Journal of Information Systems Engineering and Management
2024, 9 (4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Skin Cancer Detection Using Machine Learning

Dr. D. R. V. A. Sharath Kumar[1*], K. Narasimha Rao[2], V. Sundara Siva Kumar[3], Dr. V. Vijayasri Bolisetty[4]

[1*]Associate Professor Maturi Venkata Subba Rao (MVSR) Engineering College, Hyderabad. drvask_ece@mvsrec.edu.in
[2]Assistant Professor, Maturi Venkata Subba Rao (M.V.S.R) Engineering College
[3]Assistant Professor, Maturi Venkata Subba Rao (MVSR) Engineering College Hyderabad
[4]Associate Professor Department of ECE Aditya University Surampalem. vasudha.tweety@gmail.com

ARTICLE INFO

Received: 11 Oct 2024

Revised: 04 Nov 2024

Accepted: 02 Dec 2024

ABSTRACT

There is a growing need for early diagnosis of skin cancer because of the rapid growth rate of melanoma skin cancer, its high treatment costs and high mortality rate. The detection of skin cancer cells was usually done manually, and most cases require a lengthy cure. Currently the main problem in skin cancer detection is high misclassification rate and low accuracy. This paper provides a technique based on deep learning techniques to detect the cancer from skin images. Convolutional neural network-based model consisting of six layers with hidden layers is used in this work. The problem of low accuracy is addressed with the help of regularization technique and features are selected with the help of convolution method. To improve the accuracy of the model hyper parameter tuning along with model parameter tuning are performed. Publicly available dataset is used in the research which contains images with cancer and normal instances. The major step in this work includes data collection, preprocessing, data cleaning, visualization, and model development. At the end a comparative analysis is performed with state-of-the-art techniques. The proposed model achieved good accuracy of 88% on HAM dataset as compared to state of the art techniques.

Keywords: Deep learning, CNN, Skin Cancer, Melanoma, Detection, Diagnosis

1. INTRODUCTION

The skin is one of the largest organs in the human body and regulates body temperature in addition to shielding the body from extreme heat and light. Additionally, it is used to store water and fat. Skin cancer can also develop in cases when skin cells are damaged, such as by excessive ultraviolet (UV) exposure. ln nations like Canada, the USA, and Australia, skin cancer is rising quickly and at a higher rate. The risk of skin cancer infection is one of the most significant issues with skin in the body. Skin cells, the primary building blocks of the skin, are where skin cancer starts. Skin cells develop and divide to create new cells as they expand. Every day, when skin cells age, they die and are replaced by new cells that take their place. This methodical approach may occasionally go wrong. When the skin may not require them, new cells are created, and when there is no need, existing cells may degenerate. These kinds of excess cells produce tumour, which are masses of tissue. Melanoma is the most common and dangerous type of skin cancer and the one that causes the most mortality out of all the different subtypes. Although the exact origins of melanoma are still unknown. There are several contributing variables, including parental heredity and ultraviolet light. Despite the great likelihood of recovery from this illness, it is nonetheless seen as a major problem due to its pervasive nature. Research suggests that if we can identify this cancer at an early stage, it may assist to lower the likelihood of mortality. Therefore implementing a method that can automate the diagnosis process will be useful and eliminate manual errors as a result. According to numerous studies conducted over the past few years, the use of computer vision techniques is highly commendable and growing at an exponential rate. Therefore, employing these techniques can speed up diagnostics and also reduce errors. Artificial neural network, has recently gained popularity in domains including computer vision, digital image processing, and image categorization techniques. Artificial neural networks are brain inspired with several neuron layers. These have been extended with convolutional neural networks (CNNs), which are producing notable results even for broad and challenging tasks like object detection. Several applications in various medical imaging techniques, such as lesion classification, MR image fusion, breast cancer and tumor identification, and panoptic analysis make use of CNNs due to their high and efficient performance. CNN model based pertained models like Reset and VGGNET are equally good for image processing tasks. This paper provides a CNN based model consisting of six layers with hidden layers. The problem of low accuracy is addressed with the help of regularization technique. To improve the accuracy of the model hyper parameter tuning along with model parameter tuning

Journal of Information Systems Engineering and Management
2024, 9 (4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

are performed. Publicly available dataset is used in the research which contains images with cancer and normal instances. The step in the research includes data collection, Preprocessing, data cleaning, visualization, and model development. For data preprocessing several techniques are used like standardization, normalization, and feature scaling. The major contributions are as follows: • Propose an efficient technique which can classify cancer images from benign images. • To improve the performance of the CNN low accuracy is addressed with the help of regularization technique. To improve the accuracy of the model hyper parameter tuning along with model parameter tuning are performed. Publicly available dataset is used in the research which contains images with cancer and normal instances. The step in the research includes data collection, Preprocessing, data cleaning, visualization, and model development. For data preprocessing several techniques are used like standardization, normalization, and feature scaling. The major contributions are as follows:

- Propose an efficient technique which can classify cancer images from benign images.
- To improve the performance of the CNN model by hyper parameter tuning of the model.

LITERATURE SURVEY

1. The paper titled Skin Cancer Detection Using SVM algorithm by Brindha et al in the classification of ISIC image dataset, showing a considerable increase in accuracy from 61% to 83%.
2. The paper titled Skin Cancer Detection Using Transfer Learning methods, namely, Reznet50, and InceptionV3 by Pham and his colleagues were also able to achieve accuracy of 79.5% and 87%, respectively.
3. The paper titled Skin Cancer Detection Using three different architectures; namely, VGG19, ResNet, and Inception V3 by Mijwil exploited and compared by using ISIC archives between 2019 and 2020 with a considerable number of more than 24000 images for skin cancer detection and present the accuracy of each architecture to be 73.11%, and the best 86.9%, respectively.
4. . The paper titled Skin Cancer Detection Using region-based CNN technique with support vector Machine (SVM) classifier by Nawaz et al. employed a dataset based on ISIC-2016 with an increased number of images by using data augmentation step with more than 7000 images for Melanoma classification and obtained accuracy of 89.1%.
5. The paper titled Skin Cancer Detection Using deep learning method by Ashraf et al. considered examination of skin lesion images including a region of interest segmentation preprocessing and also image augmentation. The result without region of interest segmentation and augmentation was about 81.3%.
6. The paper titled Skin Cancer Detection Using deep learning method by Alzubaidi and his colleagues were able to reach 87.5% accuracy to classify skin lesion images by implementing a multiphase training scenario and multistage CNN model with the aim to overpass limited number of labelled data for medical application.
7. The paper titled Skin Cancer Detection Using CNN architecture by Lafraxo and coworkers to recognize Malignant among dermoscopic image, in which, they employed regularization, and particularly, geometric and color augmentations to enlarge the datasets; namely, ISBI (to 18000 image), PH2 (to 2880 images), and MED-NODE (to 1800 images), and were able to achieve the accuracies of 81.44%, 87.39%, and 87.77%, respectively.
8. The paper titled Skin Cancer Detection Using ResNet50 and Adam optimizer by Hassan et al. achieved a superior accuracy of 77.47% for the classification of retinal optical coherence tomography images with 84495 total number of images.
9. The paper titled Skin Cancer Detection Using CNN/transformer coupled network by Alahmadi and coworkers. They acquired 85.51% and 84.11% accuracy for ISIC and PH2 datasets images.

EXISTING SYSTEM

Despite being one of the worst kinds of cancer, skin cancer deaths have risen rapidly in recent years. Lack of education about the disease's warning signals and the Identifying cancer early, when it's still treatable, is crucial to preventing its spread. Melanoma, basal cell carcinoma, and squamous cell carcinoma are deadly skin cancers. Atypical basal cell carcinoma and squamous cell carcinoma are other skin cancers. This study uses machine learning and image processing to classify skin cancers. Before preprocessing, skin pictures are entered. After removing unwanted hair with a dull razor, a Gaussian filter is used to smooth the image. The median filter filters noise and maintains lesion margins. In the segmentation step, colour-based k-means clustering is used because colour is an important factor in determining malignancy. ABCD and Gray Level Cooccurrence Matrix extract statistical and textural characteristics. Asymmetry, border colour, and diameter (GLCM). The ISIC 2019 Challenge dataset contains eight kinds of skin pictures. For classification, a Multi-class Support Vector Machine (MSVM) was built with 96.25 percent accuracy.

Journal of Information Systems Engineering and Management
2024, 9 (4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

PROPOSED SYSTEM

The proposed model used in the research consists of five layers of CNN. Image consists of R, G and B channels of dimensions 75x100. The image is then passed to the convolutional layer with filter size of 16 and kernel size of 3x3. Reactivation function is used. After two layers max pooling function is applied. This process is repeated three times with number of filters increased or decreased. After this process the resulting vector is flattened with the help of fully connected layers. This is classification problem and output is in the form of probabilities. Sigmoid function is used at the output layer which clamps the output between 0 and 1. Normalization technique is used to normalize the outputs. In the research there are seven units used at the output layer because there are seven classes in the dataset.

BLOCK DIAGRAM AND ITS DESCRIPTION

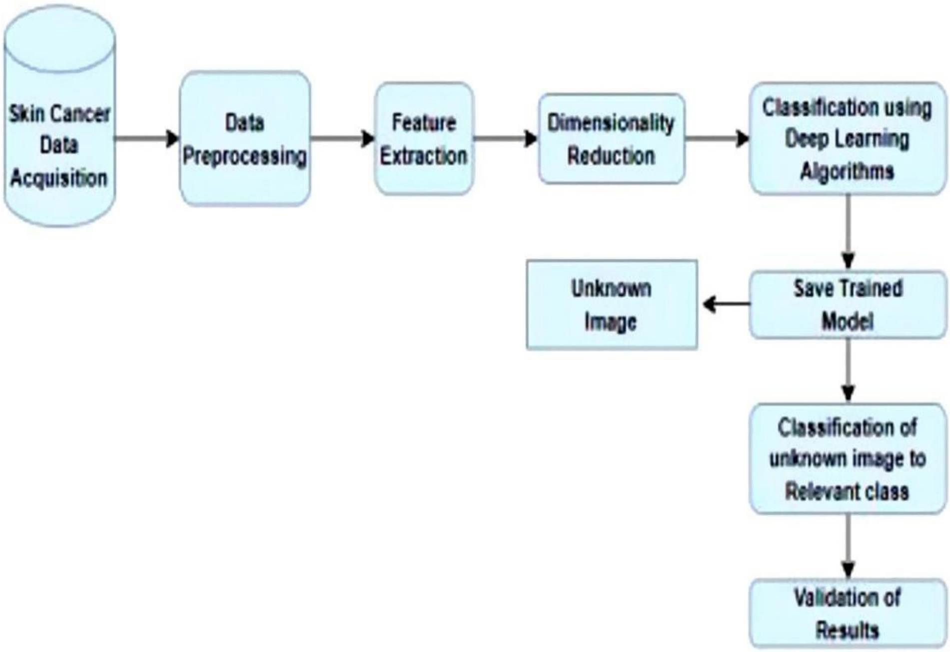


Fig.2.1. BLOCK DIAGRAM OF SKIN CANCER DETECTION

Data Acquisition

• This block is the first step in the process and involves gathering images of skin lesions. These images can be obtained from a variety of sources, such as dermatologists' offices, online databases, or clinical trials.
• It is important to ensure that the data is representative of the real-world population of skin lesions. This means that the data should include lesions of all types, including benign and malignant lesions.
• The data should also be labelled by dermatologists so that the open learning algorithms can be trained to correctly classify new lesions

Data Preprocessing

• Once the data has been acquired, it needs to be preprocessed before it can be used by the open learning algorithms.
• Preprocessing typically involves tasks such as resizing the images, normalization, and data augmentation.
• Resizing the images ensures that all of the images have the same dimensions, which is important for training the algorithms.
• Normalization scales the pixel values of the images to a common range. This can help to improve the performance of the algorithms.
• Data augmentation is a technique that can be used to artificially increase the size of the dataset. This is done by creating new images from the existing images, such as by rotating or flipping the images.

Feature Extraction

• The next step is to extract features from the preprocessed images.
• Features are characteristics of the images that can be used to distinguish between different types of skin

Journal of Information Systems Engineering and Management
2024, 9 (4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

lesions.

• In the context of skin cancer classification, some common features that might be extracted include the shape, color, and texture of the lesion.

• Feature extraction can be done manually by a dermatologist or automatically using a computer vision algorithm.

Dimensionality Reduction

• After the features have been extracted, it is important to reduce the dimensionality of the data.

• Dimensionality reduction refers to the process of reducing the number of features in the data.

• This is important because high-dimensional data can be difficult for open learning algorithms to learn from.

• There are a number of dimensionality reduction techniques that can be used, such as principal component analysis (PCA) and linear discriminant analysis (LDA).

Classification using Deep Learning Algorithms

• Once the data has been preprocessed, features have been extracted, and dimensionality reduction has been performed, the data is ready to be used by the open learning algorithms.

• In the context of skin cancer classification, deep learning algorithms are often used.

• Deep learning algorithms are a type of machine learning algorithm that is inspired by the structure of the human brain.

• Deep learning algorithms have been shown to be very effective for image classification tasks.

Saving Trained Model

• Once the open learning algorithms have been trained and validated, the model can be saved for future use.

• The saved model can then be used to classify new images of skin lesions.

Classification of Unknown Image to Relevant Class

• This block refers to using the trained model to classify a new image.

• The new image is first pre-processed and then features are extracted from it.

• The features are then fed into the trained model, which outputs a classification.

• The classification is the model's prediction of the class of the new image.

Validation of Results

• After the open learning algorithms have been trained, it is important to validate their performance.

• This is done by testing the algorithms on a dataset of unseen images.

• The performance of the algorithms is typically measured by metrics such as accuracy, precision, and recall.

OPERATION OF THE PROJECT

Data Acquisition: The process starts with a vast dataset of dermoscopic images. These images capture close-up views of skin lesions at a magnified level, revealing details invisible to the naked eye. Image Preprocessing: Tensor Flow helps pre-process these images. This might involve resizing them to a standard size, adjusting brightness and contrast for consistency, and potentially removing noise or irrelevant background information.

Convolutional Neural Networks (CNNs): Tensor Flow excels at building and training CNNs. These are a type of deep learning architecture particularly adept at image recognition. CNNs work by learning patterns from the image data. Imagine the CNN dissecting the image into smaller squares (filters) and analyzing the features within each square. As it progresses through layers, it combines these features to recognize more complex patterns, ultimately distinguishing cancerous from benign lesions.

Transfer Learning: Tensor Flow allows leveraging pre-trained models on massive image datasets (like Image Net). These models have already learned powerful image recognition capabilities. By "transferring" this knowledge as a starting point, we can fine-tune the model specifically for skin cancer classification, significantly reducing training time and effort.

Training the Model: Tensor Flow facilitates the training process. Here, the pre-processed images are fed into the CNN architecture. The model analyzes these images, comparing them to labeled data (where each image is classified as cancerous or benign). Through an iterative process (backpropagation), the model adjusts its internal weights and biases to minimize classification errors.

Evaluation and Refinement: Tensor Flow helps evaluate the model's performance. A separate dataset (not used for training) is used to test the model's accuracy. Metrics like sensitivity (correctly identifying cancerous lesions) and specificity (correctly identifying benign lesions) are assessed. Based on the results, the

Journal of Information Systems Engineering and Management
2024, 9 (4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

model architecture or training process might be refined for better performance.
Deployment: Once satisfied with the model's accuracy, Tensor Flow can be used to deploy it for real-world use. This could involve integrating it into a medical imaging system or even developing a mobile app for accessible skin cancer screening.
Benefits: Tensor Flow-powered skin cancer detection offers several advantages. It can analyze large numbers of images quickly, potentially assisting dermatologists in identifying suspicious lesions they might miss. It can also be used in remote areas where access to specialists is limited.
Limitations: It's important to remember that these models are tools to aid diagnosis, not replacements for a doctor's expertise. Factors like a patient's medical history and a physical examination are crucial for a definitive diagnosis.
Future Advancements: Research is ongoing to improve the accuracy and efficiency of these models. Additionally, integration with other technologies like telemedicine holds promise for even more accessible and effective skin cancer screening. Tensor flow is a powerful framework for building machine learning models, including those for skin cancer detection. Here's a breakdown of the process:

DATA COLLECTION

Data collection means pooling data by scraping, capturing, and loading it from multiple sources, including offline and online sources. High volumes of data collection or data creation can be the hardest part of a machine learning project, especially at scale. Data collection allows you to capture a record of past events so that we can use data analysis to find recurring patterns. From those patterns, you build predictive models using machine learning algorithms that look for trends and predict future changes. Predictive models are only as good as the data from which they are built, so good data collection practices are crucial to developing high-performing models. The data needs to be error-free and contain relevant information for the task at hand. For example, a loan default model would not benefit from tiger population sizes but could benefit from gas prices over time.

Test Dataset

Once we train the model with the training dataset, it's time to test the model with the test dataset. This dataset evaluates the performance of the model and ensures that the model can generalize well with the new or unseen dataset. The test dataset is another subset of original data, which is independent of the training dataset. However, it has some similar types of features and class probability distribution and uses it as a benchmark for model evaluation once the model training is completed. Test data is a well-organized dataset that contains data for each type of scenario for a given problem that the model would be facing when used in the real world. Usually, the test dataset is approximately 20-25% of the total original data for an ML project.

CONVOLUTIONAL NEURAL NETWORK (ALGORITHM)

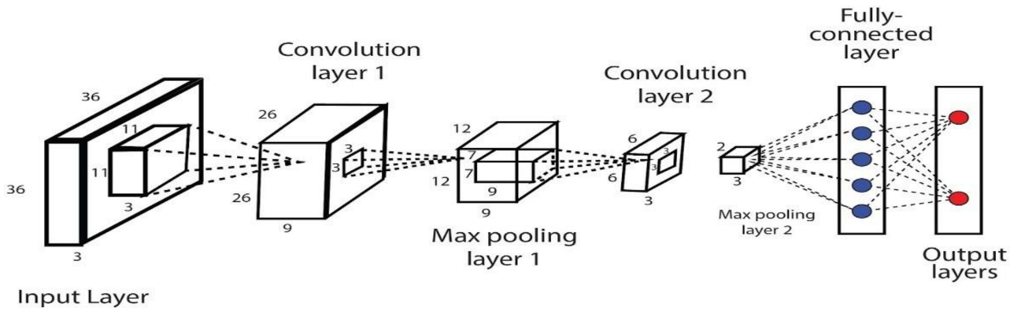


Fig.2.3. CNN ALGORITHM

Convolutional Layer 1

- This is the first layer of the CNN. It takes the input image, which is typically a 3D array of pixels (width, height, and color channels).
- The layer applies filters (or kernels) to the input image. These filters are small matrices of weights that are learned during training.
- The filters are applied to the image in a sliding window fashion, producing a feature map. The feature map shows how well the filter matches the input image at each location.
- In the image, the filter size is not specified, but there are 36 output channels, which means there are 36

Journal of Information Systems Engineering and Management
2024, 9 (4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

different filters being applied to the input image.

Max Pooling Layer 1

• This layer performs down sampling on the output of the convolutional layer.
• Down sampling reduces the dimensionality of the data, which can help to improve the performance of the CNN and reduce overfitting.
• There are a variety of pooling operations, but max pooling is a common choice. Max pooling takes the maximum value from a rectangular region of the input feature map. The size of this region is called the pooling window.
• In the image, the pooling window size is 2x2, and there is a stride of 2. This means that the pooling window is moved two pixels at a time across the input feature map.

Convolutional Layer 2

• This layer works similarly to convolutional layer 1, but it takes the output of the pooling layer as input.
• The number of filters used in this layer is typically smaller than the number of filters used in the first layer. This is because the pooling layer has already reduced the dimensionality of the data.
• In the image, there are 26 output channels, which mean there are 26 different filters being applied to the output of the pooling layer.

Max Pooling Layer 2

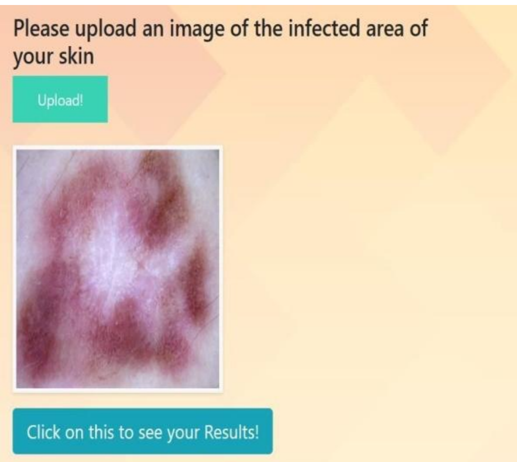
• This layer works similarly to pooling layer 1, but it takes the output of convolutional layer 2 as input.
• The pooling window size and stride are also typically the same as in pooling layer 1.
• In the image, the pooling window size is not specified, but the stride is 2.

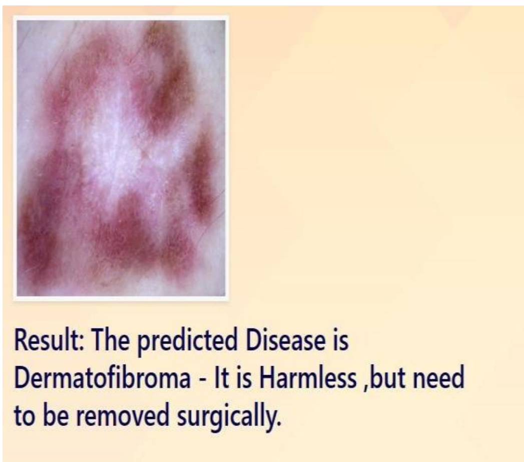
Fully-connected layer

• The image you sent does not contain a fully-connected layer. Fully-connected layers are typically used in CNNs after the convolutional and pooling layers.
• Fully-connected layers are layers where every neuron in the layer is connected to every neuron in the previous layer.
• This is in contrast to convolutional layers, where only a small subset of neurons in the layer is connected to each neuron in the previous layer.
• Fully-connected layers are used to learn more complex relationships between the features extracted by the convolutional layers.

Results:

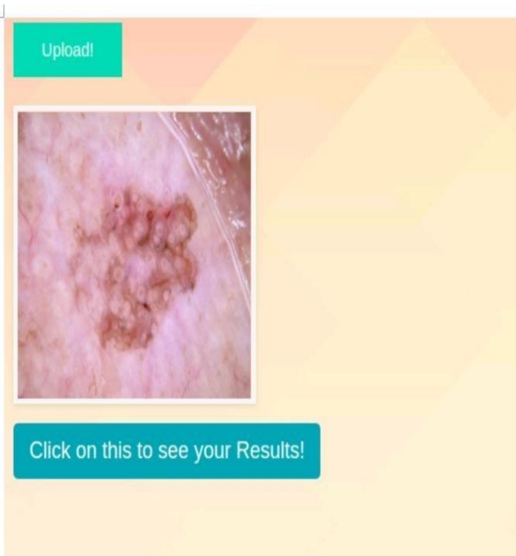
UPLOADED IMAGE:

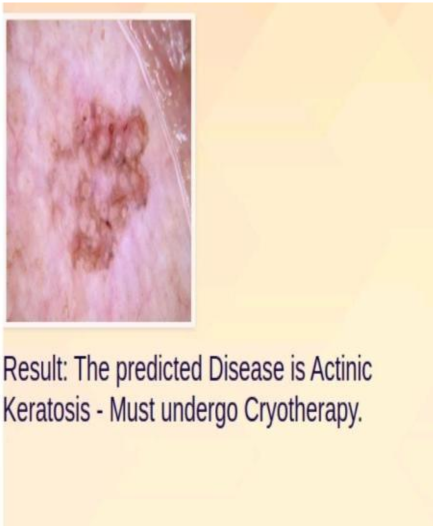




Journal of Information Systems Engineering and Management
2024, 9 (4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

UPLOADED IMAGE





PREDICTED DISEASE OF UPLOADED IMAGE

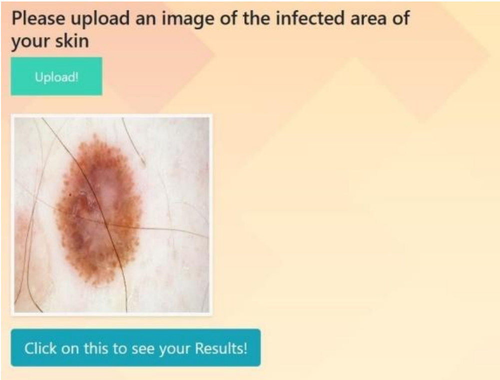
UPLOADED IMAGE:

UPLOADED IMAGE

PREDICTED DISEASE OF UPLOADED IMAGE:

Journal of Information Systems Engineering and Management
2024, 9 (4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

PREDICTED DISEASE
UPLOADED IMAGE:



UPLOADED IMAGE
PREDICTED DISEASE OF UPLOADED IMAGE:

PREDICTED DISEASE

Test Results: All the test cases mentioned above passed successfully. No defects encountered. User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.
CNNs (Convolutional Neural Networks) are particularly adept at image analysis tasks like skin cancer detection. In this study, a CNN-based model is proposed to distinguish cancerous lesions from benign ones. The model incorporates regularization techniques and hyper parameter tuning to optimize its performance. Regularization helps prevent overfitting, a phenomenon where the model performs well on training data but poorly on unseen data. Hyper parameter tuning involves adjusting the model's parameters to achieve the best possible accuracy. By combining these techniques, the researchers aim to create a robust and accurate skin cancer detection system.

CONCLUSION

An approach based on Convolutional Neural Networks has been proposed in this paper for melanoma classification. This system is designed to assist patients and doctors in identifying benign or malignant skin cancer classes. Based on the experimental and evaluation sections, the model can be considered as a benchmark for the detection of skin cancer by healthcare professionals. Taking some random images can help any doctor identify accurate results, but the traditional approach takes too much time to detect cases correctly. CNN model in the research achieves 88% of accuracy with low false positive rates. Many researchers used pertained models like Alex net, VGGNET 16 and Mobile Net but they all are based on pertained library without any intervention of data modeling. In this research CNN is used with 6 layers and trained on our own infrastructure. 100 epochs are used in the project however if more than 100 are used then accuracy is increased to a certain level.

Journal of Information Systems Engineering and Management
2024, 9 (4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

REFERENCES

[1] R. Ashraf et al., "Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection," IEEE Access, vol. 8, pp. 147858-147871, 2020, doi: JO.I 109/ACCESS.2020.3014701.
[2] M. Dildar et al., "Skin cancer detection: A review using deep learning techniques," Int. J. Environ. Res. Public Health, vol. 18, no. 10, 2021, doi: 10.3390/ijerphl8105479.
[3] R. Kaur, H. Gholamhosseini, R. Sinha, and M. Linden, "Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopic Images," Sensors, vol. 22, no. 3, pp. 1-15, 2022, doi:
[4] K. Aljohani, "Automatic Classification of Melanoma Skin Cancer with Deep Convolutional Neural Networks," pp. 512-525, 2022.
[5] N. Kausar et al., "Multiclass skin cancer classification using ensemble of fine-tuned deep learning models," Appl. Sci., vol. 11, no. 22, pp. 1-20, 2021, doi: 10.3390/appl 12210593.
[6] S. Bechelli and J. Delhommelle, "Machine Leaming and Deep learning Algorithms for Skin Cancer Classification from Dennoscopic Images," Bioengineering, vol. 9, no. 3, pp. 1-18, 2022, doi: I 0.3390/bioengineering9030097.
[7] 0. Sharif, M. M. Hoque, A. S. M. Kayes, R. Nowrozy, and I. H. Sarker, "applied sciences Learning Techniques," pp. 1-23, 2020.
[8] I. Kousis, I. Perikos, I. Hatzilygeroudis, and M. Virvou, "Deep learning Methods for Accurate Skin Cancer Recognition and Mobile Application," Electron., vol. 11, no. 9, pp. 1- 19, 2022, doi: I0.3390/electronicsl I091294.
[9] M. Arif, F. M. Philip, F. Ajesh, D. Izdrui, M. D. Craciun, and 0. Geman, "Automated Detection of Nonmelanoma Skin Cancer Based on Deep Convolutional Neural Network," J. Health. Eng., vol. 2022, 2022, doi: I0. l I55/2022/6952304.
[10] M. M. Mijwil, "Skin cancer disease images classification using deep learning solutions," Multimedia. Tools Appl., vol. 80, no. 17, pp. 26255-26271, 2021, doi: I0.1007/s11042-021-!0952-7.