IJCRT.ORG





INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AN AUTOMATIC DETECTION OF SKIN CANCER USING CNN

¹Shabnam Ara Chowdhary, ²Er.Mandeep Kaur ¹M.Tech Student, ²Assistant Professor, ¹Computer Science Engineering, ¹Desh Bhagat University, Mandi Gobindgarh, Fatehgarh Sahib, Punjab- 147301, India

Abstract: The depleting ozone capacity causes more Ultra Violet (UV) rays towards the earth and can cause skin cancer. World Health Organization (WHO) states that people with skin cancer continues to increase from year to year in each decade. There are about 2-3 million non-melanoma skin cancer patients and 2 million patients with the disease melanoma skin cancer. Therefore it is necessary to detect skin cancer early to reduce the number of skin cancer sufferers. Early detection of skin cancer can be done by doctors by looking directly based on the ABCD feature, but this method is less efficient because visual response can be different for different persons, so Computer-Aided Diagnosis (CAD) was made to help parties in detecting skin cancer. The process carried out is pre-processing by using an image or color augmentation filtering, using noise removal filtering, brightness, contrast enhancement using histogram equalization and resizing. Furthermore, the classification process is carried out using Convolutional Neural Network (CNN) to classify skin cancer. The study was conducted on 2808 benign skin cancer data and 2394 non-cancer malignant data totaling 5202 data. Furthermore, the data will be broken down with the distribution and divided into 80% training data and 20% testing data. The best results obtained by the proposed model using maximum of 1024 hidden layers with an accuracy value of 94.85%, and a loss of 22.96%.

Index Terms –Skin Cancer, Deep Learning, Artificial Neural Networks, Convolutional Neural Networks.

I. INTRODUCTION

Skin disease is the most common disease and infects the human body the fastest. In World, the incidence of skin diseases is ranked 4th after upper respiratory tract infection and diarrhea [1]. This happens because the skin is the first organ to receive external stimuli in the form of touch, temperature and other stimuli. Skin diseases can be caused by microorganisms, climatic influences, environmental conditions and unhealthy lifestyles. Especially in tropical areas like India, apart from attacking adults this disease also often occurs in children or toddlers [1].

Microorganisms will more easily infect the body if there is a lack of awareness of body hygiene. Skin disease is common because it can affect everyone without any age and education level restrictions. With a variety of infectious microorganisms such as fungi, bacteria and viruses, skin diseases have varying types both in terms of appearance, color, shape and texture on the skin. Skin diseases can be classified by identifying the color, texture and shape that appears based on the sense of sight [2]. Identification will be more optimal if you use digital techniques to process images of disease images that are taken through a camera. Research in identifying skin diseases has been carried out using the Deep Learning methods. The types of diseases that were the object of his research were scabies, measles, acne, chicken pox, and dermatitis [3,4].

NU

The level of accuracy obtained are 80%, and 85%. The author uses images by analyzing skin color and texture. Identification of skin diseases has also been carried out with a local binary pattern using a Support Vector Machine and Genetic Algorithm [5]. Color and texture analysis used Local Binary Pattern (LBP), Gray Level Co-Occurrence Matrix (GLCM), Discrete Cosine Transform (DCT) and Discrete Fourier Transform (DFT). There are 3 types of diseases tested. The level of accuracy obtained is 89.65% [6]. There are many methods that can be applied to identify one of them, namely the Extreme Learning Machine (ELM).

The ELM method has been applied in research to identify plant species based on leaf edge images. The leaf image undergoes preprocessing first and then features extraction the experimental results also showed that ELM is able to achieve good generalization performances independently from the feature extraction process employed by the data sets, avoiding data set bias problem [7]. The feature extraction used is the Gray Level Co-occurrence Matrix. ELM is also applied in research to classify the types of Parkinson's disease. The test results of 195 data showed a learning accuracy value of 88.72%. Feature extraction used is Particle Swarm Optimization (PSO) [8]. Based on the description above, the authors make a system by utilizing a digital image processing system. The system created is in the form of an application where the data input is a photo image of the skin disease. The application will recognize the type of skin disease after several processes. The process carried out is image input, image pre-processing, image analysis based on texture, then image classification to determine the type of disease and how to treat it. The technique used is GLCM for feature extraction and ELM artificial neural network for identification from images. The reason the authors use ELM is because this neural network has high accuracy in decision making [9].

Melanoma skin cancer is the most deadly type of skin cancer. Melanoma starts in the skin cells and can spread to other organs in the body through the blood. The appearance of a new mole or a change in an existing mole are common signs of melanoma symptoms [10-13]. The main cause of melanoma because the skin is often exposed to ultra violet rays. Generally doctors use the biopsy method to diagnose melanoma. Biopsy method is the taking of small pieces of body tissue for diagnostic samples. By cutting or scraping a small piece of tissue or by using a syringe to take a sample, then the sample is tested by several trained doctors for diagnosis. This biopsy process is painful and takes a long time. From the research that has been done, early diagnosis of melanoma shows that more than 90% can be cured and if it is late to diagnose melanoma, it shows that it is less than 50% curable. Therefore, if melanoma is not diagnosed at an early stage, it will cause death in the sufferer. So detecting skin cancer from an early stage is inevitable [10-13].

II. LITERATURE REVIEW

2.1. SKIN

The skin is a part of the body that forms a protective wall that surrounds the entire body and has a function as temperature control, glandular secretions and sensory dependence on the external environment [14]. Every structure of the skin has the potential for disease.

Skin disease is defined as impaired function which is restricted or dominant on the surface of the skin. Skin disease caused by fungi, bacteria and viruses. Skin disease can attack someone who has a low level of immunity [14]. Skin diseases are grouped according to the type of disease and the level of infection because there are types of skin diseases that are not dangerous and there are also types of skin diseases that are so dangerous that they can cause death. There are 2 types of skin diseases according to the level of treatment, namely skin diseases that can be cured and skin diseases that are difficult or not to be cured.

The following types of skin diseases will be identified:

2.1.1 Ringworm

Ringwormis a disease that infects the skin and is caused by a fungus. This disease infects the nails, upper body of the head, groin and feet. Ringworm or what is often called 'tinea' is a common disease and can affect anyone, especially children. Examples of physical images of ringworm skin disease [15] is shown in Figure 2.1.



Figure 2.1.: Ringworm disease

The main characteristic of ringworm that is easily recognized is the appearance of a red, ring-shaped rash that is irregular in shape [15,16]. This rash is itchy, scaly, and can spread all over the body. If it is severe, the rash will swell and start to blister. Ringworms on the face and neck may not look like rings but will still feel itchy and blistered. However, these parts of the body can become scaly and crusty. In parts of the body such as the hands, this disease causes the skin between the two fingers and the palms to thicken. Usually this disease only infects one hand, but it does not rule out the possibility of infecting both palms.

The cause of ringworm is dermatophyte fungi, which are herbs that can survive on human skin. This fungus lives in the part of the skin called keratin, which is the basic building block for human skin, nails and hair. This fungus reproduces and spreads through the box directly with humans, animals, objects, and even soil [15,16]. There are several groups that are more susceptible to infection with this disease such as children, the elderly, diabetics, people who have atherosclerosis, people who have had previous fungal infections, people with poor blood circulation, and people with decreased immune systems such as people with HIV or users. steroid tablets.

Treatment can be done by providing anti-fungal drugs in the form of ointments. But if it doesn't get better or gets worse, sufferers can take anti-fungal tablets. Especially for the symptoms of this disease that are on the head, sufferers are advised to use anti-fungal shampoo while bathing so that the spread of fungus does not spread more widely.

Ringworm disease is a disease that can spread widely and quickly through direct contact through the skin and the use of personal items such as soap towels, clothes and others. Thus, the most effective way to prevent the transmission of this disease is to maintain daily body hygiene such as regularly bathing every day and maintaining the cleanliness of personal items.

2.1.2 Melanoma

Melanoma is a disease that develops in the skin structure, namely melanocytes[17,18]. Melanocytes function as a producer of melanin which functions to absorb ultraviolet rays and protect the skin from damage. Melanoma is a type of skin disease that is rare and very dangerous for the body. Melanoma usually affects the skin and over time it can attack other organs of the body. A common sign of melanoma is the appearance of a new mole or a change in an existing mole on the body. Examples of physical images of melanoma skin disease are shown in Figure 2.2.



Figure 2.2. Melanoma disease

Melanoma occurs when the skin's pigment cells cannot develop normally. Until now, there is no known cause of melanoma. However, some researchers think that melanoma can occur because a person is exposed to too much ultraviolet radiation, both natural and artificial [17,18]. Until now, researchers are still looking for the exact cause of this disease. There are several factors that are known to increase a person's risk of

developing melanoma. Chances are you have lots of moles and spots on your skin, have pale skin and burn easily, have a family with melanoma and have red and blonde hair.

Symptoms of melanoma are usually characterized by the appearance of a mole and changes in existing moles. Usually normal moles have one color and are round or oval in shape and less than 6 millimeters [17,18]. While melanoma usually has the following characteristics:

- Has more than one color a.
- The shape is irregular b.
- The diameter is greater than 6 mm c.
- Feels itchy and may bleed d.

Prevention that can be done is to avoid the skin from direct exposure to sunlight for a long time. This prevention also applies to people who have had melanoma and other skin diseases. This is done to prevent the appearance of disease cells that have been removed during treatment.

2.1.3 Hemangioma

Hemangioma is a disease that has symptoms in the presence of birthmarks that appear on the skin due to the presence of more or abnormal blood vessels in the form of red bumps. This disease is classified as a type of blood vessel tumor. Hemangiomas can appear on the inner skin and make the skin look blue or purple. Hemangiomas usually appear on the scalp, back, chest, face, and can sometimes appear anywhere on the body. This disease can shrink as the sufferer gets older [19,20]. The depreciation process usually occurs at the age of 10 years.

The initial symptom of a hemangioma is usually a flat red mark that appears on the body and multiplies rapidly and makes the skin stand out. But after that, the hemangioma will enter an inactive phase and disappear slowly. Hemangioma can be there at birth, but can also appear only after the child is a few months old. A physical example of a hemangioma skin disease image is shown in Figure 2.3.



The difference in skin color will remain or be permanent, but the color of the birthmark will be faded or not as bright as when the hemangioma first appeared. About 50 percent of hemangiomas have faded by the time a child is 5 years old. See a doctor if the hemangioma is infected, bleeding or has an open wound.

This disease is a type of blood vessel tumor that is not malignant and rarely causes complications. The appearance of hemangiomas usually occurs several months after the baby is born. About 50 percent of hemangiomas shrink when the child is 5 years old and eventually fade after 10 years of age [19,20]. Hemangiomas generally don't need treatment, unless the growth is very large and bothersome.

The initial symptom of a hemangioma appears as a red mark on the skin that can grow or develop rapidly so that it can be seen protruding from the surface of the skin. But after that, the hemangioma will enter an inactive phase, then slowly disappear. Even if it disappears, the hemangioma will leave a permanent difference in skin color, although not as bright as it was when it first appeared.

Additional blood vessels formed due to an abnormal growth will collect and form a hemangioma. The cause of this abnormal growth of blood vessels is not yet known with certainty.

Diagnosis of a hemangioma can be made only through a physical examination. To support the diagnosis, a Dopler ultrasound examination can be performed to see blood circulation through the hemangioma area to distinguish it from other rashes that have the same characteristics (for example rubella, measles, or acrodermatitis). This examination can also help the doctor determine whether a hemangioma is increase in size, settle or shrink. If the hemangoma growth looks unusual or causes scarring, a blood test or skin biopsy



may be performed. Most cases of hemangiomas do not require special treatment because they do not cause physical problems and will fade slowly. However, hemangiomas can cause psychological and social problems. That reason makes some people feel the need to treat it.

Another reason hemangiomas are treated is if their growth interferes with vision. In such cases, treatment can be given by administering drugs or laser surgery [19,20]. Medicines that can be given are:

• Corticosteroids. This drug can be given orally, topically, or injected at the site where the hemangioma appears. Side effects that may occur after using corticosteroids include growth disorders, high blood sugar or blood pressure levels, and cataracts.

• Beta blockers. In mild cases, timolol can be given in gel form. As for severe cases, given oral propranolol. Side effects that can arise from using this group of drugs are wheezing, increased blood sugar, and increased blood pressure.

• Vincristine. This medicine is given if the hemangioma has impaired vision or breathing. Vincristine is administered by injection every month.

Laser treatment methods can be used to stop the growth of the hemangioma and treat the pain that appears. Laser surgery can also reduce skin discoloration after the hemangioma has healed. Some of the possible side effects of treating hemangiomas through laser surgery are pain, bleeding, scarring, and discoloration of the skin.

In general, hemangiomas are harmless and painless, but in some cases they can cause bleeding, infection, or become painful open sores. In rare cases, hemangiomas can cause disturbances in breathing, vision, hearing, or interfering with the flow of urination and bowel movements, depending on the location and size of the hemangioma

2.1.4 Eczema

Eczemaor what is also called atopic dermatitis is a condition in which the skin becomes dry, cracked, itchy, and reddish in color. Atopic eczema is a type of long-term (chronic) disease and the symptoms can improve and then become severe again [21,22]. This usually happens two or three times a month. Atopic eczema is the most common form of eczema. Eczema itself is an inflammatory condition in the skin characterized by dry and reddish skin, while the word atopic refers to people that are likely to have certain allergies. Atopic eczema sufferers are at risk for other atopic conditions, such as asthma and hay fever. Examples of physical images of skin diseases eczema like picture 2.4.



Figure 2.4. Eczema disease

Most of the atopic eczema sufferers experience this condition before the age of five years and there are some that persist into adolescence and adulthood [21,22].

The following are symptoms of atopic dermatitis that often occur:

- 1. Itching especially at night.
- 2. Rash red and spread from scratching.
- 3. Blisters that can ooze and can become open sores when scratched.
- 4. Cracked, dry, scaly skin
- 5. Blisters that ooze and leave a mark after scratching

No	Researcher / Year	Research Title	Information	
1	Eric Vander Putten, Ameer Kambod and Mobeen Kambod (Year 2018) [23]	Deep Residual Neural Networks for Automated Basal Cell Carcinoma Detection	In this study, the Deep Residual Neural Network method was used to analyze the KSB dermoscopy images. This study succeeded in achieving 93% accuracy in the detection of KSB.	
2	Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M., Swetter, Helen M. Blau, Sebastian Thrun (Year 2017)[24]	Dermatologist level classification of skin cancer with deep neural networks	In this study, testing was carried out by comparing the system performance with 21 certified dermatologists from clinical images with two use case classifications, namely kerationocyte carcinoma versus benign seborrheic keratosis and malignant melanoma versus benign nevi. This study managed to achieve an accuracy of 93.3%.	
3	Suleiman Mustafa, Ali Baba Dauda, Mohammed Dauda (Year 2017)[25]	Image Processing and SVM Classification for Melanoma Detection	In this study the authors uses geometric features of the ABCD melanoma rule to train SVM machine learning algorithms. This study managed to achieve an accuracy of 88.0%.	
4	Aswin.R.B,J. Abdul Jaleel,Sibi Salim (Year 2014) [26]	Hybrid Genetic Algorithm - Artificial Neural Network Classifier for Skin Cancer Detection	In this study, the feature extraction technique used here is GLCM (Gray Level Co- occurrence Matrix) and RGB color features, and genetic algorithms are used to improve classification accuracy. This study managed to achieve an accuracy of 88.0%.	
6	Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton (Year 2012)[27]	ImageNet Classification with Deep Convolutional Neural Networks	In this study, a Deep Convolutional Neural Network was used to classify ImageNet, which is more than 15 million labeled high- resolution images, consisting of about 22,000 categories. The best model learning accuracy obtained in this study is 78.1%.	
7	Joseph Supriya, R Panicker Janu (Year 2016) [50]	Skin Lesion Analysis System for Melanoma Detection with an Effective Hair Segmentation Method	This study uses the thresholding method of segmentation, hair removal (preprocessing), and SVM (classification).	
8	M. Toğaçar, Z. Cömert, B. Ergen and Ü. Budak (Year 2019) [28]	Brain Hemorrhage Detection based on Heat Maps, Autoencoder and CNN Architecture	Grayscale, Scaling, Extreme Learning Machine (CNN) with Autoencoder with the success rate in the classification as 98.57%	

III. PROPOSED METHODOLOGY 3.1 DATA SET

The data used for this research is a dermoscopy image obtained from https://www.kaggle.com. The data set consists of 5202 images for training data and 30 images for data testing. The training data consisted of 2808 melanoma images and 2394 non-melanoma images. Test data consisted of 4 images of melanoma and 4 images of non melanoma.

3.2 SYSTEM ANALYSIS

The flow of this research begins by examining and looking for several literature reviews on skin cancer detection using several methods used in classifying types of skin cancer and differentiating skin cancer (Malignant) from non-cancer (Benign) using the classification method. Furthermore, after references and literature studies have been obtained, the next process is to find the data needed in the study. The data used can be found on the official website of the International Skin Imaging Collaboration Archive which contains data on skin cancer and its label. The data to be used is primary data, so preprocessing is necessary to get a good image. After preprocessing, the next step is to extract the features from the preprocessing results to get the statistical features of the image. In the next stage, the statistical features will be used as input for the classification process. The system built will have several stages. The first stage will start from assessing the dermoscopy image which will be used as training data and testing system data to be created. The next stage is preprocessing using the image adjustment method to increase the light intensity of the image, then using the thresholding method to create a black and white image on the image to be stopped, then the classification stage using the Convolutional Neural Network method. The general architecture of the system can be seen in Figure 3.1



Figure 3.1: Proposed Architecture or System.

3.2.1 Dermoscopy Image Input

The dermoscopy image that will be used is a stored image of the skin on the desktop. The image to be processed is an image in .jpg form. Images as example which will be used can be seen in Figure 3.2(a&b).



3.2.2. Preprocessing

This stage is an image processing stage that aims to produce better image to place the next stage. This preprocessing stage consists of resize and saturation, brightness and contrast processes.

Change size

Resize Image is the stage in which the process of changing a size the image becomes larger or smaller than the initial image size with the existing size predetermined. The process aims to reduce the burden on the device because the object detection method is very complex and takes up a lot of space large storage. However, the first stage is resizing dimensions of the resulting image whereas dimensions of the skin cancer detection are not always the same. This stage is needed for assistance pixels from the image to be processed at the testing stage. In this study, it has different pixel sizes. Therefore need to be resized so that each training data and testing data has a size the same dimensions and value ranges. The larger the number of pixels it will be the more time it takes in the image processing process. Example Images that go through the resizing process can be seen in the equation below:

$$[(n + 2p)x (n + 2p)image] * [(f x f)filter] \rightarrow [(n x n)image]$$

• Saturation

Saturation states the level of purity of light color, which indicates how much white color is given to the color. For example, red is a 100% saturated color, while pink is red with a very low saturation level (because there is a white color in it). So, if the saturation represents the actual color, S_v then the saturation states how deep the *c* color is with vectors *v* in equation below:-

$$S_v = \begin{cases} 0 & if \ V = 0 \\ \frac{c}{V}, otherwise \end{cases}$$

IJCRT2308416 International Journal of Creative Research Thoughts (IJCRT) www.ijcrt.org d886

• Brightness

Brightness aims to change the pixel size into M x N size pixels. This is done because each image that is processed does not necessarily have the same color. Brightness is also used to reduce digital images so that the number of pixels to be processed is not too much. The more the number of pixels, the more data input. So that causes longer computation time.

$$B=\frac{L}{4\pi D^2}$$

Here the denominator is just the area of a sphere of radius D. All that this formula says is that brightness is the luminosity divided by the area which is illuminated. Because the area of a sphere increases as the square of its radius, it's the square of D which appears in the denominator.

• Contrast or Contrast Limited Adaptive Histogram Equalization

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a method of repair from Adaptive Histogram Equalization (AHE) for the contrast enhancement process on imagery. CLAHE is the process of adding contrast by providing a limit value (clip limit) on the image so that the image looks clearer and there is no increase in contrast excessive. CLAHE is generally used for low contrast medical image enhancement. In Red's image, Blue, Green (RGB), CLAHE can be formulated as follows:



3.3 THRESHOLDING

Thresholding is the process of changing an image with a gray degree into a binary or black and white image so that it can be seen which areas are included in the object and the background of the image clearly. The input for the thresholding process is a Grayscale Image. The output of this process is a Binary Image. Thresholding is used to separate the part of the image that is needed to remove the part of the image that is not needed. In its implementation, Thresholding requires a value that is used as a limiting value between the intensity of the main object and the background, and this value is called Threshold. Thresholding is used to partition the image by setting the intensity value of all pixels greater than the Threshold T value as the foreground and smaller than the Threshold T value as the background or vice versa.

$$g(x, y) = \begin{cases} 1, if f(x, y) \ge T \\ 0, if f(x, y) < T \end{cases}$$

3.4 CLASSIFICATION

In this study, the method used for the classification process is Convolutional Neural Network. The Convolutional Neural Network Architecture for melanoma identification can be seen in Figure 3.3.



Figure 3.3: Neural Network Architecture for Melanoma Identification

The description of the Convolutional Neural Network architectural image for melanoma identification can be seen in Table 3.1.

LAYER	NUMBE <mark>R OF NODES</mark>	DESCRIPTION
INPUT	22 <mark>4 x 224</mark>	THE NUMBER OF EACH IMAGE PIXEL
HIDDEN	30	NODES
OUTPUT	2	MELANOMA AND NON MELANOMA

Table 3.1: Convolutional Neural Network Architectural Image for Melanoma Identification

The architecture that will be used in this network consists of 3 layers namely the Input Layer, Hidden Layer, and Output Layer. Input data as much as 224 x 224 nodes, Hidden as many as 30 nodes and Output consists of 2 nodes (Non Melanoma and Melanoma). Hidden Node is determined randomly. Thereinafter the scheme will determined through several trials against system requirements. Selection of Hidden 30 Nodes are a good weight in resulting in high accuracy in melanoma identification systems and not takes a lot of time in the image processing. Data to be entered in the Input Layer will be transformed first. Training conducted in order to find the optimal weight and bias or suitable for use on Testing process. Training steps for Epoch to be processed are as follows:-

Feed Forward Backpropagation

Training Feed forward backpropagation training includes 3 phases. The first phase is the forward phase. Pattern The input is calculated forward from the input screen to the output screen using the activation function which is determined. The second phase is the reverse phase. The difference between network output and target desired is an error that occurs. The error propagates backwards, starts of the line directly related to the units on the output screen. The third phase is weight modification to reduce errors that occur.

• Phase I: Forward Propagation

During forward propagation, the input signal $(= x_i)$ is propagated to the hidden layer use the specified activation function. Output of each hidden display unit $(= z_j)$ it is then propagated forward again to the hidden layer above it using the function activation specified. And so on until it produces network output(= y_k).

Next, the network output $(= y_k)$ is compared with the target to be achieved $(= t_k)$. The difference from t_k to y_k that is $(t_k - y_k)$ an error that occurs. If this error is over smaller than the specified tolerance limit, then the iteration is terminated. However, if it is a mistake still greater than the tolerance limit, the weight of each line in the network will be modified to reduce errors that occur.

Phase 2 Back Propagation

Based on the error $t_k - y_k$ the factor $\delta_k (k = 1, 2, ..., m)$ that is used for distributes faults across units y_k to all hidden units directly connected to y_k . δ_k also used to change line weights that are directly related to the unit of output.

In the same way, the factor δ_i (j = 1, 2, ..., p) is calculated in each unit on the screen hidden as the basis for changing the weight of all lines originating from hidden units on the screen in beneath it. And so on until all the factors δ in the hidden unit are related directly with the input unit calculated.

Phase 3: Weight Change

After all δ factors are calculated, the weights of all lines are modified together. The change in weight of a line is based on the factor $\boldsymbol{\delta}$ the neurons on the screen above it. For example, the change in the weight of the line leading to the output screen is based on the δ_k that is in the output unit.

These three phases are repeated continuously until the stopping conditions are met. Generally the stop condition that is often used is the number of iterations or errors. The iteration will be terminated if the number of iterations performed exceeds the set maximum number of iterations, or if the error that occurs is less than the allowable tolerance limit.

The training algorithm for a network with one hidden layer (with a binary sigmoid activation function) is as follows:

Step 0: Initialize all weights with small random numbers.

Step 1: If the termination conditions are not met, perform steps 2 - 8.

Step 2: For each pair of training data, perform steps 3 - 8.

Phase I: forward propagation

Step 3: Each input unit receives a signal and passes it to the hidden unit above it. JCR

Step 4: Calculate all outputs in hidden units $z_i (j = 1, 2 \dots p)$

$$\begin{aligned} \mathbf{z}_{-net_j} &= \mathbf{v}_{jo} + \sum_{i=1}^{n} x_i \mathbf{v}_{ji} \\ \mathbf{z}_j &= f'\left(\mathbf{z}_{-net_j}\right) = \frac{1}{1 + e^{\mathbf{z}_{-net_j}}} \end{aligned}$$

Step 5: Calculate all network outputs in units $y_k (k = 1, 2 \dots m)$

$$y_{-net_k} = w_{ko} + \sum_{j=1}^{p} z_j w_{kj}$$
$$y_k = f' \left(y_{-net_k} \right) = \frac{1}{1 + e^{y_{-net_k}}}$$

> Phase II: Backward propagation

Step 6: Calculate the output unit factor based on the error in each unit of output y_k ($k = 1, 2 \dots m$)

$$\delta_k = (t_k - y_k) f'(y_{-net_k}) = (t_k - y_k) y_k (1 - y_k) (eq.3.13)$$

 δ_k is the error unit to be used in changing the weight of the layer below it (*step 7*).

Calculate the weight w_{ki} change term (which will be used later to convert the weight w_{ki}) with the rate as:-

$$\alpha \cdot \Delta w_{ki} = \alpha \delta_k Z_i; k = 1, 2, \dots, m; j = 0, 1, \dots, p$$

Step 7: Calculate the hidden unit factor based on the error in each hidden unit as Z_i (i = 1, 2, ..., p)

$$\delta_{-net_j=} \sum_{k=1}^{m} \delta_k w_{kj}$$

Hidden unit δ factor: $\delta_j = \delta_n net_j f'(z_n net_j) = \delta_n net_j Z_j (1 - Z_j)$ Calculate the term change in the weight of V_{ji} (which will be used later to change the weight of V_{ji}). $\Delta v_{ji} = \alpha \, \delta_j \, x_i \, ; \, j = 1, 2, ..., p \, ; \, I = 0, 1, ..., n$

> Phase III: Change in weight

Step 8: Calculate all changes in weight. Change in line weight leading to the output unit:

$$W_{kj} (new) = W_{kj} (old) + \Delta W_{kj} (k = 1, 2, ..., m; j = 0, 1, ..., p)$$

Change in line weight leading to hidden units:

$$V_{ji} (new) = V_{ji} (old) + \Delta V_{ji} (j = 1, 2, ..., p; I = 0, 1, ..., n)$$

Once done, the network can be used for pattern recognition. In this case, only forward propagation (steps 4 and 5) is used to determine the network output. If the activation function used is not binary sigmoid, then steps 4 and 5 must be adjusted accordingly. Likewise the derivatives in steps 6 and 7.

However under the backpropagation where the global error E on the network output layer feedforward neural are:

$$\mathbf{E} = \frac{1}{2} \sum_{j}^{\mathbf{N}} (\mathbf{d}_{j} \mathbf{o}_{j})^{2}$$

where: $\mathbf{d}_{\mathbf{i}}$ = expected output and $\mathbf{o}_{\mathbf{i}}$ = neural network output.

This error is the core of the system where the iteration that occurs tries to fulfill the desired global error as quickly as possible. The basic equation of weight correction develops by laying the basis of the analysis towards the momentum term, weight decay and batch learning. The basic equation is as follows:-

$$W_{ji}(n+1) = W_{ji}(n) + \Delta W_{ji}$$

Where n is the iteration generation (epoch). Fixed weight change ΔW_{ii} proportional to negative gradient on equations down as follows:

$$\Delta W_{ji} = -\eta \frac{\partial E}{\partial W_{ji}}$$

The proportional factor η is the learning rate which defines the depth of the iteration step. Selection of the learning rate parameter that is too large makes the resulting weight overshoot the global minimum error. The learning rate is too small to make the convergence process slow. The learning rate parameter error will be resulting in an oscillation around the minimum global error value. 120,

IV. RESULT AND SIMULATION

4.1 SYSTEM IMPLEMENTATION

At the implementation stage of the melanoma identification system using the CNN method requires supporting hardware and software, including:

Hardware and Software Specifications

Hardware and software specifications used to create the program to identify the type of hurting the skin is as follows:

- 1. PROCESSOR AMD E1-2100 APU with Radeon (TM) HD Graphics (4 CPUs), ~ 2.0 GHz
- 2. Memory (RANDOM ACCESS MEMORY) 8 GB
- 3. Open Source Linux Centos 7.0 64-bit based operating system
- 4. One Tera Byte HDD capacity
- 5. Python

4.2 SIMULATION

The data used for this study consisted of two type of data, namely data training used in the learning process, data validation used in the process of testing the results of learning when doing training (validation) to do testing after training. The data used is benchmark data in the form of images with the already JPEG (Joint Photographic Experts Group) format has been used for training and testing data available in challenges in representation learning: for skin cancer detection.

4.2.1. Data Training

Training data are data used to carry out the learning process facial expressions. The data used for the training process were 5202 images for four categories.

4.2.2. Data Validation

Data Validation is data used to test current accuracy carry out the training process. The data used for the validation process are as large as 2808 positive and 2394 negative images for benign an malignant categories respectively.

4.2.3. Data Testing

Data Testing is data that is used to carry out the testing process for the results of learning and detecting skin cancer that have been stored in a model. The data used is in the form of images from dermoscopy image expressions that are different from the images that have been used in the training and validation processes. The testing data used has different color and size modes.

```
1 import pandas as pd
                        2 import numpy as np
                        3 import tensorflow as tf
                        4 import matplotlib.pyplot as plt
                        5
                        6
                        7 %matplotlib inline
                   Figure 4.1: Python Machine Learning Libraries
 1 from os import listdir
 2 from os.path import isfile, join
 3 # Load Dataset:
4 # from the folder of images create dataframe with file-labels
 5 train_benign = [['/content/drive/MyDrive/SkinCancer/train/benign/{}'.format(f), 1 ]
                  for f in listdir('/content/drive/MyDrive/SkinCancer/train/benign/
 6
                  if isfile(join('/content/drive/MyDrive/SkinCancer/train/benign/', f))]
7
 8 train_malignant = [['/content/drive/MyDrive/SkinCancer/train/malignant/{}'.format(f), 0 ]
                     for f in listdir('/content/drive/MyDrive/SkinCancer/train/malignant/')
9
10
                     if isfile(join('/content/drive/MyDrive/SkinCancer/train/malignant/', f))]
11 test_benign = [['/content/drive/MyDrive/SkinCancer/test/benign/{}'.format(f), 1 ]
                 for f in listdir('/content/drive/MyDrive/SkinCancer/test/benign/')
12
13
                 if isfile(join('/content/drive/MyDrive/SkinCancer/test/benign/', f))]
14 test_malignant = [['/content/drive/MyDrive/SkinCancer/test/malignant/{}'.format(f), 0 ]
                    for f in listdir('/content/drive/MyDrive/SkinCancer/test/malignant/')
15
                    if isfile(join('/content/drive/MyDrive/SkinCancer/test/malignant', f))]
16
17 files = pd.DataFrame(train_benign+train_malignant+test_benign+test_malignant,
                 columns=['img', 'binary label'])#.sample(frac=1, random state=1869097)
18
19 files.head()
20
```

Figure 4.2: Loading Training and Test Data from Files System to Memory Stream for Processing





test len 781 [0.46 0.54]

Figure 4.4: Splitting the Trained Data Thereafter Validating and Testing Data

```
1 import tensorflow as tf
2 train_dataset = tf.data.Dataset.from_tensor_slices((X_train, y_train))
3 val_dataset = tf.data.Dataset.from_tensor_slices((X_val, y_val))
4 test_dataset = tf.data.Dataset.from_tensor_slices((X_test, y_test))
5 for xb, yb in train_dataset.batch(4).take(1):
6 print(xb)
7 print(yb)
tf.Tensor(
[b'/content/drive/MyDrive/SkinCancer/train/benign/1101.jpg'
b'/content/drive/MyDrive/SkinCancer/train/malignant/180.jpg'
b'/content/drive/MyDrive/SkinCancer/train/malignant/95.jpg'
b'/content/drive/MyDrive/SkinCancer/test/malignant/2.jpg'], shape=(4,), dtype=string)
tf.Tensor([1 0 0 0], shape=(4,), dtype=int32)
```

Figure 4.5: Splitting the Trained Data Thereafter Validating and Testing Data

4.2.4 Pre-Processing

This step is an image dispensation stage that aims to manufacture a enhanced image for advance processing. This preprocessing stage consists of resize and color augmentation processes.

```
# data augmentation for an image
def augment(image, label):
    # randomly Rotate
    image = tf.image.rot90(image, k=4)
    # random flip left-rigt and up-down
    image = tf.image.random_flip_left_right(image)
    image = tf.image.random_flip_up_down(image)
    # image = tf.image.random_crop(image,
    # size= [image.shape[0], image.shape[1], 3])
    # color augmentation
    image = tf.image.random_hue(image, 0.05)
    image = tf.image.random_saturation(image, 0.8, 1.3)
    image = tf.image.random_brightness(image, 0.09)
    image = tf.image.random_contrast(image, 0.9, 1.1)
    return image, label
```

Figure 4.6: Image Augmentation Process



4.3 CLASSIFICATION

In this study, the method used for the classification process is Convolutional Neural Network.

1. Determination of filter, pool size, stride and padding The filter used is 3 x 3, while the pool size used is 2 x 2.

The Stride and Padding settings are set to default, namely 1 and 0.

2. Determination of the activation function The calculation results between input, weight and bias will be calculated again with the equation of the activation function to get the output of each layer. In this study, the authors used the ReLu on activation function convolutional layer and softmax activation function on the output layer for get a result which is categorical data. Systematically, function ReLu activation can be seen in equation below.

$$\sigma(a) = max(0, a)$$

Where a is the calculation result in the layer. The softmax function can be seen from equation below

$$\sigma(a_j) = \frac{\exp(a_j)}{\sum_{k=1}^m \exp(a_k)}$$

3. Determination of the optimizer: Optimizer is an algorithm for determining the optimal weight In this research, the optimizer used is Adam.



4. Determination of the batch size: The batch size is used to determine the number of 16observations made before making changes in weight, which is determined relative based on computer specifications. In this study, the authors used several Batch size is the batch size with a value of 128 and the batch size with a value of 1024.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 244, 3)]	0
conv2d (Conv2D)	(None, 112, 122, 32)	2432
batch_normalization (BatchNo	(None, 112, 122, 32)	128
activation (Activation)	(None, 112, 122, 32)	0
conv2d_1 (Conv2D)	(None, 56, 61, 64)	51264
batch_normalization_1 (Batch	(None, 56, 61, 64)	256
activation_1 (Activation)	(None, 56, 61, 64)	0
conv2d_2 (Conv2D)	(None, 28, 31, 128)	204928
batch_normalization_2 (Batch	(None, 28, 31, 128)	512
activation_2 (Activation)	(None, 28, 31, 128)	0
conv2d_3 (Conv2D)	(None, 14, 16, 192)	614592
batch_normalization_3 (Batch	(None, 14, 16, 192)	768
activation_3 (Activation)	(None, 14, 16, 192)	0
conv2d_4 (Conv2D)	(None, 7, 8, 256)	1229056
	(None, 7, 8, 256)	1024
activation_4 (Activation)	(None, 7, 8, 256)	0
flatten (Flatten)	(None, 14336)	0

dense (Dense)	(None,	1028)	14738436
dense_1 (Dense)	(None,	516)	530964
dropout (Dropout)	(None,	516)	0
dense_2 (Dense)	(None,	256)	132352
dropout_1 (Dropout)	(None,	256)	0
dense_3 (Dense)	(None,	1)	257
Total params: 17,506,969 Trainable params: 17,505,625 Non-trainable params: 1,344			

Figure 4.8 Determining the Batch Size

In the training process, the input image measuring $224 \times 244 \times 3$ will first be processed by Convolutional Layer First with a 32 kernel sized filter $112 \times 122 \times 32$ with the results of the size 56 x 61 x 64. In the first pooling layer, size Dimensions were reduced from 224 to 56. At each pooling layer, parameters do not change. Then, the result from the First Convolutional Layer measuring $112 \times 122 \times 32$ will be at process by Convolutional Layer Second with 128 kernels filter 28 x 31 x 128 in size with the results of the size 28 x 31 x 128. In the second pooling layer, the dimensions are reduced from 56 to 28.

The results from the Second Convolutional Layer with size $28 \times 31 \times 128$ will be processed by Convolutional Layer Third with a filter 256 kernels of 512 batch normalization size with the results of the size $14 \times x$ 192. In the third pooling layer, measure dimensions reduced from 28 to 14. The results from the Second Convolutional Layer with batch size 128 will be processed by Convolutional Layer Third with a filter 192 kernels of 14 x 16 x 192 size with the results of the size 7 x 8 x 256. In the third pooling layer, measure dimensions reduced from 14 to 7. After the Convolutional Layer process, 2 fully connected layers are generated which has 256 neurons in the first layer and 3 neurons in the second layer. The Training Process for Each Layer uses Deep Convolutional Neural Network can be seen in figure 4.8.

5. Determination of epoch: Epoch is literally the amount used to repeat the process learning. The larger the number of epochs the higher the yield rate learning. There are several epoch values used in this research. The number of epochs used in this study is 80 and saving the as comma separated value ".esv" extension.

```
1 #### training: Done using colab
  3 from tensorflow.keras.callbacks import ReduceLROnPlateau
  4 learning control = ReduceLROnPlateau(monitor='val accuracy',
                patience=3, verbose=1, factor=.5, min_lr=0.00001)
  5
   6 #### Save the model with best weights (best result w.r.t. loss metrics)
   7 from tensorflow.keras.callbacks import ModelCheckpoint
   8 chkpoint_path= '/content/drive/MyDrive/SkinCancer/epochs:{epoch:03d}-val_acc:{val_accuracy:.3f}.hdf5'
  9 checkpointer = ModelCheckpoint(chkpoint_path, verbose=1,save_best_only=True)
  10 history = net.fit(train_dataset.map(preprocess).map(augment).shuffle(100).batch(64),
  11
                 epochs=80.
  12
                 validation_data=val_dataset.map(preprocess).map(augment).batch(64),
  13
                callbacks=[learning_control, checkpointer])
  14
  15
  16 #### Save history.history dictionary as a pandas DataFrame
  17 csv path = '/content/drive/MyDrive/SkinCancer/'
  18 pd.DataFrame.from_dict(history.history).to_csv(csv_path+'base_model.csv', index=False)
Epoch 00075: val_loss did not improve from 0.21317
Epoch 76/80
Epoch 00076: val_loss did not improve from 0.21317
Epoch 77/80
Epoch 00077: val_loss did not improve from 0.21317
Epoch 78/80
59/59 [======
          Epoch 79/80
Epoch 00079: val_loss did not improve from 0.21246
Epoch 80/80
59/59 [-----
                 =====] - ETA: 0s - loss: 0.1196 - accuracy: 0.9752
Epoch 00080: val_loss did not improve from 0.21246
                   ===] - 390s 7s/step - loss: 0.1196 - accuracy: 0.9752 - val_loss: 0.2216 - val_accuracy: 0.9428
59/59 [====
```

Figure 4.9 Determining Training Model using Epoch



In the above testing model the configuration of the training process that displays a graph of the results with maximum accuracy of 94.85%, 22.96% loss and confusion matrix of the training process in Figure 4.10 and Figure 4.11 respectively.





Figure 4.11: Model Loss using CNN Classification for Skin Cancer Detection

V. CONCLUSION

The conclusions that can be drawn are based on the results of the skin cancer recognition system with training and testing expressions using the Convolutional Neural Network (CNN) method is able to classify expressions faces quite well and has an accuracy of up to 84.85% of the 5202 skin cancer (malignant and benign) images. In the training process, data augmentation is needed to reproduce data in order to avoid overfitting that affects data accuracy practice and test data. Selection of batch size, number of convolutional layers, optimizers and epoch also affects the accuracy obtained. After going through several tests. The batch size used must match the specifications of the device used, the amount of epoch used must be adjusted to other parameters amount. More convolutional layers can make the process more accurate learning is getting better. A good activation function is used in research is ReLu.

References

- [1] https://www.healthline.com/health/skin-disorders
- [2] Patterson MJ. Streptococcus. In: Baron S, editor. Medical Microbiology. 4th ed. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 13. PMID: 21413248.
- [3] Pal, Anabik & Ray, Sounak & Garain, Utpal. (2018). Skin disease identification from dermoscopy images using deep convolutional neural network.
- [4] L. -F. Li, X. Wang, W. -J. Hu, N. N. Xiong, Y. -X. Du and B. -S. Li, "Deep Learning in Skin Disease Image Recognition: A Review," in IEEE Access, vol. 8, pp. 208264-208280, 2020, doi: 10.1109/ACCESS.2020.3037258.
- [5] Wei, L., Gan, Q., & Ji, T. (2018). Skin Disease Recognition Method Based on Image Color and Texture Features. Computational and Mathematical Methods in Medicine, 2018.
- [6] Das, Nibaran & Pal, Anabik & Mazumder, Sanjoy & Sarkar, Somenath & Gangopadhyay, Dwijendranath & Nasipuri, Mita. (2013). An SVM Based Skin Disease Identification Using Local Binary Patterns. Proceedings - 2013 3rd International Conference on Advances in Computing and Communications, ICACC 2013. 208-211. 10.1109/ICACC.2013.48.
- [7] Luciano D. S. Pacífico, Larissa F. S. Britto, Emilia Oliveira, Teresa Bernarda Ludermir: Automatic Classification of Medicinal Plant Species Based on Color and Texture Features. BRACIS 2019: 741-746
- [8] Shahsavari, Marziye & Rashidi, Hassan & Bakhsh, Hajar. (2016). Efficient classification of Parkinson's disease using extreme learning machine and hybrid particle swarm optimization. 148-154. 10.1109/ICCIAutom.2016.7483152.
- [9] Kang, Jaeyong & Ullah, Zahid & Gwak, Jeonghwan. (2021). MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers. Sensors. 21. 2222. 10.3390/s21062222.
- [10] Scott Litin, M.D., Mayo Clinic Family Health Book, Mayo Clinic 200 First Street SW Rochester, MN 55905 - All Rights Reserved
- [11] Alquran, Hiam & Abuqasmieh, Isam & Alqudah, Ali & Alhammouri, Sajidah & Alawneh, Israa & Abughazaleh, Ammar & Hasayen, Firas. (2017). The Melanoma Skin Cancer Detection and Classification Using Support Vector Machine. 10.1109/AEECT.2017.8257738.
- [12] Ahmed, Hosameldin & Nandi, Asoke. (2019). Support Vector Machines (SVMs). 10.1002/9781119544678.ch13.
- [13] Darshini, Mogana & Mohd Taib, Nur Aishah & Har, Yip & Lio, Pietro & Dhillon, Sarinder Kaur.
 (2019). Predicting factors for survival of breast cancer patients using machine learning techniques. BMC Medical Informatics and Decision Making. 19. 48. 10.1186/s12911-019-0801-4.
- [14] https://www.msdmanuals.com/home/skin-disorders/biology-of-the-skin/structure-and-function-of-the-skin
- [15] Ilag, Leodevico. (2018). Ringworm Disease- Causes, Diagnosis and Treatment: AMYCOT®, a Novel Natural Treatment for Ringworm and other Tinea Infections.
- [16] Mukhopadhyay AK. A historical note on the evolution of "ringworm". Indian J Dermatol Venereol Leprol 2019;85:125-128
- [17] Kasparian, Nadine & Bartula, Iris. (2021). Melanoma. 10.1093/med/9780190097653.003.0022.
- [18] Chong, Wei-Lynn & Pitney, Thomas & Sinnott, Michael. (2021). Melanoma Screening. 10.1007/978-3-030-67123-5_14.
- [19] Mocellin, Simone. (2021). Hemangioma. 10.1007/978-3-030-58710-9_117.

IJCRT2308416 International Journal of Creative Research Thoughts (IJCRT) www.ijcrt.org d897

- [20] Mocellin, Simone. (2021). Congenital Hemangioma. 10.1007/978-3-030-58710-9_55.
- [21] Rigopoulos, Dimitris & Baran, Robert. (2021). Eczema. 10.1201/9781003159117-7-7.
- [22] Nilsson, Eskil. (2020). Eczema. 10.1201/9781003068617-13.
- [23] Putten, E. Vander, Kambod, A., & Kambod, M. (2018). Deep Residual Neural Networks for Automated Basal Cell Carcinoma Detection, (March ,2018)
- [24] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115– 118
- [25] Mustafa, Suleiman & Dauda, Ali & Dauda, Mohammed. (2017). Image processing and SVM classification for melanoma detection. 1-5. 10.1109/ICCNI.2017.8123777.
- [26] R. B. Aswin, J. A. Jaleel and S. Salim, Hybrid genetic algorithm Artificial neural network classifier for skin cancer detection, 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), Kanyakumari, 2014, pp. 1304-1309, doi: 10.1109/ICCICCT.2014.6993162.
- [27] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf
- [28] M. Toğaçar, Z. Cömert, B. Ergen and Ü. Budak, "Brain Hemorrhage Detection based on Heat Maps, Autoencoder and CNN Architecture," 2019 1st International Informatics and Software Engineering Conference (UBMYK), Ankara, Turkey, 2019, pp. 1-5, doi: 10.1109/UBMYK48245.2019.8965576.

