International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Feature Set Clustering and Classification for Melanoma Detection using Enhanced K Nearest Neighbour

[1]Deepthi Rapeti, [2]Dr. D. Vivekananda Reddy

Submitted: 28/12/2023 **Revised**: 04/02/2024 **Accepted**: 12/02/2024

Abstract: A correct diagnosis is one of the most important parts of medical care. In terms of diagnostic instability and difficulty, dermatology ranks high among medical specialties. In order to confirm a correct diagnosis, dermatologists frequently ask for further information, such as the patient's medical history and results of additional tests. Consequently, it is critical to discover a way that can ensure a correct and trustworthy diagnosis in a timely manner. Over the years, a number of methods have been created to aid in the machine learning-based diagnosis. But there are features, like great accuracy, that the traditional systems don't have. The skin is the primary organ of the human body, and skin cancer is the most common form that affects a large number of people annually. The goal of current research is to lower the death rate from skin cancer. It is simple to cure malignant melanoma when detected in its early stages. The early stages of extinction suggest a high chance of survival, thus prompt diagnosis is essential. This research describes a system for clinical decision-making that uses a image set of the area of the skin that needs to be diagnosed as input for the diagnosis of melanoma. The system determines the features that reflect the extent of damage by analyzing an image sequence to identify the afflicted area. Based on these features, the system produces a determination. The process of creating a model of classification for the accurate identification of melanoma that is malignant, a severe form of skin cancer, is discussed in this work. This research proposes a Feature Set Clustering and Classification for Melanoma Detection using Enhanced K Nearest Neighbour (FSCC-MD-EKNN) for accurate classification of melanoma. The proposed model when compared with the existing models achieves 98.4% accuracy in classification.

Keywords: *Feature Set, Clustering, Classification, Melanoma Detection, K Nearest Neighbour, Early Detection, Prediction Accuracy.*

1. Introduction

Based on statistics from the World Health Organization (WHO), cancer is one of the most concerning diseases, taking the lives of around 12 million people annually [1]. Skin cancer, especially deadly melanoma, is a severe variety with a recent marked increase in incidence rate, despite the fact that it can take many different forms. If detected early, melanomas have a 93.5% survival rate under treatment [2]. However, this rate may decrease to as low as 8% if cancer is discovered in its later stages. The aging of the world population and increased Ultraviolet (UV) radiation exposure are the key factors behind the growth in melanoma cases. Studies have shown that light-skinned individuals have a higher chance of developing skin cancer because their higher skin layers' pigmentation offers less protection [3].

Over the past few decades, there has been an increase in the number of cases of melanoma skin cancer. The rate of melanoma has increased dramatically since the early 1970s, it has climbed by an average of 4% every year [4]. There are about 139,000 cases of melanoma skin cancer annually worldwide. Of all the skin cancers, it is most deadly. While skin type and other hereditary variables may also play a role, prolonged exposure to UV radiation is the primary cause of melanoma [4]. When melanoma is discovered in its early stages, the best course of action is to remove it immediately. In other situations, life expectancy is shortened to less than a year if the condition is not identified in a timely manner [5]. It is crucial to differentiate between benign lesions versus melanoma as soon as feasible. Dermatologists utilize various experience-driven diagnostic techniques, such as the Menzies method, the rule of seven factors, and the acronym Asymmetry, Border, Color, Diameter and Evolving (ABCDE) to determine the total dermatoscopy score [6]. Using dermoscopy images, these techniques enable the identification of symptoms indicative of a cancerous lesion through the examination of a set of features. Nevertheless, in certain situations, it could be challenging to interpret these features visually and, as a result, arrive at the correct diagnosis [7].

Melanoma, which makes up only 7% of all cancers of the skin, is a particularly deadly form of the disease, responsible for 75% of skin cancer deaths. If melanoma is discovered and treated early on, it can be cured; but, if it is discovered later, it can spread deeply into the skin and other body parts. It is harmful if it spreads to other bodily parts, as it is hard to treat [8]. Melanoma is a result of continuous contact with UV light on the skin

[1]Research scholar Department of Computer Science and Engineering, SVUCE, Sri Venkateswara University, Andhra Pradesh, India deepthirapeti1@gmail.com
[2]Assistant Professor Dept of Computer Science and Engineering, SVUCE, Sri Venkateswara University, Tirupati, Andhra Pradesh, India. svuvivek@gmail.com

and the presence of melanocytes in certain body parts. Melanoma is a type of cancer that almost always starts in pigment cells and manifests as a dark lesion. In certain cases, the lesion can appear pink, white, or tan. In addition to color, melanomas differ from benign lesions in a number of other ways, including texture and shape [9]. Melanoma diagnosis is a critical procedure that has error margins and inaccuracies. Therefore, a better detection strategy is required for the diagnosis, necessitating the application of image processing principles [10].

Pre-processing, segmentation, feature extraction, and classification are typically included. First, linear filters are used for pre-processing [11]. These filters ignore picture qualities and noise that are distributed throughout the image, softening the contrast and the edge region of the image. Nonlinear filters, such as an adaptive mean filtering and a median filter are effective at reducing speckle noise. However, these filters have the disadvantage of losing information because they blur important and edge areas of the image [12]. These filters cannot effectively eliminate noise as a result. Several filters, such as an Anisotropy Diffusion (AD) filter, are recommended to address this issue [13]. The AD filter looks in local directions for boundary preservation and efficient noise removal. On the contrary, the AD filter's ability to reduce noise varies depending on a number of variables that can be changed at different resolutions. The general process of melanoma classification is shown in Figure 1.

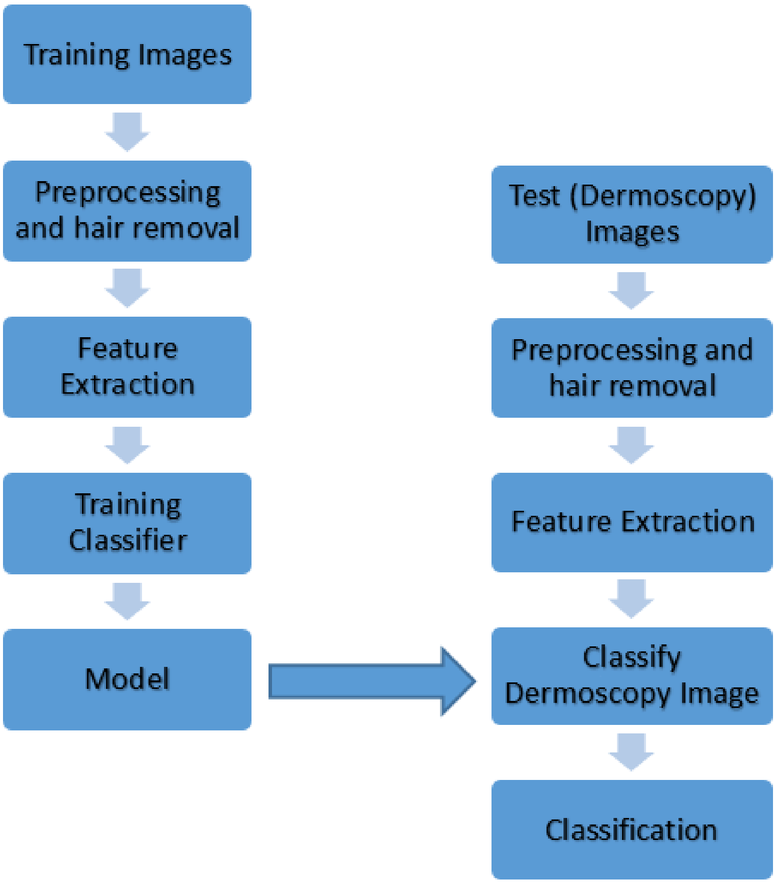

Fig 1: Melanoma Detection General Procedure

In order to find significant patterns in a variety of data representations, methods based on machine learning are frequently employed. Medical imaging applications have effectively employed machine learning (ML) algorithms through the use of certain feature selection techniques as well as transfer learning methods [14]. These methods can be crucial in the early stages of detection for assessing the possibility of melanomas that in pigmented lesions of the skin because visual diagnostics are non-invasive. To be more precise, medical practitioners can use diagnostic tools that use AI models to analyze medical images and monitor modifications to skin lesions as time passes [15]. This can help detect and treat melanomas early and effectively, which also means that medical costs will be more economically managed. It has been demonstrated that these AI systems can identify skin cancer with an accuracy comparable to that of human specialists [16]. More accurate image categorization research is to be conducted in the field of skin. Many studies are being conducted to locate and categorize tumours utilizing a range of automated techniques since the process of scanning and uploading images to a computer for diagnosis has gotten easier recently [17]. This method uses a enhanced K-nearest neighbor (KNN) to improve picture classification with a high degree of accuracy. Region- and contour-based techniques are effective in extracting features, and KNN approaches are effective in classifying data. The KNN process is shown in Figure 2.

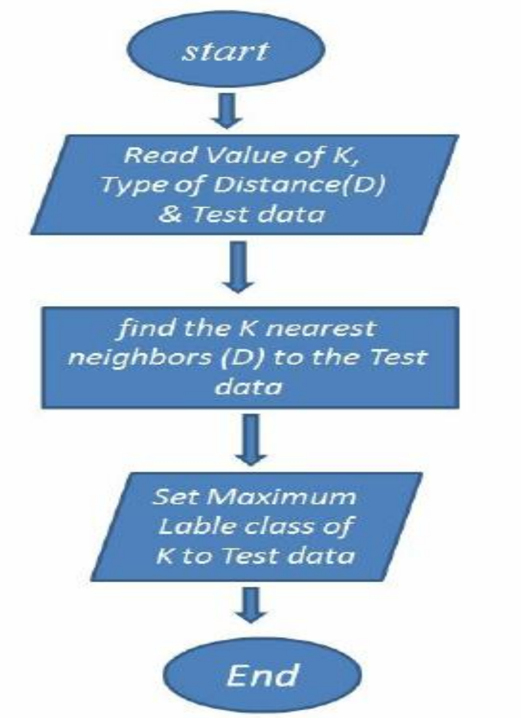

Fig 2: KNN Process

Melanoma is a type of tumor that mostly affects the skin, retinas, nerve centers, including meninges. It starts in cells that produce melanin. Melanoma is a type of tumor that mostly affects the skin, eyes, neurological centers, & meninges [18]. It starts in cells that produce melanin. A largely supervised learning technique is used for cancer prediction, using algorithms for categorization based on probabilities or conditional judgments. Decision trees, neural networks with convolution (CNN), support vector machines (SVM), & k-nearest neighbors (KNN) are among the most widely used techniques or approaches [19]. KNN is regarded as a nonlinear classifier and is used for regression and classification. The centroid is thought of as the center of a particular cluster, and there is no need for a training phase. The data under this approach is divided into training and test sets [20].

Based on the subsequent equation, each test set rows of the k neighbors that is closest based on the distance calculated by Euclid is noticed. Melanoma, squamous cell carcinoma skin cancer (SCC), as well as base cell skin cancer (BCC) are the three kinds of skin cancer. Rarely do the first two forms of skin cancer result in mortality; they are categorized as non-melanoma. Melanoma is the most deadly kind of skin cancer. The primary cause of significant skin cancers is UV exposure from the sun. Because pigmentation in the skin's layers that face the sun offers superior protection, fair-skinned races are more likely to develop skin cancer versus dark-skinned races. This research proposes a Feature Set Clustering and Classification for Melanoma Detection using Enhanced K Nearest Neighbour for accurate classification of melanoma.

2. Literature Survey

One of the most deadly forms of skin cancer is melanoma. However, because of their similar symptoms and outward look, it can occasionally be difficult to differentiate it from nevus. This disease has a greater death rate than any other consolidated malignancy associated to the skin. The number of young people who have instances is rising, although the survival rates increase dramatically if the disease is discovered early. It takes a great deal of money and effort for medical professionals to identify every patient with melanoma. In this study, Khan et al. [1] offered an intelligent system that uses cutting-edge image processing techniques to detect and differentiate nevus from melanoma. The obtained photos' skin lesion is first cleaned up of noise using a Gaussian filter, and then the lesion is segmented out using enhanced K-mean clustering. Textural and color features are extracted from the lesion to produce a unique hybrid super feature vector. Skin cancer is classified as either nevus or melanoma using support vector machines (SVMs).

An unsupervised learning technique called clustering data tries to separate a set of points of data into several groups. It is an important but challenging topic in data mining and machine learning. It has been successfully applied in many different fields. Nonetheless, in certain situations, traditional clustering methods require balance significance to be taken into account. Thus, using entropy-aware similarity which is the degree of balances, Son et al. [2] tackled the problem of imbalanced clustering and introduces a novel technique for balanced clustering. The author introduced a new approach to balanced clustering: entropy-aware similarities for balanced clustering (EASB). This method maximizes balance by complementarily clustering imbalanced data and utilizing entropy in a new similarity formula that takes distances and angular differences into account.

These days, computer vision is crucial to patient risk assessment, computer-aided diagnostics, and disease detection. This is particularly true regarding skin cancer, which, if not detected in its early stages, can be lethal. Numerous computer-aided diagnosis and detection methods have been developed in the past for this reason. The complex visual properties of skin lesion images, such as inhomogeneous features and fuzzy borders, hindered their performance. In this work, two strategies for identifying and categorizing malignant and benign tumors from dermoscopic pictures were presented by Magdy et al. [3]. The first approach uses trained deep artificial neural networks as feature extractors and uses KNN to operate as classifier. The second one uses the Grey Wolf Optimizer in conjunction with AlexNet to maximize performance by optimizing the network's hyperparameters. Additionally, the author investigated two methods for categorizing photos of skin cancer: deep learning (DL) and machine learning (ML). Artificial neural networks, KNNs, support vector machines, Naïve Bayes, as well as decision trees are the techniques utilized in machine learning. Convolutional neural networks and pretrained DL networks, such as AlexNet, the VGG-16, VGG-19, EfficientNet-b0, ResNet-18, ResNet-50, ResNet-101, DenseNet-201, Inception-v3, and MobileNet-v2, are included in the DL approach that we employed. The ISIC archive collection contains 4000 photos that we use for training and testing our studies.

Due to the poor contrast of lesions of the skin, the large intraclass variance of melanomas, the substantial amount of visual similarity among melanoma and melanoma lesions, as well as the presence of numerous artifacts in the image, automated melanoma identification in images from dermoscopy is an extremely difficult task. To address these issues, Yu et al. [4] suggested a unique approach that uses extremely deep CNNs for melanoma detection. The far deeper network may develop richer and more discriminating characteristics for better recognition when compared to current methods that use either low-level handmade characteristics or CNNs with shorter architectures. The author offered a series of strategies to assure efficient learning and instruction under sparse training material in order to fully use very deep networks. Initially, the author utilized residual learning to address degradation and overfitting issues that arise as a network gets deeper. By using this method, the author considered that the speed improvements brought about by deeper networks trickle down to our networks. Next, to achieve precise skin lesion segmentation, we build an entirely convolutional residual networks (FCRN) and augment its performance with a multi-scale contextual data integrating strategy.

The automatic identification of lesions in dermoscopy images presents several obstacles due to the intricate lesion features and detection backdrop. There is a dearth of research on major intra-class differences as well as inter-class similarities of lesion features, and the prior solutions primarily concentrate on employing larger and more complicated models to increase the detection accuracy. The greater model size also presents difficulties for future algorithm applications; in this research, Wei et al. [5] proposed a lightweight model with features discrimination based on the fine-grained classification principle for skin cancer recognition. The proposed model consists of a feature discriminating network and two identical feature extraction units of the lesion classification network. First, the recognition model's lightweight CNN feature extraction module receives two sets of training samples positive and negative sample pairs.

Melanoma is a prevalent type of cancer of the skin that poses a serious risk to a large number of individuals worldwide. Dermatologists may make mistakes when detecting melanoma with their unaided eyes. Consequently, the use of artificial intelligence-enabled image processing technologies can assist dermatologists in examination and decision-making. Ichim et al. [6] chosen empirically and took into account the texture, shape, color, size, as well as convolutional pixel connections of melanomas due to the different properties of this type of lesions as well as the noises and artifacts in the photos and the AlexNet. Based on the learning-adjusted weight as well as the choices made at the first level, a single classifier determines whether the lesions is a melanoma at the second level. The ultimate determination, whether the condition is melanoma or not, is made at the second level using a back-propagation perceptron. Subjective as well as objective levels are trained in two different periods.

The most deadly type of skin cancer is melanoma. It has proven to be difficult to distinguish melanoma tumors from non-melanoma lesions, nonetheless. For this purpose, numerous computer-aided diagnosis as well as detection methods have been constructed created in the past. Their efficacy has been hindered by the intricate visual aspects of the skin disease images, which include fuzzy boundaries and non-uniform features. In this study, Adegun et al. [7] presented a deep learning-based approach to automatically detect and separate melanoma lesions that gets beyond these drawbacks. For effective learning and feature extraction, an improved encoder-decoder network is suggested, with both of these sub-networks connected via a number of skip routes. This approach increases the meaning of the encoder map features nearer to that of the decoder's feature maps. The system classifies melanoma lesions pixel-by-pixel using a softmax classifier and a multi-stage, multi-scale methodology. The author developed a novel technique termed Lesion-classifier that classifies skin lesions between melanoma as well as non-melanoma categories according to the outcomes of pixel-wise classification.

Pereira et al. [8] contributed to the advancement of existing DL based solutions for the melanoma detection challenge. The goal of this topic of active research is to provide non-invasive means of detecting and classifying melanoma, the most deadly form of skin cancer. Beyond the typically used color properties of dermoscopic pictures, the suggested approach takes advantage of both 2D and 3D aspects of the skin disease surface. Utilizing an uncertainty-aware decision function, two rival classification techniques, Multiple Instance Learning (MIL) and DL are integrated. While MIL extracts 3D features, chooses the most important set, and performs categorization at two distinct learning instances, the DL approach performs classification using RGB data. The application of dense light-fields for identifying skin lesions and DL uncertainty evaluation algorithms in conjunction using MIL to train a robust group classifier are the innovative components of this work. The ensemble model attains a cross-validated cancer classification precision of 84.00% while training against nevus lesions and 90.82% when distinguishing against each of the lesion types, despite the significant class imbalance that is frequently found in medical picture datasets.

For the purpose of melanoma screening, short-term tracking of lesion alterations has become a well recognized clinical recommendation. A melanocytic lesion will be removed to rule out melanoma if there is a notable change in the lesion after three months. Nonetheless, the choice to make a modification or not is mostly based on the subjective experience and prejudice of certain therapists. Zhang et al. [9] developed a revolutionary deep learning based approach for the first time to automatically detect changes in short-term lesions in melanoma screening. The job of lesion change detection involves comparing two dermoscopy images of the same lesion in a brief amount of time. To make the choice whether the lesion has changed or not, a unique deep network based on Siamese structure is proposed. In addition to the deep convolutional features, a new structure called Tensorial Regression Process is developed under the Siamese framework to extract broad characteristics of lesion images. To emulate the way clinicians make decisions when comparing two lesion images, they tend to pay greater attention to regions that exhibit particular patterns. To this end, a segmentation loss, or SegLoss, is further developed and added as a term for regularization to the proposed network.

Dermatologists frequently use the follow-up dermoscopic pictures of skin lesions to identify or rule out potential melanoma. Nevertheless, single time-point photographs of lesions are used in the development of current algorithms for the early identification of melanoma. In borderline circumstances, disregarding the temporal, morphological alterations of lesions can result in a misdiagnosis. In this work, Yu et al. [10] used consecutive dermoscopic pictures to provide an automated approach for early melanoma diagnosis. The author build this approach in three steps to do this. Using estimated the Euclidean transformations, we first align sequential dermoscopic photographs of skin lesions. Next, the author computed image variations among the consecutive images to identify the lesion growth region. Finally, the author suggested a spatio-temporal network for capturing the dermoscopic shifts from the aligned lesion pictures and the associated difference images. In conclusion, the author created a module for early diagnosis that calculates the likelihood of malignancy in lesion pictures over time.

3. Proposed Model

This work is primarily concerned with identifying melanoma with accurate classification. This raises the option of creating other clustering categories by combining the remaining data in diverse ways, with the goal of improving the effectiveness of the melanoma classification [21]. The images were run through the basic model to extract the data of the penultimate layer, which contains 2304 features, and then analyzed in a comparison manner to ascertain which will perform best with the base model in order to acquire valuable clustering information [22]. All characteristics were rescaled utilizing unity-based normalisation prior to clustering [23]. The main goal of this research is to use perform accurate clustering to increase the melanoma classification rate. To get important information about the possible categories, it was thought important to also look at a non-algorithmic clustering technique [24]. According to this clustered model was created using the relationship among each class as well as the importance of getting proper diagnosis. The acquired classes were divided into three categories: non-suspicious lesions B and C, which require medium priority, and suspicious lesions A, that require prompt medical referral [25].

An abnormal growth or change in the appearance of the skin in comparison to the surrounding skin is called a skin lesion. There are two main categories of skin lesions: primary and secondary. A primary skin lesion is an abnormality of the skin that can either be present at birth or develop over time. It is possible for primary skin lesions to become worse or changed, leading to the development of secondary skin lesions [26]. A secondary skin lesion originates from the crust that forms when a mole is scraped until it bleeds. Dermatologists can recommend home care, medication, or surgery as treatments for affected skin, depending on the type of lesion. Some skin lesions, no matter how harmless they seem, pose a significant risk to patients since they can signal the existence of cancer and necessitate surgical removal. The most fatal form of skin cancer, melanoma, is treatable when caught early but becomes uncontrollable after it has spread. Protecting patients' growths and ensuring they receive prompt treatment2 depend on a correct diagnosis of skin patches.

Clinicians often screen for skin cancer through visual inspection, which has the drawbacks of being subjective, time-consuming, and error-prone. Dermoscopy is a non-invasive imaging method that can photograph skin lesions more clearly by removing surface reflection and capturing images under illumination and magnification. Nevertheless, in everyday clinical situations, the dermatologist's accuracy in melanoma identification with dermoscopy pictures was below 80%. Clinicians need an automated diagnosis system to help them make better decisions, and skin cancer detection has to be more efficient and effective. Melanoma and non-melanoma are typically classified using classic ML techniques for creating automated diagnostic tools. substantial diagnostic performance is challenging to attain, however, due to the fact that ML algorithms necessitate manually constructed features and dermoscopy images exhibit low inter-class variability and substantial intra-class variation. The proposed model framework is shown in Figure 3.

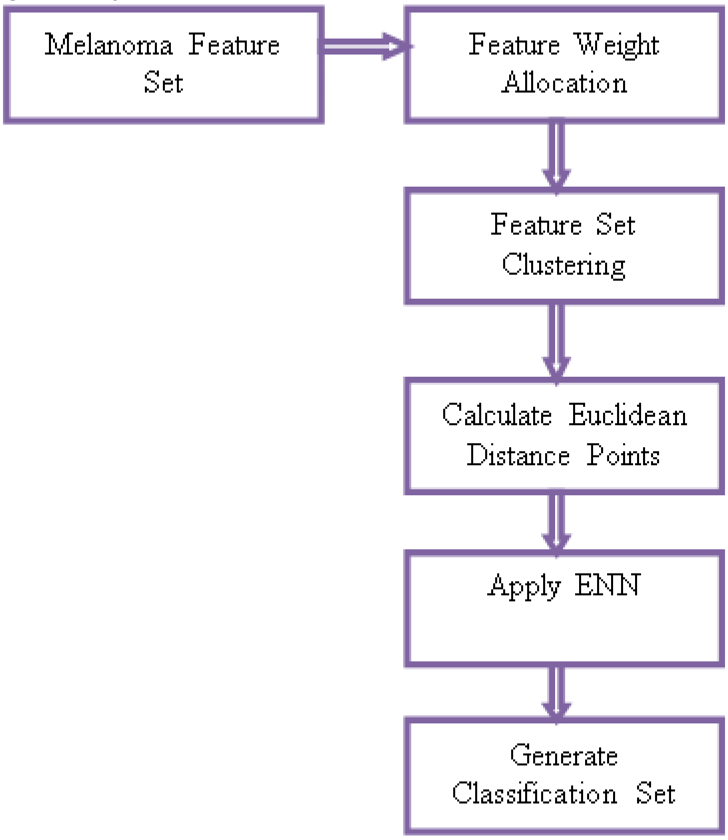

Fig 3: Proposed Model Framework

Among the several image space classification methods, nearest neighbor classification is among the most straightforward. In this way, the test image is tagged with the label of the learning set's nearest point based on the image's spaces. As a default, EKNN typically assigns a distance to each pixel in an image based on the distance between several data points inside the image, using the Euclidean distance measurement and then allocates the weights to the data points. The distance is the standard unit of measurement for the geometric separation of any two pixels.

In order to divide the features into two categories, classifiers or machine learning techniques are fed with the data straight after extraction. There are two parts to the whole process. During the training phase, the classifiers were given patterns of benign and cancerous images based on characteristics and class labels. To train the model, the previously derived feature characteristics of normal skin (type=0) and images with melanoma (type = 1) are provided. The feature space was then used to depict these data points. Utilizing training data, an unknown test pattern was input, and subsequently, it was labelled and shown in the feature space once more. Each sample image is represented as a point in an n-dimensional space with its attributes in a feature space, an abstract space. The size of the patterns is determined by the amount of features utilized to describe them.

The application of machine learning techniques to automate the analysis could pave the way for a medical framework and system that helps with disease detection, lowers healthcare costs, improves clinical reliability, helps doctors communicate objectively, reduces errors caused by human fatigue, and more. One way to accomplish these aims is to develop an AI system that can distinguish between benign and malignant pigmented skin lesions. Dermoscopy images of pigmented skin lesions are analyzed using Machine Learning technique in order to identify potentially cancerous skin lesions at an early stage. The EKNN algorithm is used to perform the classification. By adjusting the viewing and light orientations on the globe of all conceivable directions, the skin texture data are acquired. This research proposes a Feature Set Clustering and Classification for Melanoma Detection using Enhanced K Nearest Neighbour for accurate classification of melanoma.

Algorithm FSCC-MD-EKNN

{

Input: Feature Selected Set {FSset}

Output: Classification Set {Cset}

Step-1: The feature set is considered as input and the features are analyzed. The features analysis is performed that is used to consider the minimum and maximum feature ranges for processing. The feature analysis helps in detection of features that are used for training the model. The feature analysis is performed as

$$minRange[N] = \sum_{f=1}^{N} getattr\big(FSset(f)\big) + \min\left(getattr\big(FSset(f, f+1)\big)\right)$$

$$maxRange[N] = \sum_{f=1}^{N} getattr\big(FSset(f)\big) + \max\left(getattr\big(FSset(f, f+1)\big)\right)$$

Step-2: The features are analyzed and the ranges of each feature are considered. The weight allocation is performed to each and every feature by considering the feature dependency level. The weight allocation helps in choosing the most useful features to train the classifier. The weight allocation is performed as

$$corr[N] = \sum_{f=1}^{N} \frac{\sum_{f=1}\left(f(i) - \overline{f(i)}\right) * \left(f(i+1) - \overline{f(i+1)}\right)}{\sqrt{\sum\left(f(i) - \overline{f(i)}\right)^2 * \sum\left(f(i+1) - \overline{f(i+1)}\right)^2}}$$

$$Walloc[N] = \prod_{f=1}^{N} \frac{maxRange(f)}{minRange(f)} + getattr\big(FSset(\max(f, f+1))\big) + corr(FSset(f, f+1))$$

$$W \leftarrow \sum_{f=1}^{N} setRand(f)$$

$$Walloc[N] \leftarrow \begin{cases} W+1 & if\ corr(f, f+1) < cTh \\ 0 & Otherwise \end{cases}$$

Here f(i) is the current feature, $\overline{f(i)}$ is the mean of the attributes of the feature, cTh is the correlation threshold. W is the weight of the feature.

Step-3: The feature weight allocation is used to remove the low weighted features and the remaining high weighted features are considered. Feature clustering is performed that is used to group the features that are having similar attribute range. The feature cluster set is generated as

$$Fcluster[N] = \sum_{f=1}^{N} \frac{\max\left(Walloc\big(FSset(f, f+1)\big)\right)}{len(Walloc)} + simm\big(FSset(f, f+1)\big) + \max\left(simm\big(Walloc(f, f+1)\big)\right) \begin{cases} Fcluster(f) \leftarrow FSset(f) if\ simm\big(\max\big(Walloc(f, f+1)\big)\big) > Th \\ continue & Otherwise \end{cases}$$

Step-4: Using the Euclidean distance, EKNN can find out how similar or different a new instance is from an old training instance. Equal weights are assigned to all instance features, regardless of their relevance and then based on correlation, weights are updated. When calculating the distance between two instances, the EKNN model gives different weights to the characteristics. The process of EKNN is applied as

The nearest neighbour detection is performed by calculating the distance points and considering the feature points that are nearer attributes are considered and weight updation is performed as

$$Nneigh[N] = \sum_{f=1}^{N} \frac{\min\,(dist(f, f+1))}{size(Fcluster)} + \min\,(getattr(Walloc(FSset(f)))$$

$$Wupdateset[N] = \sum_{f=1}^{N} \frac{\max\,(Nneigh(f, f+1))}{\min\,(Nneigh(f, f+1))} + \max\,(Fcluster(f, f+1))$$

Step-5: Classification task is labelling dataset instances with class labels derived from their attributes. In classification, the objective is to construct a model that, given a set of attributes, can reliably forecast which class future instances will fall into. The proposed model generates the classes for the features that is performed as

$$Cset[N] = \sum_{f=1}^{N} \frac{\sum_{f=1} \max\big(Nneigh(f, f+1)\big) + \min\big(dist(f)\big) + \max\big(Wupdateset(f, f+1)\big)}{\max\big(Fcluster(f, f+1)\big)} \begin{cases} Cset(f) \leftarrow 1 \; if \max\Big(Wupdate\big(FSset(f)\big)\Big) > wTh \\ Cset(f) \leftarrow 0 \qquad Otherwise \end{cases}$$

}

4. Results

The dangerous diseases have also been on the rise alongside global warming. One of these is malignant melanoma. Cancer of the skin affects a large percentage of the population every year, despite the fact that the skin is the body's principal organ. Scientists have been trying to find ways to lower the skin cancer mortality rate today. Malignant melanoma is highly curable when caught early using stage prediction. The leading cause of mortality in humans is cutaneous malignancy. The condition, which is characterized by the uncontrolled proliferation of cells, can manifest anywhere in the body but most commonly on sun-exposed areas. When caught early, most tumors of this sort are treatable. Many lives could be saved if skin cancer could be predicted and detected quickly. Derma cancer can now be detected in its early stages because to new, cutting-edge technology. The biopsy technique is the gold standard for diagnosing skin cancer. The procedure involves removing dead skin cells and sending a sample to a lab for analysis. The procedure is more time-consuming and painful.

This research developed a skin cancer detection system that makes use of KNN for the purpose of early disease detection. People will find it more helpful. In order to analyze the nonparameterized EKNN classification method, the Euclidean distance is utilized and weight allocation is performed to data points. The technique of the EKNN approach is the form of nearest neighbour pixel values. To finish the skin cancer categorization, image recognition on both the training and test samples are performed. By picking the k nearest points and reporting the majority sign having highest weighted data points, the EKNN classifier enhances the strategy to diagnose melanoma. The algorithm selects k values that are unique to each classification. The k-output value is assessed using cross-validation from various subsets. It does a better job of controlling the amount of noise introduced by different pixel rates in the training data set. A set of criteria was used to identify and categorize skin lesions. This research proposes a Feature Set Clustering and Classification for Melanoma Detection using Enhanced K Nearest Neighbour (FSCC-MD-EKNN) for accurate classification of melanoma. The proposed model is compared with the traditional Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks (AMRDI-DRN) and Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network (ASCD-DIELDLN). The results represent that the proposed model performance in cancer classification is high than the traditional models.

The proposed model considers the extracted features and then all features are allocated with the weights. The weights are allocated based on the feature dependency. The features which has highest weight with in a specific threshold are considered for training the model. The Feature Weight Allocation Accuracy Levels of the existing and proposed models are indicated in Table 1 and Figure 4.

Table 1: Feature Weight Allocation Accuracy Levels

Images Considered	Models Considered		
	FSCC-MD-EKNN Model	AMRDI-DRN Model	ASCD-DIELDLN Model
10000	97.5	94.5	91.7
20000	97.7	94.8	91.9
30000	97.9	94.9	92.0
40000	98.2	95.1	92.3

50000	98.4	95.3	92.5
60000	98.6	95.4	92.7

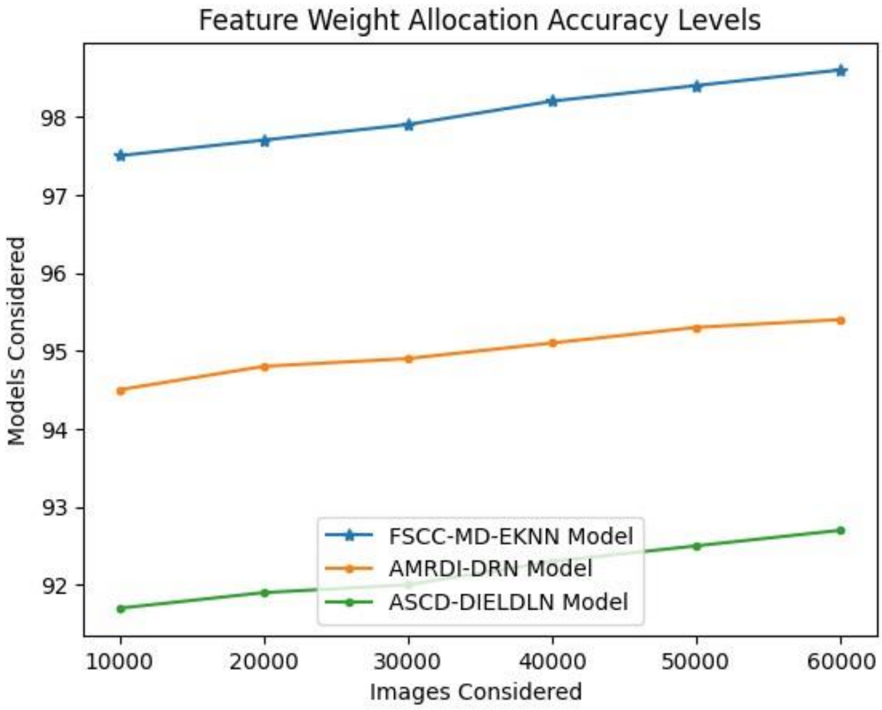

Fig 4: Feature Weight Allocation Accuracy Levels

Feature selection is a method for selecting useful, consistent, and non-redundant features for use in building models. As both the quantity and diversity of datasets increase, it is crucial to systematically decrease their sizes. The primary objective of feature selection is to decrease the computational cost of modeling while simultaneously improving the performance of a predictive model. The Table 2 and Figure 5 represent the Feature Selection Time Levels of the existing and proposed models.

Table 2: Feature Selection Time Levels

Images Considered	Models Considered		
	FSCC-MD-EKNN Model	AMRDI-DRN Model	ASCD-DIELDLN Model
10000	15.0	22.0	27.3
20000	15.3	22.3	27.5
30000	15.5	22.5	27.6
40000	15.7	22.7	27.8
50000	15.9	22.8	27.9
60000	16	23	28

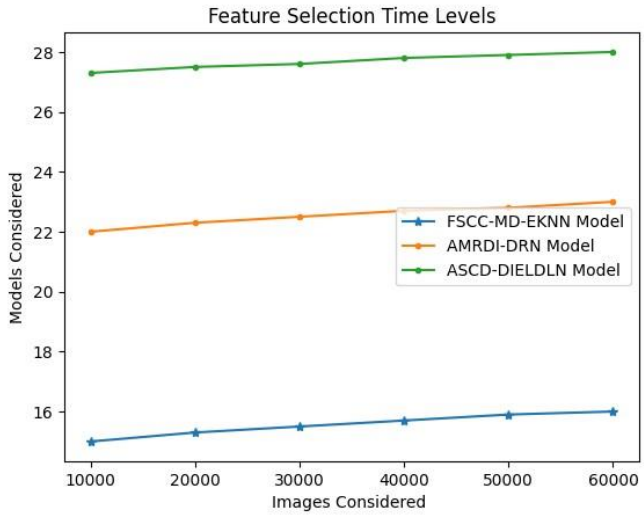

Fig 5: Feature Selection Time Levels

The goal of clustering is to group similar data points of features together and separate out unusual ones from the rest of the unlabeled data in a way that makes sense in prediction. To put it simply, clustering is a method for identifying and then grouping data points that share common characteristics. The Clustering Accuracy Levels of the proposed and existing models are shown in Table 3 and Figure 6.

Table 3: Clustering Accuracy Levels

Images Considered	Models Considered		
	FSCC-MD-EKNN Model	AMRDI-DRN Model	ASCD-DIELDLN Model
10000	97.3	93	92.0
20000	97.5	93.1	92.4
30000	97.6	93.4	92.6
40000	97.8	93.7	92.8
50000	98.0	93.9	92.9
60000	98.2	94	93

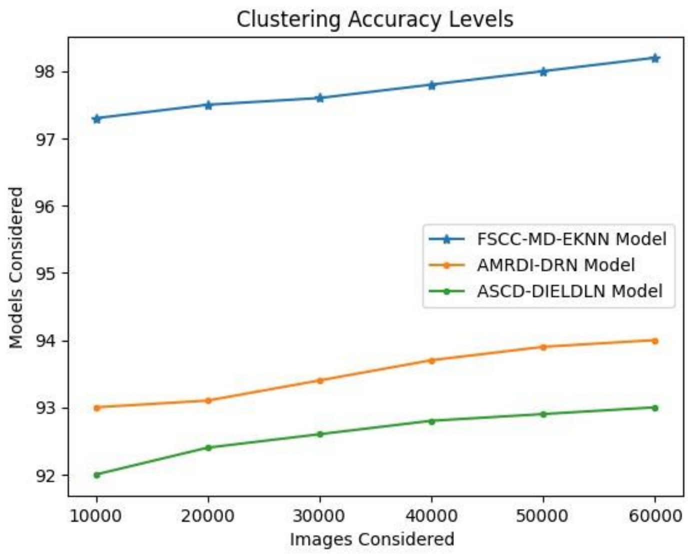

Fig 6: Clustering Accuracy Levels

When it comes to regression and classification, many learning experts turn to the KNN algorithm. It is based on the premise that values or labels assigned to similar data points are more likely to be consistent. The KNN algorithm keeps the whole training dataset in its memory during training. In order to make predictions, it uses a distance metric, like the geometric distance, to determine how far away the input data point is from all the training instances. The program then uses the distances between the input data point and its K closest neighbors to determine who they are. For classification purposes, the algorithm predicts the input data point's label based on the most common class label among its K neighbors. Regression uses a weighted average or average of the K neighbors' goal values to forecast the input data point's value. The KNN Nearest Set Calculation Accuracy Levels of the proposed and existing models are represented in Table 4 and Figure 7.

Table 4: KNN Nearest Set Calculation Accuracy Levels

Images Considered	Models Considered		
	FSCC-MD-EKNN Model	AMRDI-DRN Model	ASCD-DIELDLN Model
10000	97.7	94.9	91.8
20000	97.9	95.0	92.0
30000	98.1	95.2	92.4
40000	98.3	95.4	92.7
50000	98.5	95.7	92.9
60000	98.6	95.8	93

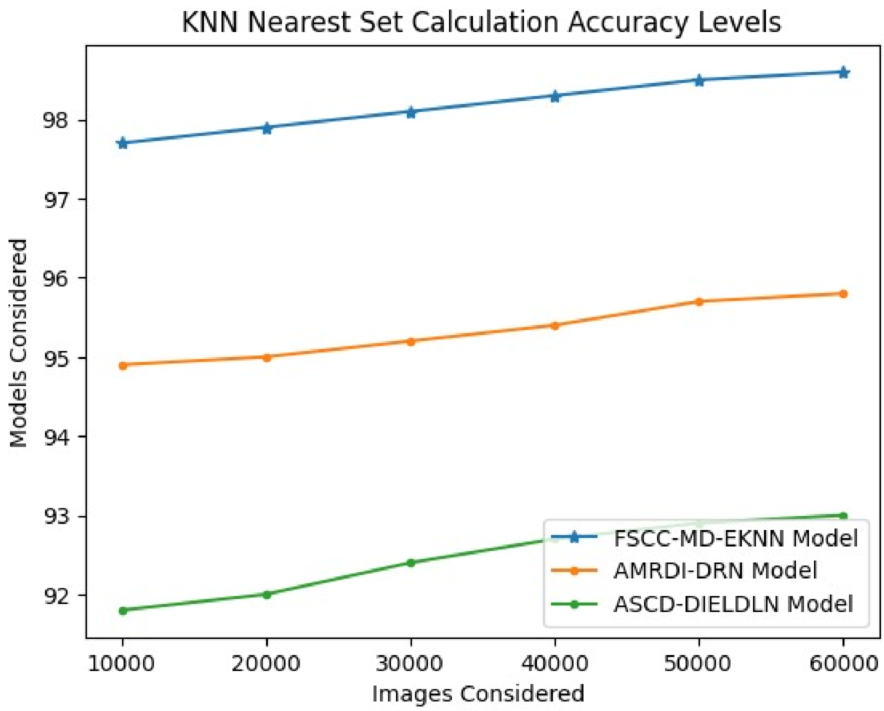

Fig 7: KNN Nearest Set Calculation Accuracy Levels

A comparable process to categorization is classification, which entails recognizing, differentiating, and understanding ideas and objects. When facts are grouped together into classes, it can be considered that they are classified. It might also mean a procedure that groups similar items together and separates those that are different. Recognizing, comprehending, and arranging objects and thoughts into preset groups, also called sub-populations, is what is known as classification. Classification in algorithms for machine learning use a variety of algorithms trained on these pre-categorized datasets for training to assign appropriate labels to new datasets. The Classification Time Levels of the proposed and existing models are shown in Table 5 and Figure 8.

Table 5: Classification Time Levels

Images Considered	Models Considered		
	FSCC-MD-EKNN Model	AMRDI-DRN Model	ASCD-DIELDLN Model
10000	17.1	25.2	28.1
20000	17.4	25.4	28.3
30000	17.5	25.6	28.5
40000	17.7	25.7	28.6
50000	17.9	25.9	28.7
60000	18	26	29

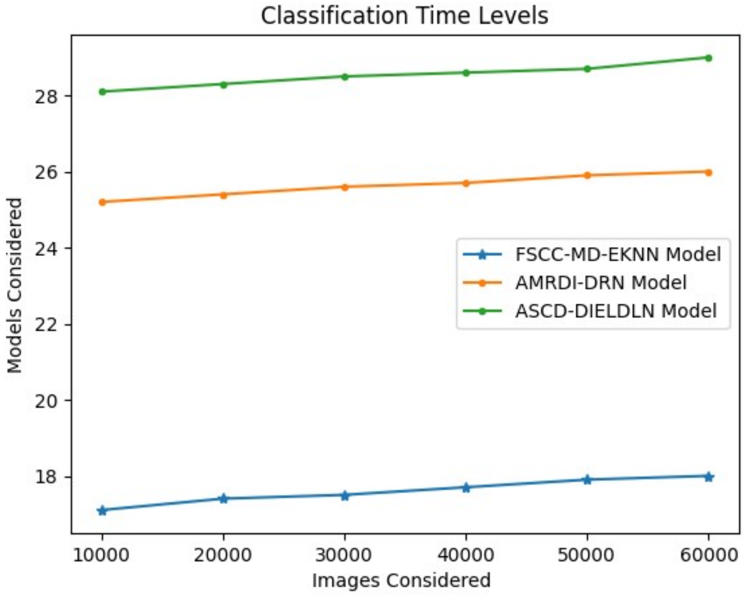

Fig 8: Classification Time Levels

The goal of the supervised machine learning technique known as classification is to have the model attempt to guess the right label from a set of input data. Before making predictions on previously unknown data, a classification model is first thoroughly trained using the training data. It is then tested on test data. Classification rules are a kind of classifier that predict the problem's class by testing value restrictions on the characteristics. The Table 6 and Figure 9 represents the Classification Accuracy Levels of the existing and proposed models.

Table 6: Classification Accuracy Levels

Images Considered	Models Considered		
	FSCC-MD-EKNN Model	AMRDI-DRN Model	ASCD-DIELDLN Model
10000	97.6	92.5	94.1
20000	97.9	92.6	94.3
30000	98.0	92.9	94.5
40000	98.2	93.0	94.7
50000	98.4	93.2	94.9
60000	98.6	93.4	95

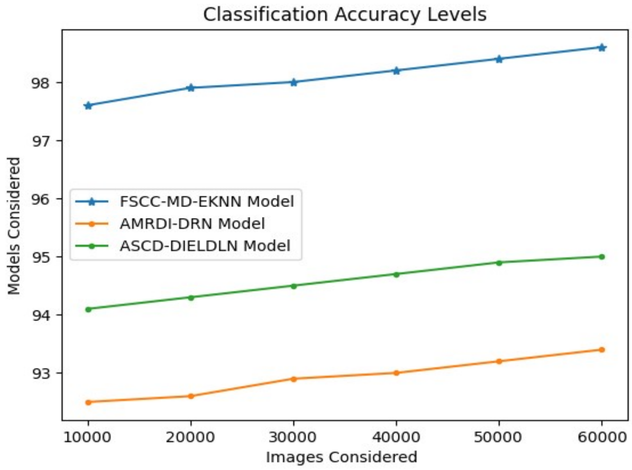

Fig 9: Classification Accuracy Levels

5. Conclusion

Quickly spreading to other parts of the body, melanoma skin cancer is one of the worst forms of the disease. Early detection increases the likelihood that melanoma can be cured. Methods from both the modern and traditional eras have been considered. This work provides a thorough and critical analysis of the most recent machine learning algorithms that have been utilized to distinguish between benign and malignant melanoma cells. While there are over a hundred different kinds of cancer, the important aspect is that treatment has a better chance of curing any kind of cancer if caught early enough. It is common for skin cancer symptoms to not manifest until the disease has spread. This is why skin cancer is so dangerous. Because the disease has progressed to this point, there is a high risk of death and serious complications from any treatment that aims to remove the cancerous tissue. Patients at high risk of developing skin cancer have routine exams to detect potentially malignant cells; the goal is to eliminate them before they may metastasize. However, these precancerous cells are invisible to the human eye and naturally disperse in an unpredictable manner. For this reason, it is both inefficient and prone to error to rely on the age-old method of randomly collecting biopsies in the hopes of detecting cancer cells. Skin pathology diagnosis is a time-consuming and intricate process that can easily be misdiagnosed. To aid doctors in accurately categorizing skin lesions, automated methods built on machine learning are crucial. Images are pre-processed, segmented, features are extracted, and classification is the last stage of the proposed system. This research proposes a Feature Set Clustering and Classification for Melanoma Detection using Enhanced K Nearest Neighbour for accurate classification of melanoma. The proposed method achieved a 98.6% accuracy rate throughout testing. To get even better accuracy and quicker outcomes, the system can be enhanced in the future with evolutionary algorithms or ensemble learning techniques. Optimization techniques can be applied with the machine learning models for improving the accuracy rate.

References

[1] M. Q. Khan et al., "Classification of Melanoma and Nevus in Digital Images for Diagnosis of Skin Cancer," in IEEE Access, vol. 7, pp. 90132-90144, 2019, doi: 10.1109/ACCESS.2019.2926837.

[2] S. B. Son, S. Park and J. Kim, "Entropy-Aware Similarity for Balanced Clustering: A Case Study With Melanoma Detection," in IEEE Access, vol. 11, pp. 46892-46902, 2023, doi: 10.1109/ACCESS.2023.3275561.

[3] Magdy, H. Hussein, R. F. Abdel-Kader and K. A. E. Salam, "Performance Enhancement of Skin Cancer Classification Using Computer Vision," in IEEE Access, vol. 11, pp. 72120-72133, 2023, doi: 10.1109/ACCESS.2023.3294974.

[4] L. Yu, H. Chen, Q. Dou, J. Qin and P. -A. Heng, "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks," in IEEE Transactions on Medical Imaging, vol. 36, no. 4, pp. 994-1004, April 2017, doi: 10.1109/TMI.2016.2642839.

[5] L. Wei, K. Ding and H. Hu, "Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network," in IEEE Access, vol. 8, pp. 99633-99647, 2020, doi: 10.1109/ACCESS.2020.2997710.

[6] L. Ichim and D. Popescu, "Melanoma Detection Using an Objective System Based on Multiple Connected Neural Networks," in IEEE Access, vol. 8, pp. 179189-179202, 2020, doi: 10.1109/ACCESS.2020.3028248.

[7] Adegun and S. Viriri, "Deep Learning-Based System for Automatic Melanoma Detection," in

IEEE Access, vol. 8, pp. 7160-7172, 2020, doi: 10.1109/ACCESS.2019.2962812.

[8] P. M. M. Pereira et al., "Multiple Instance Learning Using 3D Features for Melanoma Detection," in IEEE Access, vol. 10, pp. 76296-76309, 2022, doi: 10.1109/ACCESS.2022.3192444.

[9] B. Zhang et al., "Short-Term Lesion Change Detection for Melanoma Screening With Novel Siamese Neural Network," in IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 840-851, March 2021, doi: 10.1109/TMI.2020.3037761.

[10] Z. Yu et al., "Early Melanoma Diagnosis With Sequential Dermoscopic Images," in IEEE Transactions on Medical Imaging, vol. 41, no. 3, pp. 633-646, March 2022, doi: 10.1109/TMI.2021.3120091.

[11] L. D. Biasi, A. A. Citarella, M. Risi and G. Tortora, "A Cloud Approach for Melanoma Detection Based on Deep Learning Networks," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 3, pp. 962-972, March 2022, doi: 10.1109/JBHI.2021.3113609.

[12] R. Rastghalam, H. Danyali, M. S. Helfroush, M. E. Celebi and M. Mokhtari, "Skin Melanoma Detection in Microscopic Images Using HMM-Based Asymmetric Analysis and Expectation Maximization," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3486-3497, Sept. 2021, doi: 10.1109/JBHI.2021.3081185.

[13] A. A. Adegun and S. Viriri, "Deep Learning-Based System for Automatic Melanoma Detection," in IEEE Access, vol. 8, pp. 7160-7172, 2020, doi: 10.1109/ACCESS.2019.2962812.

[14] P. A. Dyachenko et al., "Detection of Melanoma Cells in Whole Blood Samples Using Spectral Imaging and Optical Clearing," in IEEE Journal of Selected Topics in Quantum Electronics, vol. 27, no. 4, pp. 1-11, July-Aug. 2021, Art no. 7200711, doi: 10.1109/JSTQE.2020.3047437.

[15] P. A. Dyachenko et al., "Detection of Melanoma Cells in Whole Blood Samples Using Spectral Imaging and Optical Clearing," in IEEE Journal of Selected Topics in Quantum Electronics, vol. 27, no. 4, pp. 1-11, July-Aug. 2021, Art no. 7200711, doi: 10.1109/JSTQE.2020.3047437.

[16] Q. U. Ain, H. Al-Sahaf, B. Xue and M. Zhang, "Generating Knowledge-Guided Discriminative Features Using Genetic Programming for Melanoma Detection," in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 5, no. 4, pp. 554-569, Aug. 2021, doi: 10.1109/TETCI.2020.2983426.

[17] S. Albahli, N. Nida, A. Irtaza, M. H. Yousaf and M. T. Mahmood, "Melanoma Lesion Detection and Segmentation Using YOLOv4-DarkNet and Active Contour," in IEEE Access, vol. 8, pp. 198403-198414, 2020, doi: 10.1109/ACCESS.2020.3035345.

[18] R. Ashraf et al., "Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection," in IEEE Access, vol. 8, pp. 147858-147871, 2020, doi: 10.1109/ACCESS.2020.3014701.

[19] Davis, L.E.; Shalin, S.C.; Tackett, A.J. Current state of melanoma diagnosis and treatment. Cancer Biol. Ther. 2019, 20, 1366–1379. [CrossRef] [PubMed]

[20] MacGill, M. What to Know about Melanoma. Medical News Today. Available online: https://www.medicalnewstoday.com/ articles/154322 (accessed on 15 September 2022).

[21] Heistein, J.B.; Archarya, U. Malignant Melanoma. StatPearls. Available online: https://www.statpearls.com/ArticleLibrary/ viewarticle/24678 (accessed on 15 September 2022).

[22] Blundo, A.; Cignoni, A.; Banfi, T.; Ciuti, G. Comparative Analysis of Diagnostic Techniques for Melanoma Detection: A Systematic Review of Diagnostic Test Accuracy Studies and Meta-Analysis. Front. Med. 2021, 8, 637069. [CrossRef] [PubMed]

[23] Silva, T.A.E.d.; Silva, L.F.d.; Muchaluat-Saade, D.C.; Conci, A.A. Computational Method to Assist the Diagnosis of Breast Disease Using Dynamic Thermography. Sensors 2020, 20, 3866. [CrossRef] [PubMed]

[24] Fanizzi, A.; Losurdo, L.; Basile, T.M.A.; Bellotti, R.; Bottigli, U.; Delogu, P.; Diacono, D.; Didonna, V.; Fausto, A.; Lombardi, A.; et al. Fully Automated Support System for Diagnosis of Breast Cancer in Contrast-Enhanced Spectral Mammography Images. J. Clin. Med. 2019, 8, 891. [CrossRef] [PubMed]

[25] Hasan, M.J.; Shon, D.; Im, K.; Choi, H.-K.; Yoo, D.-S.; Kim, J.-M. Sleep State Classification Using Power Spectral Density and Residual Neural Network with Multichannel EEG Signals. Appl. Sci. 2020, 10, 7639. [CrossRef]

[26] Hasan, M.J.; Kim, J.-M. A Hybrid Feature Pool-Based Emotional Stress State Detection Algorithm Using EEG Signals. Brain Sci. 2019, 9, 376. [CrossRef] .