Skin Disease Prediction Using Dynamic Testing in Machine Learning

[1] Prof. T.A.Dhumal , [2]Miss. Sakshi Shivsharan,[3]Miss. Aparna Vasmale

[4]Miss. Pallavi Jadhav,[5]Miss. Shruti Adat,[6]Miss. Anushka Chavan

SVERI's COLLEGE OF ENGINEERING,PANDHARPUR,MAHARASHTRA,INDIA.

Abstract:

One of the most prevalent illnesses in the modern world is skin disease. A small circular or haphazardly shaped area on the patient's skin can be seen as a result of skin disease. When this illness develops into skin cancer, it can occasionally be very dangerous. This article discussed a few deep learning-based methods for extracting features from various skin cancer images, which can then be used with machine learning classifiers to identify the type of skin disease. A transfer learning model is created for our experiments in which a support vector machine is used for classification and the VGG- 16 layer CNN architecture is used to extract 1000 features from the input image.

Keywords: Convolutional Neural Network, Deep Learning, Melanoma, Skin Lesion, Support Vector Machine, Transfer Learning.

1.INTRODUCTION:

Skin cancer is one of the most common types of cancer. It is found in nearly half of all cancer patients. Skin cancer is typically found and classified into several types, including Melanoma, Basal, Squamous, Benign, and others. Melanoma is one of the scariest types of cancer found and detected to date, with approximately 7300 deaths per year in the United States. For skin cancer detection, two types of images are available. The skin image is captured by a specialised dedicated system in the pathological centre with a focus on the region of interest and a high zoom (E.g. 20x), which requires a skilled dermatologist to determine whether the image is positive or negative . A computerised semi-automated system can be used to feed this kind of image into it for classification. However, with today's technology, the victim must always visit a pathological centre and seek the advice of a qualified dermatologist The victim can perform the test whenever they want, even at home, if there is computer software that can automatically detect skin cancer from a digital image captured by any digital image capturing system with little focus on the region of interest.

2. LITERATURE SURVEY:

1. Title: Skin Cancer Diagnosis Prediction System Using Data Mining Classification Methods

Authors include Dr. G. Narsimha, Dr. N. Subhash Chandra, and V. Krishna.

Both in terms of mortality rates for men and women, cancer ranks first.

Early cancer detection can aid in the complete eradication of the disease.

Therefore, there is a growing need for methods to identify cancer nodule 5 at an early stage. Lung cancer is a condition that is frequently misdiagnosed. Early detection of lung cancer saves many lives; if not, it may result in other serious issues that hasten death. The likelihood of cure and prognosis are largely dependent on the disease's early discovery and diagnosis. A diagnostic error is one of the most frequent types of medical misconduct worldwide. Information seeking and Data mining has many uses in both the corporate and scientific worlds. The application of data mining techniques in the healthcare system can yield useful knowledge. In this paper, we briefly explore the application of rule-based, decision tree, naive bayes, and artificial neural network classification-based data mining techniques to enormous volume.

2. Title: A Fully Automated Method for Detecting Skin Nodules in Postero-Anterior Chest Radiographs

Paola Campadelli, Elena Casiraghi, and Diana Artioli are the authors.

Description:

A significant amount of scientific effort has been committed in recent decades to the development of systems that could increase radiologists' accuracy in detecting skin nodules. Despite significant efforts, the situation remains unresolved.

We offer a completely automated approach for processing digital poster-anterior (PA) chest radiographs that begins with an accurate segmentation of the skin field region. The segmented lung area contains lungs concealed behind the heart, spine, and diaphragm, which are typically excluded from methodologies provided in the literature. This choice is driven by the fact that nodules may also be present in these regions. The segmented area is processed using a straightforward multiscale technique to make the nodules more visible, and an extraction scheme is then used to choose prospective nodules. Cost sensitive Convolution neural networks (CNNs) are trained to identify the genuine nodules in order to decrease the high number of false positives extracted. On two independent data sets that were produced using different methods, various learning experiments were run.

3. Title: A Method for Feature Selection and Discretization of Continuous-Valued Attributes in Medical Images for Classification Learning

Madhu Kumari and Tajinder Singh, authors

Description:

A discrete feature space is necessary for many supervised machine learning techniques. In this article, we examine earlier work on continuous feature discretization and pinpoint its distinguishing features. Then, we suggest a novel supervised method that chooses the most pertinent features for classification by combining discretization with feature selection. Associative Classifiers are to be employed as the classification method. Horlick Texture features that were taken from MRI images are the features that are employed. The findings demonstrate how effective and appropriate the suggested strategy is for pre-processing continuous valued features.

3. METHODOLOGY:

3.1 Dataset Used

Many academics use online-available data sets like the ISIC, PH2, and EDRA datasets in their work. Here, the ICIS public dataset was used for tests. Downloading the dataset is possible at https://www.isic-archive.com. More than 10,000 photos of benign illness and melanoma are included in the data collection. The fact that the data sets have an imbalance problem indicates that data balancing should be used whereas benign images correspond to skin cancer at an earlier stage. A subset of the dataset is formed using 20% random photos of both the class and the dataset for experimental reasons, and the experimental dataset contains 1000 images of both classes.

3.2 Method and Algorithms

The benign and malignant types of images are shown as two input classes in Figure 1, and these images are then fed into the CNN deep learning model for feature extraction. The CNN model 2016 is used to extract 1000 features from a single image, after which all of the images were manually labelled and saved as text files in the following phase. Finally, the feature with labels in this text file makes it simple to apply any supervised type of machine learning algorithm to the classification problem. Support vector machines, decision tree linear regressions, and the K-Nearest Neighbour algorithm are a few examples of machine learning techniques that are used for categorization. The confusion matrix and ROC curve are used to evaluate the performance of each classifier.

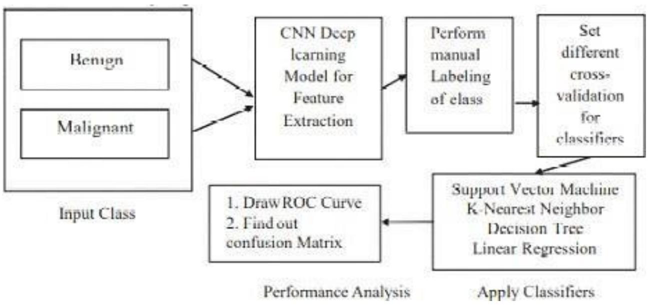


Fig 1. Flowchart of the Proposed Methodology

3.3 Algorithm SPBL Algorithm

Input: Training dataset (x1, y1), , (xn, yn) Output: Classifier parameter w

1: Initialize the model with a pretrained CNN and classifier parameters w;

2: Initialize the SP-regularize f, latent weight variables v and pace parameter λ;

3: Predetermine the initial curriculum Φ

4: repeat

5: Update w via Eq. 12;

6: Update v via Eq. 4, and then get the curriculum Φ;

7: Update the complexity level of each class via Eq. 7;

8: Update penalty weight parameter ω via Eq. 10;

9: Update reconstructed curriculum Φ∗ and weight ω∗ via Algorithm 1;

10: Tune the CNN model and extract features;

11: In every T epochs:

12: Augment λ;

13: until Model converge

14: return w

3.4 Dataset

This dataset consists of 10000 images of malignant and benign kind diseases, which were formed from TheHam that is Human against machine dataset We grouped images according to the ISIC classification, and all subsets were divided into the same number of images, except for malignancies and moles, whose images are slightly more prominent.

The data set contains the following diseases:

- basal cell carcinoma

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

- melanoma
- squamous cell carcinoma
- benign lesion
- vascular keratosis

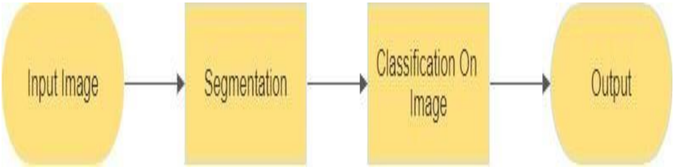


Fig 2. Dataflow Diagram

A. Input Image:

An input image was obtained from different and well-known databases. The input image contains basically two types of skin diseases first benign and another malignant. The benign type of image shows the Cancer disease at an early stage whereas the malignant type of image shows cancer in a later stage. The data set contains thousands of images of both the classes.

B. Convolution Neural Network Working:

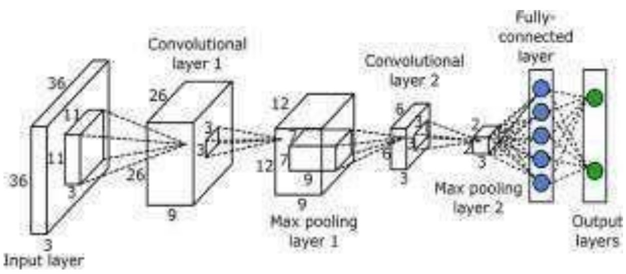


Fig 3. Convolutional neural network

A convolutional neural network is the deep learning algorithm that simulates the visual cortex of the human brain. The idea of the convolution neural network is that filters the image before training the deep neural network. It typically uses four layers i.e convolution, pooling, ReLU, and fully connected layer.

1. Convolution Layer of Deep Learning:

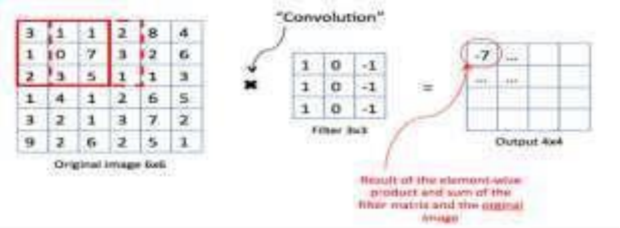


Fig 4. Convolution Layer

3.5 Diagram

2. Strides

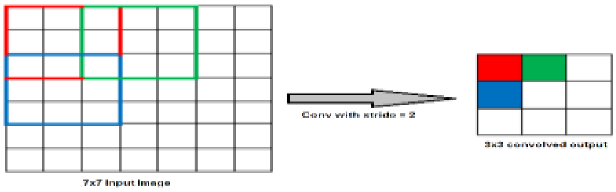


Fig 5. Strides

3. Padding Working

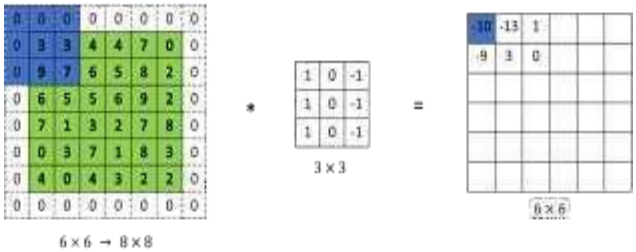


Fig 6. Padding

4. Rectified Linear Unit(ReLU)

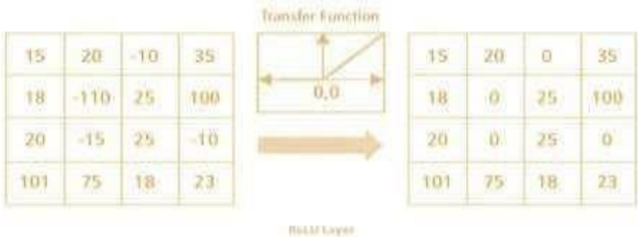


Fig 7. Rectified Linear Unit Function

5. Pooling

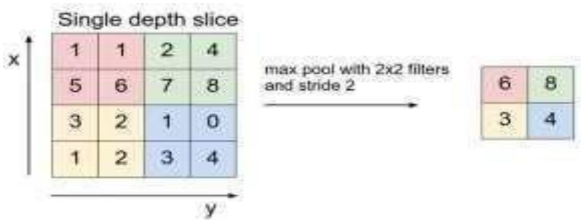


Fig 8. Pooling

C. Feature Extraction Using CNN Model:

For feature extraction, during classification, the convolutional neural network model are used in which is a well- known model for feature extraction from the image input. Here, both the input image uses the VGG 16

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

to make 16 layer architecture. Each convolution layer contains filters of size 3* 3 at each layer. The pooling layer is used to reduce the feature extracted from the convolution layer. At the first stage examine the lowlevel features of the image and at a later stage, these filters are used to extract the high- level and middlelevel features from the image.The architecture of VGG 16 is shown in figure 4

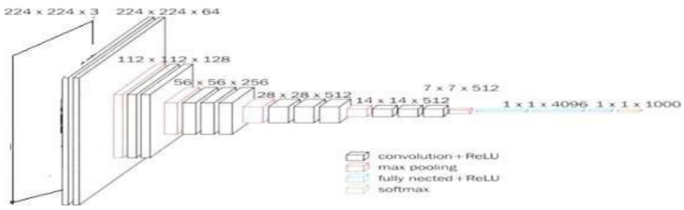


Fig 9. Architecture of VGG-16 CNN Model

D. Transfer Learning:

For incomplete data sets, different supervised and semi-supervised learning techniques have been investigated. However, the majority of them make the assumption that the distributions of the labelled and unlabeled data are equivalent. The domains, tasks, and distributions utilised for training and testing might vary with transfer learning, though. In this instance, there are two stages of setup that use the transfer learning paradigm. The VGG-16 model was used for deep learning in the first setup to extract the low-level, middle-level, and high-level features from the input dataset. The labels are then added to the features that Satge1 extracted from the multiple input photos on a second level. Finally, all the features are sent to the supervised machine learning model to perform the classification as shown in figure 5. The classifiers like support vector machine, decision tree, KNN, and Linear Discriminate analyzer are used for classification because they are suitable for binary and linear classification.

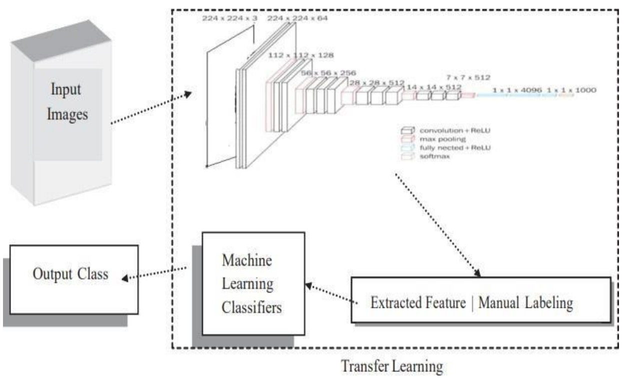
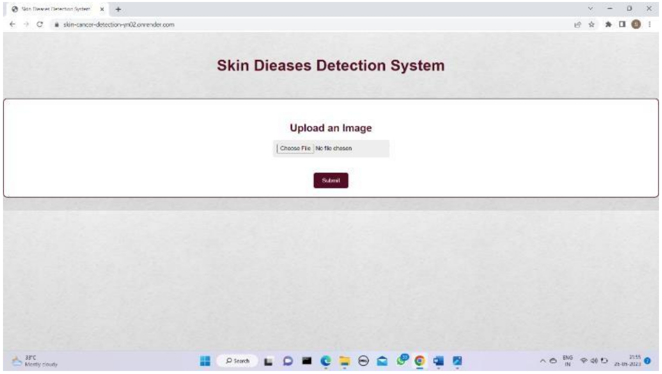
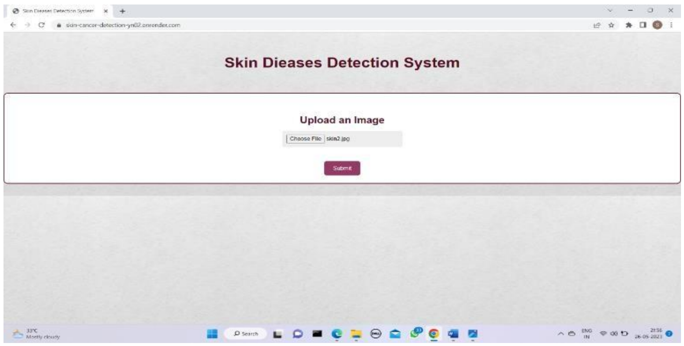
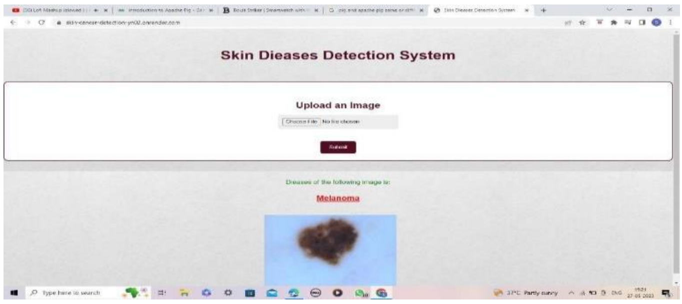


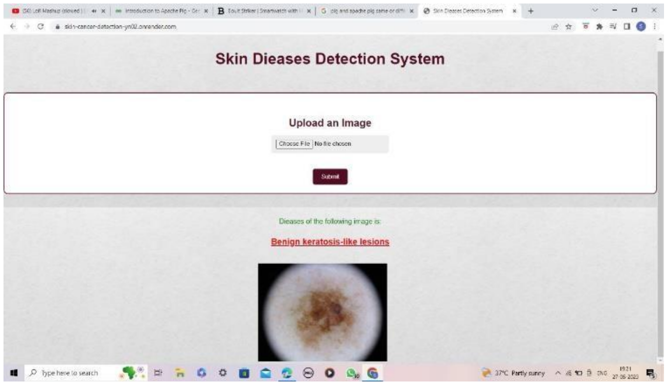
Fig 10. Proposed Transfer learning model for classification

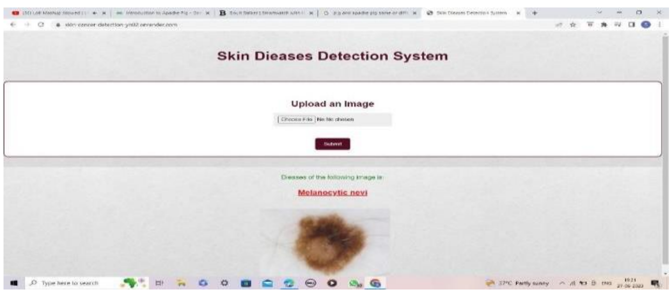
4.RESULTS











5. CONCLUSION

The proposed work may improve classification accuracy up to the signature mark through the use of the transfer learning approach. The accuracy was marginally lower than the transfer learning model when the same data was applied to a pure deep learning model. The transfer learning model requires some human intervention during the labelling of data and has a longer observed execution time than deep learning models. According to experimental findings, some models, such as decision trees and K nearest neighbours, are accurate in more than 99% of iterations.

While less than 50% of complex models, such as ensemble learning with boosted trees, perform as expected. Therefore, the aforementioned dataset is suitable for linear binary classifiers.

6. REFERENCES

1. Maron, Roman C., et al. "Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks." European Journal of Cancer 119 (2019): 57-65.

2. Tschandl, Philipp, et al. "Expert -level diagnosis of nonpigmented skin cancer by combined convolutional neural networks." JAMA dermatology 155.1 (2019): 58-65.

3. Xue, Cheng, et al. "Robust Learning at Noisy Labeled Medical Images: Applied to Skin Lesion Classification." arXiv preprint arXiv:1901.07759 (2019).

4. He, Xinzi, et al. "Dense deconvolution net: Multi path fusion and dense deconvolution for high resolution skin lesion segmentation." Technology and Health Care 26.S1 (2018): 307-316.

5. Burdick, Jack, et al. "Rethinking skin lesion segmentation in a convolutional classifier." Journal of digital imaging 31.4 (2018): 435-440.

6. Brinker, Titus J., et al. "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task." European Journal of Cancer 113 (2019): 47-54.

7. Zhang, Xinyuan, et al. "Towards improving diagnosis ofskin diseases by combining deep neural network and human knowledge." BMC medical informatics and decision making 18.2 (2018): 59.

8. Hekler, Achim, et al. "Pathologist-level classification of histopathological melanoma images with deep neural networks." European Journal of Cancer 115 (2019): 79-83

9. Tschandl, Philipp, Christoph Sinz, and Harald Kittler. "Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation." Computers in biology and medicine 104 (2019): 111-116.