UDC 61:004

AN INNOVATIVE TECHNIQUE FOR CLASSIFICATION AND DETECTION OF SKIN CANCER USING IMAGE PROCESSING AND SVM ALGORITHM

Dr. Vijayalakshmi D.[1], Karun M.[2], Anand N.[3], Rajendran R.[4]

[1,2,3,4]Bangalore Institute of Technology, Bangalore, India
ORCID: [1]http://orcid.org/0000-0002-0908-6643, [2]http://orcid.org/0009-0009-1269-3231, [3]http://orcid.org/0009-0009-7349-0614, [4]http://orcid.org/0009-0004-3912-2301
Emails:[1]lakshmi@bit-bangalore.edu.in, [2]manan3@mail.com, [3]namratha141anand@gmail.com, [4]ramitharajendrann@gmail.com

DOI: https://doi.org/10.15673/atbp.v15i4.2714

Abstract. *Skin cancer is a serious health issue that affects millions of people worldwide. One in every three cancers is a skin cancer and, most people fail to identify and get a diagnosis which is why early detection is essential for effective treatment and improving patient outcomes. In recent years, computer vision and machine learning have become important tools for the automatic detection of skin cancer. One of the commonly used machine learning algorithms for this task is Support Vector Machines (SVMs). SVMs are a type of supervised learning algorithm that is used for classification tasks. In skin cancer detection, the SVM classifier is trained on a dataset of dermatoscopic images of skin lesions. The first step in the process is feature extraction, where relevant information is extracted from the images to serve as input to the SVM classifier. This information can include color, texture, and shape features, among others.*

The training and testing of the SVM classifier is then performed using a portion of the dataset, with the remainder being used to evaluate its performance. During the testing phase, the SVM classifier is used to predict the class label of each image, which can be malignant or benign. The accuracy of the classifier is evaluated by comparing its predictions to the actual class labels of the images in the evaluation dataset.

The results of using SVMs in skin cancer detection have been promising, with high accuracy rates being achieved. This highlights the potential of SVMs as a useful tool for skin cancer screening and early detection. In conclusion, the use of SVMs in skin cancer detection provides a fast, automatic, and reliable method for detecting skin cancer, which can help to improve patient outcomes. Currently it is really very important to watch and analyse the cancer disease automatically at intervals the first stages. Irregular streaks square measure one in every of the foremost very important features (included in most of dermoscopy algorithms) that show high association with carcinoma and basal cell malignant growth malady. The diagnostic test technique for the detection is most painful and harmful so we have a tendency to tend to square measure going for the machine-driven detection. Here we have a tendency to tend to square measure practice the GLCM choices for the detection the choices of skin lesions square measure extracted normalized symmetrical grey Level Co-occurrence Matrices GLCM. GLCM based texture choices square measure extracted from each of the four classes and given as input to the Multi-Class Support vector machine that's utilized for classification purpose.

Анотація. *Рак шкіри є серйозною проблемою зі здоров'ям, яка вражає мільйони людей у всьому світі. Кожен третій вид раку є раком шкіри, і більшість людей не в змозі визначити та поставити діагноз, тому раннє виявлення має важливе значення для ефективного лікування та покращення результатів пацієнтів. В останні роки комп'ютерний зір і машинне навчання стали важливими інструментами для автоматичного виявлення раку шкіри. Одним із часто використовуваних алгоритмів машинного навчання для цього завдання є Support Vector Machines (SVM). SVM – це тип алгоритму навчання під наглядом, який використовується для класифікаційних завдань. Під час виявлення раку шкіри класифікатор SVM навчається на наборі даних дерматоскопічних зображень уражень шкіри. Першим кроком у цьому процесі є виділення ознак, де відповідна інформація витягується із зображень, щоб служити вхідними даними для класифікатора SVM. Ця інформація може включати особливості кольору, текстури та форми, серед іншого.*

Навчання та тестування класифікатора SVM потім виконується з використанням частини набору даних, а решта використовується для оцінки його продуктивності. На етапі тестування класифікатор SVM використовується для прогнозування мітки класу кожного зображення, яке може бути злоякісним або доброякісним. Точність класифікатора оцінюється шляхом порівняння його прогнозів із фактичними мітками класів зображень у наборі даних оцінки.

Результати використання SVM у виявленні раку шкіри були багатообіцяючими, з високими показниками точності. Це підкреслює потенціал SVM як корисного інструменту для скринінгу та раннього виявлення раку шкіри. Підсумовуючи, використання SVM для виявлення раку шкіри забезпечує швидкий, автоматичний і надійний метод виявлення раку шкіри, який може допомогти покращити результати лікування пацієнтів. В даний час дійсно дуже важливо спостерігати і аналізувати ракове захворювання в автоматичному режимі на перших стадіях. Нерегулярні смуги квадратного розміру мають один квадратний розмір у кожній із головних дуже важливих ознак (включених у більшість алгоритмів дермоскопії), які показують високий зв'язок із карциномою та базально-клітинним захворюванням злоякісного росту. Техніка діагностичного тестування для виявлення є найболючішою та шкідливою, тому ми маємо тенденцію схилятися до квадратної міри, йдучи до виявлення, керованого машиною. Тут ми маємо тенденцію до квадратичного вимірювання на практиці вибору GLCM для виявлення вибору уражень шкіри квадратного виміру, витягнутого нормалізованого симетричного рівня сірого, матриці спільного розташування GLCM. Квадратна міра вибору текстури на основі GLCM витягується з кожного з чотирьох класів і надається як вхідні дані для векторної машини підтримки кількох класів, яка використовується для цілей класифікації.

Ключові слова: *машинне навчання, алгоритм SVM, обробка зображень, рак шкіри*
Keywords: Machine Learning, SVM Algorithm, Image Processing, Skin Cancer

PURPOSE AND OBJECTIVE

- Skin cancers are the most common and deadly form of cancers in the contemporary world
- The reported cancer incidence in India in 2022 is estimated to be approximately 1.9 to 2 million and nearly 10 million across the world.
- Majority of patients suffering from skin cancer fail to diagnose themselves until their illness becomes terminal
- Our proposed idea aims at overcoming this problem by developing an Android application that assists in early skin cancer detection and classifies it into one of the five deadly types – Actinic Keratosis, Basal Cell Carcinoma, Cherry Nevus, Dermatofibroma, and Melanoma using image processing and various machine learning classification algorithms like SVM
- The application also provides contacts of the nearest oncologists and diagnoses centers.

INTRODUCTION

Skin cancer is the unusual development of skin cells which frequently develops on skin exposed to the sun. However, having referenced that, it might likewise happen on territories of the skin not exposed to the daylight. Among three sorts of skin disease, namely, Squamous Cell Carcinoma (SCC), Melanoma and Basal Cell Carcinoma (BCC). Melanoma is generally perilous in which endurance rate is extremely low. The other two go under the non-melanoma malignancy.

In spite of the fact that melanoma is a hazardous illness around the world, early discovery of melanoma alongside other skin cancer types can build the opportunity of endurance of the person in question. In the event that an individual has dubious spots or developments on the skin, the specialist will inspect the skin or allude to an expert for determination.

The expert will look at the shape, size, surface and shade of the dubious region on the skin. In the event that it is suspected to be dangerous, a skin biopsy method might be performed, in which they will eliminate the dubious zone or a part of it to ship off to a lab for testing which is costly, excruciating and requires a long time to think of the outcome. The result of a biopsy usually takes about 2 to 3 weeks to determine whether the lesion is cancerous and if it is, what type of skin cancer it is. While there are several forms of skin cancer, the most common ones, namely- Basal cell carcinoma (BCC), Squamous cell carcinoma (SCC) (actinic keratosis), and Melanoma (the most serious because it has a tendency to spread) are easily identifiable.

Consequently, there is an increasing requirement for programmed skin disease discovery framework with high exactness. The identification of the skin malignant growth consistently accompanies the cycle of biopsy that should be performed by the dermatologist.

CANCER TYPES

1. Actinic Keratosis

Fig. 1 – Actinic Keratosis

- A rough, scaly patch on the skin caused by years of sun exposure.
- Actinic Keratoses usually affects older adults.
- Reducing sun exposure can help reduce the risk.

Fig. 2 – Basal Cell Carcinom

2. **Basal Cell Carcinoma**

- It is the most common form of skin cancers.
- A type of skin cancer that begins in the basal cells.
- Basal cells produce new skin cells as old ones die.
- Limiting sun exposure can help prevent these cells from becoming cancerous.

Fig. 3 – Cherry Nevus

3. **Cherry Nevus**

- A small red dome-shaped bump on the skin ranging between 0.5 – 6 mm in diameter.
- Usually several are present, typically on the chest and arms and increasing in number with age.
- This type of skin growth is not a cause for concern unless it bleeds often or changes in size, shape or color.

Fig. 4 – Dermatofibroma

4. **Dermatofibroma**

- It is referred to as benign, fibrous histiocytomas of the skin.
- It is a commonly occurring cutaneous entity usually centred within the skin's dermis.
- The growths often develop after some type of small trauma to the skin, including a puncture from a splinter or bug bite.

Fig. 5 – Melanoma

5. **Melanoma**

- It is the most serious type of skin cancers.
- It occurs when the pigment-producing cells that give colour to the skin become cancerous.
- Symptoms might include a new, unusual growth or a change in an existing mole.
- Melanomas can occur anywhere on the body.

BACKGROUND WORK

- The following images shows the statistics of skin cancer affected in India as well as across the world.

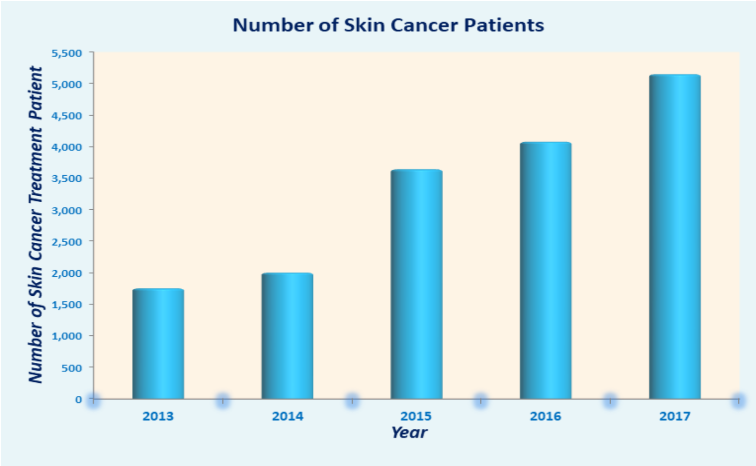

Fig. 5 – Number of Skin Cancer Patients

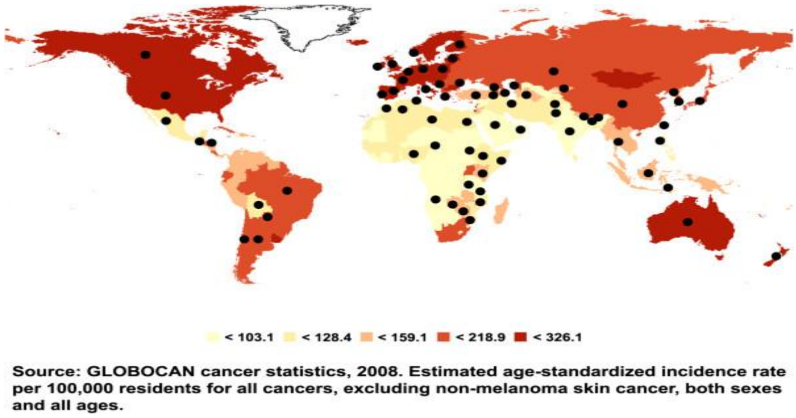

Fig. 6 – Globocan cancer statistics. 2008. Estimated age-standartized incidence rate per 100 000 residents for all cancers, excluding non-melanoma skin cancer, both sexes and all ages

- The statistics shows the need for the diagnosis and treatment for the disease caused.
- The traditional methods followed since years is described below.

Diagnosis

To diagnose skin cancer, the doctor may:

- **Examine your skin:** Your doctor may look at your skin to determine whether your skin changes are likely to be skin cancer. Further testing may be needed to confirm that diagnosis.
- **Remove a sample of suspicious skin for testing (skin biopsy):** Your doctor may remove the suspicious-looking skin for lab testing. A biopsy can determine whether you have skin cancer and, if so, what type of skin cancer you have.

Determining the extent of the skin cancer

If your doctor determines you have skin cancer, you may have additional tests to determine the extent (stage) of the skin cancer.

Because superficial skin cancers such as basal cell carcinoma rarely spread, a biopsy that removes the entire growth often is the only test needed to determine the cancer stage. But if you have a large squamous cell carcinoma, Merkel cell carcinoma or melanoma, your doctor may recommend further tests to determine the extent of the cancer.

Additional tests might include imaging tests to examine the nearby lymph nodes for signs of cancer or a procedure to remove a nearby lymph node and test it for signs of cancer (sentinel lymph node biopsy).

Doctors use the Roman numerals I through IV to indicate a cancer's stage. Stage I cancers are small and limited to the area where they began. Stage IV indicates advanced cancer that has spread to other areas of the body. The skin cancer's stage helps determine which treatment options will be most effective.

Treatment

Your treatment options for skin cancer and the pre-cancerous skin lesions known as actinic keratoses will vary, depending on the size, type, depth and location of the lesions. Small skin cancers limited to the surface of the skin may not require treatment beyond an initial skin biopsy that removes the entire growth.

If additional treatment is needed, options may include:

- **Freezing:** Your doctor may destroy actinic keratoses and some small, early skin cancers by freezing them with liquid nitrogen (cryosurgery). The dead tissue sloughs off when it thaws.
- **Excisional surgery:** This type of treatment may be appropriate for any type of skin cancer. Your doctor cuts out

(excises) the cancerous tissue and a surrounding margin of healthy skin. A wide excision — removing extra normal skin around the tumor — may be recommended in some cases.

- **Mohs surgery:** This procedure is for larger, recurring or difficult-to-treat skin cancers, which may include both basal and squamous cell carcinomas. It's often used in areas where it's necessary to conserve as much skin as possible, such as on the nose.

During Mohs surgery, your doctor removes the skin growth layer by layer, examining each layer under the microscope, until no abnormal cells remain. This procedure allows cancerous cells to be removed without taking an excessive amount of surrounding healthy skin.

- **Curettage and electrodesiccation or cryotherapy:** After removing most of a growth, your doctor scrapes away layers of cancer cells using a device with a circular blade (curet). An electric needle destroys any remaining cancer cells. In a variation of this procedure, liquid nitrogen can be used to freeze the base and edges of the treated area.

These simple, quick procedures may be used to treat basal cell cancers or thin squamous cell cancers.

- **Radiation therapy:** Radiation therapy uses high-powered energy beams, such as X-rays, to kill cancer cells. Radiation therapy may be an option when cancer can't be completely removed during surgery.
- **Chemotherapy:** In chemotherapy, drugs are used to kill cancer cells. For cancers limited to the top layer of skin, creams or lotions containing anti-cancer agents may be applied directly to the skin. Systemic chemotherapy can be used to treat skin cancers that have spread to other parts of the body.
- **Photodynamic therapy:** This treatment destroys skin cancer cells with a combination of laser light and drugs that makes cancer cells sensitive to light.
- **Biological therapy:** Biological therapy uses your body's immune system to kill cancer cells.

RESEARCH DESIGN AND METHODOLGY

OVERVIEW

Our proposed method aims to provide the most suitable and the efficient solution for detecting skin cancer at early stages.

As mentioned earlier, this project proposes an idea to develop an app which takes an image as an input, processes it and classifies it into one of the 5 major classes of cancer namely - Actinic Keratosis, Basal cell Carcinoma, Cherry Nevus, Dermatofibroma and Melanoma

The design for this project constitutes of two major parts-

- The mobile application part
- The image processing and classification

The methodology for the above mentioned parts is explained in detail in the following sections.

Requirements:

Hardware requirements:

1) Camera capable of producing colour pictures and picture resolution of 640x480

Software requirements:

1) MATLAB

MATLAB is a proprietary multi-paradigm programming language and numeric computing environment developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages.

MATLAB IN IMAGE PROCESSING:

Use MATLAB and Simulink to gain insight into your image and video data, develop algorithms, and explore implementation tradeoffs.

- Design vision solutions with a comprehensive set of reference-standard algorithms for image processing, computer vision, and deep learning.
- Collaborate with teams using OpenCV, Python, and C/C++ using interoperable APIs and integration tools.
- Use workflow apps to automate common tasks and accelerate algorithm exploration.
- Accelerate algorithms on NVIDIA GPUs, cloud, and datacenter resources without specialised programming or IT knowledge.
- Deploy algorithms to embedded devices, including NVIDIA GPUs, Intel processors and FPGAs, and ARM-based embedded processes.

Image Processing toolbox in MATLAB:

Image Processing Toolbox provides a comprehensive set of reference-standard algorithms and workflow apps for image processing, analysis, visualisation, and algorithm development. You can perform image segmentation, image enhancement, noise reduction, geometric transformations, and image registration using deep learning and traditional image processing techniques. The toolbox supports processing of 2D, 3D, and arbitrarily large images.

Image Processing Toolbox apps let you automate common image processing workflows. You can interactively segment image data, compare image registration techniques, and batch-process large datasets. Visualisation functions and apps let you explore images, 3D volumes, and videos; adjust contrast; create histograms; and manipulate regions of interest (ROIs).

You can accelerate your algorithms by running them on multicore processors and GPUs. Many toolbox functions support C/C++ code generation for desktop prototyping and embedded vision system deployment.

2) **Android studio IDE**

Android Studio is the official integrated development environment (IDE) for Google's Android operating system, built on JetBrains' IntelliJ IDEA software and designed specifically for Android development. It is available for download on Windows, macOS and Linux based operating systems.It is a replacement for the Eclipse Android Development Tools (E-ADT) as the primary IDE for native Android application development.

Android Studio was announced on May 16, 2013, at the Google I/O conference. It was in the early access preview stage starting from version 0.1 in May 2013, then entered beta stage starting from version 0.8 which was released in June 2014. The first stable build was released in December 2014, starting from version 1.0. At the end of 2015, Google dropped support for Eclipse ADT, making Android Studio the only officially supported IDE for Android development.

On May 7, 2019, Kotlin replaced Java as Google's preferred language for Android app development.Java is still supported, as is C++.

3) **SQLite**

SQLite is an open source SQL database that stores data to a text file on a device. Android comes in with built in SQLite database implementation.SQLite supports all the relational database features. In order to access this database, you don't need to establish any kind of connections for it like JDBC,ODBC e.t.c

The main package is android.database.sqlite that contains the classes to manage your own databases

Working methodology of image processing and classification

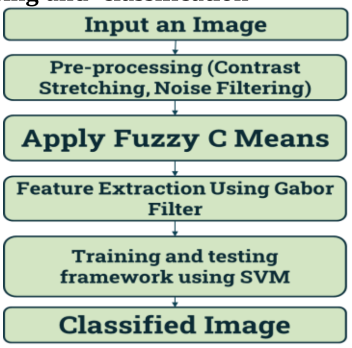

Fig. 7 – Flowchart representing the working methodology is as shown above

DETAILED IMPLEMENTATION:

1) Input an image:

A colour image of the suspicious skin of appropriate resolution (640x480 or higher) is fed as input to the application.

The app can be programmed to ask for additional details like age, gender,and other symptoms to help produce a more accurate prediction.

2) Preprocessing of the image

The image by default is not ready for classification. The image contains noise and other undesirable components like body hair that has to be removed for a more efficient classification.

Pre-processing is a common name for operations with images at the lowest level of abstraction — both input and output are intensity images. These iconic images are of the same kind as the original data captured by the sensor, with an intensity image usually represented by a matrix of image function values (brightnesses). The aim of pre-processing is an improvement of the image data that suppresses unwilling distortions or enhances some image features important for further processing, although geometric transformations of images (e.g. rotation, scaling, translation) are classified among pre-processing methods here since similar techniques are used.

The enhancement techniques are employed in order to increase the contrast of an image. Generally, an image can be enhanced by spreading out the range of scene illumination. This procedure is called contrast stretching. Contrast stretching (often called normalisation) is a simple image enhancement technique that attempts to improve the contrast in an image by 'stretching' the range of intensity values it contains to span a desired range of values, the full range of pixel values that the image type concerned allows. Contrast stretching changes the distribution and range of the digital numbers assigned to each pixel in an image. This is normally done to accent image details that may be difficult for the human viewer to observe.

Contrast enhancement is a procedure to improve the representation of pictures. For skin cancer pictures,it is once in a while important to upgrade the Contrast in the area of interest.

Histogram equalisation is one of the most usually utilised procedures for contrast enhancement.

Noise is always present in digital images during image acquisition, coding, transmission, and processing steps.

It is very difficult to remove noise from the digital images without the prior knowledge of filtering techniques. In this article, a brief overview of various noise filtering techniques.

These filters can be selected by analysis of the noise behaviour. In this way, a complete and quantitative analysis of noise and their best suited filters will be presented over here.

Filtering image data is a standard process used in almost every image processing system. Filters are used for this purpose. They remove noise from images by preserving the details of the same. The choice of filter depends on the filter behaviour and type of data.

Median filtering is a common image enhancement technique for removing salt and pepper noise. Because this filtering is less sensitive than linear techniques to extreme changes in pixel values, it can remove salt and pepper noise without significantly reducing the sharpness of an image. In this topic, you use the Median Filter block to remove salt and pepper noise from an intensity image.

Median filter is one of the types of filters which is used for reducing noise from the skin images. Image noise is frequently created by the picture sensor and hardware of a scanner or camera. It is an undesirable by-product of image catch that impedes the predetermined data. In this, we use the Image processing toolbox is the one which provides a set of standard reference algorithms that can perform image segmentation. image enhancement, noise reduction, etc.

3) Image segmentation using FUZZY C Means

The next stage is image segmentation. which is a technique used in image processing to separate an image into numerous parts, depending on the qualities of the pixels in an image. Image segmentation includes detecting and segmenting the region of interest, based on whether they are identical in colour or shape.

After the pre-processing technique is applied to an image which makes it free from all kinds of impurities, we need to segment that image

focusing on the region of interest. to make the classification easy. An image segmentation technique called clustering, is an approach which is used to separate a group of elements in a region.

Image segmentation is a commonly used technique in digital image processing and analysis to partition an image into multiple parts or regions, often based on the characteristics of the pixels in the image. Image segmentation could involve separating foreground from background, or clustering regions of pixels based on similarities in colour or shape. For example, a common application of image segmentation in medical imaging is to detect and label pixels in an image or voxels of a 3D volume that represent a tumour in a patient's brain or other organs.

During medical diagnosis for cancer, pathologists stain body tissue with hematoxylin and eosin (H&E) to distinguish between tissue types. They then use an image segmentation technique called *clustering* to identify those tissue types in their images. Clustering is a method to separate groups of objects in a scene. The K-means and fuzzy c means are commonly used clustering techniques..

Clustering methods consist of defining groups of pixels. Therefore, all the pixels in the same group define a class in the segmented image.

Clustering of data is a method by which large sets of data are grouped into clusters of smaller sets of similar data.

Fuzzy c-means (FCM) clustering algorithm is one of the most commonly used unsupervised clustering techniques in the field of medical imaging. Medical image segmentation refers to the segmentation of known anatomic structures from medical images. Fuzzy C-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. Fuzzy logic is a multi-valued logic derived from fuzzy set theory. FCM is popularly used for soft segmentations like brain tissue models. And also FCM can provide better results than other clustering algorithms like KM, EM, and KNN. In this paper we presented the medical image segmentation techniques based on various types of FCM algorithms.As one kind of image segmentation algorithm, fuzzy C-means clustering is an effective and concise segmentation algorithm. However, the drawback of FCM is that it is sensitive to image noise. To solve the problem, this paper designs a novel fuzzy C-mean clustering algorithm based on multi-objective optimization. We add a parameter λ to the fuzzy distance measurement formula to improve the multi-objective optimization. The parameter λ can adjust the weights of the pixel local information. In the algorithm, the local correlation of neighbouring pixels is added to the improved multi-objective mathematical model to optimise the clustering.

The results of the significance test between the two algorithms. While the clustering method with the c-mean fuzzy algorithm can provide a significantly better performance, compared to k-means.

Clustering-based segmentation performance with the median threshold determination method can provide better performance. The performance of the fuzzy c-means algorithm gives better performance than k-mean, both when using thresholding with mean and median methods.

4) Feature extraction

Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data.Feature extraction can be accomplished manually or automatically.

In this process, various features are extracted from an image and converted into numerical data while the original information remains unchanged in the dataset. After the region of interest is partitioned in the image from the previous process, we move onto feature extraction where many features like texture, colour, diameter, size etc are extracted which is later useful for the classification process.

The texture analysis is done with the help of the GLCM (Gray Level Co-occurrence Matrix) features and Gabor Filter, and the texture element is extracted from an RGB coloured picture.

Feature extraction identifies the most discriminating characteristics in signals, which a machine learning or a deep learning algorithm can more easily consume.

Training machine learning or deep learning directly with raw signals often yields poor results because of the high data rate and information redundancy.

Fig. 8 – Schematic process for applying feature extraction to signals and time series data for a machine learning classifier

Feature extraction using Gabor filter:

A Gabor filter is a linear filter used in image processing for edge detection, texture classification, feature extraction and disparity estimation. It is a bandpass filter, i.e. it passes frequencies in a certain band and attenuates the other frequencies outside the band.

Gabor filter is the implementation of the Gabor transform which is a short term Fourier transformation with a Gaussian window for analysis in the spatial domain. The distortion information of content adaptive image steganography incorporates the texture information of the image. On embedding there causes texture anomaly in an image.

This texture anomaly can be characterised by the Gabor output obtained by 2D Gabor filtering.

The two dimensional Gabor filter represents the texture information because of its spatial selectivity and orientation.

We have opted for Gabor filters as they are the filters which are mostly utilised for the analysis of texture. They are linear filters and perform this by focusing on the region of interest and analyse if there is any frequency content around it. Most people use this as a model for recognizing texture because the methodology of this filter is similar to how people differentiate texture usually. The Gabor filter has a Gaussian function with variance values and centre frequencies.

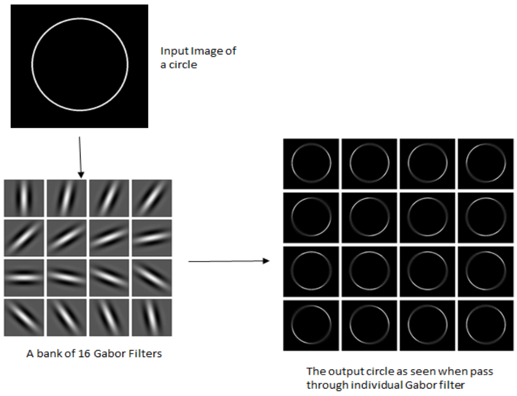

Fig. 9 - The corresponding oriented edge features are detected when passed through individual-oriented Gabor filters

Gray-Level Co-Occurrence Matrix (GLCM):

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM). The GLCM functions characterise the texture of an image by calculating how often pairs of pixels with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. (The texture filter functions, described in Calculate Statistical Measures of Texture cannot provide information about the spatial relationships of pixels in an image.)

The GLCM is utilised for analysing the texture feature. Here, the spatial relation that each pixel has with other pixels is considered for viewing the texture. They compute how often a pair of pixels with certain values in a spatial relation arise in an image thereby forming a matrix and obtaining information from them.They assist in characterising the texture present in an image. A gray co-matrix function is used to create a GLCM. The following statistics give some details about the texture extracted from an image.

1)Contrast- measures intensity between two pixels.

2)Correlation-is the joint correlation or probability of any two pixels in an image.

3)Energy-is the addition of the squared components in the Gray Level Co-occurrence Matrix.

4) Homogeneity- it shows how close the distributed objects are present in the Gray Level Co-occurrence Matrix.

Texture feature	Equation
Contrast	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2P(i,j)$
Entropy	$-\sum_{i=1}^{N}\sum_{j=1}^{N}P(i,j)\,lgP(i,j)$
Correlation	$\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(i-\bar{x})(j-\bar{y})P(i,j)}{\sigma_x\sigma_y}$
Energy	$\sum_{i=1}^{N}\sum_{j=1}^{N}P(i,j)^2$

Fig. 10 - Formulas to calculate texture features based on GLCM

Colour extraction using RGB model:

Any image consists of pixels, each pixel represents a dot in an image. A pixel contains three values and each value ranges between 0 to 255 , representing the amount of red,green and blue components. The combination of these forms an actual colour of the pixel.

The human visual system can distinguish hundreds of thousands of different colour shades and intensities, but only around 100 shades of grey. Therefore, in an image, a great deal of extra information may be contained in the colour, and this extra information can then be used to simplify image analysis, e.g. object identification and extraction based on colour.

The analysis of the texture features is done with the help of the Gabor Filter. Now, colour plays a significant part in diagnosing skin cancer. Dermatologists believe that a possible symptom of skin cancer could be the variation of the colour of the affected part in the skin region. To extract the colour features, the statistical parameters like variance, entropy, mean and standard deviation are determined.

5) Image classification

Image classification is the task of assigning a label or class to an entire image. Images are expected to have only one class for each image. Image classification models take an image as input and return a prediction about which class the image belongs to.

Classification between objects is a complex task and therefore image classification has been an important task within the field of computer vision. Image classification refers to the labelling of images into one of a number of predefined classes. There are potentially a different number of classes in which a given image can be classified. Manually checking and classifying images could be a tedious task especially when they are massive in number and therefore it will be very useful if we could automate this entire process using computer vision. The advancements in the field of autonomous driving also serve as a great example of the use of image classification in the real-world. The applications include automated image organisation, stock photography and video websites, visual search for improved product discoverability, large visual databases, image and face recognition on social networks, and many more.

There are several classifiers available, namely –

- Decision Tree.
- Naïve Bayes Classifier
- K-Nearest Classifier
- Support Vector Machines and
- Artificial Neural Networks

The classifier used in this project is the SVM Classifier. It was favoured over other algorithms because-

- Most effective in cases where the dimensions are greater than the number of samples
- More effective in high dimensional spaces
- Clear margin of separation can be used to optimize the process for accurate results

Support Vector Machines(SVM) is considered to be a classification approach but it can be employed in both types of classification and regression problems. It can easily handle multiple continuous and categorical variables. SVM constructs a hyperplane in multidimensional space to separate different classes. SVM generates optimal hyperplanes in an iterative manner, which is used to minimise an error. The core idea of SVM is to find a maximum marginal hyperplane(MMH) that best divides the dataset into classes.

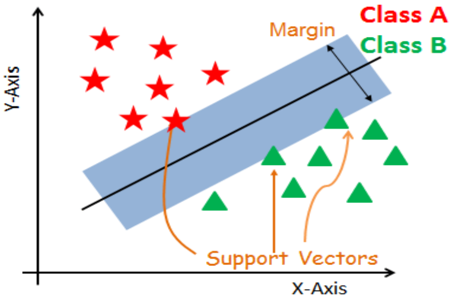

Fig. 11 - The structure of SVM

SVM is a very good algorithm for doing classification. It's a supervised learning algorithm that is mainly used to classify data into different classes. SVM trains on a set of label data. The main advantage of SVM is that it can be used for both classification and regression problems. SVM draws a decision boundary which is a hyperplane between any two classes in order to separate them or classify them. SVM is also used in Object Detection and image classification.

RESULTS

Several images were fed as inputs to the developed application and the results achieved isas follows:

1) **Input image:** An image is input to the application is as shown in the figure below. The sample image represents cancerous part of the skin infected.

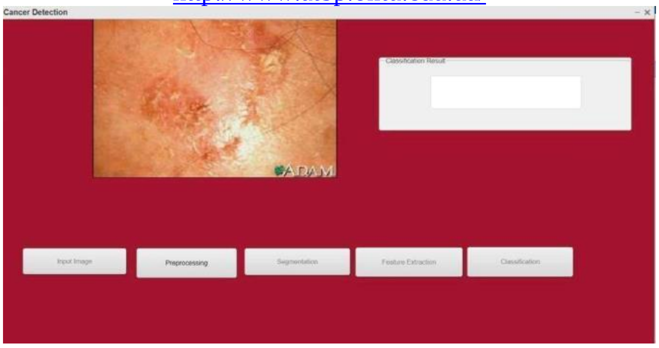

Fig. 12.1 – Input image

2) **Pre-processing stage**: For the input image, the pre-processing is performed by using methods like Median filtering for noise reduction and histogram equalization for contrast enhancement.

Fig. 12.2 – Pre-processed image

3) **Segmentation:** The image is segmented using Fuzzy C Means clustering techniqueand results are shown.

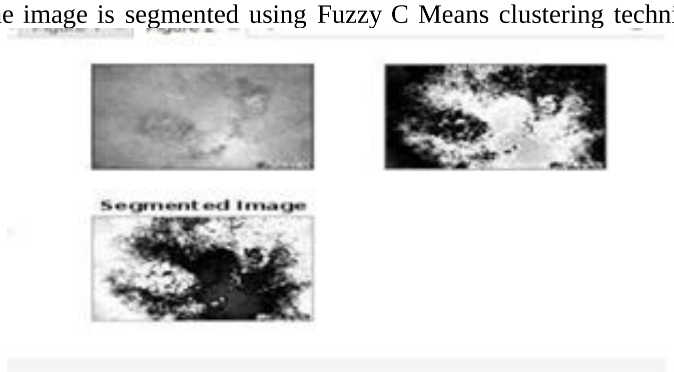

Fig. 12.3 – Segmented image

4) **Feature extraction:** Extracted features for the input image using Gabor filter andGLCM methods are shown in the fig.4 given below:

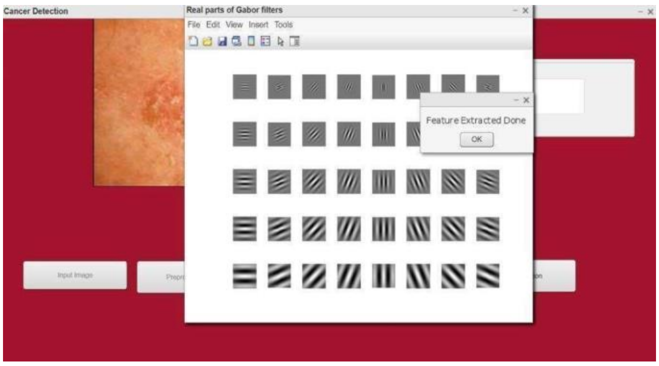

Fig. 12.4 – Feature extracted image

5) **Classification:** The image is classified using SVM algorithm into one of the fivetypes of skin cancer and result is shown as the output of the application.

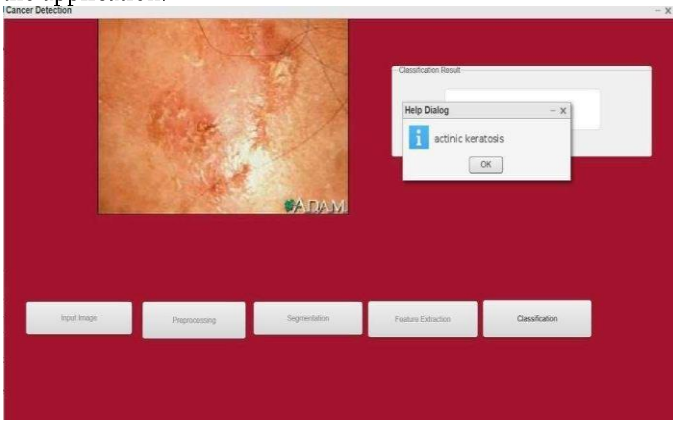

Fig. 12.5 – Classified image

Quantitatively assessment of the performance of the developed application was done byusing four parameters: Accuracy, Sensitivity, Precision and Specificity.

Sensitivity

The proportion of correctly identified positive samples among all positive samples ismeasured by sensitivity.

Sensitivity = TP/ TP + FN

Specificity

The proportion of correctly identified negative samples among all negative specificity.Specificity = TN/ TN + FP

Precision

It is the degree to which estimates from different samples agree with one another.Precision = TP/ TP+FP

Accuracy

The overall rate of perfectly identified samples is referred to as accuracy.

Accuracy=TP+TN/ TP+FP+TN+FNwhere,

TP= true positives FP= false positivesTN= true negatives FN= false negatives

A dataset of over 1000 images was considered and used to train this model and the above- mentioned parameters- True Positive(TP), True Negative(TN), False Positive(FP) and False Negative(FN) were calculated for various cancer types and the results are tabulated as follows:

Table 1.1 - Classification outcomes: TP, TN, FP, FN

Class	TP	FP	FN	TN
Actinic Keratoses	244	65	56	1135
Cherry Nevus	249	67	51	1133
Dermatofibroma	229	60	71	1140
Melanocytic Nevus	255	48	45	1151
Basal Cell Carcinoma	235	53	65	1147
Melanoma	230	65	70	1135

Based on these parameters, four evaluation metrics- Sensitivity, Specificity, Precisionand Accuracy were calculated and the results are tabulated as follows

Table 1.2 - Evaluation Metrics

Class	Sensitivity	Specificity	Precision	Accuracy
Actinic Keratoses	0.8133	0.9458	0.7896	0.9194
Cherry Nevus	0.83	0.9441	0.7879	0.9214
Dermatofibroma	0.7633	0.95	0.7923	0.9226
Melanocytic Nevus	0.85	0.9551	0.8415	0.9126
Basal Cell Carcinoma	0.7833	0.9558	0.8159	0.9213
Melanoma	0.7666	0.9458	0.7796	0.91

The overall values of these evaluation metrics were found as below-

Table 1.3 - Final Result

Metric	(%)
Accuracy	92.04
Sensitivity	80.11
Specificity	95.01
Precision	80.17

CONCLUSION

In conclusion, the use of machine learning for skin cancer detection and classification has shown promising results in recent years, with many research studies demonstrating high accuracy rates in detecting and classifying skin cancer. However, it's important to note that the use of machine learning for skin cancer detection and classification also raises several ethical considerations. It's crucial to address these ethical considerations when developing and deploying these types of models, and to work to address any potential ethical concerns. Additionally, it's also important to note that machine learning models are not perfect, they are not capable of replacing the expertise and experience of trained healthcare professionals. It's important that the technology is used in conjunction with human expertise and judgement, and that healthcare professionals are properly trained on how to use the technology and interpret its results. The use of machine learning for skin cancer detection and classification has several advantages, including:

High accuracy: Machine learning models have shown high accuracy rates in detecting and classifying skin cancer, which can lead to early diagnosis and better outcomes for patients.

Automation: Machine learning models can automate the process of detecting and classifying skin cancer, which can save time and reduce the workload for healthcare professionals.

Scalability: Machine learning models can be easily scaled to process large amounts of data, which can help to improve the detection and classification of skin cancer in large populations.

Cost-effective: Machine learning models can be cost-effective in the long run, as they can help to reduce the need for human labor, and increase the efficiency of the diagnostic process.

Consistency: Machine learning models can provide consistent and reliable results, which can help to reduce diagnostic errors and improve patient outcomes.

Personalization: Machine learning models can be trained to detect and classify skin cancer for specific populations, like people with darker skin tones, which can help to improve diagnostic accuracy for certain groups of people.

REFERENCE

1. R. A. Said et al., "Skin Cancer Detection and Classification Based on Deep Learning," 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 2022, pp. 1-11, doi: 10.1109/ICCR56254.2022.9996077

2. Fraiwan, Mohammad, and Esraa Faouri. 2022. "On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning" Sensors 22, no. 13: 4963. https://doi.org/10.3390/s22134963

3. Grignaffini, Flavia, Francesco Barbuto, Lorenzo Piazzo, Maurizio Troiano, Patrizio Simeoni, Fabio Mangini, Giovanni Pellacani, Carmen Cantisani, and Fabrizio Frezza. 2022. "Machine Learning Approaches for Skin Cancer Classification from Dermoscopic Images: A Systematic Review" Algorithms 15, no. 11: 438. https://doi.org/10.3390/a15110438

4. Pal, Osim Kumar. "Skin Disease Classification: A Comparative Analysis of K-Nearest Neighbors (KNN) and Random Forest Algorithm."

5. Das, Kinnor, Clay J. Cockerell, Anant Patil, Paweł Pietkiewicz, Mario Giulini, Stephan Grabbe, and Mohamad Goldust. 2021. "Machine Learning and Its Application in Skin Cancer" International Journal of Environmental Research and Public Health 18, no. 24: 13409. https://doi.org/10.3390/ijerph18241340

6. Naeem, Ahmad & Farooq, Shoaib & Khelifi, Adel & Abid, Adnan. (2020). Malignant Melanoma Classification Using Deep Learning: Datasets, Performance Measurements, Challenges and Opportunities. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3001507.

7. Sreedhar, Burada & B.E, Manjunath & Kumar, M.. (2020). A Comparative Study of Melanoma Skin Cancer Detection in Traditional and Current Image Processing Techniques. 654-658. 10.1109/I-SMAC49090.2020.9243501.

8. M. A. Thaajwer and U. P. Ishanka, "Melanoma Skin Cancer Detection Using Image Processing and Machine Learning Techniques," 2020 2nd International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, 2020, pp. 363-368, doi: 10.1109/ICAC51239.2020.9357309.

9. Waghulde, Minakshi & Kulkarni, Shirish & Phadke, Gargi. (2019). Detection of Skin Cancer Lesions from Digital Images with Image Processing Techniques. 1-6. 10.1109/PuneCon46936.2019.9105886.

10. Smaoui, Nadia & Derbel, Nabil. (2018). Simple But Efficient Approach for Image Based Skin Cancer Diagnosis. 274-280. 10.1109/SSD.2018.8570526.

Отримана в редакції 12.06.2023. Прийнята до друку 12.10.2023. Received 12 June 2023. Approved 12 Jctober 2023. Available in Internet 03 January 2024