SKIN CANCER PREDICTION USING DEEP LEARNING TECHNIQUES

Moturi Swathi [1],Dr. Sk Mahaboob Basha [2]

[2]Professor, Department Of CSE,Sree Dattha Institute Of Engineering And Science Sheeriguda, Ibrahimpatnam Hyderabad , Telangana,India.

[1]PG Scholar, Department Of CSE, Sree Dattha Institute Of Engineering And Science Sheeriguda, Ibrahimpatnam Hyderabad , Telangana,India.

***Abstract*:**Modern medicine has made great strides in curing cancer, a deadly disease caused by cells in the body that have gotten out of hand. Therefore, cancer's fast growth can infiltrate and destroy nearby structures and even spread to faraway places, killing those there. In addition, cancer does not discriminate based on gender; it can impact any and all human cells at any stage of life. These days, skin cancer is on the rise, and it manifests itself in a number of ways, one of which is the unchecked damage to DNA that mutations in skin cells cause. Additionally, basal cell, squamous cell, and melanomas skin cancers are the three main types of skin cancer. For efficient pre-processing, novel segmentation and feature extraction approaches have also been suggested here. In addition, successful image conversion, contour detection, and wavelet transform have all made use of the current technologies. Pure mathematics and inflated pixel estimates are common sources of error in image pre-processing when using data collected from sensors onboard satellites.

Key words: Skin cancer, basal cell, DNA.

1. INTRODUCTION

The fast increase in the number of individuals affected by cancer, which is approximately 2.5 million, has made it a deadly disease in India recently. More than 0.7 million new cases of cancer are detected annually in India, according to estimates. Conversely, cancer claims the lives of almost 0.3 million Indians annually per annum. Oral, lung, esophageal, and stomach cancers in males and cervical, breast, and oral cancers in women are the most well-known cancer websites, according to the National Cancer Registry Programme (NCRP). Cancers of the mouth and lungs account for more than half of all cancer-related deaths in India, whereas cancers of the cervix and breast account for more than half of all cancer-related deaths in females.

1.1 SKIN CANCER

Cancer of the skin occurs when abnormal skin cells proliferate uncontrollably. It manifests as cancerous tumours when DNA damage to skin cells goes unrepaired, causing mutations that set off errors in genes that cause the skin cells to grow slowly. Plus, there are three distinct subtypes of skin cancer: basal cell, squamous cell, and melanomas. Regardless, melanoma and non-melanoma are the main categories of skin cancer. Here, melanocytes (skin colour) are the product of a malignant tumour called melanoma, which can spread to other parts of the body. It can also travel to other parts of the body and eventually kill the host. In contrast, non-melanoma cancers metastasize, or spread, to other organs and tissues.

A higher death rate was caused by malignant melanoma, the worst kind of skin cancer in humans. There has been a reasonable uptick in the incidence of melanoma recently, especially among Caucasian individuals. Melanoma has risen to the position of sixth most significant cancer in Australian women and fifth most important cancer in males in North America [1,2]. Reports issued from the country where the majority of the world's population resides also play a role. In addition, melanoma is the sixth most common cancer in men and the seventh most common cancer in women [3,4]. On the other hand, if caught early enough, this form of skin cancer can be cured [5,6]. By moving in this approach, melanoma skin cancer can be detected in its early stages with a simple excision, which can minimise mortality.

2. LITERATURE REVIEW

Dermatologists utilise dermoscopy as a diagnostic method for melanoma because of its usefulness in detecting the disease at an early stage. Additionally, it successfully visualises numerous pigmented structures, such as dots, pigment networks, streaks, and blue with white patches, which are not apparent to the human eye [7,8]. Therefore, the provided lesion image cannot be identified as melanoma using only two attributes. Consequently, dermoscopy images are utilised to gain greater assurance in classifying lesions.
In the process of melanoma detection in cancer diagnosis system, the dermatologists are adopting the ABCD rule which is used for analysingthe four standard parameters such as Diameter, Asymmetry, Colours andBorder for diagnosing the melanoma at starting stage [7].
Moreover, a new 7- point checklist method is used by [8] for analysing the parameters. Moreover, the melanoma images with hair and reflection and the invisible of melanoma borders that creates a visual identification very complex task to the experts who are involving the process of skin cancer diagnosis. Not only that, but even for experienced doctors, deciphering melanoma pictures is a subjective and efficient process [9].

Computer Aided Diagnosis (CAD) became crucial in melanoma detection for handling all these issues, helping doctors understand the pictures better and making the right diagnosis. Moreover, this kinds of diagnosis systems used for reducing the time taken for diagnosing the disease and it also used for improving the detection accuracy [10].

3. METHODOLOGY

3.1 Image database

The selected melanoma skin lesion images that are available in the skin cancer image database.

3.2 User interface module

The different kinds of skin cancer images such as melanoma skin lesion and other type of skin cancer images have been collected by this module from the skin cancer image database that has the reasonable volume of pixels in the input melanoma skin cancer images. Moreover, all these skin cancer images have been forwarded into the next component called image data pre-processor that is used to pre-process the input skin melanoma images for performing the classification.

3.3 Data pre-processor

Image conversion, contour detection, wavelet transform, ABCD parameters, segmentation, and feature extraction are the six subcomponents used in this data pre-process.
Here, the proposed model uses separate new and existing methods for performing image conversion, contour detection, wavelet transformation, for framing ABCD rules, segmenting the input melanoma images and extract the necessary features from input image. This data pre-processor is sent to the next module called classification module to perform image classification process.

Image Conversion: Many file formats are used to store the medical images along with various conversion problems due to the presence of different image formats. The Image conversion techniques are necessary to perform for fast image conversion. Here, the input images can be converted from any format to the standard formats such as JPG, PNG and BMP formats. It can be rotated the input images in 90 degree increments and also rotates automatically for compensating with the orientation of EXIF. Moreover, it also can be resized or enlarged the input medical images by a number of images or to a fixed size which is retaining the original ratio optionally. The file type of input skin images are also added as a command in the context menu which is available in the menu bar.

Contour Detection: Contour detection is playing major role in image processing. In this scenario, split the input images by using classification and the segmentation techniques into many parts which is related to the detection of the connected contours that are also separating these various parts. Here, image input detection is an easy when detects the local image edges by using the various image analysis techniques. On the other hand, the image edge detection of continuous contours that is very difficult task and also need in detailed analysis about the input images.

Wavelet Transform: This sub component is responsible for performing wavelet transform by applying the standard techniques that are available in the literature. Generally, it is an important task in image compression application. It is more suitable method when it is compared to the Fourier transform. Moreover, the Fourier transform is not to calculate the spectral data value practically. Because of, it needs the necessary information about the previous and future of the signal over the entire time. In addition, it is not able to observe the frequency values which are dynamic with time because the resulting function after performing the Fourier transform in different independent time. Moreover, the wavelet transforms are working based on the wavelets with different frequency in certain time duration.

Image Segmentation: The suggested prediction model's data pre-processor uses this sub-component to carry out the picture segmentation task. Partitioning an original skin image into several segments, such as sets of pixels and as super-pixels, is achieved through the image segmentation method. A less complicated and more easily analysed representation of an input skin image is the primary goal of this segmentation procedure. Additionally, it is utilised to insert the numerous picture objects and their respective borders (e.g., lines, curves, etc.) into the input photos. Lastly, it covers the whole input skin image and assigns new labels to every pixel.

Feature Extraction: This sub component is responsible to extract the necessary features which can contribute more in the image classification process in terms of decision making in the proposed prediction model. Here, it starts with a set of measured data and also built the derived features that must be useful, informative and non-redundant. Generally, it performs a dimensionality reduction in this data pre-processor.

3.4 Clustering module

This module is responsible for grouping the segmented images which are having more relevant pixels and super pixels. The clustering module is applied an existing fuzzy c-means clustering method that is used to group the pre-processed skin melanoma images effectively.

3.5 Classification module

The classification module is consists of four additional components namely SVM Classifier, EMSVM Classifier, ANN-BPN Classifier and FTCM Classifier. These four sub components are also used different classification algorithms such as SVM Classification algorithm, Enhanced Multiclass SVM classification algorithm, Artificial Neural Network based Back Propagation Network and Fuzzy Temporal Cognitive Map classification algorithms for classifying the melanoma images.

ANN-BPN Classifier: This sub component is used to classify the skin melanoma images effectively by using the existing neural network classifier called ANN-BPN classification algorithm. The three layers of the neural network here are the input, output, and hidden layers. A single process layer, the hidden layer, connects the input and output layers. This hidden layer compresses data by making use of a smaller number of neurons than are present in the input and output layers combined. In addition, the training activities over the neural network are used to accomplish the image compression operation. To build a hierarchical neural network with two additional process layers, the back-propagation neural network is further extended.

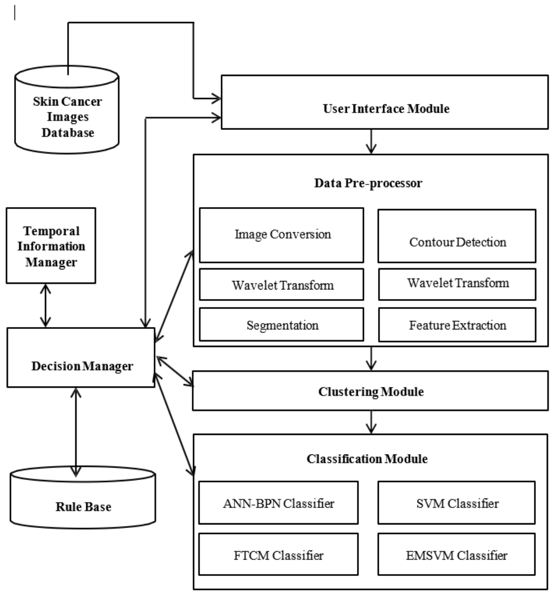

Figure 3.1 System Architecture

4. RESULTS AND ANALYSIS:

In the performance analysis, skin cancer datasets is consider as it predicts the malignant and benign cancer.

```
Downloading skin-cancer-malignant-vs-benign, 340467838 bytes compressed
[==============================================] 340467838 bytes downloaded
Downloaded and uncompressed: skin-cancer-malignant-vs-benign
Data source import complete.
```

Figure.

To perform the prediction and classification of skin cancer, the following packages are imported as mentioned as urlopen, unquote, urlparse, HTTPError, etc.

```
import os
import sys
from tempfile import NamedTemporaryFile
from urllib.request import urlopen
from urllib.parse import unquote, urlparse
from urllib.error import HTTPError
from zipfile import ZipFile
import tarfile
import shutil

CHUNK_SIZE = 40960
```

The basic library files of pandas and numpy are imported and it uses matplotlib to visualize the images and graphs. Seaborn of SNS and gloab are used to retrieve the images of skin cancer along with random function of seed library.

```
# importing basic libraries

import pandas as pd
import numpy as np
import os

import matplotlib.pyplot as plt
import seaborn as sns

from glob import glob ## glob is used to retrieve files

# set seed
np.random.seed(21)
```

```
# Loading train images
img_benign_train = [read(os.path.join(directory_benign_train, filename)) for filename in os.listdir(directory_benign_train)]
img_malignant_train = [read(os.path.join(directory_malignant_train, filename)) for filename in os.listdir(directory_malignant_train)]

# Loading test images
img_benign_test = [read(os.path.join(directory_benign_test, filename)) for filename in os.listdir(directory_benign_test)]
img_malignant_test = [read(os.path.join(directory_malignant_test, filename)) for filename in os.listdir(directory_malignant_test)]

#img_benign_train
type(img_benign_train)
```

- A numpy array uses considerably less memory than a list.
- A variety of mathematical procedures that this list cannot handle are supported by Numpy arrays. While element-wise operations are feasible in np arrays, they are not feasible in lists. (Np arrays contain only homogeneous elements, whereas lists can contain heterogeneous elements). Because of their homogeneity, calculations are substantially

more convenient.

- They are capable of executing operations considerably quicker than lists.

Then data labels classify the skin cancer as benign ‘0’ and malignant ‘1’.

```
y_benign_train = np.zeros(X_benign_train.shape[0])
y_malignant_train = np.ones(X_malignant_train.shape[0])

y_benign_test = np.zeros(X_benign_test.shape[0])
y_malignant_test = np.ones(X_malignant_test.shape[0])

y_malignant_train
```

array([1., 1., 1., ..., 1., 1., 1.])

```
Shape of X_train:  (2637, 224, 224, 3)
Shape of y_train:  (2637,)
Shape of X_test:  (660, 224, 224, 3)
Shape of y_test:  (660,)
array([0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.])
```

Data shuffles are crucial because we want our samples to be independent of one another and not to be impacted by samples that came before or after. Since the dataset's creator and sorting order are unresolved it's best to shift the training and test sets to get biased-free independent points.

Shuffling while producing epochs also ensures that the risk of creating batches that are not representative of the whole dataset is reduced.

Based on the training datasets, the following images are displayed in figure 4.1.

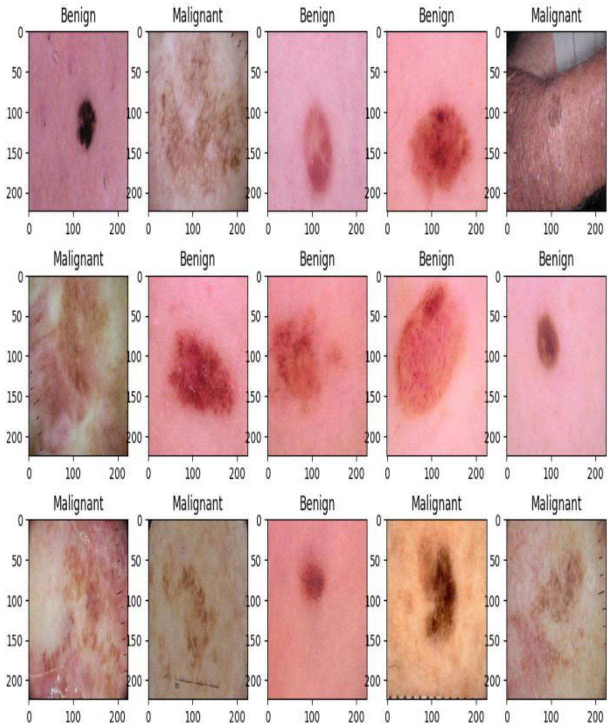

Figure. 4.1 Training datasets.

- Convolutional Layer: Image transformation is achieved through the employment of filters and feature maps. We refer it this as the Convolutional Layer. One application of the Pooling Layer's Max Pooling is down sampling. Both processing expenses and overfitting are somewhat decreased by it.
- Dropout: During training, a regularization technique that sets some nodes' weights to 0 at random drops some nodes. This compels the network to acquire features through dispersed means. Enhances generalization and avoids over fitting.
- The flatten layer is employed to transform feature maps into one-dimensional vectors, enabling their utilisation in prediction tasks. Simple ANNs with non-linear Relu activation functions are referred described as "dense layers with Relu."
- Dense layer with Softmax: An ANN layer with the binary activation function Softmax for final classification.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 224, 224, 64)      1792

 max_pooling2d (MaxPooling2  (None, 112, 112, 64)      0
 D)

 dropout (Dropout)           (None, 112, 112, 64)      0

 conv2d_1 (Conv2D)           (None, 112, 112, 64)      36928

 max_pooling2d_1 (MaxPoolin  (None, 56, 56, 64)        0
 g2D)

 dropout_1 (Dropout)         (None, 56, 56, 64)        0

 flatten (Flatten)           (None, 200704)            0

 dense (Dense)               (None, 128)               25690240

 dense_1 (Dense)             (None, 128)               16512

 dense_2 (Dense)             (None, 2)                 258

=================================================================
Total params: 25745730 (98.21 MB)
Trainable params: 25745730 (98.21 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Rate of learning An annealer is used to reduce the learning rate by a specific percentage after a certain number of training iterations/epochs.

```
0.5264                                                    33/33
[==============================] - ETA: 0s - loss: 1.9614 - accuracy:
0.5270                                                    33/33
[==============================] - 23s 503ms/step - loss: 1.9372 - accuracy: 0.5276 - val_loss: 0.6671 -
val_accuracy: 0.5455
```

```
dict_keys(['loss', 'accuracy', 'val_loss', 'val_accuracy', 'lr'])
```

```
0.8136363636363636
```

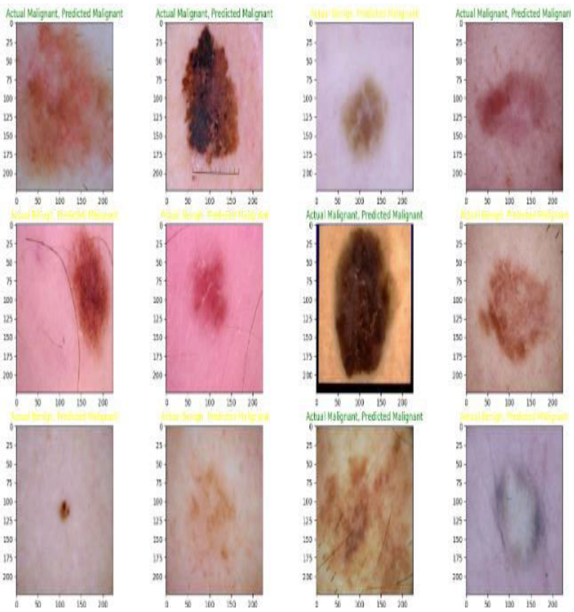

Figure 4.2. Training datasets

A green title indicates an accurate forecast. The yellow title indicates that a benign cancer was mistakenly diagnosed as malignant (but this is still acceptable because medical professionals would still closely monitor it). Red title signifies inaccurate prediction of Malignant Cancer as Benign (This is the most serious circumstance since we do not want a Malignant Cancer to go unreported or receive less care).

CONCLUSION

In this research work, the overall system architecture has been designed for the proposed prediction system and explained. This overall system architecture provides the various components of the proposed prediction model and the functionalities of each module. All the modules present in the overall system architecture which cooperates and also to provide an effective environment for predicting the cancer diseases. This system architecture considered an image database which is standard and worldwide accepted for performing the pre-processing and the classification activities in the proposed model.

REFERENCES

1. Erdmann F, Lortet-Tieulent J, Schüz J, et al. . International trends in the incidence of malignant melanoma 1953-2008—are recent generations at higher or lower risk? Int J Cancer. 2013;132(2):385-400. doi:10.1002/ijc.27616

2. Arnold M, Holterhues C, Hollestein LM, et al. . Trends in incidence and predictions of cutaneous melanoma across Europe up to 2015. J Eur Acad Dermatol Venereol. 2014;28(9):1170-1178. doi:10.1111/jdv.12236
3. El Ghissassi F, Baan R, Straif K, et al. ; WHO International Agency for Research on Cancer Monograph Working Group . A review of human carcinogens—part D: radiation. Lancet Oncol. 2009;10(8):751-752. doi:10.1016/S1470-2045(09)70213-X

4. International Agency for Research on Cancer. Radiation: Volume 100 D: A Review of Human Carcinogens: IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. IARC Publications; 2012.

5. Arnold M, de Vries E, Whiteman DC, et al. . Global burden of cutaneous melanoma attributable to ultraviolet radiation in 2012. Int J Cancer. 2018;143(6):1305-1314. doi:10.1002/ijc.31527

6. Fidler MM, Gupta S, Soerjomataram I, Ferlay J, Steliarova-Foucher E, Bray F. Cancer incidence and mortality among young adults aged 20-39 years worldwide in 2012: a population-based study. Lancet Oncol. 2017;18(12):1579-1589. doi:10.1016/S1470-2045(17)30677-0

7. Whiteman DC, Green AC, Olsen CM. The growing burden of invasive melanoma: projections of incidence rates and numbers of new cases in six susceptible populations through 2031. J Invest Dermatol. 2016;136(6):1161-1171. doi:10.1016/j.jid.2016.01.035

8. Garbe C, Keim U, Gandini S, et al. . Epidemiology of cutaneous melanoma and keratinocyte cancer in white populations 1943-2036. Eur J Cancer. 2021;152:18-25. doi:10.1016/j.ejca.2021.04.029

9. Paulson KG, Gupta D, Kim TS, et al. . Age-specific incidence of melanoma in the United States. JAMA Dermatol. 2020;156(1):57-64. doi:10.1001/jamadermatol.2019.3353

10. Baade PD, Whiteman DC, Janda M, et al. . Long-term deaths from melanoma according to tumor thickness at diagnosis. Int J Cancer. 2020;147(5):1391-1396. doi:10.1002/ijc.32930