International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

# Design and Development of Skin Cancer Prediction Using Machine Learning Over Live Cloud Infrastructure

**Dr. Prasanna Lakshmi G S**

Associate Professor, Department, of CSE, Nagarjuna College of Engineering and Technology, Devanahalli, Bengaluru
Email: *prasannalakshmigs[at]gmail.com*

**Abstract:** ***According to world cancer Research fund 30, 000 people are affected by skin cancer per year. Skin cancer the abnormal growth of skin cells, most often develops on skin exposed to the sun. But this common form of cancer can also occur on areas of your skin not ordinarily exposed to sunlight. There are two major types of skin cancer are Melanoma, Benign. Automated diagnosis of various skin lesion diseases through medical dermoscopy images is still a very challenging task. In this paper, an integrated model for segmentation of skin lesion boundaries and classification of skin lesions is proposed by cascading novel deep learning networks. In the first stage, a novel full resolution convolutional networks (FrCN) is utilized to segment the boundaries of skin lesions from dermoscopy images. Then, the segmented lesions are passed into a deep residual network for classification.***

**Keywords:** Computer Aided Diagnosis, support vector machine

## 1. Introduction

Technology plays a central role in our everyday life. Technology assistance at various stages of skin lesion processes can significantly enhance the segmentation and classification using deep learning. Proper learning of deep learning networks demands a huge number of training samples. However, a limit on the size of medical image dataset, especially a limit on reliable annotated ground - truths, is one of the challenges in adopting such deep learning approaches. We have applied different augmentation operations to the training datasets to enlarge the amount of training data, reduce the over fitting problem, and accelerate the convergence. Skin Lesion Boundary Segmentation via FrCN and Skin Lesion

Classification via ResNet - 50 are the two new approaches for segmentation and classification of skin lesion respectively, the full resolution convolutional network (FrCN) is an end - to - end supervised deep network which is trained via mapping the entire input image to its corresponding ground - truth masks with no loss, leading to result better segmentation performance of skin lesion boundaries. A deep residual network (i. e., ResNet) is one of the deep learning classification models which has been used in many image recognition applications. ResNet has the ability to address the vanishing gradient problem when deep networks go deeper (i. e., increasing network depth by stacking layers). In other words, instead of passing the learned features directly through the stacked layers, ResNet enables these layers to fit a residual mapping leading to optimize the network easier than the unreferenced mapping.

**Advantages**

- Solutions is been made available over the cloud using as - a - service model thus extending the availability of the solution across the globe.
- Most accurate
- Simple and computationally light weight thus saving time and server memory

## 2. Related Work

The current system from last two decades have witnessed a lot of efforts that attempt to provide CAD systems capable of distinguishing between melanoma and non - melanoma. However, these early investigations relied on applying low - level hand - crafted features including color, shape, and texture representations.

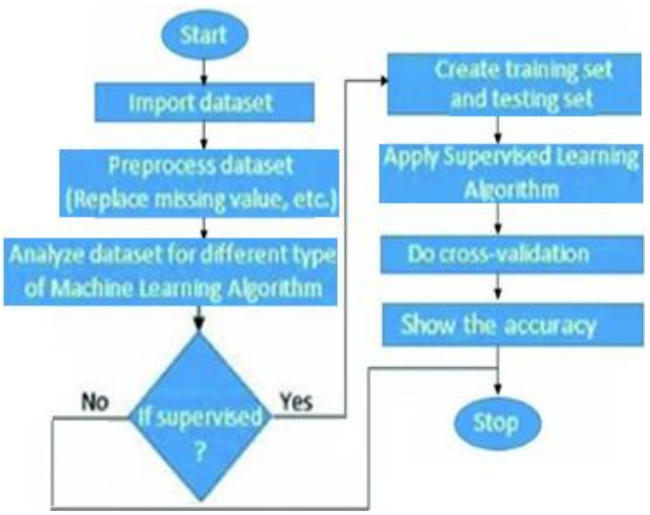
Recently, deep learning convolution neural networks (CNNs) have been getting significant consideration in the domain of medical image diagnostics and particularly in dermoscopy image analysis towards melanoma recognition, and they presented a hybrid approach for melanoma recognition which combined of sparse coding, deep learning, and support vector machine (SVM). The deep descriptors of Res - Net were aggregated with the statistical fisher vector to generate more global representations which were utilized to classify the skin lesions using SVM. This methd showed an improvement in the skin lesion diagnosis with an overall accuracy of 86.81%.

To enable early detection of skin lesion and to increase the chances of survival, and to determine certain conditions or health risk much faster with higher accuracy and to reduce chances of misdiagnosis and to capture unforeseen patterns with complex datasets and to create health risk predictions based on existing data and analyzing the data and loop it back in real time to aid the doctors in detecting medical condition faster and use this as tool to improve ongoing care.

- Implement the machine learning model using Python programming language for analyzing the skin lesions
- Train the model using thousands of training data for enabling the model to self - learn
- Test the model on the given data for the accuracy rate
- Improve the model to achieve more accuracy than any of the existing systems

- Implement the Web services using Java/J2EE to make this solution integra table with other applications
- Deploy the web services over the public cloud to make the solution available to the public across the world.
- Fisher Vector Encoding Strategy: Each image contains a set of deep features, we propose to aggregate these local deep representations into a single image representation (FV representation) using FV encoding method. Kernel based classification: The classification of the FV representations, we train an SVM classifier with Chi - squared (chi2) kernel. Although linear kernels are efficient for the classification, non - linear kernels tend to yield better performance and empirical studies have demonstrated the superiority of the chi2 kernel for image classification. Generator the bio - codes that is biometric key is generated. The extracted features key for encoding and decoding the data for transmission.

## 3. Methodology




Machine learning is a branch of artificial intelligence that aims at solving real life engineering problems. It provides the opportunity to learn without being explicitly programmed and it is based on the concept of learning from data. It is so much ubiquitously used dozen a times a day that we may not even know it. The advantage of machine learning (ML) methods is that it uses mathematical models, heuristic learning, knowledge acquisitions and decision trees for decision making. Thus, it provides controllability, observability and stability. It updates easily by adding a new patient‘s record.

The application of machine learning models on human disease diagnosis aids medical experts based on the symptoms at an early stage, even though some diseases exhibit similar symptoms. One of the important problems in multivariate techniques is to select relevant features from the available set of attributes. The common feature selection techniques include wrapper subset evaluation, filtering and embedded models. Embedded models use classifiers to construct ensembles, the wrapper subset evaluation method provides ranks to features based on their importance and filter methods rank the features based on statistical measurements.

**Computer aided diagnosis systems**

A computer aided medical diagnosis system generally consists of a knowledge base and a method for solving an intended problem. On the basis of the query posted to the system, it provides assistance to the physicians in diagnosing the patients accurately. The knowledge base of such medical systems relies on the inputs that spring up from the clinical experience of field experts. Knowledge acquisition is the process of transforming human expert knowledge and skills acquired through clinical practice to software. It is quite time consuming and labor intensive task. Common methods like Case Based Reasoning (CBR) solves the knowledge acquisition problem to some extent because the past records are maintained in a database, including possible remedies, past clinical decisions, preventive measures and expected diagnostic outcome measures. During patient diagnosis, the clinical database is matched for analogous past patient‘s record for taking suitable decisions.

Some of the major problems faced during the development of an expert diagnosis system are: medical experts are less interested to share their knowledge with others, experience knowledge (called common sense) is practically impossible to be separated and designing a unique expert system for diagnosing all diseases is difficult.

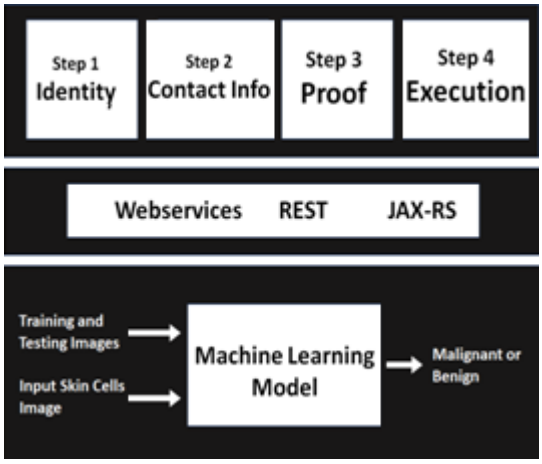
**Software reliability**

Software reliability is defined as the probability that a system will not have a failure over a specified period of time under specific conditions. The knowledge of software reliability is very vital in critical systems because it indicates the design perfection. In this work, the primary aim is to enhance the software reliability of the computer aided diagnosis systems using machine learning algorithms. To provide quality treatment and prevent misdiagnosis are the prime motivations for developing a medical diagnosis system. Diagnosing a disease of a patient accurately is a great challenge in medical field. A huge amount is spent on advanced primary health care devices based on software reliability research as they are considered as critical systems. There are several software reliability models available in the literature; however, none of the models are perfect. An important research issue is choosing a suitable estimation model based on a specific application. One advantage of software reliability over hardware reliability is that a mechanical part surely undergoes ageing; suffer from wear and tear problem over time and usage; however software do not rust or wear out. Software reliability is a vital parameter for software quality, functionality and performance. Some common software reliability models are prediction and estimation models like bathtub curve, exponential, Putnam etc.

**Supervised learning**

Supervised learning is the most common form of machine learning scheme used in solving the engineering problems. It can be thought as the most appropriate way of mapping a set of input variables with a set of output variables. The system learns to infer a function from a collection of labeled training data. The training dataset contains a set of input features and several instance values for respective features. The predictive performance accuracy of a machine learning algorithm depends on the supervised learning scheme. The

aim of the inferred function may be to solve a regression or classification problem. There are several metrics used in the measurement of the learning task like accuracy, sensitivity, specificity, kappa value, area under the curve etc. In this work, the aim is to classify the patients as healthy or ill based on the past medical records. Before solving any engineering problem, it is vital that it is necessary to choose a suitable algorithm for the training purpose based on the type of the data. The selection of a method depends primarily on the type of the data as the field of machine learning is data driven. The next important aspect is the optimization of the chosen machine learning algorithms.




**Classification task**

Classification task is a classical problem in the field of data mining which deals with assigning a pre - specified class to an unknown data. A learning model is built based on the relationship between the predictor attribute values and the value of the target. The challenge is to correctly predict the class based on learning of past data. In machine learning, this kind of classification problems are referred to as supervised learning. Hence, we need to provide a data set containing instances with known classes and a test data set for which the class has to be determined. The success of the classification ability largely depends on the quality of data provided for learning and also the type of machine learning algorithm used. For example, the classification techniques can be used to predict the fraud customers in a bank who apply for a loan or classify mangoes whether they are good or bad and lots of other real time applications. The most common type of classification problem is binary classification, where the target has two possible values like good or bad, yes or no etc. There are several methods for measuring the classification performance like confusion matrix, lift curve, receiver operator characteristics etc.

## 4. Implementation

**Account Operations**

Account operations module provides the following functionalities to the end users

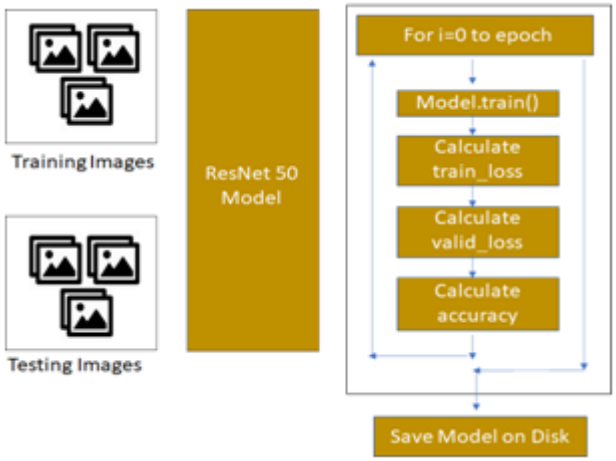
- Register a new seller/ buyer account
- Login to an existing account
- Logout from the session
- Edit the existing Profile
- Change Password for security issues
- Forgot Password and receive the current password over an email
- Delete an existing Account

Account operations module will be re - using the DAO layer to provide the above functionalities.

**Implementation of Sequential Model Algorithm**

The sequential model (also known as the KNF model) is a theory that describes cooperativity of protein subunits. It postulates that a protein's conformation changes with each binding of a ligand, thus sequentially changing its affinity for the ligand at neighboring binding sites.

This model for allosteric regulation of enzymes suggests that the subunits of multimeric proteins have two conformational states. The binding of the ligand causes conformational change in the other subunits of the multimeric protein. Although the subunits go through conformational changes independently (as opposed to in the MWC model), the switch of one subunit makes the other subunits more likely to change, by reducing the energy needed for subsequent subunits to undergo the same conformational change. In elaboration, the binding of a ligand to one subunit changes the protein's shape, thereby making it more thermodynamically favorable for the other subunits to switch conformation to the high affinity state. Ligand binding may also result in negative cooperativity, or a reduced affinity for the ligand at the next binding site, a feature that makes the KNF model distinct from the MWC model, which suggests only positive cooperativity. It is named KNF after Koshland, Némethy and Filmer, who first suggested the model.



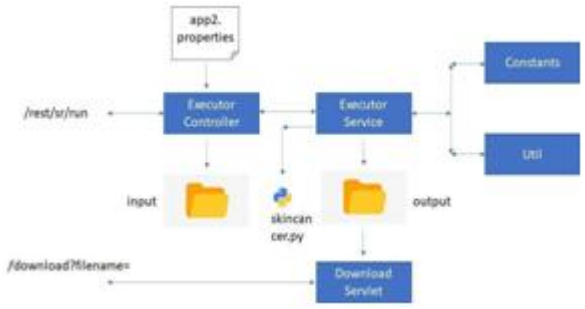

This module implements the Sequential algorithm and ResNet50libraries for developing the machine learning model to predict the skin cancer from a given input image of a cell. This model uses numerous training images for learning purpose.

ResNet, short for Residual Networks is a classic neural network used as a backbone for many computer vision tasks. This model was the winner of ImageNet challenge in 2015. The fundamental breakthrough with ResNet was it allowed us to train extremely deep neural networks with 150+layers successfully. Prior to ResNet training very deep neural networks was difficult due to the problem of vanishing gradients.

**Webservice Implementation**

In this module, we implement the web services to expose the model to the outside world. We expose an HTTP post API against which the user can upload an input image and request for executing the model.

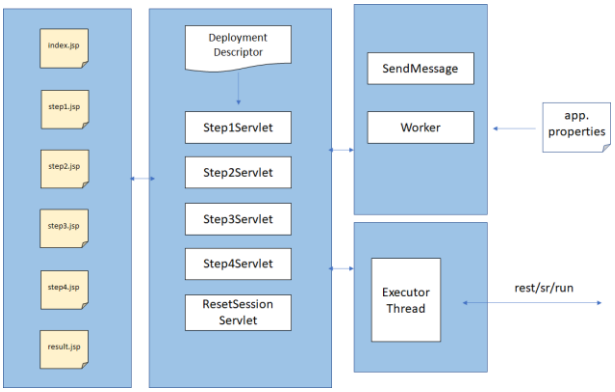
The webservice API upon receiving the request from the client, will store the uploaded image inside the '/home/ubuntu/input' location of the amazon EC2 machine and then it invokes the skincancer. py program by specifying this input folder.




The output image will be stored inside '/home/ubunut/output' location inside amazon EC2 machine. To download this image into user's machine, the webservice will provide another URL as a response by clicking on which the image gets downloaded to the client's machine.

**Third party application**

In this module, the sample third party application has been implemented to demonstrate the usage of the web services to the customers. In this application, we implement four steps




Step 1: User Identity: We collect the user's first name and the last name
Step 2: Contact Information: We collect the email ID and mobile number of the client
Step 3: Proof: We will send an OTP to customer and ask them to enter it to prove the identity
Step 4: Execution: User uploads an input image here and clicking on Run button will invoke the web service implemented in the previous module. The downloadable image link will be displayed back to the client once the result is available

## 5. Conclusion

Technology plays a central role in our everyday life. Technology assistance at various stages of skin lesion processes can significantly enhance the segmentation and classification using deep learning. Proper learning of deep learning networks demands a huge number of training samples. However, a limit on the size of medical image dataset, especially a limit on reliable annotated ground - truths, is one of the challenges in adopting such deep learning approaches.

Automated diagnosis of various skin lesion diseases through medical dermoscopy images is still a very challenging task. In this work, an integrated model for segmentation of skin lesion boundaries and classification of skin lesions is proposed by cascading novel deep learning networks. a novel full resolution convolutional networks (FrCN) is utilized to segment the boundaries of skin lesions from dermoscopy images. We conclude that segmented lesions are passed into a deep residual network for classification.

## References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018, " CA Cancer Journal for Clinicians, vol.68, no.1, pp.7 - 30, Jan, 2018.
[2] N. K. Mishra, and M. E. Celebi., "An overview of melanoma detection in dermoscopy images using image processing and machine learning, " arXiv preprint arXiv: 1601.07843, 2016.
[3] [3] M. A. Al - Masni, M. A. Al - Antari, J. M. Park, G. Gi, T. Y. Kim, P. Rivera, E. Valarezo, S. - M. Han, and T. - S. Kim, "Detection and classification of the breast abnormalities" Jeju Island, Republic of Korea, 2017, pp.1230 - 1233.
[4] L. Q. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated Melanoma Recognition in DermoscopyImages via Very Deep Residual Networks, " IEEE Transactions on Medical Imaging, vol.36, no.4, pp.994 - 1004, Apr, 2017.
[5] Z. Yu, X. Jiang, F. Zhou, J. Qin, D. Ni, S. Chen, B. Lei, and T. Wang, "Melanoma Recognition in Dermoscopy Images via Aggregated Deep Convolutional Features, " IEEE Transactions on Biomedical Engineering, Aug 20, 2018.
[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition, " in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp.770 - 778.
[7] M. Lin, Q. Chen, and S. Yan, "Network In Network, " in rXiv preprint arXiv: 1312.4400, pp.1 - 10, 2016. [8] M. A. Al - Masni, M. A. Al - antari, M. T. Choi, S. M. Han, and T. S. Kim, "Skin Lesion Segmentation in Dermoscopy Images via Deep Full Resolution Convolutional Networks, " Computer Methods and Programs in Biomedicine, vol.162, pp.221 - 231, Aug, 2018.
[8] D. de Godoy, B. Islam, S. Xia, M. T. Islam, R. Chandrasekaran, Y. Chen, S. Nirjon, P. R. Kinget, and X. Jiang, "Paws: A wearable acoustic system for pedestrian safety, " in 2018 IEEE/ACM Third International Conference on Internet - of - Things Design and Implementation (IoTDI), April 2018, pp.237–248.
[9] S. Li, X. Fan, Y. Zhang, W. Trappe, J. Lindqvist, and

R. E. Howard, "Auto++: Detecting cars using embedded microphones in real - time, " Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiqui - tous Technologies, vol.1, no.3, p.70, 2017.

[10] R. S. S. Molau, M. Pitz and H. Ney, "Computing mel - frequency cepstral coefficients on the power spectrum, " in Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01).2001 IEEE International Conference on. IEEE, 2001.

[11] S. Wolff, J. Thomas, and T. Williams, "The polarity - coincidence correla - tor: A nonparametric detection device, " IRE Transactions on Information Theory, vol.8, no.1, pp.5–9, January 19