DETECTIONOFSKINCANCERUSINGDEEP LEARNING AND IMAGE PROCESSINGTECHNIQUE

Prof. Yogesh J. Gaikwad4
Department of Information Technology
Dr. VishwnathKarad MIT World Peace University,
Pune, India.
yogesh.gaikwad@mitwpu.edu.in

2nd Vismay G. Nair
Department of Information Technology
MAEER'S MIT Polytechnic,
Pune, India.
nairvismay9@gmail.com

3rd Khushi N. Roham
Department of Information Technology
MAEER'S MIT Polytechnic,
Pune, India.
knroha@gmail.com

4th Vaishnavi N. Shinde
Department of Information Technology
MAEER'S MIT Polytechnic,
Pune, India
manushinde1815@gmail.com

5th Akshay S. Chavan
Department of Information Technology
MAEER'S MIT Polytechnic,
Pune, India
akshaychavan6905@gmail.com

***Abstract*—Disease comparable to dermatologists and could enable lifesaving and quick decisions, even outside the clinic, through the development of applications on cell phones. There is no audit of the ebb and flow work in this examination region, as far as anyone is concerned. This investigation provides an orderly audit of the most recent research on characterizing skin sores with CNNs. Our audit is limited to skin injury classifiers. Strategies that use a CNN solely for division or the order of dermoscopic designs are specifically excluded. Furthermore, this investigation discusses why the equivalence of the introduced methodology is exceptionally difficult and which difficulties should be addressed later on. Google Scholar, PubMed, Medline, Science Direct, and Web of Science were all searched of Science data bases for systematic surveys and unique examination articles distributed in English For this survey, only papers that announced adequate logical procedures are remembered. We discovered 13 papers that used CNNs to classify skin sores. Characterization strategies can be divided into three categories on a basic level. Approaches that utilise a CNN previously prepared through another huge dataset and after that streamline its boundaries to the grouping of skin sores are the most widely used and show the best exhibition with the currently accessible restricted datasets. CNN's show outperforms cutting-edge skin ailment classifiers. Surprisingly, it is difficult to consider various arrangement strategies because certain methodologies use nonpublic datasets for preparation and testing, making reproducibility difficult. Future To allow equivalence, distributions should use publicly available benchmarks and fully disclose the techniques used for preparation. (*Abstract*)**

I. INTRODUCTION

Theskin is the body's outermost layer, and it is vulnerable to environmental exposure, including dust, pollution, microorganisms, and UV radiation. These are possible causes of any type of skin disease. Furthermore, skin diseases are caused by genetic instability, making the skin more complex skin diseases. Each year, approximately 5.4 million new cases of skin cancer are reported in the United States alone. Global statistics are equally concerning. According to recent reports, the number of new cases of melanoma diagnosed each year increased by 53% between 2008 and 2018. This disease's death rate is expected to rise over the next decade. The rate of survival is less than14% if diagnosed at a later stage However, if skin cancer is detected early, the survival rate is nearly 97 percent. This necessitates early detection of skin cancer. This article addresses the issue of improved accuracy in early diagnosis.

The victimization of images is defined as the image process of mathematical operations for the abuse of any type of Signal processing in image science As an input to these signal procedures, an image, a sequence of images, or even a movie that resembles a video frame or a snapshot can be used. Image processing can produce an image, a collection of features, or an image associated with settings. Image processing techniques include analyzing individual color planes in an image and processing their two-dimensional signal. These signals are then applied using standard signal processing techniques. When the dimensions in the coordinate axes or in time are considered, when these images are processed as three-dimensional signals The term "image processing" typically refers to digital image processing; however, optical and analogue imaging processes are also possible. This chapter discusses common methods for all or a portion of them.

This research has four major components. Section 2 describes the research strategy for conducting an effective review of deep learning algorithms for skin cancer (CS) diagnosis. It includes a research area description, search strings, search criteria, information sources, information extraction methods, and selection criteria. Section 3 provides a comprehensive overview of CS detection strategies and reviews selected research publications. Section 4 includes a study summary as well as a brief conclusion.

LITERATURESURVEY

Vijayalakshmi M presented a methodology on Dermatological Diseases in 2019, which are one of the most significant clinical issues in the twenty-first century due to their exceptional mind-boggling and costly determination with challenges and subjectivity of human translation. In cases of potentially fatal infections, such as Melanoma, early detection is critical.plays a critical role in determining the likelihood of recovery. [1] Jaworek-Korjakowska et al. (2017) proposed a new method for dealing with the location and arrangement of boundary anomaly, one of the significant boundaries in an ABCD-based indicative algorithm that is widely used. In 2017, Yogendra Kumar Jain and Megha Jain proposed "[2] Skin malignant growth location and arrangement utilizing wavelet Transform and probabilistic neural organization." This paper describes a simple and effective technique for identifying and classifying skin cancerous growths. When compared to previous methods proposed in a similar space, this is a significant improvement. PNN outperforms other types of Artificial Neural Networks (ANNs) and has demonstrated phenomenal characterization execution in other applications.

[3Skin Cancer Classification Using an Automatic Lesion Detection System (ALDS) In 2016, Muhammad Ali Farooq, Muhammad AatifMobeenAzhar, and RanaHammadRaza proposed "Using SVM in addition to Neural Classifiers." Chang et alall-encompassing's work is the Automatic Lesion Detection System (ALDS) for skin, malignant growth characterization. Initially, an honing channel is used, and then hair evacuation is performed with dull razor programming, resulting in more refined results. Dynamic forms and watershed approaches are used to naturally portion out the dangerous region from the dataset image with increased efficiency, whereas disease mole grouping using SVM was worked on using Chang et AL research'sdiscoveries. [4].

II. SCOPE

This proposed system will generate a mechanized image of skin damage cells.

To recognize the type of skin-threatening development, the structure first and foremost requires an educational assortment of various skin cancer-causing cells.

It is based on the squamous cell carcinoma (SCC), basal cell carcinoma (BCC), and melanoma types of skin cancer.

IV. OBJECTIVE

- The primary goals of this application are ,
- To comprehend the new technology used for the detection of skin cancer.
- To concentrate on developing a programmed framework for skin disease discovery from advanced images utilizing image preparation and AI.

Sr NO.	Task	Estimated KLOC
1.	Registration and Login	0.2K
2.	Pre-Processing	0.1K
3.	Training of dataset	0.6K
4.	Classification using Convolutional Neural Network	0.9K
5.	Image attributes detection	0.8K
6.	Prediction	0.5K
	Total	3.1K

Diagnosis is performed in the early stages to allow for further advancement in the biomedical field.

VI. SYSTEM MODE

ProcessModel:

Theproposed deep learning model was fed images from the training set that had already been processed. To extract features from the image, a series of convolution, pooling, and Re LU layers were used. The proposed neural network has five hidden layers. When the input image passes through these levels, the features are extracted one by one and then passed on to the next layer. The image is convoluted, max pooled, and then convoluted again before being average pooled. There are two dense layers at the end. The global average pooling layer reduces overfitting by reducing the number of parameters in the model. The thick layers are the fully linked layers that determine whether the images are benign or malignant.

In addition, an artificial neural network is made up ofThe brain's simplification The brain's simplification of neurons inspired this network of nodes. When thousands of these neurons are linked together, they form the basis of a neural network, in which the connection between neurons attempts to capture pattern invariance in the face of distorted or shifted input data. A Deep CNN model has three types of layers: convolutional layers, pooling layers, and dense layers.

Testing:

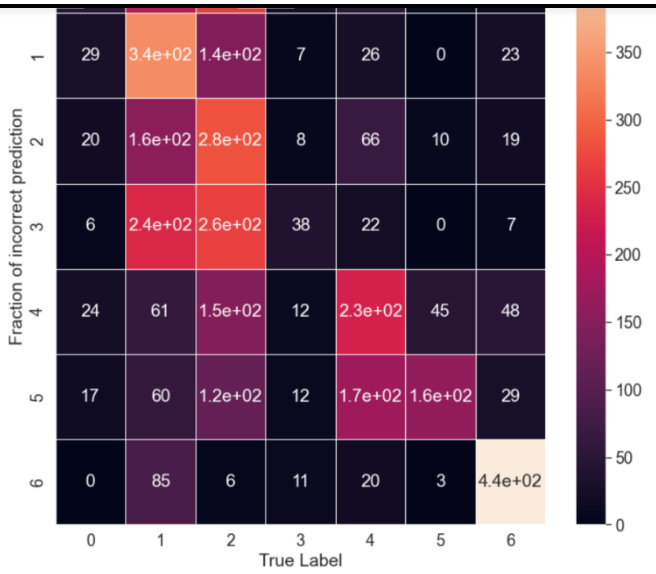
The model testing step includes inserting the test data, preparing it, and sending it to the trained CNN model. Each layer of the test image is examined for the presence of disease-related features. If it proves to be intelligent, the system generates a result based on the knowledge gained during the training phase. Otherwise, the system produces a neutral result.

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        from glob import glob
        from PIL import Image
        import os
        import shutil
        import seaborn as sns

In [2]: # Dump all the images into a folder and specify the path

        data_dir = os.getcwd() + '/all_images/'

In [3]: # Path to the destination directory where we want subfolder

        dest_dir = os.getcwd() + '/reorganized/'

In [4]: # Read the csv file containing image named & corresponding labels

        skin_df = pd.read_csv('HAM10000/HAM10000_metadata.csv')
        skin_df
```

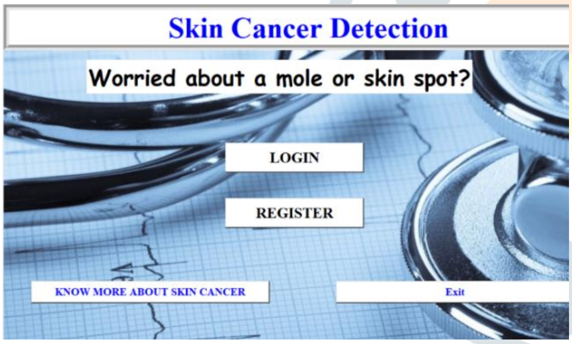
```
num_classes = 7
model = Sequential()
model.add(Conv2D(256,(3,3),activation='relu',input_shape=(SIZE,SIZE,3)))
#model.add(BatchNormalization)
model.add(MaxPool2D(pool_size=(2,2)))
model.add(Dropout(0.3))
model.add(Conv2D(128,(3,3),activation='relu'))
model.add(MaxPool2D(pool_size=(2,2)))
model.add(Dropout(0.3))
model.add(Conv2D(64,(3,3),activation='relu'))
model.add(MaxPool2D(pool_size=(2,2)))
model.add(Dropout(0.3))
model.add(Flatten())
model.add(Dense(32))
model.add(Dense(7,activation='softmax'))
model.summary()
model.compile(loss='categorical_crossentropy',optimizer='Adam',metrics=['acc'])
batch_size = 16
epochs = 50
history = model.fit(
    x_train, y_train,
    epochs=epochs,
    batch_size = batch_size,
    validation_data=(x_test, y_test),
    verbose=2)
score = model.evaluate(x_test,y_test)
print("Test accuracy",score[1])
loss = history.history['loss']
val_loss = history.history['val_loss']
epochs = range(1,len(loss) + 1)
plt.plot(epochs,loss,'y',label = 'Training loss')
plt.plot(epochs,val_loss,'r',label = 'Validation loss')
plt.title('Training and Validation loss')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()
plt.show()
```

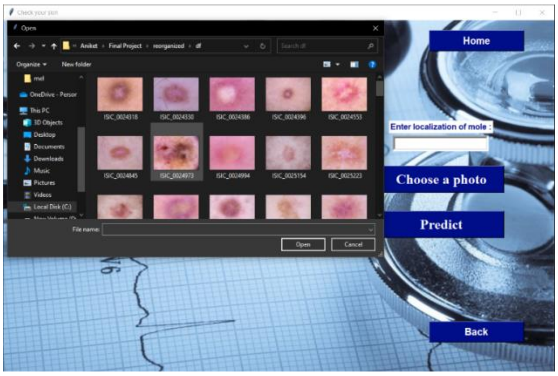
```
# Data distribution visualization

fig = plt.figure(figsize=(15,10))
ax1 = fig.add_subplot(221)
skin_df['dx'].value_counts().plot(kind='bar', ax=ax1)
ax1.set_ylabel('Count')
ax1.set_title('Cell Type')

ax2 = fig.add_subplot(222)
skin_df['sex'].value_counts().plot(kind='bar', ax=ax2)
ax2.set_ylabel('Count', size=15)
ax2.set_title('Sex')

ax3 = fig.add_subplot(223)
skin_df['localization'].value_counts().plot(kind='bar')
ax3.set_ylabel('Count',size=12)
ax3.set_title('Localization')

ax4 = fig.add_subplot(224)
sample_age = skin_df[pd.notnull(skin_df['age'])]
sns.distplot(sample_age['age'], fit=stats.norm, color='red');
ax4.set_title('Age')

plt.tight_layout()
plt.show()
```

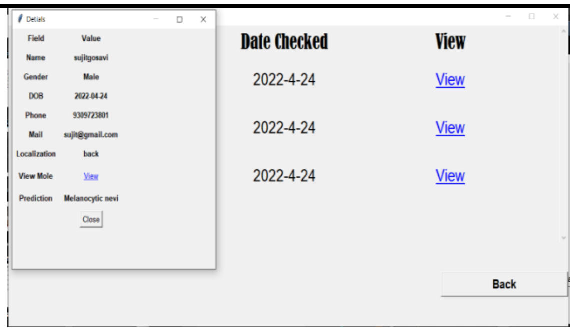

```
# Data distribution visualization

fig = plt.figure(figsize=(15,10))
ax1 = fig.add_subplot(221)
skin_df['dx'].value_counts().plot(kind='bar', ax=ax1)
ax1.set_ylabel('Count')
ax1.set_title('Cell Type')

ax2 = fig.add_subplot(222)
skin_df['sex'].value_counts().plot(kind='bar', ax=ax2)
ax2.set_ylabel('Count', size=15)
ax2.set_title('Sex')

ax3 = fig.add_subplot(223)
skin_df['localization'].value_counts().plot(kind='bar')
ax3.set_ylabel('Count',size=12)
ax3.set_title('Localization')

ax4 = fig.add_subplot(224)
sample_age = skin_df[pd.notnull(skin_df['age'])]
sns.distplot(sample_age['age'], fit=stats.norm, color='red');
ax4.set_title('Age')

plt.tight_layout()
plt.show()
```




```
# Prediction on test data

y_pred = model.predict(x_test)

# Convert prediction classes to one hot vectors

y_pred_classes = np.argmax(y_pred,axis=1)
y_true = np.argmax(y_test,axis=1)

# Print confusion matrix

cm = confusion_matrix(y_true,y_pred_classes)

fig,ax = plt.subplots(figsize =(12,12))
sns.set(font_scale = 1.6)
sns.heatmap(cm, annot=True,linewidths = .5, ax=ax)

# Plot fractional incorrect misclassification

incorr_fraction = 1 - np.diag(cm) / np.sum(cm,axis=1)
plt.bar(np.arange(7),incorr_fraction)
plt.xlabel('True Label')
plt.ylabel('Fraction of incorrect prediction')

Text(76.5, 0.5, 'Fraction of incorrect prediction')
```



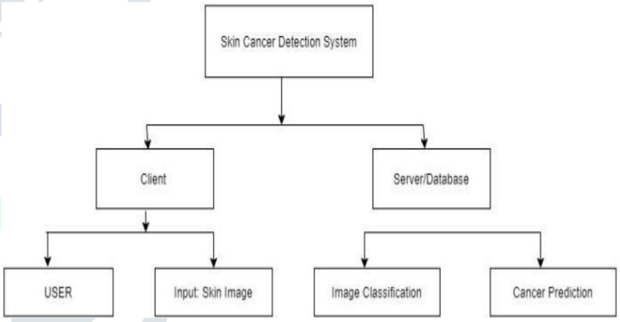




Implementation:

With the addition of an interface, the model becomes much more accessible to everyone. It consists of a sophisticated assisting website that allowed the individual/patient to submit the photograph in real time. As long as the image file is in.jpeg format, the image that must be uploaded can come from any imaging equipment that focuses on the patch/lesion on the skin. The uploaded image is first sent to the model for pre-processing before being sent to the CNN architecture. The system's information gained during the training phase is used to generate the results. The output is then sent to the GUI, where it is displayed as follows: 'Nothing to Worry About!' if the image submitted shows no signs of cancer.If the image is predicted to be benign, and 'The image submitted indicates some cancer indications!' If cancer is suspected in the image, you should seek professional help.

VII. SYSTEMBREAKDOWNREQUIREMENT



V. REQUIREDANALYSIS

RequirementAnalysis:

Skin cancer, particularly melanoma, is a growing public health concern. Deep learning (DL) algorithms may play a diagnostic role in identifying SC at various sensitivities, according to experimental studies. Previously, it was demonstrated that adding a signification (data to sound waves conversion) layer to DL algorithms improves dermoscopy diagnostics. The study's goal was to see how image quality affected the accuracy of diagnosis by signification using a simple skin magnifier with polarized light (SMP).

VIII. PROJECTESTIMATION

KLOC estimation:

Assessment is the process of determining a gauge, or estimation, which is a value that can be used for some reason regardless of whether the input data is insufficient, questionable, or shaky. Assessment determines how much money, effort, assets, and time will be required to construct a specific framework or item.

IX. SYSTEMDESIGN

ProjectSchedulingandTracking:

Project Scheduling and Tracking is critical because, in order to build a complex framework, numerous programming tasks must be completed concurrently, and the outcome of work completed during one task may have a significant impact on work directed in another errand. Without an itemized plan, these bury conditions are extremely difficult to comprehend.

Module Details:

The modules that will be implemented in this system are listed below.

1. User
2. Input:Skinimage
3. Imageclassification
4. Cancerprediction

X. ANALYSISMODELLING

Class graphs, grouping or coordinated effort outlines, and state outline charts comprise the framework examination model. They create an intelligent, execution-free perspective on the PC framework that includes a detailed meaning of each useful component. The following displays are included in the examination model:

1. Behavioralmodeling.
2. Functionalmodeling.
3. Architecturalmodeling.

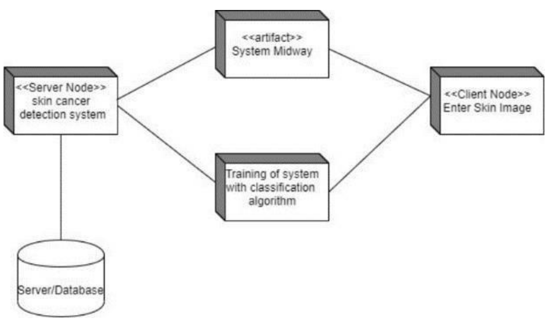
The examination demonstration makes use of a combination of text and diagrammatic structure to depict the requirement for information, capability, and conduct in a way that is generally simple and, more importantly, direct to audit for accuracy, culmination, and consistency.

XI. DEPLOYMENT

The Deployment Diagram also aids in demonstrating the actual component of an Object-Oriented programming framework. It depicts the run-time design in a static view and imagines how parts are distributed in an application. It usually entails demonstrating the equipment setups as well as the product segments that lived on.

XIV. RISKMANAGEMENT

The identification, assessment, and prioritization of risk is followed by the simple and cost-effective use of assets to limit, screen, and control the likelihood or impact of unfortunate events or to improve the recognition of risks. Dangers can arise from a variety of sources, including vulnerability for financial business sectors, risks from project disappointments (at any stage in the plan, advancement, creation, or sustainment life-cycles), legal liabilities, and intentional attack from a foe, or events of the unsure or unusual main driver. When the dangers are identified, the danger chief plans for themminimizing or eliminating the negative effects of adversity Depending on the type of risk and business examined in the following section, a variety of techniques are available



XIII.CONCLUSION

We've discussed a PC-backed conclusion framework for melanoma skin disease. It is concluded from the findings that the proposed framework can be effectively used by patients and doctors to analyze skin cancer more precisely. This instrument is more useful in rural areas where clinical specialists may be difficult to find. Because the apparatus is designed to be simple to use and robust for images obtained under any conditions, it can fill the need for pre-programmed skin cancer diagnostics. The procedures and techniques that are useful in the process were referenced in each progression. The robotized skin disease framework can be designed as a replacement for the clinician in melanoma analysis.

REFERENCE

[1]VijayalakshmiMM"MelanomaSkinCancerDetection using Image Processing and Machine Learning"PublishedinInternationalJournalofTrendinScientificResearchand Development(ijtsrd),ISSN:2456-6470, Volume-3—Issue-4,June2019

[2]W. Zhou, M. Yang, H. Li, X. Wang, Y. Lin, and Q.Tian,"Towardscodebook-free:Scalablecascadedhashingformobileimagesearch,"IEEETransactionsonMultimedia,pp.601–611,April 2014.

[3]M.A.Farooq,M.A.M.Azhar,andR.H.Raza,AutomaticLesionDetectionSystem(ALDS)forSkinCancerClassification UsingSVMandNeuralClassifiers.Oct2016

[4]InternationalResearchJournalofEngineeringandTechnology(IRJET)e-ISSN:2395-0056Volume:04Issue: 04—Apr-2017www.irjet.net

[5]R. Suganya, An automated computer-aided diagnosisof skin lesions detection and classification for dermoscopyimages. April2016.

[6]M.A.Farooq,M.A.M.Azhar,andR.H.Raza,AutomaticLesionDetectionSystem(ALDS)forSkinCancerClassification UsingSVMandNeuralClassifiers.Oct2016.