



A REVIEW ON SKIN CANCER DETECTION MODELS USING MACHINE LEARNING

¹Afroj Ahmad, ²Mr. Ramesh Vaish, ³Ajit Priy Tripathi, ⁴Ajeet Yadav, ⁵Aman Verma

^{1,3,4,5}Student, ²Assistant Professor

Department Of Computer Science & Engineering,

Babu Banarasi Das Institute of Technology and Management, Lucknow, Uttar Pradesh, India.

Abstract: This review paper explores the use of machine learning algorithms, specifically Support Vector Machines (SVM), Linear Regression, and Convolutional Neural Networks (CNN), for skin cancer detection. The paper discusses the current challenges in skin cancer detection and how machine learning can be used to address these challenges. It provides an overview of the various approaches and techniques used in skin cancer detection using machine learning, with a focus on SVM, Linear Regression, and CNN. The paper also compares the performance of these algorithms and discusses their strengths and weaknesses in skin cancer detection. SVM and linear regression models offer a simple and interpretable approach to skin cancer detection, while CNNs provide more sophisticated and accurate results but require a larger dataset and computational resources.

IndexTerms: Analysis, CNN, Machine Learning, Melanoma, Prediction Models, Skin Cancer.

1. INTRODUCTION

Skin cancer is the uncontrollable growth of damaged cells in the outer most layer of the skin. This is because of damage in DNA sequence due to the environmental factors like cigarette smoke and exposure to Ultra Violet (UV) light. DNA damage triggers mutation which leads to rapid multiplication of skin cells that forms malignant tumours. Skin cancer is classified into Melanoma, Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC). Melanoma is the most dangerous type of cancer which leads to death that usually appears on the moles and the areas on the skin which is exposed to sunlight as well as not exposed to sunlight. The affected part of the skin contains melanocytes that spread to other parts of the body. BCC is the most laggard growing and never be large in size. It appears on the skin exposed areas such as hand, face, leg, ears and scalp. It usually matures as an ulcer and does not improve. The early detection of this can be curable. Some are hostile and cannot be treated because it spreads to the deeper cells of the tissue. SQC appears on the sun exposed part and on the incurable inflammation part of the body and occurs in the person who has low immune power. It is large, appears in incurable scars and in lips. The early detection is possible. Benign is a non-cancerous which does not spread to any other parts of the body. It is caused due to exposure of sunlight, inflammation of skin, infections, and genetics. Melanoma mostly occurs in the skin rarely in the mouth and intestines with the abnormal cells that contain melanocytes which control the pigment in our skin. For women, melanoma mostly occurs on the legs and for men on the back. They usually develop from the mole with abnormal changes as an increase in size, changes in the colour, causes itches or skin breakdown. It can occur in the areas between fingernails, palms, toenails and eyes. Benign usually appears on the skin which is highly exposed to sunlight such as face, shoulders, neck, hand and leg. This appears as lump and looks like patches which continues after a week and develops expert knowledge and experience. It is a very laborious and time-consuming procedure.

Skin cancer is the common type of cancer in worldwide and especially in US. By the age of 70, skin cancer will be developed by 1 in 5 Americans. In every hour, more than 2 people die because of skin cancer. Risk for melanoma will be doubled while exposure of sunburns is more than 5 in number. Early detection helps to survive for 5 years and the survival rate is 99 percent. At least 40% of cases have skin cancer when globally accounting for common cancer. Non-melanoma skin cancer occurs 2 to 3 million people per year. Globally in 2012, 232,000 people were in skin cancer, and 55,000 people died. According to the survey of last 20 to 40 years, Australia (white people), New Zealand and South Africa People have the highest rate of Skin cancer in the world[1].

2. LITERATURE REVIEW

There have been related works done in the past to detect skin diseases using machine learning and deep learning. Aleem et al.[2] published an article introducing a mobile-enabled cancer detection system for early. melanoma skin cancer using a support vector machine (SVM). The proposed system can be identified as three main steps: pre-processing, segmentation, and classification. In the pre-processing step, image quality was improved by removing noise using the Gaussian function. In the segmentation step, the grab cut technique was used to split the image. In the feature extraction and classification step, meaningful featurisation mean, standard deviation, and perimeter were extracted. They mainly choose histogram and ABCD features proposed by the ABCD rule The SVM algorithm was applied as a classification technique. SVM algorithm provides good classification results in real-time smartphones. Even though model has been only applied for skin melanoma, This application can be extended to other skin diseases (eczema and skin rashes). It sensitivity and specificity rates are 80% and 70%. However, it would be worthwhile to evaluate the proposed system with a different algorithm such CNN.

Leonhardt L. et al. [3] proposed a KNN-based skin cancer detection system. The proposed system processed synchronous fluorescence spectra of melanoma, nevus, and normal skin samples for neural network training. A fluorescence spectrophotometer was used to measure the fluorescence spectra of the samples, whereas samples were collected from human patients immediately after surgical resection. The dimensionality of measured spectra was reduced with the PCA technique. Both KNN and ANN were trained, and their performance for melanoma detection was compared. On the test dataset, the classification error of KNN was 2–3%, while the classification error for ANN lay in the range of 3% to 4%.

A technique for the classification of four different types of skin lesion images was proposed by Dorj et al. [4]. A pre-trained deep CNN named Alex-Net was used for feature extraction, after which error-correcting output coding SVM worked as a classifier. The proposed system produced the highest scores of the average sensitivity, specificity, and accuracy for SCC, actinic keratosis (AK), and BCC: 95.1%, 98.9%, and 94.17%, respectively.

A combination of self-organizing NN and radial basis function (RBF) neural network was proposed to diagnose three different types of skin cancer, such as BCC, melanoma, and SCC [5]. The proposed system extracted colour, GLCM, and morphological features of lesion images, after which the classification model used those features as input. Furthermore, the classification performance of the proposed system was compared with k-nearest neighbour, ANN, and naïve Bayes classifiers. The proposed system achieved 93.150685% accuracy while k-nearest neighbour showed 71.232877%, ANN showed 63.013699%, and naïve Bayes showed 56.164384% accuracy scores. Another KNN-based automated skin cancer diagnostic system was proposed by Sajid P. M. [5] The proposed system employed a median filter as a noise removal technique. Then filtered images were segmented with a statistical region growing and merging technique. In this system, a collection of textual and statistical features was used. Statistical features were extracted from lesion images, whereas textual features were extracted from a curvelet domain. Finally, the proposed system classified the input images into cancerous or noncancerous with 98.3% accuracy. In this work, other classifiers such as SVM, BPN, and 3-layer NN were also implemented, and their performance was compared with the proposed system's classification performance. SVM produced 91.1% accuracy, BPN showed 90.4% accuracy, 3-layer NN showed 90.5%, whereas the proposed system achieved the highest accuracy of 98.3% for skin cancer diagnosis.

DeVries and Ramachandram [6] proposed a multi-scale CNN using an inception v3 deep neural network that was trained on an ImageNet dataset. For skin cancer classification, the pre-trained inception v3 was further fine-tuned on two resolution scales of input lesion images: coarse-scale and finer scale. The coarse-scale was used to capture shape characteristics as well as overall contextual information of lesions. In contrast, the finer scale gathered textual detail of lesion for differentiation between various types of skin lesions.

Mahbod [7] proposed a technique to extract deep features from various well-established and 6 pre-trained deep CNNs for skin lesions classification. Pretrained AlexNet, ResNet-18 and VGG16 were used as deep-feature generators, then a multi-class SVM classifier was trained on these generated features. Finally, the classifier results were fused to perform classification. The proposed system was evaluated on the ISIC 2017 dataset and showed 97.55% and 83.83% area under the curve (AUC) performance for seborrheic keratosis (SK) and melanoma classification

Viswanatha Reddy Allugunti [8] “A machine learning model for skin disease classification using convolution neural network” a Convolutional Neural Network (CNN) model for the diagnosis of skin cancer was created, constructed, and evaluated using a well-known melanoma dataset. His proposed method, which is a two-stage learning platform, has great-predicted accuracy at each stage, as demonstrated by its overall accuracy of 88.83 percent.

Aswin R. B. and Jaleel [9] described a new method for skin cancer detection based on a genetic algorithm (GA) and ANN algorithms. Images were pre-processed for hair removal with medical imaging software named Dull-Rozar and region of interest (ROI) and were extracted with the Otsu thresholding method. Furthermore, the GLCM technique was employed to extract unique features of the segmented images. Subsequently, a hybrid ANN and GA classifier was used for the classification of lesion images into cancerous and noncancerous classes. The proposed system achieved an overall accuracy score of 88%.

Md. Mahbubur Rahman et al.[10] purposed hybrid fusion technique which use modified anisotropic filtering technique is used in dermatology test image to diverge speckle from noisy images. The proposed fused vector with CNN and HFF has proven 99.49% accuracy. The best performing CNN classifier used to detect whether it's melanoma or non-melanoma skin cancer and also established 99.85% accuracy.

Ionis Kousis et al.[11] trained a model using DenseNet169 The average accuracy of the model was 92.25%, which is higher than the accuracy of other state-of-the-art models. This also means that it does better than specialist dermatologists. They built mobile application using this model which help to identify the lesion and earliest diagnosis. They also built a two-class DenseNet169 mapping model, which achieves very good results, too (an accuracy of 91.10%).

3. METHODOLOGY

i. Collection of the data set

The first stage of this project is the collection of data set. There are multiple open source data set present like HAM-1000 (human-against-machine dataset) [12] with 10,000 training images It is the latest publicly available skin lesions dataset, and it overcomes the problem of the lack of diversity second publicly available data is Derm Quest [13] which contained 22,082 dermoscopic images. These data will be used to train the Machine Learning model. Through these data it will predict whether the cancer is malignant or non-malignant.

ii. Data cleaning

Data cleaning is the first and an essential step while preparing the data for predictive model and analysis. Through data cleaning we remove duplicate, incorrect, incorrectly formatted, corrupted and incomplete data. With incorrect data the prediction becomes unreliable.

iii. Data Pre-processing

Data is arranged so that it can fit into a machine learning model by data pre-processing, which transforms it from its unorganized state. This process involves importing data, loading essential libraries, encoding categorical data, dividing data into training and test sets, and scaling features.

iv. Data visualization

Data visualization is the visual and graphical representation of the data. It can be visualized through graphs, charts and maps, it can be done through different visualization tools which makes it easier to understand the trends patterns and outliers of data. Data visualization also helps the non-technical people to understand the data without any complexity.

v. Comparison

In this stage, different Machine Learning algorithms are compared based on their working and accuracy. Some algorithm gives the direct probabilistic result like logistic regression and other like SVM we have to calculate its accuracy and precision through confusion matrix. The technique which works best for the particular project work is selected. Generally, Support Vector Machine (SVM), Classification technique, Random Forest, Decision Trees, logistic Regression are used.

vi. Training the model

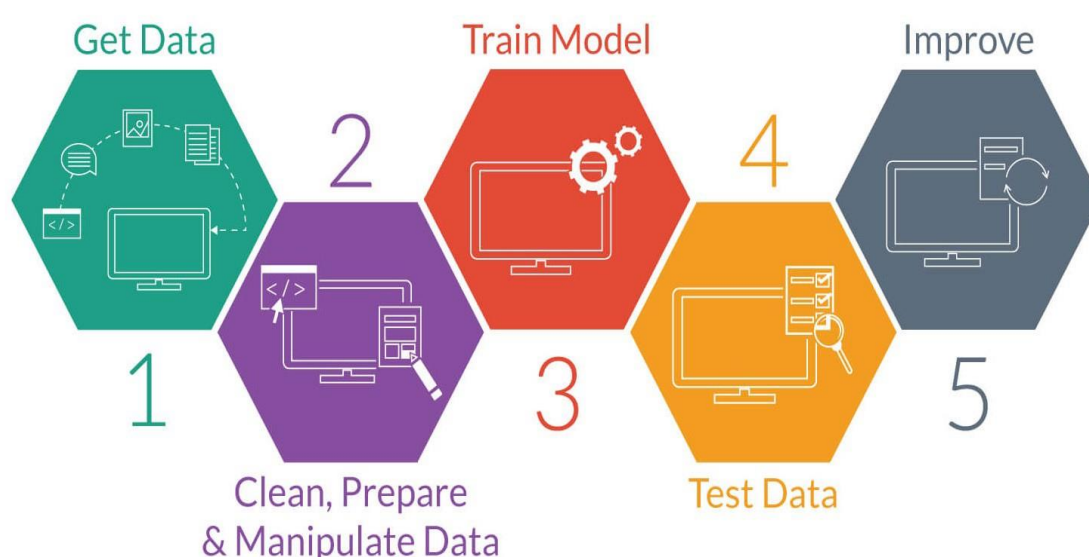
Training the model basically means determining or learning the good values for the bias and weights from labelled data. With training of models there can be a minimization of loss, loss is defined as a number which indicates how bad the predicted outcome on a single example is. The training data must have a correct value or answer, known as target attribute or simply target.

vii. Testing

In this phase it is ensured that all the different parts of the Machine Learning system are working in an effective manner to achieve the desired quality of results. If there is any fault in our data, code or model it can be reviewed and fixed.

viii. Improve

As newer data keep getting present the model needs to improve so that it provide more accurate results with newer available data.



[This Photo](#) by Unknown Author is licensed under [CC BY-SA-NC](#)

Figure 1 Methodology Diagram

4. ISSUES AND CHALLENGES IN PROJECTS

Finding right Machine Learning Technique

There are various machine learning techniques but finding the right one for the dataset is hard. Some of the Machine Techniques which are generally used are SVM, Random Forest, Decision trees, ANN, CNN etc., each having its own qualities and benefits e.g., CNN needs more data and hardware requirement and consume lots of time to train it while logistic regression, SVM and other classification algorithm takes less data hardware requirement but produces slightly less accurate results.

Unbalanced dataset

Real-world datasets used for skin cancer diagnosis are highly unbalanced [14]. Unbalanced datasets contain a very different number of images for each type of skin cancer. For example, they contain hundreds of images of common skin cancer types but only a few images for the uncommon types, making it difficult to draw generalizations from the visual features of the dermoscopic images.

High Dimensionality

High dimensionality happens when the number of features is very large compared to the samples To tackle this problem different algorithms are designed on of the algorithm is Kohonen self-organizing map Neural Network (KNN) is used as dimensionality reducer . However, there are still no perfect solutions which can handle all types of domains and datasets.

Lack of Availability of powerful hardware

Powerful hardware resources with high graphical processing unit (GPU) power are required for the NN software to be able to extract the unique features of a lesion's image, which is critical for achieving better skin cancer detection. The lack of availability of high computing power is a major challenge in deep learning-based skin cancer detection training.

Model Scalability

Pre-processing and detection of lesion edges are very crucial steps in the automated detection of skin cancer. Various optimization algorithms such as artificial the bee colony algorithm [15], ant colony optimization [16], social spider optimization [17], and particle swarm optimization [18] can be explored to increase the performance of automated skin cancer diagnostic systems.

Analysis of Genetic and Environmental Factors

Researchers have identified various genetic risk factors for melanoma, such as fair skin, light coloured eyes, red hair, a large number of moles on the body, and a family history of skin cancer. When these genetic risk factors are combined with environmental risks such as high ultraviolet light exposure, the chances of developing skin cancer become very high [19]. These factors can be combined with existing deep learning approaches for better performance.

Accessibility

One of the major problems with any project is determining how to reach the target audience. India has a wide range of cultural background and socioeconomic background which leads to difficulties such as identifying and communicating the problem with someone. This lack of information and education can cause problems for them to use the software with ease. So, the goal is to make software which is easy to use and accessible software which can be used by every community.

5. CONCLUSION

This review paper has discussed various neural network techniques for skin cancer detection and classification. All of these techniques are non-invasive. Skin cancer detection requires multiple stages, such as pre-processing and image segmentation, followed by feature extraction and classification. This review focused on ANNs, CNNs, logistic regression and SVM for classification of lesion images. Each algorithm has its advantages and disadvantages. Proper selection of the classification technique is the core point for best results. However, CNN gives better results than other types of a neural networks when classifying image data because it is more closely related to computer vision than others.

References

- [1] Z. Apalla, A. Lallas, E. Sotiriou, E. Lazaridou, and D. Ioannides, 'Epidemiological trends in skin cancer', *Dermatol. Pract. Concept.*, vol. 7, no. 2, pp. 1–6, Apr. 2017, doi: 10.5826/dpc.0702a01.
- [2] 'm-Skin Doctor: A Mobile Enabled System for Early Melanoma Skin Cancer Detection Using Support Vector Machine | SpringerLink'. https://link.springer.com/chapter/10.1007/978-3-319-49655-9_57 (accessed Mar. 24, 2023).
- [3] 'Artificial neural networks for processing fluorescence spectroscopy data in skin cancer diagnostics - IOPscience'. Accessed: Mar. 24, 2023. [Online]. Available: <https://iopscience.iop.org/article/10.1088/0031-8949/2013/T157/014057/meta>
- [4] A. D. Mengistu and D. M. Alemayehu, 'Computer Vision for Skin Cancer Diagnosis and Recognition using RBF and SOM', 2015.
- [5] P. M. Sajid and D. A. Rajesh, 'Performance evaluation of classifiers for automatic early detection of skin cancer', *J Adv Res Dyn Control Syst*, vol. 10, pp. 454–461, 2018.
- [6] T. DeVries and D. Ramachandram, 'Skin Lesion Classification Using Deep Multi-scale Convolutional Neural Networks'. arXiv, Mar. 04, 2017. doi: 10.48550/arXiv.1703.01402.
- [7] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, 'Skin Lesion Classification Using Hybrid Deep Neural Networks', in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 1229–1233. doi: 10.1109/ICASSP.2019.8683352.
- [8] V. R. Allugunti, 'A machine learning model for skin disease classification using convolution neural network', *Int. J. Comput. Program. Database Manag.*, vol. 3, no. 1, pp. 141–147, 2022, Accessed: Mar. 28, 2023. [Online]. Available: <https://www.computersciencejournals.com/ijcpdm/archives/2022.v3.i1.B.53>
- [9] R. B. Aswin, J. A. Jaleel, and S. Salim, 'Hybrid genetic algorithm — Artificial neural network classifier for skin cancer detection', in *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, Jul. 2014, pp. 1304–1309. doi: 10.1109/ICCICCT.2014.6993162.
- [10] M. M. Rahman *et al.*, 'Hybrid Feature Fusion and Machine Learning Approaches for Melanoma Skin Cancer Detection'. Preprints, Jan. 18, 2022. doi: 10.20944/preprints202201.0258.v1.
- [11] I. Kousis, I. Perikos, I. Hatzilygeroudis, and M. Virvou, 'Deep Learning Methods for Accurate Skin Cancer Recognition and Mobile Application', *Electronics*, vol. 11, no. 9, Art. no. 9, Jan. 2022, doi: 10.3390/electronics11091294.
- [12] P. Tschandl, C. Rosendahl, and H. Kittler, 'The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions', *Sci. Data*, vol. 5, no. 1, Art. no. 1, Aug. 2018, doi: 10.1038/sdata.2018.161.
- [13] A. Bøer and K. C. Nischal, 'Get set for the net - www.derm101.com: A growing online resource for learning dermatology and dermatopathology', May 2007, Accessed: Mar. 29, 2023. [Online]. Available: <https://tspace.library.utoronto.ca/handle/1807/47896>
- [14] M. A. A. Milton, 'Automated Skin Lesion Classification Using Ensemble of Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge'. arXiv, Jan. 30, 2019. doi: 10.48550/arXiv.1901.10802.
- [15] 'Fahad: Ant colony optimization-based streaming feature... - Google Scholar'. https://scholar.google.com/scholar_lookup?journal=Sci.+Program.&title=Ant+Colony+Optimization-Based+Streaming+Feature+Selection:+An+Application+to+the+Medical+Image+Diagnosis&author=L.G.+Fahad&author=S.F.+Tahir&author=W.+Shahzad&author=M.+Hassan&author=H.+Alquhayz&volume=2020&publication_year=2020&pages=1-10&doi=10.1155/2020/1064934& (accessed Mar. 29, 2023).
- [16] 'Ant Colony Optimization-Based Streaming Feature Selection: An Application to the Medical Image Diagnosis'. <https://www.hindawi.com/journals/sp/2020/1064934/> (accessed Mar. 29, 2023).
- [17] A. Y. Husodo, G. Jati, A. Octavian, and W. Jatmiko, 'Enhanced Social Spider Optimization Algorithm for Increasing Performance of Multiple Pursuer Drones in Neutralizing Attacks From Multiple Evader Drones', *IEEE Access*, vol. 8, pp. 22145–22161, 2020, doi: 10.1109/ACCESS.2020.2969021.
- [18] 'Particle swarm optimization with deep learning for human action recognition | SpringerLink'. <https://link.springer.com/article/10.1007/s11042-020-08704-0> (accessed Mar. 29, 2023).

[19] R. A. Sturm, 'Skin colour and skin cancer –MC1R, the genetic link', *Melanoma Res.*, vol. 12, no. 5, p. 405, Sep. 2002, Accessed: Mar. 29, 2023. [Online]. Available: https://journals.lww.com/melanomaresearch/Abstract/2002/09000/Skin_colour_and_skin_cancer__MC1R,_the_genetic.1.aspx
