sup>2</sup>, MD

<sup>1</sup>Del Norte High School, San Diego, CA, United States

<sup>2</sup>Department of Internal Medicine, University of California, Los Angeles, Los Angeles, CA, United States

**Corresponding Author:**

Nithika Vivek  
Del Norte High School  
16601 Nighthawk Ln  
San Diego, CA 92127  
United States  
Phone: 1 619 458 5059  
Email: [nithika.vivek@gmail.com](mailto:nithika.vivek@gmail.com)

## Abstract

**Background:** The visual similarity of melanoma and seborrheic keratosis has made it difficult for older patients with disabilities to know when to seek medical attention, contributing to the metastasis of melanoma.

**Objective:** This study aimed to present a novel multimodal deep learning-based technique to distinguish between melanoma and seborrheic keratosis.

**Methods:** Our strategy is three-fold: (1) use patient image data to train and test three deep learning models using transfer learning (ResNet50, InceptionV3, and VGG16) and one author-designed model, (2) use patient metadata to train and test a deep learning model, and (3) combine the predictions of the image model with the best accuracy and the metadata model, using nonlinear least squares regression to specify ideal weights to each model for a combined prediction.

**Results:** The accuracy of the combined model was 88% (195/221 classified correctly) on test data from the HAM10000 dataset. Model reliability was assessed by visualizing the output activation map of each model and comparing the diagnosis patterns to that of dermatologists. The addition of metadata to the image dataset was key to reducing the false-negative and false-positive rates simultaneously, thereby producing better metrics and improving overall model accuracy.

**Conclusions:** Results from this experiment could be used to eliminate late diagnosis of melanoma via easy access to an app. Future experiments can use text data (subjective data pertaining to how the patient felt over a certain period of time) to allow this model to reflect the real hospital setting to a greater extent.

*JMIR AI 2025;4:e66561*; doi: [10.2196/66561](https://doi.org/10.2196/66561)

**Keywords:** melanoma; dermatology; artificial intelligence; deep learning; multi modal; geriatric; metastasis; seborrheic keratosis; patient image data; accuracy; computer vision

## Introduction

Incidence rates of melanoma have been on an increase since 1999, with 15.1 per 100,000 in 1999 and rising to 23.0 per 100,000 in 2021 [1]. In contrast, seborrheic keratosis is a benign skin appearance that commonly occurs in older adults. While the pathology, epidemiology, and histology of melanoma and seborrheic keratosis are well understood [2, 3, 4, 5], on a surface level, these 2 lesions can seem almost identical to the untrained eye, making it difficult for individuals to know when to seek care [6]. Delayed care can allow a malignant lesion to progress into metastatic

melanoma. As the stage of melanoma progresses, the survival rate can decrease as much as 67% [7]. Thus, timely diagnosis and treatment are paramount.

The current diagnostic paradigm has not significantly advanced despite staggering technological leaps. A typical process involves a patient visiting a primary care clinic, followed by a referral to a dermatologist if there are any unusual skin lesions [8]. The dermatologist repeats the skin exam, and then further performs biopsies or excision as required [9]. These samples are sent for pathology, which makes the final diagnosis. This process requires an iterative process involving appropriate presentation to a primary care

provider, appropriate referral, appropriate visual analysis, appropriate surgical excision, all before a diagnosis can be made [10].

Deep learning models have commonly been used to encourage at-home, self-diagnosis, or easier physician diagnosis of melanoma [11, 12]. One such experiment used an Iterative Dichotomiser 3 (ID3) algorithm to learn rules from image data using texture patterns, a method known as automatic induction [13]. Another method employed transfer learning [14] and used ResNet152 to develop a binary classifier between benign and malignant skin lesions [15].

**Table 1.** Comparison of machine learning approaches taken in 65 dermatological applications across the internet, with analysis on either the HAM10000 dataset, ISIC dataset, DFUC dataset, or other common datasets ranging from 46-33126 data points for evaluation [14].

Model	Accuracy (area under the curve)
ResNet-50	71.620
VGG	68.408
InceptionV3	74.311
ResNet-50 and Inception V3	85.977
ResNet-50 and VGG	83.065
ID3 <sup>a</sup>	71.000
BottleNeckCSP	81.000

<sup>a</sup>ID3: Iterative Dichotomiser 3.

Tabular data has been extensively used in various health applications, serving as the basis of many prediction algorithms and machine learning models [16]. One relevant dermatological example used clinical features to represent the redness, flakiness, definite border extent, and other qualities to classify 6 types of erythematous-squamous skin diseases using the UCI Dermatology dataset [17]. Past tabular metadata for health applications have been used to diagnose other, nondermatological-related diseases. A Dual Bayesian ResNet50 model was used to train metadata regarding heart murmurs using XGBoost [18]. Broader applications of tabular metadata have been used through a method called MediTab, in which diverse, out-of-sample data is consolidated and aligned to improve prediction accuracy [19]. Time progression tabular deep learning was used for hypercholesterolemia, in which a multistage deep learning architecture was used to analyze familial hypercholesterolemia [20]. However, this method was not integrated with image data and was purely reliant on tabular data.

Image and tabular predictions can be combined into a hybrid model by using nonlinear least squares regression (NLS) by incorporating both image and tabular predictions in a unified regression model. Past studies have found NLS useful for fusion of heterogeneous sources of data due to its ability to model complex, nonlinear relationships inherent in such data [21]. NLS is a common technique used to fit a model to data by minimizing the square sum of residuals or the squared differences between observed data points and values predicted by the nonlinear model. Minimizing

Table 1 describes the average area under the curves (AUCs) for medical imaging segmentation for various dermatological machine learning models proposed in literature. The best-performing model was a combination of ResNet-50 and InceptionV3, with an accuracy of 80%. Most of these approaches aim to optimize models through transfer learning and various preprocessing techniques in an attempt to increase accuracy.

this difference allows for the predictions to more accurately reflect the true value. NLS has been used in pharmacokinetics to understand drug absorption, distribution, metabolism, and excretion [22]. Other applications of NLS appear in tumor growth analysis and medical imaging to enhance image quality [23,24].

Because melanoma prevalence can vary among different demographics, image inputs or metadata inputs alone may not be sufficient in formulating an accurate diagnosis [25]. This paper aims to build on the previous experiments stated and incorporate metadata into the model inputs. While NLS regression has been previously commonly used on raw medical data, this application of NLS leverages its square residuals minimizing abilities to determine ideal weights for the combination of tabular and image data at the output. Finally, providing the model with multiple input modalities helps capture heterogeneous factors that decrease the chances of the model formulating false patterns during classification.

## Methods

### Overview

Our multimodal deep learning architecture assembly is threefold: (1) use patient image data to train and test three deep learning models using transfer learning (ResNet50, InceptionV3, and VGG16) and one author-designed model, (2) use patient metadata to train and test a deep learning model, and (3) combine the predictions of the image model with the best accuracy and the metadata model, using

nonlinear least squares regression to specify ideal weights to each model for a combined prediction.

## **Dataset Analysis**

The data used in this experiment was obtained from the HAM10000 dataset [26]. 2259 images were taken from the practice of Cliff Rosendahl consecutively starting 2008 until 2017. 7756 images were taken from the University of Vienna in 1988. Because images were collected from different time periods, some were preprocessed with enhanced contrast and zoom while others were not. While all types of skin conditions were captured in the dataset, for the purposes of this analysis, those images not classed as seborrheic keratosis or melanoma were removed. There were a total of 2210 images, with 50% (1105 images) belonging to melanoma and 50% (1105 images) belonging to seborrheic keratosis.

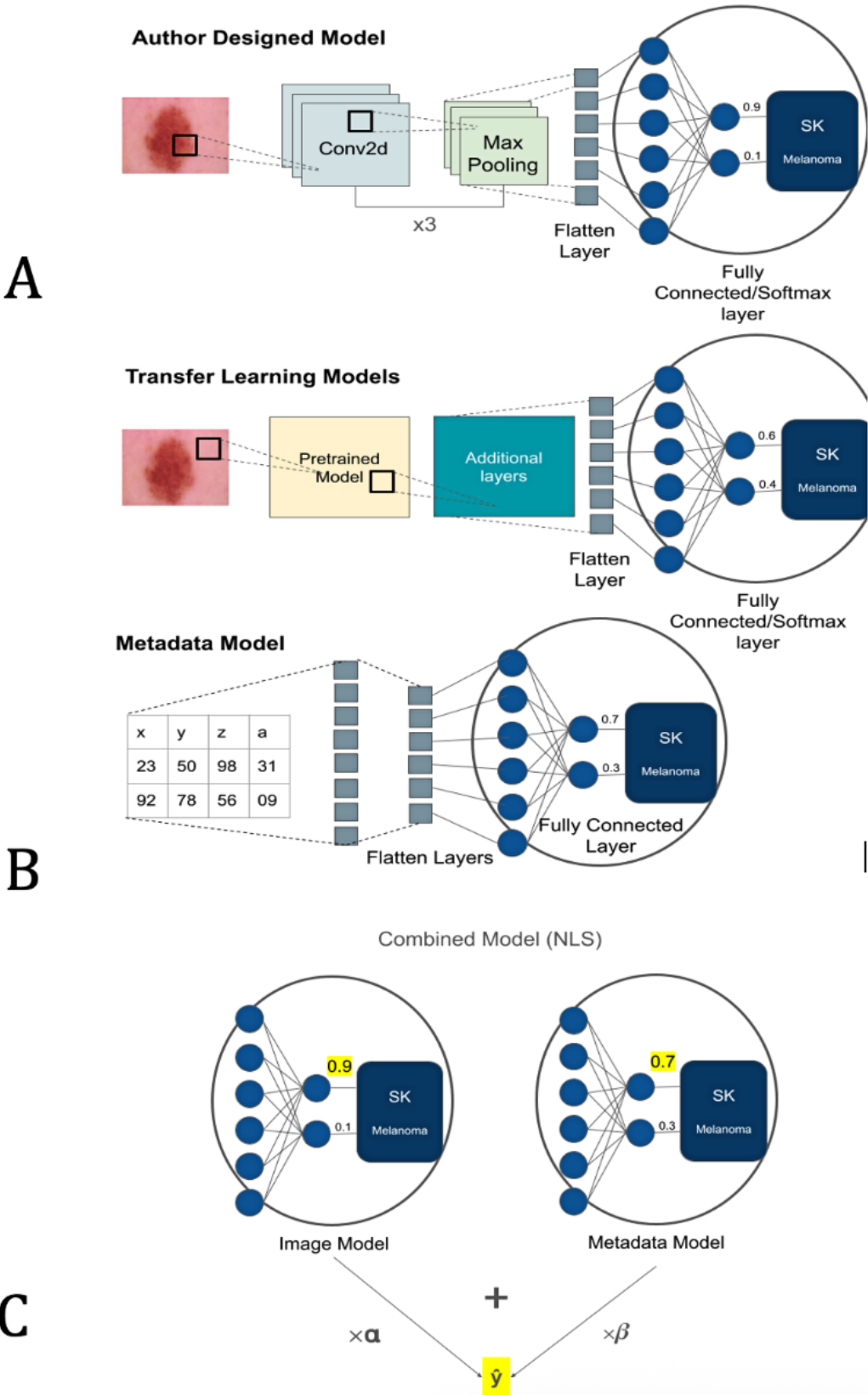
## **Data Preparation and Cleaning**

Deduplication based on lesion ID was performed to prevent train and test overlap due to the presence of preaugmented images. Using the Python package TensorFlow, the data was split into train (70% or 1547/2210 images), test (10% or 221/2210 images), and validation (20% or 442/2210 images) and then into batches to allow for parallel processing. All splits of data were then augmented and normalized to reduce overfitting and ensure equal scaling of pixel values.

## **Build and Train Image Models**

Four image models were developed as depicted in [Figure 1](#): an author-designed model and 3 transfer learning models. The author-designed model contained 3 convolutional layers with max pooling layers following each one, one flatten, and 2 dense layers. Convolutional layers help with extracting features from the image by applying certain weights to them, and max pooling layers assist in this by performing dimensionality reduction on the convolution layer output. Flatten layers once again change the dimensions, and dense layers help with forming global connections between the learned input. The output of this model was determined by the SoftMax layer, which generates a probability of the input belonging to the malignant class. The transfer learning models include pretrained ResNet50, InceptionV3, and VGG16, which were frozen to keep existing memory, and additional trainable layers were added to fine-tune the overall system. Dropouts of 0.3 and L2 weights of 0.01 were used to attempt to mitigate overfitting. All models were run for the same number of epochs, and the run time per epoch was recorded. A larger time was spent on training the transfer learning models because they have more convolutional layers and therefore take longer to output a feature map from each layer.

**Figure 1.** Architecture of the author designed and transfer learning models. (A) describes the architecture of the image model with three convolutional layers and transfer learning layers, (B) describes the metadata model for processing structured data, and (C) outlines the NLS method used to combine the predictions from each model.

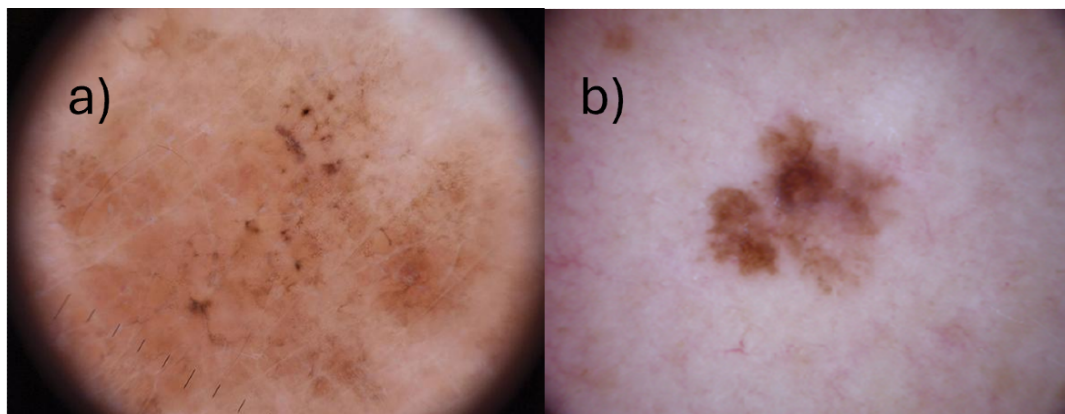


## Improving Image Model Accuracy

To improve model accuracy, further data cleaning was performed. Train, test, and validation datasets were manually parsed through with the following metrics in mind: <72 DPI

and <600 x 800 px with visuals depicted in Figure 2. 8.3% (183/2210) of the data was eliminated this way and rerun with the same model structure to analyze the effect of image quality on model accuracy.

**Figure 2.** Examples of faulty and good images. Specific metrics were used for data cleaning. (A) Faulty images, with dots per inch (DPI) <72 and approximate zoom <600x800px were removed. (B) Good images, with DP>72 and approximate zoom 600 x 800px were kept. 200 out of 2400 faulty images were removed from the dataset using these specifications, 100 from each class.



## Metadata Cleaning and Run

After optimizing and validating the image model, the metadata was cleaned and split similar to the image data. A train, test, and validation dataset was built that matched that of the images using matching image IDs to ensure controlled training. Categorical columns were made numerical through manual mapping, and the data was standardized using a built-in package called StandardScaler. A simple model architecture with only dense layers was used as visual patterns are not necessary for structured data. However, even without convolutional layers, global knowledge pattern formation was achieved through dense (fully connected) layers that connected each “node,” or learned pattern, to each other.

## Combining the Two: Non-Linear Least Squares (NLS) Regression

The image and metadata model output SoftMax probabilities for each class (melanoma and seborrheic keratosis). The NLS regression method was applied to determine optimal weights for combining each model’s prediction. The coefficients were determined through analysis of image and metadata outputs for the training dataset.

$$0.75x_1 + 0.25x_2 = \hat{y} \quad (1)$$

The above equation, outputted from the NLS function, describes the weights applied to both image ( $x_1$ ) and metadata

( $x_2$ ) model outputs to achieve an ideal accuracy.  $\hat{y}$  represents the combined prediction, with values >0.5 being classified as malignant (melanoma) and values <0.5 being classified as seborrheic keratosis.

## Ethical Considerations

No human participants were involved in this research. All data used in this research was obtained from the HAM10000 dataset, an open source and publicly available dataset. The authors of the HAM10000 dataset state that data sources were approved by the ethics committee at the Medical University of Vienna (Protocol No. 1804/2017) and the institutional ethics board at the University of Queensland (Protocol No. 2017001223).

## Results

### Comparing Model Accuracies

The simple model had the highest accuracy of 83.4% (369/442 images classified correctly) on validation data. All transfer learning models had high training accuracy but low validation accuracies, showing signs of overfitting. With the number of epochs in training constant, the transfer learning models show significantly more training time than the self-built model, as well as depicted in Table 2.



**Table 2.** Comparing model accuracies.

Model name	Training accuracy, N=1547, n (%)	Validation accuracy, N=442, n (%)	Number of epochs	Run time per epoch
ResNet50	1526 (98.65)	240 (54.29)	500	229 seconds
InceptionV3	1512 (97.75)	296 (67.04)	500	315 seconds
VGG16	1524 (98.52)	270 (61.13)	500	401 seconds
Self-Built Model (pre-data cleaning)	1242 (80.27)	348 (78.62)	500	2 seconds
Self-Built Model (post-data cleaning)	1488 (96.2)	369 (83.4)	500	2 seconds

## ROC Curves

ROC (receiver operating characteristic) curves were plotted as another method of showcasing the accuracy of each model. The self-built model had the highest AUC of 83% (369/442 images classified correctly) on validation data, consistent with the self-built model accuracy from the validation data. This model reaches its highest true-positive rate while achieving lower false-positive rates than the transfer learning models. The transfer learning models had significantly lower AUCs with ResNet50 approaching the random guess line.

## Validating Image Model

Saliency maps on test data illustrate the region of interest identified by different convolutional neural network architectures, allowing for greater model reliability and interpretability. They were generated from the last convolution layer, to help visualize which regions of an image are important for final classification. Each model demonstrates varying focus patterns, reflecting differences in feature extraction and attention and accounting for varying accuracies across all models.

## Combined Model: Confusion Matrices

The classification performance of the image, metadata, and combined models was evaluated through confusion matrices reflecting sensitivity and specificity. The image-based model shows a balanced distribution of correct classifications, achieving a true-negative rate of 42% (93/221) and a true-positive rate of 41% (91/221) on test data. The metadata-based model exhibited lower overall performance. When both image and metadata inputs were integrated, better performance was achieved across all metrics.

## Discussion

### Comparing Model Accuracies

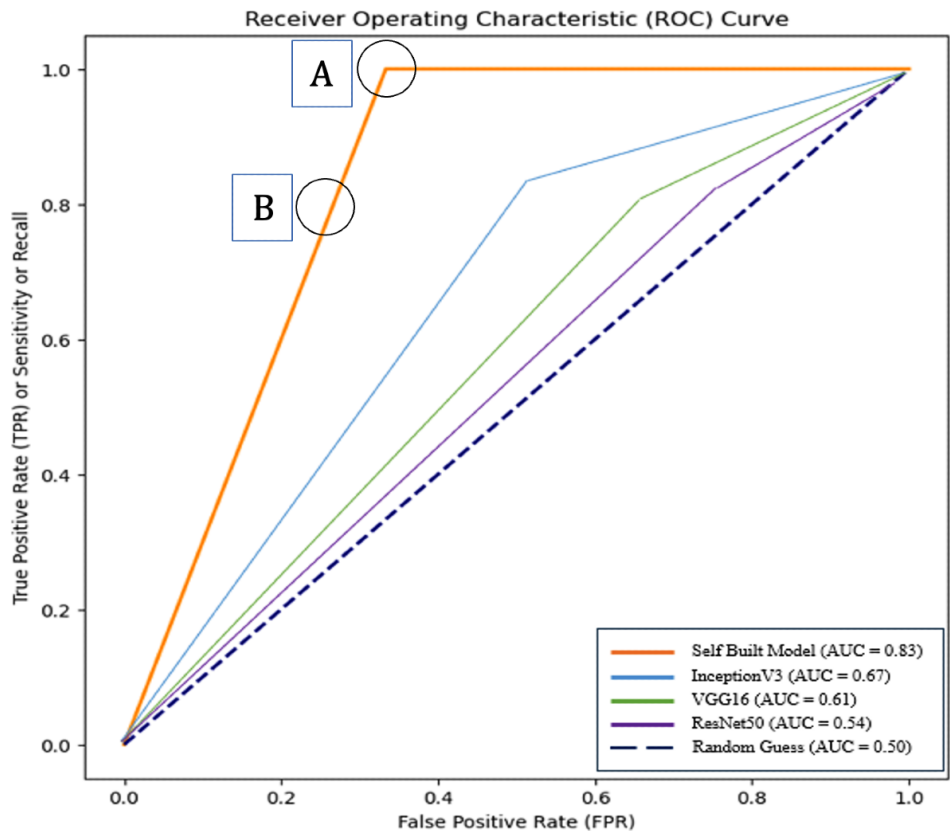
Contrary to what was expected, the transfer learning models appear to perform worse than the author-designed model. The differences in model accuracy can be attributed to model architecture, particularly the number of convolutional layers. Transfer learning models have far more convolutional layers

than the self-built model (ie, ResNet50 has 50 convolutional layers while the self-built model has only 3). As the number of convolutional layers increases, the ability of the model to detect more complex and finer features increases. Therefore, the transfer learning models are more susceptible to overfitting as they can detect more minute details like hair and wrinkles. This accounts for the overfitting occurring in the transfer learning models as seen in the large difference between training and validation accuracy.

ResNet50 differs from the author-constructed model as it contains a residual layer that directly connects the output layers to the input layers as opposed to “stacking” them. The author constructed model optimizes the accuracy by using backpropagation, where the gradients used to determine the minimum loss value are calculated using the chain rule. Rather than using the chain rule, ResNet avoids the subsequent derivation between each layer and instead connects each output to the input. While this is important for models with a large number of convolutional layers, the author constructed model only contains 3 convolutional layers, so the effect of chain rule is less amplified, deeming the residual layer unnecessary. Inception V3 differs from the author-constructed model as it uses parallel convolutional layers to analyze a wider feature range in the input images. However, because melanoma is often centered in one specific region and is attributed with a set of consistent features defined by the ABCDE rule, the detection of too many features is harmful. VGG16 specializes in using smaller kernel strides to center on more minute features, which can lead to overfitting in this situation as small details in the skin are not vital and sometimes confusing in making a classification. While past studies have shown that ResNet50 and InceptionV3 perform well in these applications, the ability of the simple model to generalize to this particular problem makes it better compared to these previous approaches.

In real-world deployment, frontend image capture tools [27] will ensure image inputs conform to these predetermined metrics as shown in Figure 3, thereby increasing usability of the model. Upon deployment of this model to the primary care office, physicians are further advised to take good quality images of their patients' lesions to ensure accurate diagnosis.

**Figure 3.** ROC curves: receiver operating characteristic curves plot the false-positive rate versus the true-positive rate. The ideal point on the curve is (0,1). FPR: false-positive rate; ROC: receiver operating characteristic; TPR: true-positive rate.



ROC Curves

ROC curves in Figure 3 are used to determine a cutoff point that optimizes the sensitivity and specificity of a specific test [28]. In medical applications, this is especially important since false-negative results could be life-threatening. As the false-negative rate is a direct function of the true-positive rate, in order to lower the false-negative rate, the true-positive rate must be increased, even if it comes at the expense of the

false-positive rate. Consequently, point A would be preferred to point B.

In addition, ROC curves can also be a measure of accuracy through the AUC depicted in the key shown in Figure 3. As the models get worse (as shown by the accuracies in Table 3), the ROC curve moves further away from the ideal point (0,1) and towards the random guess line [29].

**Table 3.** The final testing accuracy of the combined model is significantly higher than the existing accuracies from the literature review, which averaged around 70%.

Model type	Sensitivity	Specificity	Testing accuracy
Image	0.82	0.84	0.83
Metadata	0.76	0.52	0.64
Combined	0.875	0.875	0.875

Heatmaps

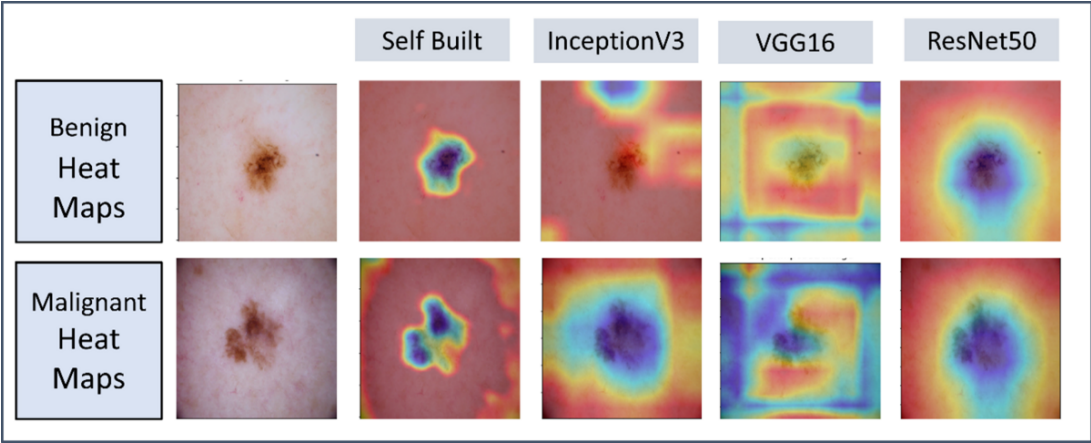
A major problem in artificial intelligence models today is lack of interpretability [30]. Artificial intelligence is often referred to as a “black box” with limited explainability regarding its decisions [31]. However, Figure 4 allows users to “see through the eyes” of the model through heatmaps.

The author-designed model has a more “fixed” area of concentration as opposed to the other three transfer learning models. However, unlike InceptionV3, ResNet50 offers human interpretability and appears to follow the pattern

presented in the author-designed model to a limited extent. However, it fails to capture differences between benign and malignant lesions as shown in the similar weight distributions between the 2 classes.

As shown in Figure 4, the author-designed model that performed the best appears to primarily look at the differences in border between the two lesions, connecting back to the ABCDE method used by dermatologists for clinical diagnosis [32]. This gives the model more reliability, as it is dissecting the image similar to how a dermatologist would.

**Figure 4.** Comparison of Grad-CAM heat maps for benign and malignant skin lesions using different CNN architectures (Self-Built, InceptionV3, VGG16, ResNet50) for one example from each class.

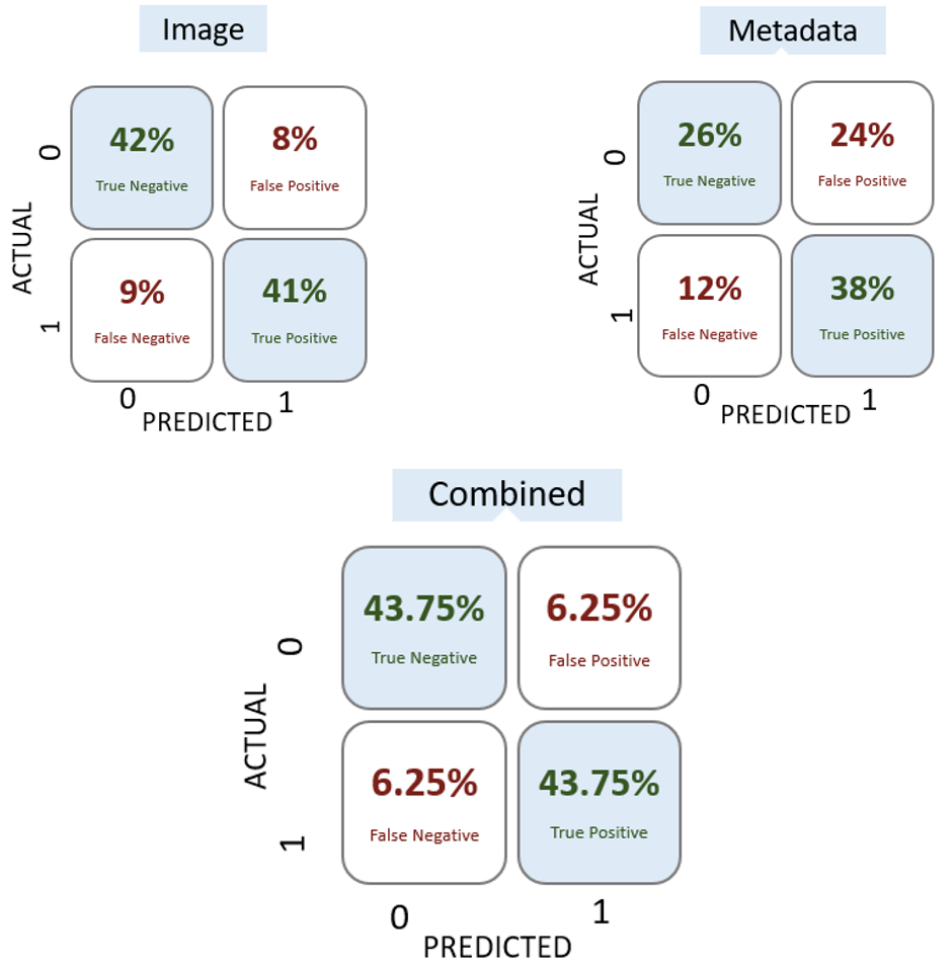


**Confusion Matrices**

Referring back to Figure 5, as the true-positive rate increases, it does so at the expense of the false-positive rate until a certain saturation point (A). Therefore, the 9% false-negative rate shown on the image confusion matrix (top left of Figure 5) can only be reduced at the expense of increasing the false-positive rate. The incorporation of the metadata adds

critical heterogeneous information enabling the joint system to achieve a higher true-positive rate (lower false-negative rate) while simultaneously lowering the false-positive rate as shown in Figure 5 below. This thereby allows for the significant improvement in overall model accuracy as shown in Table 2.

**Figure 5.** Confusion matrices depicting various metrics (false-positive and false-negative rates) as related to the ROC curve. A cutoff of 0.5 was used for prediction, with predictions over 0.5 being classified as melanoma. ROC: receiver operating characteristic.





## Comparison to Past Studies

Past work on dermatological applications of machine learning is compiled in [Table 1](#), showing an accuracy ranging around 75%. The AUC of the transfer learning models (ResNet50, InceptionV3, and VGG16) matches that of past experiments. This study showcases an improvement of overall accuracy through the incorporation of additional metadata as well as constructing a simple model with fewer convolutional layers. These two approaches were successful in increasing the overall accuracy to 87.5% (194/221), showing promising implications for a multimodality approach to deep learning in dermatology.

## Applications and Improvements for Future Studies

Out of sample testing will be used through the deployment of this model in local hospital settings in cases with known diagnoses to ensure model feasibility and usability outside the controlled environment of HAM10000. To achieve this, this model will be employed in local dermatological centers and results will be compared against dermatologist-determined diagnosis to determine out-of-sample accuracy.

Cross-validation using different train-test-validation splits will be tested to increase the confidence of the model with access to more storage and compute units. To make this possible, a resource-efficient approach to training a convolutional neural network is necessary as images occupy a large amount of storage space.

Currently, the model does poorly when presented with patients aged 40 and younger as well as lesions present on curved areas of the body such as the eyelids. This is due to the lack of data from these demographics and areas, forcing the model to use generalized patterns to predict on these data points. Access to more granular metadata from younger patients and certain areas of the model can help address this issue. However, given the predominance of melanoma in older age groups, the authors believe this to be a natural obstacle of diagnosis in unusual populations.

As machine learning is a rapidly growing field, many new techniques can be used to improve the accuracy of the model. Combining metadata and image model predictions can be done through deep learning rather than regression, thereby enabling end-to-end joint training of the system to improve accuracy. Alternate architecture designs that combine image and metadata at the input or intermediate layers can also be explored. Additionally, using more granular metadata with less repetitions and more variations (eg, more data on different ages) can decrease the possibility of overfitting.

Using text data can also be a major change to this experiment. While this study only used structured data (patient metadata) and image data, in the real hospital setting, anecdotes, pain scale, lesion progression, and other descriptive factors can greatly influence a doctor when making a diagnostic decision. Using these records and combining them into the deep learning network through natural language processing can improve robustness and applicability of this model to the real world.

In order to make the application useful to a wider range of common citizens, making the model more robust by supporting a multi-way classification will allow older patients to use it in the home setting. Training the model on multiple types of lesions will motivate a more patient-friendly output as simply differentiating between benign and malignant eliminates the need to narrow down lesion possibilities.

## Conclusion

In this manuscript, we introduce a multimodal technique that employs heterogeneous forms of data to produce a probability of the lesion belonging to either class. The model expands upon current model architectures and is adapted and trained for the specific problem at hand. This strategy can be applied to a multitude of medical applications in addition to current studies to provide a more comprehensive diagnosis of a certain disease through the addition of multiple data modalities.

## Acknowledgments

We gratefully acknowledge all data contributors, ie, the Authors and Compilers of the HAM10000 dataset, and the submitting institutions that made this data publicly available.

## Data Availability

All data generated or analyzed during this study are included in this published article [18].

## Authors' Contributions

NV performed data curation, formal analysis, investigation, methodology, software, validation, visualization, and writing the original draft. NV and KR were involved in conceptualization and reviewed the manuscript.

## Conflicts of Interest

None declared.

## References

1. Okobi OE, Abreo E, Sams NP, et al. Trends in Melanoma incidence, prevalence, stage at diagnosis, and survival: an analysis of the United States Cancer Statistics (USCS) Database. *Cureus*. Oct 2024;16(10):e70697. [doi: [10.7759/cureus.70697](https://doi.org/10.7759/cureus.70697)] [Medline: [39493095](https://pubmed.ncbi.nlm.nih.gov/39493095/)]

2. Waseh S, Lee JB. Advances in melanoma: epidemiology, diagnosis, and prognosis. *Front Med (Lausanne)*. 2023;10:1268479. [doi: [10.3389/fmed.2023.1268479](https://doi.org/10.3389/fmed.2023.1268479)] [Medline: [38076247](https://pubmed.ncbi.nlm.nih.gov/38076247/)]
3. Ye Q, Chen KJ, Jia M, Fang S. Clinical and histopathological characteristics of tumors arising in seborrheic keratosis: a study of 1365 cases. *Ther Clin Risk Manag*. 2021;17:1135-1143. [doi: [10.2147/TCRM.S316988](https://doi.org/10.2147/TCRM.S316988)] [Medline: [34737570](https://pubmed.ncbi.nlm.nih.gov/34737570/)]
4. Wollina U. Recent advances in managing and understanding seborrheic keratosis. *F1000Res*. 2019;8:1520. [doi: [10.12688/f1000research.18983.1](https://doi.org/10.12688/f1000research.18983.1)] [Medline: [31508199](https://pubmed.ncbi.nlm.nih.gov/31508199/)]
5. Roh NK, Hahn HJ, Lee YW, Choe YB, Ahn KJ. Clinical and histopathological investigation of seborrheic keratosis. *Ann Dermatol*. Apr 2016;28(2):152-158. [doi: [10.5021/ad.2016.28.2.152](https://doi.org/10.5021/ad.2016.28.2.152)] [Medline: [27081260](https://pubmed.ncbi.nlm.nih.gov/27081260/)]
6. Moscarella E, Brancaccio G, Briatico G, Ronchi A, Piana S, Argenziano G. Differential diagnosis and management on seborrheic keratosis in elderly patients. *Clin Cosmet Investig Dermatol*. 2021;14:395-406. [doi: [10.2147/CCID.S267246](https://doi.org/10.2147/CCID.S267246)] [Medline: [33953590](https://pubmed.ncbi.nlm.nih.gov/33953590/)]
7. Heistein JB, Acharya U, Mukkamalla SKR. Malignant Melanoma. StatPearls Publishing; 2024. URL: <https://www.ncbi.nlm.nih.gov/books/NBK470409> [Accessed 2025-08-08]
8. Lowell BA, Froelich CW, Federman DG, Kirsner RS. Dermatology in primary care: Prevalence and patient disposition. *J Am Acad Dermatol*. Aug 2001;45(2):250-255. [doi: [10.1067/mjd.2001.114598](https://doi.org/10.1067/mjd.2001.114598)] [Medline: [11464187](https://pubmed.ncbi.nlm.nih.gov/11464187/)]
9. Scolyer RA, Rawson RV, Gershenwald JE, Ferguson PM, Prieto VG. Melanoma pathology reporting and staging. *Mod Pathol*. Jan 2020;33(Suppl 1):15-24. [doi: [10.1038/s41379-019-0402-x](https://doi.org/10.1038/s41379-019-0402-x)] [Medline: [31758078](https://pubmed.ncbi.nlm.nih.gov/31758078/)]
10. Davis LE, Shalin SC, Tackett AJ. Current state of melanoma diagnosis and treatment. *Cancer Biol Ther*. Nov 2, 2019;20(11):1366-1379. [doi: [10.1080/15384047.2019.1640032](https://doi.org/10.1080/15384047.2019.1640032)]
11. Liutkus J, Kriukas A, Stragyte D, et al. Accuracy of a smartphone-based artificial intelligence application for classification of melanomas, melanocytic nevi, and seborrheic keratoses. *Diagnostics (Basel)*. Jun 21, 2023;13(13):2139. [doi: [10.3390/diagnostics13132139](https://doi.org/10.3390/diagnostics13132139)] [Medline: [37443533](https://pubmed.ncbi.nlm.nih.gov/37443533/)]
12. Wei ML, Tada M, So A, Torres R. Artificial intelligence and skin cancer. *Front Med (Lausanne)*. 2024;11:1331895. [doi: [10.3389/fmed.2024.1331895](https://doi.org/10.3389/fmed.2024.1331895)] [Medline: [38566925](https://pubmed.ncbi.nlm.nih.gov/38566925/)]
13. Deshabhoina SV, Umbaugh SE, Stoecker WV, Moss RH, Srinivasan SK. Melanoma and seborrheic keratosis differentiation using texture features. *Skin Res Technol*. Nov 2003;9(4):348-356. [doi: [10.1034/j.1600-0846.2003.00044.x](https://doi.org/10.1034/j.1600-0846.2003.00044.x)] [Medline: [14641886](https://pubmed.ncbi.nlm.nih.gov/14641886/)]
14. Jeong HK, Park C, Henao R, Kheterpal M. Deep learning in dermatology: a systematic review of current approaches, outcomes, and limitations. *JID Innov*. Jan 2023;3(1):100150. [doi: [10.1016/j.xjidi.2022.100150](https://doi.org/10.1016/j.xjidi.2022.100150)] [Medline: [36655135](https://pubmed.ncbi.nlm.nih.gov/36655135/)]
15. Jojoa Acosta MF, Caballero Tovar LY, Garcia-Zapirain MB, Percybrooks WS. Melanoma diagnosis using deep learning techniques on dermatoscopic images. *BMC Med Imaging*. Dec 2021;21(1). [doi: [10.1186/s12880-020-00534-8](https://doi.org/10.1186/s12880-020-00534-8)]
16. Hollmann N, Müller S, Purucker L, et al. Accurate predictions on small data with a tabular foundation model. *Nature New Biol*. Jan 2025;637(8045):319-326. [doi: [10.1038/s41586-024-08328-6](https://doi.org/10.1038/s41586-024-08328-6)] [Medline: [39780007](https://pubmed.ncbi.nlm.nih.gov/39780007/)]
17. Ahmed A, Ahmad H, Khurshid M, Abid K. Classification of skin disease using machine learning. *VFAST trans softw eng*. 2023;11(1):109-122. [doi: [10.21015/vtse.v11i1.1204](https://doi.org/10.21015/vtse.v11i1.1204)]
18. Krones F, Walker B, Mahdi A, Kiskin I, Lyons T, Parsons G. Dual bayesian resnet: a deep learning approach to heart murmur detection. *Comput Cardiol Conf*. 2022;49. [doi: [10.22489/CinC.2022.355](https://doi.org/10.22489/CinC.2022.355)]
19. Wang Z, Gao C, Xiao C, Sun JM. MediTab: scaling medical tabular data predictors via data consolidation, enrichment, and refinement. Presented at: Thirty-Third International Joint Conference on Artificial Intelligence {IJCAI-24}; Aug 3-9, 2024; Jeju, South Korea. URL: <https://www.ijcai.org/proceedings/2024> [doi: [10.24963/ijcai.2024/670](https://doi.org/10.24963/ijcai.2024/670)]
20. Khademi S, Hajiakhondi-Meybodi Z, Vaseghi G, Sarrafzadegan N, Mohammadi A. FH-tabnet: multi-class familial hypercholesterolemia detection via a multi-stage tabular deep learning network. Presented at: 2024 32nd European Signal Processing Conference (EUSIPCO); Aug 26-30, 2024; Lyon, France. URL: <https://eurasip.org/Proceedings/Eusipco/Eusipco2024/pdfs/0001416.pdf> [doi: [10.23919/EUSIPCO63174.2024.10715254](https://doi.org/10.23919/EUSIPCO63174.2024.10715254)]
21. Gahrooei MR, Yan H, Paynabar K, Shi J. Multiple tensor-on-tensor regression: an approach for modeling processes with heterogeneous sources of data. *Technometrics*. Apr 3, 2021;63(2):147-159. [doi: [10.1080/00401706.2019.1708463](https://doi.org/10.1080/00401706.2019.1708463)]
22. Aoki Y, Hayami K, Toshimoto K, Sugiyama Y. Cluster Gauss-Newton method for finding multiple approximate minimisers of nonlinear least squares problems with applications to parameter estimation of pharmacokinetic models. National Institute of Informatics; 2020. URL: [https://www.nii.ac.jp/TechReports/public\\_html/20-002E.pdf](https://www.nii.ac.jp/TechReports/public_html/20-002E.pdf) [Accessed 2024-04-12]
23. Tabatabai MA, Kengwoung-Keumo JJ, Eby WM, et al. A New Robust Method for Nonlinear Regression. *J Biom Biostat*. 2014;5(5):211. [doi: [10.4172/2155-6180.1000211](https://doi.org/10.4172/2155-6180.1000211)] [Medline: [26185732](https://pubmed.ncbi.nlm.nih.gov/26185732/)]
24. Xia Z, Yao Z, Wu Y, et al. Comparative between linear least-squares and nonlinear least-squares computation method for regional and voxelized quantitative analysis in total-body dynamic 18F-FDG PET. *J Nucl Med*. 2024;65(supplement 2):241042-241042. URL: [https://jnm.snmjournals.org/content/65/supplement\\_2/241042](https://jnm.snmjournals.org/content/65/supplement_2/241042) [Accessed 2025-08-08]

25. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature New Biol.* Feb 2, 2017;542(7639):115-118. [doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)]
26. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data.* 2018;5(1). [doi: [10.1038/sdata.2018.161](https://doi.org/10.1038/sdata.2018.161)]
27. Goh HA, Ho CK, Abas FS. Front-end deep learning web apps development and deployment: a review. *Appl Intell.* Jun 2023;53(12):15923-15945. [doi: [10.1007/s10489-022-04278-6](https://doi.org/10.1007/s10489-022-04278-6)]
28. Unal I. Defining an optimal cut-point value in roc analysis: an alternative approach. *Comput Math Methods Med.* 2017;2017:3762651. [doi: [10.1155/2017/3762651](https://doi.org/10.1155/2017/3762651)] [Medline: [28642804](https://pubmed.ncbi.nlm.nih.gov/28642804/)]
29. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J Anesthesiol.* Feb 2022;75(1):25-36. [doi: [10.4097/kja.21209](https://doi.org/10.4097/kja.21209)] [Medline: [35124947](https://pubmed.ncbi.nlm.nih.gov/35124947/)]
30. Ennab M, McHeick H. Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions. *Front Robot AI.* 2024;11:1444763. [doi: [10.3389/frobt.2024.1444763](https://doi.org/10.3389/frobt.2024.1444763)] [Medline: [39677978](https://pubmed.ncbi.nlm.nih.gov/39677978/)]
31. A. S, R. S. A systematic review of explainable artificial intelligence models and applications: recent developments and future trends. *Decision Analytics Journal.* Jun 2023;7:100230. [doi: [10.1016/j.dajour.2023.100230](https://doi.org/10.1016/j.dajour.2023.100230)]
32. Duarte AF, Sousa-Pinto B, Azevedo LF, et al. Clinical ABCDE rule for early melanoma detection. *Eur J Dermatol.* Dec 1, 2021;31(6):771-778. [doi: [10.1684/ejd.2021.4171](https://doi.org/10.1684/ejd.2021.4171)] [Medline: [35107069](https://pubmed.ncbi.nlm.nih.gov/35107069/)]

## Abbreviations

**AUC:** area under the curves  
**ID3:** Iterative Dichotomiser 3  
**NLS:** nonlinear least squares regression  
**ROC:** receiver operating characteristic

*Edited by Khaled El Emam; peer-reviewed by Anil Kumar Vadathya, Mirsaeed Abdollahi; submitted 16.09.2024; final revised version received 21.05.2025; accepted 05.07.2025; published 13.08.2025*

### Please cite as:

Vivek N, Ramesh K

*Deep Learning Multi-Modal Melanoma Detection: Algorithm Development and Validation*

*JMIR AI* 2025;4:e66561

URL: <https://ai.jmir.org/2025/1/e66561>

doi: [10.2196/66561](https://doi.org/10.2196/66561)

© Nithika Vivek, Karthik Ramesh. Originally published in *JMIR AI* (<https://ai.jmir.org>), 13.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR AI*, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.ai.jmir.org/>, as well as this copyright and license information must be included.