A Comprehensive Review of Artificial Intelligence Algorithms and Applications in Melanoma Diagnosis

Doruk Yalçın
Department of Computer Engineering
Koç University
Istanbul, Turkiye
dyalcin21@ku.edu.tr

***Abstract*—Melanoma, a lethal form of skin cancer, poses a significant health risk worldwide with rising incident rates. The usage of Artificial Intelligence (AI) tools in dermatology for melanoma detection can help curb the demand for accurate and efficient diagnosis of the disease. This review examines the current state of AI, Machine Learning (ML), and Deep Learning (DL) applications in the identification of melanomas through the analysis of various studies that have demonstrated the potential of these technologies that could outperform traditional methods and provide life-saving diagnoses. The primary usage of Convolutional Neural Networks (CNNs) has the potential to completely revolutionize the field of dermatological diagnosis.**

Keywords—AI in dermatology, Convoluted neural networks, Deep learning, Machine learning, Melanoma, Predictive modeling

I. Artificial Intelligence Terminology And Important Models

A. Artificial Intelligence

Artificial Intelligence is the broadest definition of a computer mimicking human behavior. Any technique or model that allows for human-like decision-making can be defined as Artificial Intelligence. Techniques such as "If-Then" rules and "decision trees" allow for the algorithms to perform tasks that would otherwise need human intelligence.

B. Machine Learning

Machine Learning is a subset of artificial intelligence. Machine learning focuses on learning from data and identifying patterns exhibited in data distributions. The interpretation of data allows for algorithms to make decisions and predictions. Machine learning techniques include Linear regression, Support Vector Machines (SVM), classification, clustering, and more.

1) Linear Regression: Linear Regression is the statistical process of modeling the relationship between two variables, one dependent and one independent. Linear regression can also be used to model the relationship between multiple independent variables. *[1]*

2) Decision Trees: A decision tree is a hierarchical model where through internal decision nodes and terminal leaves, data points are labeled into discrete outcomes. Decision trees can be used for both classification and regression. Given an input, a test is applied to the input and a branch is selected depending on the outcome until the input reaches a leaf node. *[2]*

3) Random Forest: A decision forest is the combination of multiple decision trees. The training of multiple decision trees on a random subset of the given training data and then combining the predictions of these decision trees results in the random forest model. *[3]*

4) Support Vector Machines: Support Vector Machines (SVM) is a method used for classification, regression, and outlier detection. SVMs are used to find the hyperplane that best divides sets of data into classes. SVMs can be used for both linear and non-linear data models.

5) k-Nearest Neighbor: The k-Nearest Neighbors (k-NN) algorithm is a supervised learning method used primarily for classification. An input sample is classified by its proximity to a plurality of neighbors with the sample assigned to the class most common amongst its closest neighbors.

6) k-Means Clustering: The k-Means Clustering algorithm is an unsupervised learning method used for classification. n data points are clustered into k distinct clusters based on their proximity to the nearest mean of neighboring clusters. The process is repeated until all data points are assigned to stable clusters.

7) Neural Networks: Neural Networks are models inspired by the structure of the brain. Neural networks consist of layers of nodes similar to neurons with each node connected to another node in varying weights. The weights of nodes are adjusted according to the accuracy of the output of the model. Neural networks allow for feedback-adjustable models and form the basis of Deep Learning.

8) Principal Component Analysis: Principal Component Analysis is a statistical technique used to simplify complex, high-dimensional data while keeping data trends and patterns. Dimensions with high variance are kept while a number of dimensions with minimal variance are removed.

C. Deep Learning

Deep Deep Learning is a subset of machine learning. Through the construction of neural networks, deep learning models perform particularly well on interpreting large amounts of data. Neural networks, inspired by the functioning of the human brain, are used in image recognition, speech recognition, and natural language processing

1) Programming Convolutional Neural Networks: Python has become the preferred programming language used for the development of AI models in all areas of application. The availability of Python libraries such as TensorFlow and PyTorch provide pre-compiled routines and functions of most Machine Learning algorithms and deep learning models and allow for ease of use. *[4]* TensorFlow, developed by Google, and PyTorch, developed by Meta, are commonly used in most AI programs. The release of Keras, a TensorFlow application programming interface (API), in

Written: 15.03.2024 | Reviewed: 22.03.2024 | DOI: 10.5281/zenodo.10963190

2015 also provides an easy-to-use interface for the construction of deep learning models. Another programming language, MATLAB, is also popular amongst researchers and offers numerous toolboxes designed for data analysis and the construction of neural networks. The prototyping of algorithms that require extensive mathematical computations are often constructed in MATLAB.

2) Convolutional Neural Network Architectures: Deep Learning architectures form the basis of deep learning models and are frameworks designed to dictate how data is processed within neural networks. These architectures are categorized into several types such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory Networks. For applications in image recognition and classification in medical diagnosis, CNNs are used regularly. Deep Convolutional Neural Networks (DCNNs) are designed to analyze high-dimensional datasets by its ability to learn from raw data, eliminating the need for manual feature extraction. This ability is essential for image analysis where DCNNs can identify subtle and complex features in large amounts of images and data, such as a dataset of dermoscopic images.

a) VGG16: VGG16 is an object recognition and classification algorithm, introduced in 2014 from Oxford University, with high success rates and current popularity. *[5]*The VGG16 model includes 16 learnable parameter layers, also known as weight layers. The VGG16 model is pre-trained on the ImageNet dataset which is an image dataset including 14 million images, designated in 1000 classes. This pre-training allows for the model to be conveniently tweaked for specific usage. The implementation of VGG16 is primarily completed using Python and libraries such as TensorFlow and Keras provide pre-built functions for its implementation in projects.

b) Xception: Extreme Inception, popularly known as Xception, is a CNN architecture introduced in 2017. The model uniquely introduces the depthwise separable convolution modification which allows for spatial convolution to be separately executed for each channel of the input. Depthwise separable convolution significantly reduces the number of parameters and the computational cost of the model in comparison to other CNNs. The usage of Xception is advantages in cases with large image datasets which would otherwise require significant computation cost and time. Xception is commonly used in Python using the TenserFlow and Keras libraries; however, MATLAB has also recently integrated Xception into its system.

c) DenseNet: Densely Connected Convolutional Networks, known as DenseNet, is a unique CNN architecture introduced in 2017. DenseNet is unique from other CNNs in its architecture where each layer is directly connected to every other layer in a feed-forward model. *[6]*The DenseNet architecture contains "dense blocks" where each layer receives additional inputs from all preceding layers and each layer passes its own feature maps to proceeding layers. This model allows for effective feature propagation where emphasis on fine details is required on the analyzing of datasets. In the DenseNet model, as every layer has access to the gradient from the original loss function, the vanishing gradient problem is significantly reduced; this allows for each layer to continue its relevance in the training of the CNN model.

d) MobileNet: MobileNet is a CNN model designed for use in mobile applications and embedded visual systems. Introduced by Google in 2017, MobileNet's main goal is to provide a lightweight CNN model for use in systems with limited computational power. *[7]* Similar to Xception, MobileNet also employs the depthwise convolution method where the convolution operations are split into two layers: a depthwise convolution and a pointwise convolution. This approach reduces the total computation needed for the model to work. Code for the MobileNet architecture can be written using Python libraries such as TensorFlow and PyTorch.

II. Machine Learning Applications In Dermatology

Dermatology, the branch of medicine that focuses on the skin, hair, and nails plays a crucial role in the overall healthcare system and the wellbeing of the general populace. Skin diseases are among the most common of all health concerns and affect millions globally. These conditions can vary between mild and temporary to severe and chronic. Dermatologic concerns can significantly reduce the quality of life of patients both physically and mentally. The visibility of dermatologic diseases has the added effect of causing psychological distress, contributing to the need for effective and accessible care from reliable sources. As the number of skin conditions around the world increase due to reasons such as aging populations, environmental changes, and shifts in lifestyles, the integration of AI tools is necessary to meet the demand for dermatological care.

Consistent and reliable AI tools for the diagnosis of dermatological diseases would serve both physicians and patients. For physicians, these tools would reduce misdiagnosis and increase the rate of accurate assessments. Patterns and nuances of conditions that would elude the most experienced physicians would be identified by these tools. While increasing correct diagnosis, this capability would speed up the decision making process, allowing for rapid treatment and better patient outcomes. Patients would benefit from the rapid decision making process as it would potentially allow for the early diagnosis of their skin conditions, increasing its importance in deadly diseases such as melanoma. The integration of AI into dermatological diagnosis processes would also ensure that patients in rural areas with limited healthcare access have a precautionary approach against skin diseases. Short-term advances in AI and its adoption into regular practice would serve to lift the burden off of doctors and the healthcare system by reducing the number of healthy patients under evaluation by physicians. Machine Learning models excel in uncovering patterns and trends in diverse observations of the same condition. Certain shared factors, unknown to current medical practitioners, may be uncovered in the application of Machine Learning in medicine. While such discoveries would need clinical trials and rigorous validation, they would be beneficial in assisting physicians.

The main limitation and ethical concern in the development and implementation of AI tools in dermatology, and in medicine as a whole, is in the aspect of data collection. Medicine is a delicate and private subject in most people's lives and a subject that is considered taboo for discussion, evident by the existence of the physician-patient privilege as

a legal and ethical concept in numerous countries. Machine Learning is based in statistics and the creation of advanced AI tools is fundamentally tied to the abundance of data, regardless of how well certain AI algorithms are implemented. The US, EU, and Turkey have strict regulations on unapproved data collection regarding medical conditions, medical history, and its use [8]. Research using significantly large datasets are currently only possible with the collaboration of nationwide medical operations or via obtaining data from countries with lenient data laws; both possibilities are problematic. Research conducted only in collaboration with large institutions risks the creation of the monopolization of crucial AI tools that would benefit the general public in the hands of the few, using the data of many. Research based on datasets comprising data from certain parts of the world may result in biased datasets that predominantly reflect the demographic and clinical characteristics of specific populations. This geographic and demographic concentration may influence the representation of skin types, genetic predispositions, and prevalent conditions; inhibiting the creation of a generalizable diagnosis tool. Ethically, this also raises concerns regarding privacy and consent. While some countries impose strict limitations to data gathering, others don't; this situation leads to unequal contribution and benefiting from AI advancements.

III. Machine Learning Applications In Melanoma

Melanoma is a potentially deadly and aggressive cancer that arises from genetic and metabolic disturbances in melanocytes. The incidence rate of melanomas is on the rise in developed nations, especially among populations with fair skin and its rate of incidence is increasing faster than any other type of solid tumor. [9] Melanomas can occur anywhere on the skin and unlike other skin cancers, have a high tendency to spread to other parts of the body. Due to the deadly nature of melanomas and their susceptibility to spreading, the early detection and accurate classification of melanomas is crucial in ensuring effective treatment and improving survival rates.

The diagnosis of lesions as melanomas is aided by five criteria, labeled as the ABCDEs of melanoma diagnosis. A stands for the asymmetry of the lesion, B stands for border irregularity, C stands for color diversity and variegation, D stands for diameter, with a lesion with a diameter larger than 6mm considered as risky, and E stands for evolution, change in appearance, of the lesion. [9] Machine learning application principles excel in detecting and classifying commonalities between similar cases and would excel in detecting asymmetry, border irregularity, and color variegation in lesions. Using ML algorithms to analyze and interpret existing datasets of melanoma and nevus cells would allow for models to accurately identify the distinctions in lesions in accordance with the ABCDE criterias. It's worth noting that with current datasets machine learning models would best work on the ABCD part of the acronym as the evolution (E) of the lesion is usually not present in datasets or observed by patients over a required time frame. The absence of observational, non-static data highlights an important shortcoming and characteristic of current machine learning applications in melanoma diagnosis.

Gareau et al. aims to create a machine learning model to help differentiate nevus cells from melanoma cells. [10]Using a dataset of 349- reduced from the initial 648- images obtained from dermoscopy imaging, consisting of both nevi and melanomas, Gareau et al. have trained an image classifier named Eclass. Eclass is an alternative to the leading deep learning model with a CNN architecture. The images are evaluated on 38 statistically significant imaging biomarker cues, based on cues such as color, intensity, distribution, and the ABCDs of melanoma detection; in comparison the CNN model was built on the interpretation of raw pixels from dermoscopy imaging. Eclass trained several classifier algorithms such as decision trees, SVMs, k-NN, and random forest and combined them using the median score of the classifiers, with a split of data as 75% training and 25% test. The performance of both models, Eclass and CNN, was tested using the Area Under the Receiver Operating Characteristic (AUROC) curve; a AUROC value of 0.5 indicates no diagnostic ability while a value of 1 indicates perfect diagnostic ability. The Eclass model demonstrated an AUROC value of 0.71 with a confidence interval of 0.56 to 0.85, which suggests a high ability to differentiate between nevus and melanoma with a wide variance in success in different datasets. The leading CNN model demonstrated an AUROC value of 0.67 with a narrower confidence interval of 0.63 to 0.71, which reflects lower discrimination ability but more consistent performance on different datasets. A Monte Carlo Simulation, which is a technique of repeatedly testing models on random data samples, demonstrated that the Eclass model was superior to the CNN model 74.88% of the time. The Eclass model was also significantly superior in its time complexity in its training on datasets. The Eclass model and the study demonstrated some drawbacks. The initial 648 dermoscopy images were reduced to 349 usable images. The image recognition algorithm in Eclass was unable to process images with hair or markings on skin, while CNN showed more scope in the recognition of these images. Such characteristics are common in real world scenarios and represent a common attribute in humans. The loss of 46.14% of total imaging data is a critical area of improvement for its clinical application. The diagnostic app was also tested on 10 clinicians aged 26 to 64 and received an average rating of 2.3 out of 4. It is essential to consider this grading in a nuanced manner as a small sample size is tested and it is conceivable that there might be an inherent skepticism towards the implementation of novel tools regardless of its potential success. Future evaluation including a larger and more diverse group of clinicians would serve to better represent the medical community.

Soenksen et al. aims to create a Deep Convolutional Neural Network (DCNN) to accurately identify suspicious pigmented skin lesions (SPLs). The study aims to create a DCNN model for single-lesion classification and detection of outlier lesions, named 'ugly ducklings.' [11] The model was trained from a publicly available dataset of 33,980 images from 133 individuals in Spain. Parts of the image were separated into 6 possible classes by a consensus from a committee, 3 of which represented possible skin conditions: Nonsuspicious Pigmented Lesions Type A (NSPL-A), Nonsuspicious Pigmented Lesions Type B (NSPL-B), and Suspicious Pigmented Lesions (SPLs). Images included both dermoscopy images and nondermoscopy images to

accommodate possible real-world imaging. Soenksen, et al. integrated the CNN image recognition models of VGG16 and Xception to improve their models. The integrated model's performance was validated by the consensus of three certified dermatologists. The DCNN system's outputs demonstrated significant success with the dermatologists' assessments, especially in its classification of the ugly ducklings. The VGG16-based model showcased superior performance with an Area Under the Curve micro of 0.91 and an all-class accuracy of 79.94%, showcasing success and potential for distinguishing between suspicious and non-suspicious lesions. The study also achieved the first quantifiable definition and features of the ugly duckling, measuring lesion oddness which is a significant advancement over traditional visual assessments. The study showcases some limitations and concerns. The ground truth of the performance of the model was determined to be the board of dermatologists; such a model would benefit greatly from histopathological confirmations. The dataset would benefit greatly from the inclusion of more diverse imaging sources instead of a single source, allowing for testing and training on different demographics, cameras, and settings. The study also lacks algorithmic transparency and explainability. Given the current skepticism of AI tools in medicine, the study would benefit greatly from transparency in its AI-based decision-making tool.

Pham et al. worked on creating a deep-CNN architecture using mini-batch data samples in an attempt to create an AI diagnostic model that would outperform dermatologists in the diagnosis of melanomas in given dermoscopic images. [12] The researchers experimented with several CNN architectures such as InceptionV3, ResNet50, and DenseNet169; ultimately selecting DenseNet169 as the superior model for binary classification between melanoma and nevus lesions. With a large dataset of 17,302 images, the training sets were split into several mini-batch samples which consisted of balanced examples of nevus and melanoma images for the model to train on. A custom loss-function was also used to correct any imbalances present in the data samples. The CNN model presented an AUC value of 94.4%, a sensitivity rate of 85.0%, and a specificity rate of 95.0%. A sample of 157 doctors in Germany, from university hospitals and private practices with a variety of positions in hospital hierarchies were sampled for average dermatologist success metrics. The sample exhibited an average AUC value of 67.1%, a sensitivity rate of 74.1%, and a specificity rate of 60.0%. The CNN model outperformed the sample of doctors in the distinguishing of melanoma and nevus, as seen in AUC value, in the correct identification of lesions as melanomas, as seen in the sensitivity rating, and in the correct identification of non-melanoma cells, as seen in the specificity rating. The model exhibited outstanding success in comparison to the sample of dermatologists and existing machine learning models. This model sets new benchmarks for AI success in the diagnosis of melanomas and is promising in the clinical application of machine learning models in clinical settings. The next step for this study would be to start controlled real-world clinical applications of the model as the success rate of the model is impressive by all presented metrics.

Orhan and Yavşan's study focuses on creating a successful melanoma detection model by comparing the performances of five popular CNN architectures. Using a dataset of 8,598 images containing both nevus cells and cancerous melanoma cells, the researchers trained five CNN models: AlexNet, MobileNet, ResNet, VGG16, and VGG19. [13]Dataset collection was completed by combining several different datasets, allowing for training on several different scenarios and demographics. Training models on different datasets remains crucial as melanomas present themselves in a variety of different ways, resulting in the initial problem of correctly identifying them; training an ML model on a dataset comprising only of certain scenarios may cause the model to display shortcomings in different datasets and real-world applications. The dataset was split into 60% for training, 20% for testing, and 20% for validation. After testing the MobileNet model was selected as the superior model by the researchers because of its sustainability for binary classification problems and its lightweight structure which were essential for its deployment into a mobile application. The MobileNet model displayed an accuracy rate of 84.94%, a precision rate of 79.29%, specificity rate of 64.42%. The CNN model exhibits slightly lower success rates in comparable metrics in comparison to other recent models but also integrates crucial factors to the training of the dataset that may better reflect clinical applications. The training of the model on combined datasets is a crucial reason for the decrease in accuracy; however, it is a correct reflection of real-world applications of AI models in the diagnosis of melanomas, as different occasions result in diverse imaging in both the obtaining of images and the diversity of patients.

Cozzolino et al. aims to develop a machine-learning tool capable of predicting the overall short-term survival of patients diagnosed with cutaneous malignant melanoma (CMM). [14] The study also aimed to improve upon the limitations of the American Joint Committee on Cancer (AJCC) melanoma staging systems, adding to the predefined clinical and pathological criteria. Using the dataset from the Veneto Cancer Registry (VCR) and regional health service records, the model was trained using 2600 cases of invasive CMM diagnosis from 2015 to 2017. The team used various machine learning models such as Logistic Regression, SVM, Random Forest, Gradient Boosting, kNN, and a Deep Neural Network. The models were trained and validated using five-fold cross-validation and Grid Search optimization in predicting the 3-year mortality probability with high accuracy. The models were evaluated on balanced accuracy, precision, recall, and F1 score; the Deep Neural Network and Random Forest algorithms were the most successful. The DNN model had the highest evaluation score in all metrics, indicating a strong ability to correctly classify both positive and negative cases while maintaining a high rate of true positive cases. The Random Forest model demonstrated strong recall but had a lower precision compared to the DNN model, indicating a higher rate of false positives. The DNN and Random Forest models exhibited 91.1% and 88.0% balanced accuracies, respectively. Both models were effective in identifying the most significant predictors of short-term mortality in CMM patients, which included age, tumor stage, mitotic count, and the presence of ulceration. The study also developed a web-based application

and made the tool available to physicians for clinical usage. The findings of this study demonstrated that a machine-learning tool can outperform the traditional AJCC staging system and provide a more accurate prediction of patient outcomes.

IV. Trends And Patterns

The integration of Machine Learning applications in the diagnosis of melanomas can be characterized by its emphasis on the employment of deep learning and Deep Convolutional Neural Networks (DCNNs). Recent research in this field have employed various DCNN architectures such as VGG16, Xception, DenseNet, AlexNet, MobileNet, and Resnet with VGG16 and DenseNet being architectures employed in several studies. While there exists no single DCNN architecture that has proved to be most ideal, the preference between the models relies heavily on dataset size, the computational complexity the researchers can afford, and the planned medium for which the model is prepared for.

The AUROC curve, accuracy, precision, recall, specificity, and F1 score were the commonly used metrics for which the performance of the models were evaluated. The integration of clinicians into the evaluation process also provided insight to the analysis of the model's success in comparison to physicians. The addition of experts into the research provided comparable success rates and evaluation scores for the models which were crucial in contextualizing the accomplishments of the models.

V. Discussion

The current landscape of AI in medical research significantly lacks transparency in how the models are built. Computer Science and Engineering rely heavily on any code used being made open and publicly available but most of the research avoids this process entirely by keeping the methodologies of the research abstract. Although most of the research involves a collaboration between doctors and software engineers and common courtesy in these areas may be different, it is widely expected in programming to provide the program repositories in research; usually providing the entirety of the code used. The researchers involved in the creation of these models have not provided enough information in their methodologies to allow for significant reproduction of their research, making it difficult to reproduce the results of their research. The use of artificial intelligence in medicine is controversial as is, any attempt at veiling the creation process of these models are cause for concern.

The data in which sensitive medical images are obtained are an important topic of discussion in the advancement of AI technologies in medicine. There are several algorithms and models that researchers can use in the creation of high-performing diagnosis algorithms but the main competitive edge is procured solely on the size of the dataset used to train to model. Some researchers have used large image datasets acquired from a single geographic location with relatively homogenous demographics. The usage of images captured by a single camera in a uniform demographic can not be expected to perform adequately when faced with a different camera lens on a patient with differing skin types and genetic dispositions. Research with models trained on diverse datasets exhibit lower success rates in comparison to other research which may prove to be unappealing to researchers in their display of succession rates, leading to an incentive to work with homogenous datasets. However, the problem of acquiring the images still remains a problem and will only increase in severity if researchers set out to obtain more diverse imaging. Different countries have different data privacy laws and governmental regulations may affect the medical care inequalities of certain countries relative to others, caused by a statistical bias to overperform in demographics from which the data is obtained. The advances in AI will require lawmakers to take another look in national data privacy laws, regardless of their stances in the issue.

The use of artificial intelligence in areas where experts are highly trained and respected have resulted in both reasonable and unreasonable skepticism. Gareau et al. had their AI algorithm be evaluated by a group of physicians and obtained a subpar rating by the experts. While the subpar rating could be due to an underperforming and hard to use application, more research should be directly evaluated by a diverse group of experts to uncover any potentially unreasonable skepticism and prejudice. With the rapid development of incredibly successful AI models in all aspects of life, it remains uncertain as to whether certain people will embrace the change or attempt to hinder its development. The model designed by Pham et al. proved that the creation of an AI diagnostic tool that could outperform dermatologists is possible, such research being evaluated directly by physicians would uncover the current state of affairs and outlook regarding the trust that medical experts have in the use and integration of new AI tools as helpful tools.

References

[1] E. Alpaydın, "Introduction to Machine Learning," Boston, Massachusetts, The MIT Press, 2020, p. 77.

[2] E. Alpaydın, "Introduction to Machine Learning," Boston, Massachusetts, The MIT Press, 2020, p. 213.

[3] E. Alpaydın, "Introduction to Machine Learning," Boston, Massachusetts, The MIT Press, 2020, p. 235.

[4] "Stack OverFlow," Stack OverFlow, 2023. [Online]. Available: https://survey.stackoverflow.co/2023/#overview. [Accessed 20 March 2023].

[5] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *ICLR*, 2015.

[6] G. Huang and Z. Liu, "Densely Connected Convolutional Networks," in *EEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017.

[7] A. G. Howard and M. Zhu, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.

[8] G. Greenleaf, "Global Data Privacy Laws: 89 Countries, and Accelerating," Queen Mary University of London, School of Law, London, United Kingdom, 2012.

[9] D. J. Kerr and J. F. Thompson, "Skin cancer: melanoma," in *Oxford Textbook of Oncology*, Oxford, United Kingdom, Oxford University Press, 2016, p. 674–689.

[10] D. S. Gareau and J. Browning, "Deep learning-level melanoma detection by interpretable machine learning and imaging biomarker cues," *Journal of Biomedical Optics,* vol. 25, no. 11, 2020.

[11] L. R. Soenksen, T. Kassis and S. Conover, "Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images," *SCIENCE TRANSLATIONAL MEDICINE,* no. 13, 2021.

[12] T. Pham and C. Luong, "AI outperformed every dermatologist in dermoscopic melanoma diagnosis, using an optimized deep-CNN architecture with custom mini-batch logic and loss function," *Nature Scientific Reports,* no. 11, 2021.

[13] H. Orhan and E. Yavşan, "Artificial intelligence-assisted detection model for melanoma diagnosis using deep learning techniques," *Mathematical Modelling and Numerical Simulation with Applications,* vol. 3, no. 2, p. 159–169, 2023.

[14] C. Cozzolino and A. Buja, "Machine learning to predict overall short-term mortality in cutaneous melanoma," *Discover Oncology,* vol. 14, no. 13, 2023.

[15] E. Alpaydın, "Introduction to Machine Learning Third Edition," Boston, Massachusetts, The MIT Press, 2020, p. 77.