

SKIN CANCER DETECTION AND CLASSIFICATION USING MACHINE LEARNING ALGORITHMS

M. Ajay Krishna*¹, Dr. C. Priya*²

*¹PG Student, Department Of Computer Applications Dr. M.G.R Educational And Research Institute
Chennai, Tamilnadu, India.

*²Professor, Faculty Of Computer Applications Dr. M.G.R Educational And Research Institute
Chennai, Tamilnadu, India.

DOI : <https://www.doi.org/10.56726/IRJMETS51980>

ABSTRACT

Skin cancer stands out as the most prevalent form of cancer, and early detection is crucial for effective treatment. Detecting cancer in its early stages, however, can be a costly process. Skin cancer manifests as the abnormal growth of skin cells, and it is highly treatable when identified and addressed promptly. There are four primary types of skin cancers: Actinic Keratoses (AK), Basal cell carcinoma (BCC), Dermatofibroma, and Melanoma. Late detection can result in the cancer spreading to other organs. To address this challenge, convolutional neural networks (CNNs) can be employed to detect skin cancer from images. This implementation utilizes the ISIC image dataset and HAM10000 dataset. Transfer learning is applied to enhance the model's performance in CNNs. Pre-trained models are utilized to extract features, which are then used for classifying different types of skin cancer. Various machine learning and deep learning methods are employed in this implementation, including Random Forest. These methods contribute to the development of a comprehensive system for the early detection and classification of skin cancer, ultimately improving the chances of successful treatment.

Keywords: Random Forest, Skin Cancer, Machine Learning.

I. INTRODUCTION

Since the 1970s, skin cancer has held the title of the most prevalent disease globally. Over the previous several decades, there has been an uptick in people diagnosed with non-melanoma and melanoma skin cancers, respectively. Melanoma can be identified in that only one in three cases of cancer, as stated by the World Health Organization (WHO), and according to statistics provided by the Skin Cancer Foundation, one out of every five people in the United States will develop skin cancer at some point during their lifetime. For the past several centuries, the incidence of skin cancer has risen at a relatively constant rate, particularly in the Western hemisphere. Countries such as the United States, Canada, and Australia are just some of the places where this trend has been observed. Infectious diseases of the skin typically have the potential to have a significant detrimental effect on the overall health of people all over the world. According to one piece of a study released in 2017, multiple studies have demonstrated that skin cancer is responsible for 1.79 percent of the disease burden assessed in disability-adjusted life years on a global scale [1]. The incidence of skin cancer accounts for around 7 percent of all newly diagnosed instances of cancer worldwide [2], resulting in a loss of more than \$8 billion for the Medicare program in the United States in 2011. Clinical data suggest that there are such disparities in results based on race in the case of skin cancer: Even though people with darker skin tones are approximately 20 to 30 times as likely to develop melanoma than those who have lighter skin tones, it has been discovered that people with darker skin tones either have a higher or lower mortality risk for specific types of melanoma, depending on their skin tone.

In order to administer the appropriate treatment, it is essential to identify a skin lesion correctly. It is possible to accomplish early diagnosis of melanoma in dermoscopy photographs and pictures using this method, which improves the survival rate.

Dermatologists who have had considerable training in the many skin lesions that melanomas might cause are the most qualified to make an accurate diagnosis. Because of this, diagnosing melanoma can be a challenging task because there is no clear separation between skin lesions and the skin itself, malignant and non-melanoma skin lesions appear visually similar, and there are other factors to take into account. Therefore, creating a

trustworthy automatic detection method for skin tumors, such as a system that can automatically analyze skin lesions, will be greatly useful to pathologists. This is especially relevant in an era where knowledge is scarce.

According to the findings of this study, the classification methods of K-nearest Neighbors, Support Vector Machines, and Decision Trees all produced subpar results in terms of precision and accuracy. After conducting further research into the mathematics that underlies classification, it was found that employing Deep Learning models was the most sophisticated method for getting the desired outcomes (also known as deep learning models). We experimented with many different mathematical models, both with and without the application of Learning Algorithms. However, we concluded that the depth and quality of activation that was made available by pre-trained models did not meet up. Consequently, we merged our mathematical expertise and developed a model known as a Dense Convolutional Network, which offered an accuracy of more than 86.6 percent.

We were able to accomplish efficient and reliable picture categorization by using Deep Learning, a branch of artificial intelligence that is exceptionally robust and potent in its capabilities. The structure and operation of this artificial brain were extremely comparable to those of the human brain, with neurons activating across the brain to transfer information, categorize the data, draw conclusions, and produce consequences. Neural Networks are used in Deep Learning, a form of machine learning. A Neural Network structure is a stack of layers collectively known as a Neural Network structure. As their name suggests, neural networks can perform functions comparable to those that neurons in the brain can execute, such as recognizing patterns and generating predictions.

The project's overarching objective is to come up with a method of screening for skin cancer that is both reasonably quick and straightforward for the general public to use. In most instances, the sooner it is recognized, the better the probability that the individual will make a full recovery. According to the American Academy of Dermatology Association, most dermatological malignancies are treatable if found at an early stage. The trained model is just a preliminary step before doing a skin biopsy. Visit a dermatologist and get a skin biopsy performed if you want to know for sure if you have skin cancer or not. This is the only method to get an accurate diagnosis.

The organization of this document is broken down into its parts in the following paragraphs. In Chapter 2, "Literature Survey," there was a comprehensive summary of all the previous studies. In Section 3, you should describe the design of the proposed system (System Architecture). The experimental setup and the dataset used in the experiment are discussed in Section 4 (Results and Discussion). Section 4 also examines the results of the experiment. Finally, in the section under "Conclusion," both the conclusion reached about the system and the work that will be done in the future to improve it were discussed.

II. PAGE LAYOUT

Title: Melanoma Detection Using an Objective System Based on Multiple Connected Neural Networks

Author: Loretta Ichim, Dan Popescu

Year: 2020

Description: Melanoma is a common form of skin cancer that dangerously affects many people around the world. Detection of melanoma with the naked eye by dermatologists may be subject to errors. Therefore, the implementation of image processing devices equipped with artificial intelligence can act as a support for the dermatologist in examination and decision making. However, due to the various characteristics of this type of lesions and the presence of noises and artifacts in the images, it is difficult to distinguish melanomas from benign lesions. In this article, we propose a new type of intelligent system which is based on several neural networks connected on two levels of classification[1].

Title: Transfer Learning Based System for Melanoma Type Detection

Author: Rashmi Patil*, Sreepathi Bellary

Year: 2021

Description: Skin disease is found in different sorts, for example, basal, squamous cell carcinoma and melanoma among which melanoma is one that is very difficult to predict. Finding melanoma at an early stage is crucial. Melanomas come in many forms and may display none of the typical warning signs. Early detection can vastly increase chances for the cure. Computer vision can assume significant part in Medical Image Diagnosis

and it has been demonstrated by numerous existing frameworks. Here, we represent computer aided strategy for identification of type of melanoma utilizing the transfer learning techniques.

Title: Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning

Author: Mohamed A. Kassem, Khalid M. Hosny

Year: 2020

Description: Melanoma is a type of skin cancer with a high mortality rate. The different types of skin lesions result in an inaccurate diagnosis due to their high similarity. Accurate classification of the skin lesions in their early stages enables dermatologists to treat the patients and save their lives. This paper proposes a model for a highly accurate classification of skin lesions. The proposed model utilized the transfer learning and pre-trained model with GoogleNet.

Title: An Improved Skin Lesion Matching Scheme in Total Body Photography

Author: Konstantin Korotkov, Josep Quintana

Year: 2019

Description: Total body photography is used for early detection of malignant melanoma, primarily as a means of temporal skin surface monitoring. In a prior work, we presented a scanner with a set of algorithms to map and detect changes in pigmented skin lesions, thus demonstrating that it is possible to fully automate the process of total body image acquisition and processing. The key procedure in these algorithms is skin lesion matching that determines whether two images depict the same real lesion. In this paper, we aim to improve it with respect to false positive and negative outcomes. To this end, we developed two novel methods: one based

Title: Machine learning approach in melanoma cancer stage detection

Author: Rashmi Patil, Sreepathi Bellary

Year: 2020

Description: Melanoma is a dangerous skin cancer and spreads very fast. Hence, it is the deadliest skin cancer and causes most deaths. Classification of cancer stages is a very tedious task and very important when a patient is diagnosed. Diagnosis of cancer at the surgical treatment time mainly depends on the stage of cancer or tumor thickness. In this paper, two methods are designed to classify melanoma cancer stages. The first system classifies melanoma as stage 1 and stage 2. Second system classifies melanoma as stage 1, stage 2 or stage 3 melanoma

III. SYSTEM ANALYSIS

Proposed System

The proposed model makes a distinction among lesion maligna, superficial spreading, and nodular melanoma. This permits the early diagnosis of the virus and the quick isolation and therapy necessary to stop the transmission of infection further. Deep learning (DL) and the standard non-parametric machine learning method are exemplified in the deep layer topologies of the Long Short Term Memory (LSTM), which are neural network algorithms. The effectiveness of a LSTM classifier was evaluated using data. The outcomes of the experiments show that the proposed method is superior in terms of diagnostic accuracy compared to the methodologies that are currently considered state of the art.

Advantage

- The machine learns itself and divide the data provided into the levels of prediction and in a very short period of time gives the accurate results, thereby promoting and supporting development of Dermatology.
- Doctors ordinarily use the biopsy method for skin Disease detection. This procedure removes a sample from a suspected skin lesion for medical examination to determine whether it is diseaseous or not.
- An LSTM network enables you to input sequence data into a network, and make predictions based on the individual time steps of the sequence data.

Existing system

Skin diseases rate has been increasing for past few decades, many of these diseases are very dangerous, particularly if not treated at early stages. In Sudan skin diseases are big issue, according to the latest WHO data published in May 2014 Skin Diseases Death in Sudan reached 1,974 or 0.76% of total death. The age adjusted death rate is 9.81 per 100,000 of population, this results ranks Sudan number 8 in the world [1].

In addition, dermatologist use variety of visual clues such as color, scaling and arrangement of lesions, the body site distribution and others, when these individual components are analyzed separately, the recognition of the disease can be quite complex that requiring high level of experience. Diagnosis by humans depends on subjective judgment of the dermatologists so it's hardly reproducible, unlike computer aided diagnostic systems which are more realistic and reliable.

Disadvantage

- Skin Disease is caused by un-repaired deoxyribonucleic acid (DNA) in skin cells, which generate genetic defects or mutations on the skin.
- Melanoma is a hazardous, rare, and deadly type of skin Disease.
- According to statistics from the American Cancer Society, melanoma skin disease cases are only 1% of total cases, but they result in a higher death rate

A. Section Headings

No more than 3 levels of headings should be used. All headings must be in 10pt font. Every word in a heading must be capitalized except for short minor words as listed in Section III-B.

- 1) Level-1 Heading: A level-1 heading must be in Small Caps, centered and numbered using uppercase Roman numerals. For example, see heading "III. Page Style" of this document. The two level-1 headings which must not be numbered are "Acknowledgment" and "References".
- 2) Level-2 Heading: A level-2 heading must be in Italic, left-justified and numbered using an uppercase alphabetic letter followed by a period. For example, see heading "C. Section Headings" above.
- 3) Level-3 Heading: A level-3 heading must be indented, in Italic and numbered with an Arabic numeral followed by a right parenthesis. The level-3 heading must end with a colon. The body of the level-3 section immediately follows the level-3 heading in the same paragraph. For example, this paragraph begins with a level-3 heading.

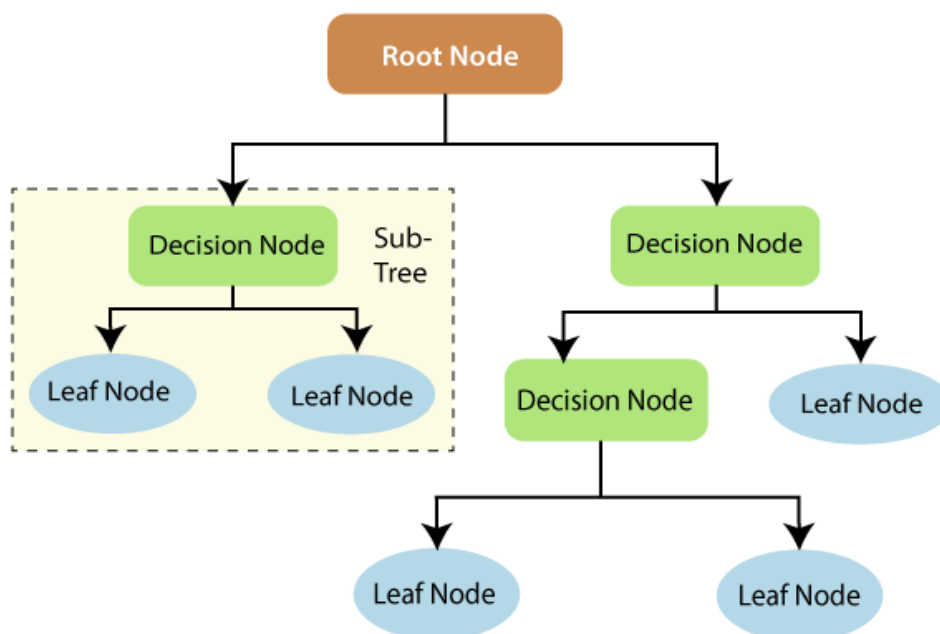


Fig. 1.

Fig. 1 A sample line graph using colors which contrast well both on screen and on a black-and-white hardcopy

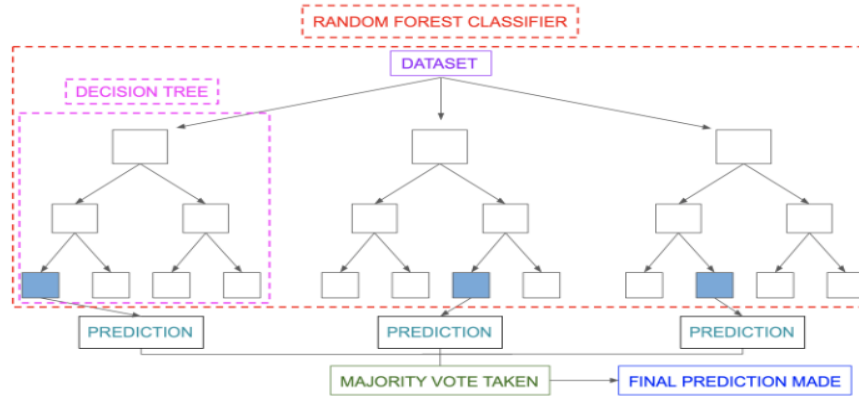


Fig. 2

B. Figure Captions

Figures must be numbered using Arabic numerals. Figure captions must be in 8 pt Regular font. Captions of a single line (e.g. Fig. 2) must be centered whereas multi-line captions must be justified (e.g. Fig. 1). Captions with figure numbers must be placed after their associated figures, as shown in Fig. 1.

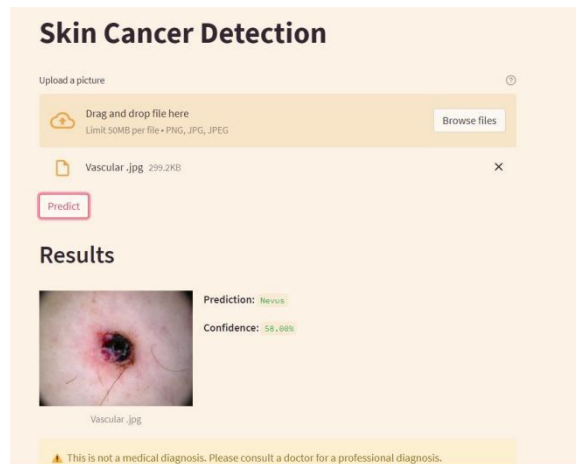


Fig. 3 Example of an unacceptable low-resolution image

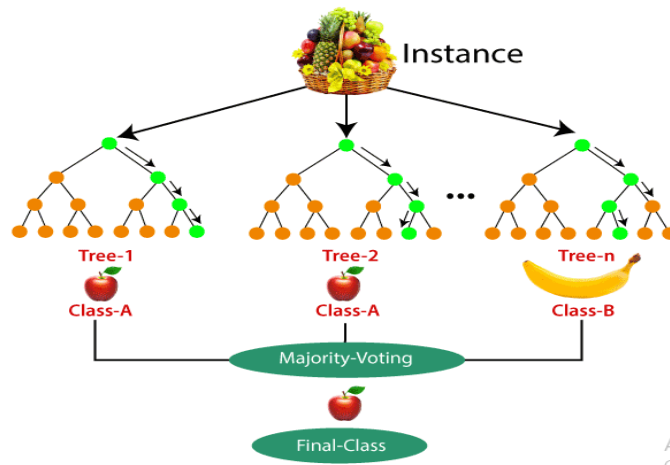


Fig. 4 Example of an image with acceptable resolution

Table Captions **Image Source: Javatpoint**

Regression in random forests

Regression is the other task performed by a random forest algorithm. A random forest regression follows the concept of simple regression. Values of dependent (features) and independent variables are passed in the random forest model.

We can run random forest regressions in various programs such as SAS, R, and python. In a random forest regression, each tree produces a specific prediction. The mean prediction of the individual trees is the output of the regression. This is contrary to random forest classification, whose output is determined by the mode of the decision trees' class.

Although random forest regression and linear regression follow the same concept, they differ in terms of functions. The function of linear regression is $y = bx + c$, where y is the dependent variable, x is the independent variable, b is the estimation parameter, and c is a constant. The function of a complex random forest regression is like a blackbox.

Applications of random forest

Some of the applications of the random forest may include:

Banking

Random forest is used in banking to predict the creditworthiness of a loan applicant. This helps the lending institution make a good decision on whether to give the customer the loan or not. Banks also use the random forest algorithm to detect fraudsters.

Health care

Health professionals use random forest systems to diagnose patients. Patients are diagnosed by assessing their previous medical history. Past medical records are reviewed to establish the right dosage for the patients.

Stock market

Financial analysts use it to identify potential markets for stocks. It also enables them to identify the behavior of stocks.

E-commerce

Through rain forest algorithms, e-commerce vendors can predict the preference of customers based on past consumption behavior.

When to avoid using random forests

Random forest algorithms are not ideal in the following situations:

Extrapolation

Random forest regression is not ideal in the extrapolation of data. Unlike linear regression, which uses existing observations to estimate values beyond the observation range. This explains why most applications of random forest relate to classification.

Sparse data

Random forest does not produce good results when the data is very sparse. In this case, the subset of features and the bootstrapped sample will produce an invariant space. This will lead to unproductive splits, which will affect the outcome.

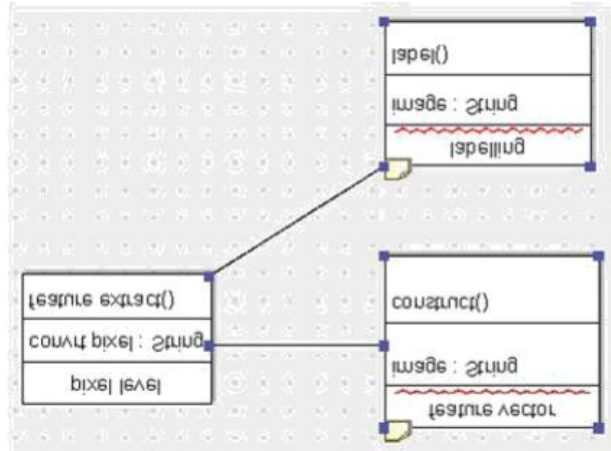
Advantages of random forest

- It can perform both regression and classification tasks.
- A random forest produces good predictions that can be understood easily.
- It can handle large datasets efficiently.
- The random forest algorithm provides a higher level of accuracy in predicting outcomes over the decision tree algorithm.

Disadvantages of random forest

- When using a random forest, more resources are required for computation.

It consumes more time compared to a decision tree algorithm



REQUIREMENT SPECIFICATION

Functional Requirements

- Graphical User interface with the User.

Software Requirements

For developing the application the following are the Software Requirements:

1. Python
2. Django
3. MySQL
4. MySQLclient
5. WampServer 2.4

Operating Systems supported

1. Windows 7
2. Windows XP
3. Windows 8

Technologies and Languages used to Develop

1. Python

Debugger and Emulator

- Any Browser (Particularly Chrome)

Hardware Requirements

For developing the application the following are the Hardware Requirements:

- Processor: Pentium IV or higher
- RAM: 256 MB
- Space on Hard Disk: minimum 512MB

IV. SYSTEM STUDY

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are,

- **ECONOMICAL FEASIBILITY**
- **TECHNICAL FEASIBILITY**
- **SOCIAL FEASIBILITY**

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

1) TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

SYSTEM TEST

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

2) TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfactory, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

- Valid Input : identified classes of valid input must be accepted.
- Invalid Input : identified classes of invalid input must be rejected.
- Functions : identified functions must be exercised.
- Output : identified classes of application outputs must be exercised.
- Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

Unit Testing

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

Acceptance Testing

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.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

V. CONCLUSION

Within the scope of this study, a Long Short Term Memory (LSTM) model for the diagnosis of skin cancer was created, constructed, and evaluated using a well-known melanoma dataset. Our proposed method, which is a

two stage learning platform, has great-predicted accuracy at each stage, as demonstrated by its overall accuracy of 88.83 percent. This is true not only for classification algorithms such as DT, RF, GBT.

The strategy that has been suggested is based on LSTM, and it is possible to think of it as an effective method of multiclass categorization. In terms of melanoma classification accuracy, the modular and hierarchical structure of our LSTM classifier not only beats state-of-the-art machine learning techniques, but it significantly minimises the amount of computational effort that is required. The fact that this strategy is only tested on a single dataset is one of the method's drawbacks.

VI. FUTURE SCOPE

Future work is dependent on the increased Long Short Term Memory accuracy in classifying skin illnesses, and SCM is used to manage the feature extraction technique.

VII. REFERENCE

- [1] Zhang Ce, Xin Pan, Huapeng Li, Gardiner A, Sargent I, Jonathon S Hare, et al. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *Isprs Journal of Photogrammetry and Remote Sensing*. 2017; 140:133-144.
- [2] Bi Lei, Jinman Kim, Euijoon Ahn, Feng D. Automatic Skin Lesion Analysis using Large-scale Dermoscopy Images and Deep Residual Networks. *ArXiv abs*. 2017; 1703:04197.
- [3] Yuan Y, Lo YC. Improving Dermoscopic Image Segmentation with Enhanced Convolutional-De convolutional Networks. *IEEE J Biomed Health Inform*. 2019 Mar; 23(2):519-526. DOI: 10.1109/JBHI.2017.2787487. Epub 2017 Dec 25. PMID: 29990146.
- [4] Feng-ying Xie, Jiawen Yang, Liu J, Zhi-guo Jiang, Yushan Zheng, Yukun Wang. Skin lesion segmentation using high-resolution convolutional neural network. *Computer methods and programs in biomedicine*. 2020; 186:105241.
- [5] Yao Y, Luo Z, Li S, Fang T, Quan L. MVSNet: Depth Inference for Unstructured Multi-view Stereo. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) *Computer Vision – ECCV 2018*. ECCV 2018. Lecture Notes in Computer Science, 2018, 11212. Springer, Cham. https://doi.org/10.1007/978-3-030-01237-3_47
- [6] Li Y, Shen L. Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network. *Sensors (Basel)*. 2018 Feb 11;18(2):556. DOI: 10.3390/s18020556. PMID: 29439500; PMCID: PMC5855504.
- [7] Rashmi Patil, Sreepathi Bellary. Machine learning approach in melanoma cancer stage detection, *Journal of King Saud University - Computer and Information Sciences*, 2020, ISSN:1319-1578, <https://doi.org/10.1016/j.jksuci.2020.09.002>.
- [8] Konstantin Korotkov, Josep Quintana, Ricard Campos, Am´erica Jes´us-Silva, Pablo Iglesias, Susana Puig, et al. An Improved Skin Lesion Matching Scheme in Total Body Photography, *IEEE Journal of Biomedical and Health Informatics*. 2019 Mar;23(2):586-598.
- [9] Kassem MA, Hosny KM, Fouad MM. Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning, in *IEEE Access*. 2020;8:114822-114832. DOI: 10.1109/ACCESS.2020.3003890.
- [10] Ichim L, Popescu D. Melanoma Detection Using an Objective System Based on Multiple Connected Neural Networks, in *IEEE Access*. 2020;8:179189-179202. DOI: 10.1109/ACCESS.2020.3028248.
- [11] Patil R, Bellary S. Transfer learning based system for melanoma type detection. *Revue d'Intelligence Artificielle*. 2021;35:(2):123-130. <https://doi.org/10.18280/ria.350203>
- [12] Ravi V, Pradeepkumar D, Deb K. Financial time series prediction using hybrids of chaos theory, multilayer perceptron and multi-objective evolutionary algorithms. *Swarm Evolut Comput*. 2017;36:136-149.