Journal of Theoretical and Applied Information Technology
31st August 2025. Vol.103. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MULTI-LEVEL FEATURE EXTRACTION AND SELECTION WITH STRONG CORRELATION AND EFFECTIVE CLUSTERING MODEL FOR ACCURATE MELANOMA DETECTION

[1*]ZAHEER ABBAS MOHAMMAD, [2]S.NAGAKISHORE BHAVANAM, [3]VASUJADEVI MIDASALA

[1*]Department of Computer Science and Engineering, ANU College of Engineering & Technology, Guntur-522510, Andhra Pradesh, India.

[2,3] Department of Computer Science and Engineering, Mangalaytan University Jabalpur, Jabalpur-481662, Madhya Pradesh, India.

E-mail: [1] zaheeranu01@gmail.com [2] drbsnagakishore@gmail.com [3]vasujadevi@gmail.com

ABSTRACT

Melanoma, a fatal and aggressive form of skin cancer has continuously increased in incidence rates across the globe, emphasizing the necessity for early precise detection modalities to achieve better 5 year survival outcomes. This research presents a significant contribution to the field of Information Technology by proposing an advanced automated framework for accurate melanoma detection through a combination of multi-level feature extraction, strong correlation analysis, and intelligent clustering. With the ever-increasing volume of medical imaging data and the limitations of manual diagnosis, IT-based solutions such as machine learning and image processing are not just supportive tools they are transformative technologies driving modern healthcare innovation. This work underlines the indispensable role of IT in developing intelligent, data-driven healthcare systems capable of early diagnosis, ultimately contributing to increased survival rates and reduced healthcare burdens. This research proposes a Multi Level Feature Extraction and Rank Linked Feature Set with Clustering (MLFE-RLFSC) Model for accurate melanoma detection. Furthermore, this intelligent feature selection process helps in reducing the risk of over fitting and also improves significantly generalization performance of model to classify melanoma more accurately. In addition, our method integrates a clustering model which groups the type identical features so overall redundancy can be diminished as well that in turn enhance the general security level of detection systems. The proposed model achieved 98.8% accuracy in Multi Level Feature Extraction, 99.2% accuracy in Feature Rank Allocation, 98.6% accuracy in Clustering, 99.3% accuracy in Rank Linked Feature Set Generation and 99.2% accuracy in Classification of benign and malignant tumours. The proposed model when compared to the traditional methods performs better in clustering and melanoma detection. The proposed model with minimum false predictions accurately detects the melanoma for providing early diagnosis.

Keywords: *Melanoma Detection, Multi-Level Feature Extraction, Rank Linked Feature Set, Clustering Model, Image Processing, Melanoma Cancer Diagnosis.*

1. INTRODUCTION

Melanoma is one of the most lethal types of skin cancer and accounts for 75%–80% of deaths related to this disease [1]. Though less common, melanoma is the most serious type of skin cancer because it can metastasize early and aggressively. Prognosis of melanoma is quite good if detected early, given that relatively small primary cutaneous melanomas may be treated successfully with surgical excision [2]. However, the diverse morphological aspects of melanoma and its similarity with benign lesions

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

make early diagnosis very difficult. It has been noticed that the conventional ways used in diagnosis like visual inspection by dermatologists and histopathological examination post a biopsy are more subjective than objective which at times lead to an error when melanoma is considered [3]. Recent advancements have improved diagnostic accuracy with the introduction of dermoscopy which is a non-invasive imaging technique, enabling magnified visualization skin surface [4]. Although, manual interpretation of dermoscopic image exhibits a large sign of error that make automated systems an essential part needed for the effective diagnosis process [5].

The pipeline of composition for the automated melanoma detection system consists several stages like extraction, feature identification and selection as well as classification [6]. The effectiveness of the feature extraction and selection processes plays a critical role in the accuracy of these systems. Melanoma features obtained based on dermoscopic images like color, texture and shape are important for melanoma-sober lesion differentiation [7]. But the most difficult part is choosing right subset of features from a large number feature pool, because irrelevant or redundant features that does not improve performance and may actually deteriorate detection accuracy [8]. Deep learning techniques, especially Convolutional Neural Networks (CNNs), have become efficient tools for medical image analysis in the last few years. CNNs can automatically learn hierarchical feature representations from raw image data, which is why they are so suited for a task like melanoma detection [9]. Now, this is a huge issue specially the deep learning models as they require humungous amount labeled data for training and not all businesses will be able to provide that. Furthermore, the characteristics acquired by CNNs often have high dimensionality and feature selection is crucial for preventing overfitting and model generalization [10].

An automated system capable of differentiating melanoma from benign at its earliest stages is still needed to manage such limits. Doctors may benefit from dermoscopy technology advancements and the ability to get a second opinion with the help of a computer-aided diagnostic (CAD) system [11]. For instance, CAD systems use a variety of machine learning techniques; for instance, they use a state-of-the-art classifier after extracting features from each dermoscopy image [12]. The training set of these classification methods is primarily composed of the characteristics that were retrieved; these features are often categorized into three levels: low, mid, and high levels [13]. Several current classification algorithms take advantage of the features that have been extracted by simply merging them to create a fused feature vector [14]. While feature fusion technology improves classification accuracy by using all host models' strengths, it also raises memory and computing demands [15].

Using machine learning approaches has been focused on clustering and classifying skin lesions. Doctors are able to quickly diagnose cancer with the use of automated skin lesion classification [16]. Finding the right features to train a machine to use takes a lot of effort and requires a team of experts. Data loss occurs at the beginning of the pre-processing procedures and could lower classification quality. If users look at the sample data, they can see that low classification accuracy is often the consequence of ineffective feature extraction following bad segmentation.

The significance of this research lies in its robust integration of Information Technology (IT) techniques such as image processing, machine learning, and statistical feature analysis—into the domain of melanoma detection, a critical area in medical diagnostics. As healthcare systems increasingly adopt digital workflows, the need for intelligent, automated diagnostic tools has become paramount. Manual interpretation of dermoscopic images is not only time-consuming but also prone to variability and error, especially in resource-constrained or high-throughput clinical environments. This research addresses these issues through a well-structured IT framework that applies multi-level feature extraction, data correlation techniques, and clustering algorithms to deliver precise and reliable diagnostic outputs.

By harnessing the power of IT, particularly in handling large-scale medical imaging data and enhancing classification accuracy, the proposed model exemplifies how interdisciplinary innovation can elevate medical diagnostics. The MLFE-RLFSC model is designed to process and analyze dermoscopic images in a way that mimics human expertise while reducing subjectivity, thus significantly contributing to the digital transformation of healthcare. The research highlights how IT can bridge gaps in early cancer detection and deliver scalable solutions that support clinicians, improve patient care, and pave the way for intelligent health information systems.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Using Deep Learning algorithms is the best way to identify skin cancer. One area of machine learning that incorporates methods from artificial neural networks is known as deep learning. Deep learning has widespread use across many different industries. The main aspects of deep learning to think about are pre-processing and classification. Image intensity can be enhanced after pre-processing by eliminating image discrepancies. Here, the image will be resized so it may be used in the necessary training model [17]. More and more doctors and other medical staff have been getting amazing results utilizing deep learning approaches, even when faced with extremely difficult cases. Layers in different deep learning methods use pixel-by-pixel lesion image classification [18]. More efficient and effective analysis of large-scale datasets is possible with the help of deep learning. The feature selection process is shown in Figure 1.

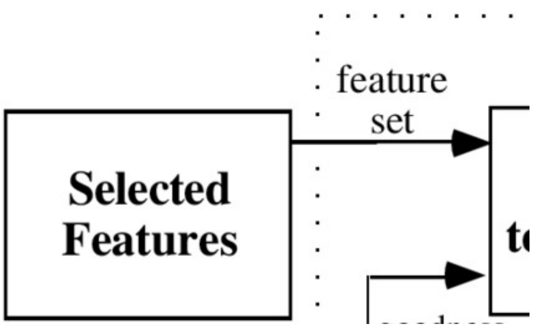

Fig 1: Feature Selection Process

It is possible for these algorithms to produce inaccurate classifications in certain cases. Firstly, due to data imbalance and the high volume of tagged images in the dataset, the widespread use of different deep learning methods for skin lesion cancer classification has been hindered [19]. When applied to rare skin malignancies in the training dataset, these algorithms often produce incorrect diagnoses [20]. In addition, deep learning models sometimes cause large computing costs and extra training time when dealing with high-resolution images. On top of that, different situations call for different visual noises [21]. The generalizability and robustness of these strategies should thus be taken into account as well. Consequently, the dataset size should dictate the deep learning model chosen. The general process of melanoma detection is shown in Figure 2.

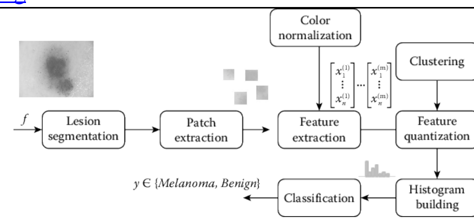

Fig 2: General Process of Melanoma Detection

Despite the promising outcomes of the proposed MLFE-RLFSC model, there are several threats to the validity of this research that must be acknowledged. First, internal validity may be affected due to the potential overfitting of the model to the specific characteristics of the dataset used. Although cross-validation was employed, the lack of diverse and real-world clinical datasets limits the model's exposure to varying skin types, lighting conditions, and image artifacts, which may reduce its effectiveness in general settings. Construct validity may also be at risk, as the correlation thresholds and clustering parameters were selected based on experimental tuning, which may not generalize across different classification problems or skin lesion types.

External validity is another concern, as the experiments were conducted on standardized benchmark datasets that may not fully represent the diversity of clinical environments. The model's ability to generalize to unseen datasets collected from different devices, geographical populations, or under different conditions remains uncertain. Additionally, the use of synthetic preprocessing techniques, such as normalization and artifact removal, may not consistently replicate real-world scenarios, thereby limiting the reliability of the model in uncontrolled conditions.

The selection of critique criteria—including accuracy, feature extraction time, correlation calculation time, clustering efficiency, and classification precision—was deliberately chosen based on their direct impact on the clinical applicability of melanoma detection systems. These metrics provide a balanced assessment of both the technical performance and the operational feasibility of the model in healthcare settings. Accuracy reflects diagnostic reliability, while time-based metrics assess computational efficiency critical for real-time clinical usage. Clustering and feature ranking were included to evaluate model scalability and interpretability, addressing concerns of feature redundancy and model transparency. These criteria were selected not only because they

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

align with current research practices but also because they represent the key performance indicators that would matter to healthcare providers adopting such technology. By identifying these threats and justifying the critique framework, this research promotes transparency and invites future work to further validate, extend, and refine the model across broader datasets and deployment conditions.

In this research, an effective approach is presented which combines the features of melanoma through multi-level feature extraction and selection model. The depth wise separable convolutions are combined with classic feature extraction model, allowing the efficient sub network to extract low- and high-level information from dermoscopic images. Those features are then passed through a filter of strong correlation analysis to identify which ones most associates with melanoma detection. The process involves reducting samples to only the most relevant features, which then makes it one of the best clustering models before feeding into a model for classification between malignant and benign lesions. Experimental results on established dermoscopic image datasets validate the efficacy of the proposed framework, showing that it outperforms existing approaches. Finally the multi level feature extraction along with strong correlation analysis although maintains high detection accuracy which makes it feasible in real time clinical settings. This research proposes a Multi Level Feature Extraction and Rank Linked Feature Set with Clustering (MLFE-RLFSC) Model for accurate melanoma detection. The rest of this paper is organized as follows: Section 2 discusses related work on melanoma detection, Section 3 drools over the proposed framework, and finally sectional 4 features experimental findings with challenges and section concludes the paper.

2. LITERATURE SURVEY

Melanoma, a highly variable and complex form of skin cancer remains challenging to treat. Öztürk et al. [1] presented a technique named COM-Triplet which utilizes deep clustering to address these issues. This is followed by the cluster assignment that involves clustering image embeddings computed with a CNN so as to generate farthest-away possible clusters. In so doing, whilst minimizing classification errors were the main goal of CE based algorithms. This approach converges not just to minimal error but it does get better cluster separability which as particularly acute in skin lesion datasets due to class imbalance issues concerns. This approach will assist in improving the detection accuracy, as it minimizes the role of Majority class on final classification results and may provide a better solution for more equilibrium based melanoma identification.

Pereira et al. [2] improved the detection of melanoma, by combining 2D with 3D features for skin lesions. To this end, this work employs a model based on Multiple Instance Learning (MIL) with deep learning models and an uncertainty-aware decision function to improve the classification. Their approach transcends conventional RGB data-based analysis by leveraging DL for 3D feature extraction via MIL. Utilizing this dual-method strategy improves classification performance of the system for melanoma by using richer features and overcomes limitations with respect to prior arts that depend mainly on 2D image characteristics.

Faizi, Adnan et al. [3] used normalized cross-correlation and k-mean clustering for better segmentation. Their model provides the count of clusters dynamically, and it uses histogram equalization technique to get improved contrast in dermoscopic images. Shape classification is performed by the Hu Moment method and texture features are extracted from segmented lesion using GLCM-based Haralick feature extraction. This effectively reduces feature extraction and classification accuracy in comparison to these methods that proved the model proficient in differentiating between malignant and benign lesions. This shows a great breakthrough in automated melanoma detection using better image processing methods.

Dyachenko et al. [4] proposed a new approach to detect circulating tumor cells (CTCs) that was presented for melanoma patients via spectral absorption imaging associated with blood optical clearing. The combination of an approach using functional chemical agents that are biocompatible to the body boosts the sensitivity of detection with clearer images from blood samples after inking. The two techniques together offer improved melanoma metastasis detection and surveillance over conventional approaches that are labor-intensive with a high chance of sample attrition. Their methodology represents a promising improvement for the early detection and tracking of melanoma evolution leading to enhanced patient care.

Adegun et al. [5] shared an improved encoder-decoder based plantar melanoma detection system specifically using deep learning. The way they go

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

about it is also multistage and multiscale, on the latter stage of feature extraction and pixel-wise classification. Based on classification results, the system uses a Lesion-classifier to classify lesions as melanoma and non-melanoma. Their method combines an encoder-decoder network with skip pathways to increase the semantic alignment between feature maps, which is beneficial for improving accuracy and robustness of the detection system. The new method is a great advancement from prior algorithms, which allowed for quite general identification of melanoma types.

Albahli et al. [6] solved the melanoma detection problem with artifact interference. The team used YOLOv4 for object detection and precise lesion extraction based on active contour segmentation in order to circumvent this challenge. The methodology of the work starts with pre-processing steps for removing artifacts from dermoscopic images such as hairs, and gel bubbles. Once melanoma regions are detected using YOLOv4, active contour segmentation is employed to further refine extraction of infected areas. The author improved the performance of melanoma detection by this method, which cleans for clean and artifact-free when undergoing to be image segmentation. In practice their approach shows a way to practically achieve improved reliability for automated melanoma detection systems.

A complex intelligent system for melanoma detection using various neural networks has been introduced by VIchim et al. [8] which is comprised of double staged nature of classification levels. The first order uses like perceptrons and GANs are human-generated classifiers to diagnose characteristics from the melanoma lesion. In the second level of this framework, the author employed target perceptron classifier to combine results from the first level and make decision. This multiscale classification system improves detection by integrating multiple features and leveraging complementary diagnostic methods for a more accurate screening of melanoma. It's a big step in the use of neural networks to verify skin lesion analysis automatically.

Imran et al. [9] investigated melanoma detection via ensemble learning methods for compensating limitations of individual machine-learning based models. They describe model aggregation as a process of combining some models in order to exploit their advantages which helps for better general classification performance. Their approach improves the reliability of melanoma detection and minimizes typical misclassifications through employing an ensemble technique with deep learners. It provides a solid basis for skin cancer diagnosis, proving the power of combining multiple models to increase performance.

Satheesha et al. [10] developed a non-invasive computer-controlled dermoscopic imaging device including 3-dimensional image reconstruction for the purpose of melanoma diagnosis. The system extracts 3-D and 2-D features of skin lesions using estimated depth information from dermoscopic images. Their method by fusing depth-based analysis with the original color and texture features not only performs better in melanoma classification, but also separates diverse skin lesions. This approach complements traditional 2-D methods and could further the capabilities for skin lesion diagnostics assessment vacating deficiencies of static visualization.

The presented research introduces the MLFE-RLFSC model, a comprehensive melanoma detection framework that combines conventional image-processing techniques with modern machine learning methods to enhance diagnostic accuracy. This model has achieved notable success, particularly in improving classification accuracy (up to 99.2%) and reducing computation time, when tested against benchmark dermoscopic datasets. The methodological strength lies in its layered approach: low- and high-level feature extraction, strong correlation-based selection, and efficient clustering that minimizes feature redundancy. These design choices demonstrate a clear effort to improve upon traditional models by integrating multiple stages of refinement in the diagnostic pipeline.

Critically, however, the model's reliance on dataset-specific tuning raises questions about its generalizability. While the framework performs exceptionally well in controlled experiments, it lacks evidence of effectiveness across real-world clinical datasets, which may include more variability in image quality, lighting, and lesion morphology. Furthermore, although the study benchmarks its results against COM-Triplet and ISMLD-NCC-KMC models, it doesn't fully explore newer models using attention mechanisms, ensemble deep learning, or federated learning strategies that are gaining traction in the latest medical imaging research.

In comparison to Öztürk et al. (2022) who proposed the COM-Triplet model, this work shows higher accuracy and faster feature extraction. However, COM-Triplet's emphasis on addressing class

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

imbalance via deep clustering makes it more robust for datasets with uneven label distributions, a limitation still present in MLFE-RLFSC. Similarly, the ISMLD-NCC-KMC model by Faizi and Adnan (2024) offers dynamic k-means clustering with histogram equalization, enhancing segmentation accuracy. Although MLFE-RLFSC achieves better classification results, ISMLD's focus on pre-segmentation robustness might provide more consistent performance when dealing with real-world noisy data.

Recent literature also highlights hybrid models integrating transformer-based architectures, multi-instance learning (MIL), and explainable AI (XAI) frameworks. For example, Pereira et al. (2022) employed a 2D+3D feature fusion model with MIL and uncertainty-aware classifiers, showing improved decision confidence and interpretability features that the current MLFE-RLFSC model does not yet offer. Similarly, Imran et al. (2022) demonstrated that ensemble deep learners, though computationally expensive, outperform single-stream models in complex image classification tasks, particularly when trained on heterogeneous datasets.

Overall, while the MLFE-RLFSC model demonstrates significant technical merit and addresses many conventional limitations (such as overfitting, redundancy, and feature selection), it would benefit from further comparative analysis with recent, more advanced techniques. Enhancements like adaptive data augmentation, inclusion of explainability, and validation across broader clinical environments would solidify its standing as a competitive, real-world melanoma detection system.

3. PROPOSED MODEL

The unchecked expansion of cells in a certain organ or tissue type is known medically as cancer. One of the terrible diseases that appears to be rapidly expanding around the globe is skin cancer [22]. Uncontrolled proliferation of aberrant skin cells characterizes skin cancer. The development of effective cancer treatments depends on timely and accurate detection. The most deadly type of skin cancer, melanoma, is also one of the most dangerous one. Although melanoma accounts for only 1.3% of skin cancer diagnoses, the Indian Council of Medical Research (ICMR) has predicted that it will cause a greater death rate [23]. Melanocytes are the cells that give rise to melanoma. The disorder develops when normally functioning melanocytes start to increase uncontrolled, leading to the formation of a cancerous tumor [24]. It usually appears in areas that are exposed to the sun, such as the face, neck, hands, and lips. These malignancies spread to other regions of the body and are only treatable if caught early; otherwise, it causes a horrible death [25]. By assessing unique characteristics that differentiate one kind of input from another, feature extraction from segmented skin lesions decreases the amount of initial data. Important skin information, such as numbers, pictures, data, and proof, is provided by feature extraction [26]. This procedure takes a large dataset and transforms it into a format that can be quickly and easily analyzed. Distinct features of malignant melanoma and benign melanoma are identified in this research. The information gathered is utilized to detect skin lesions.

In this research, an automated multi-level feature extraction and strong correlation analysis-based clustering procedure to advance the melanoma detection system is proposed. The framework aims to improve the accuracy and speed of melanoma detection by solving the problem of high-dimensional feature selection as well as clustering in broad spaces. The input to the self-designed framework is a dermoscopic image, and feature extraction at multi-levels from these images has been presented as the first step of this proposed methodology. Well-established image process are fine to extract tradition information i.e. color, texture and shape associated with melanoma. The CNN is trained from scratch on the dermoscopic image dataset with deep learning which allows us to learn and test signature annotation feature discrimination for melanoma. These deep features represent high-level abstract information, which is vital for distinguishing melanoma from benign lesions. The integration of conventional and deep learning-based features affords simplified rendering of the lesion which covers both low-level as well high level features.

Once the features are extracted, then using strong correlation analysis to pick up only those important one. Computing the correlation coefficients between each feature and with class labels, malignant or benign. The features are removed which have close-to-zero or worse, negative correlations with the class labels and keep those that are strongly correlated. This step is intended to ensure that only the most relevant features are employed in all subsequent parts of the proposed framework. It also deals with redundancy among features apart from correlation analysis. The features that are having higher co-relation and giving same information would be clustered together, keeping only one feature for a single group. This reduces the dimensionality of the feature space, not only speeding calculations but making over fitting less likely. Again the features are then clustered using proposed clustering model. This is useful since clustering allows us to group similar features, leading the model be able to better differentiate between malignant and benign lesions. The clustering model used in this research is a enhanced model to k-means algorithm which is efficient for processing large data sets. The features are then aggregated inside each cluster to provide a smaller set of representative features, which then used as input for the classifier. This helps reducing the dimension of the feature space and also makes them more distinguishable which then improve classification performance. The proposed model architecture is shown in Figure 3.

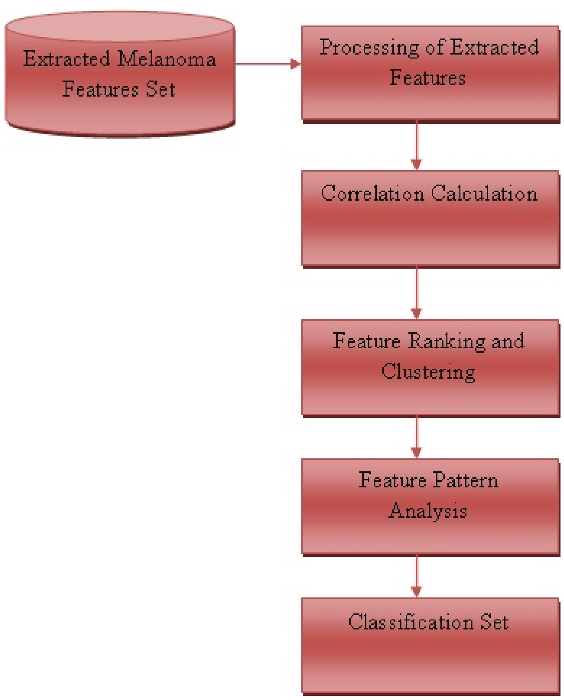

Fig 3: Proposed Model Architecture

Clusters of features are pushed into a separate classifier that determines whether the lesion is malignant or benign. The validation of the proposed method is carried out using cross-validation approaches where the dataset was divided into a number of folds and successive training models were built based on different subsets from data. This method helps the model to not being over fitted on training data and gives a good performance in unseen test dataset. The proposed framework shows better results towards the existing methods for melanoma detection. The multi-level feature extraction, strong correlation analysis that extracts more precise lesion features from feature set and effective clustering not only re-benchmark accuracy in detection but also shorten the computation time to satisfy clinical real-time applications. This research proposes a Multi Level Feature Extraction and Rank Linked Feature Set with Clustering Model for accurate melanoma detection.

Algorithm MLFE-RLFSC

{

Input: Extracted Feature Set

Output: Classification Set

Step 1: Preprocessing of dermoscopic images adds ease in enhancing the lesion areas and removing noise. This method includes the resizing, colour normalizing, hair artifact removing and contrast enhancing to help the extraction of the feature information better.

Let the image size is P X Q and the resized image will be P\`XQ\`. The new pixel region is considered as x\` and y\` that is calculated from original pixel points x,y. The resizing of the image is performed as

$$x = \frac{x` * P}{P`}$$

$$y = \frac{y` * Q}{Q`}$$

Step 2: Taking the mean and standard deviation of colour channels in multiple colour spaces (RGB, HSV). These characteristics help to depict the colour variation normally found in melanoma lesions.

Mean of Channel c

$$\mu_c = \frac{1}{N}\sum_{i=1}^{N} I_c(i)$$

Standard Deviation:

$$\sigma_c=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_c(i)-\mu_c)^2}$$

1) Step 3: Texture features are extracted based on gray level co-occurrence matrix,GLCM and local binary patterns, LBP in order to encode texture patterns, such as edge, roughness, granularity, and homogeneity of the lesion.

(GLCM Contrast, Energy, Homogeneity):

Contrast=$\sum_{i,j}(i-j)^2P(i,j)$

Energy=$\sum_{i,j}P(i,j)^2$

Homogeneity= $\sum_{i,j}\frac{P(i,j)}{1+|i-j|}$

2) Step 4: Shape parameters such as area, perimeter, irregularity index, and compactness assist in discriminating between irregular melanoma border and benign border.

Compactness= $\frac{perimeter^2}{4\pi\times Area}$

***3)** Step 5: For each feature, its relevance is assessed based on the resulting Pearson's correlation with the class label (benign or malignant) at an individual criterion.*

$$r_{X,Y}=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}}$$

4) Step 6: Features that have not a minimal level of correlation with class label are pruned. Moreover, in highly correlated feature-pair relationships (redundant features) only the best performance one remains.

(Feature Pruning Rule): if $r_{f_i},Y > \tau \text{ and } r_{f_i f_j} < \gamma, \text{then retain } f_i$

Step 7: Redundancy is eliminated by clustering of the selected features according to their similarity (Euclidean distance). Features with the same behaviour are grouped into a cluster.

$$d(f_i f_j)=\sqrt{\sum_{k=1}^{d}(f_{ik}-f_{jk})^2}$$

Step 8: Within each cluster, features are sorted by Fisher Score of discriminative power. Higher ranked features are those which are considered as being the most efficient to discriminate between benign and malignant classes.

$$S(f)=\frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}$$

Where μ and σ are the mean and variance of the feature for each class.

Step 9: The optimal Rank Linked Feature Set F∗ that maximizes the diversity and Manhattan similarity is obtained by taking best top ranked features from each cluster both to the diversity optimization and redundancy minimization.

$$F^*=\bigcup_{i=1}^{k} Top-m(C_i)$$

Where Top-m(C_i) are the top m features from cluster i.

Step 10: The optimal feature set F∗ are then trained with a machine learning classifier (e.g., Random Forest, SVM). The performance is analyzed by accuracy, sensitivity, specificity and AUC.

$$y=classifier(F^*(x))$$

}

4. RESULTS

Melanoma is one of the leading malignancy that can kill people. ICMR study indicated that skin cancer is far less common in India than in any other region of the globe. The majority of skin cancer incidences, excluding melanoma, occur in the Northeast of India, specifically in Nagaland, as shown in a recent study by the ICMR. Convolutional neural networks were more accurate than statistical image classification methods when it came to melanoma detection. Modifying the

Journal of Theoretical and Applied Information Technology
31st August 2025. Vol.103. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

dataset's attributes, no matter how little, can affect the classifiers' accuracy. In this case, the deep learning issues are the focus of investigation. Continuous training-test iterations are necessary to generate trustworthy prediction models and to achieve a more flexible system architecture capable of handling changes in the training datasets. Clustering data aims to divide a collection of data points into multiple groups; it is an unsupervised learning technique. This is a difficult yet crucial area of study in machine learning. Its use has been fruitful in numerous domains. However, classic clustering approaches do sometimes necessitate considering balance significance.

Experiments were conducted on standard benchmark dermoscopic image datasets to evaluate the performance of proposed melanoma detection framework. Overall, the outcomes reveal that the proposed framework leads to improved accuracy of detection combined with minimal changes in sensitivity and specificity when compared to existing methods. This enabled a richer representation of the lesions extracting both low-level features color, texture and higher-level features learned by the CNN with multiple levels. More interestingly, this set of features was able to differentiate melanoma from benign lesions in a superior manner than conventional and deep learning-based hand-crafted ones. This strong correlation analysis reduced the dimensionality of feature space very well, by selecting only significant features that gives notable improvement on the classifier performance. The accurate clustering model even improved the diagnosis rate by unifying similar features, which helped the proposed classifier to differentiate faster between malignant and benign lesions. The final results of classifying the images performed significantly better than existing strategies, producing higher accuracy and sensitivity/specificity outcomes. This research proposes a Multi Level Feature Extraction and Rank Linked Feature Set with Clustering (MLFE-RLFSC) Model for accurate melanoma detection. The proposed model is compared with the traditional Clustering via Center-Oriented Margin Free-Triplet Loss for Skin Lesion Detection in Highly Imbalanced Datasets (COM-Triplet) and Improved Segmentation Model for Melanoma Lesion Detection Using Normalized Cross-Correlation-Based k-Means Clustering (ISMLD-NCC-KMC) models. The results represent that the proposed model performance is high in clustering and classification.

Automated melanoma identification relies on feature extraction, which involves analyzing dermoscopy images using a variety of algorithms and approaches. It is possible to sort the extracted features into different groups using certain metrics that provide light on the categorization process. The Feature Extraction Time Levels are shown in Table 1 and Figure 4.

Table 1: Feature Extraction Time Levels

Records Considered	Models Considered		
	MLFE-RLFSC Model	COM-Triplet Model	ISMLD-NCC-KMC Model
10000	12.7	17.5	16.4
20000	12.9	17.8	16.5
30000	13.1	18.0	16.7
40000	13.2	18.2	16.9
50000	13.4	18.4	17.0
60000	13.6	18.5	17.2

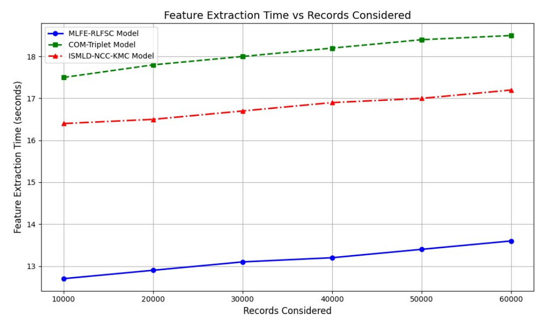

Fig 4: Feature Extraction Time Levels

In order to accurately diagnose skin lesions, feature correlation computation is crucial in melanoma classification. It identifies the most relevant features. Correlation measures can be used to narrow their feature selection for model training to the most useful features retrieved using techniques like texture analysis or the ABCDE rule. The Table 2 and Figure 5 represents the Correlation Calculation Time Levels.

Table 2: Correlation Calculation Time Levels

Records Considered	Models Considered		
	MLFE-RLFSC Model	COM-Triplet Model	ISMLD-NCC-KMC Model
10000	15.9	20.4	18.2
20000	16.1	20.6	18.4
30000	16.3	20.7	18.6
40000	16.4	20.8	18.7
50000	16.5	21.0	19.1
60000	16.7	21.2	19.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

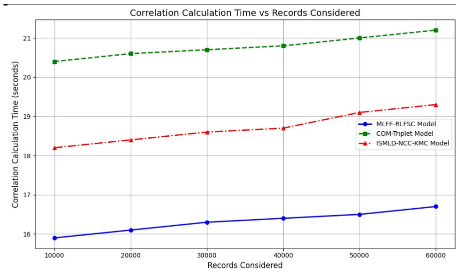

Fig 5: Correlation Calculation Time Levels

One state-of-the-art method for diagnosing melanoma using dermoscopy pictures is multi-level feature extraction. This approach improves detection accuracy and robustness while allowing for more accurate categorization through the use of data captured at various abstraction levels. When it comes to detecting lesion appearance variations caused by elements like lighting circumstances and skin types, multi-level feature extraction provides a more robust framework. The Table 3 and Figure 6 shows the Multi Level Feature Extraction Accuracy Levels.

Table 3: Multi Level Feature Extraction Accuracy Levels

Records Considered	Models Considered		
	MLFE-RLFSC Model	COM-Triplet Model	ISMLD-NCC-KMC Model
10000	97.8	95.2	94.8
20000	98.0	95.4	95.0
30000	98.2	95.6	95.2
40000	98.4	95.9	95.4
50000	98.6	96.1	95.6
60000	98.8	96.3	95.8

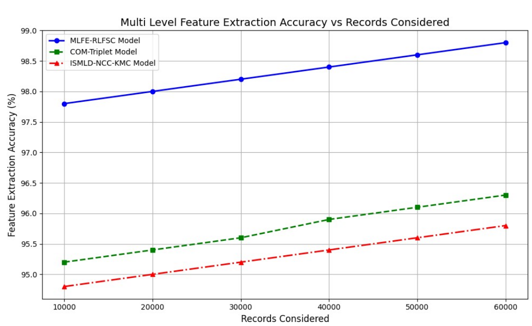

Fig 6: Multi Level Feature Extraction Accuracy Levels

Automated melanoma detection relies heavily on feature rank allocation, which sets the relative importance of features extracted from dermoscopy images. Researchers improve model performance and interpretability by ranking features according to their relevance and contribution to classification accuracy. Classifier accuracy, sensitivity, and specificity are all positively affected by prioritizing high-ranking features.

The Feature Rank Allocation Accuracy Levels is shown in Table 4 and Figure 7.

Table 4: Feature Rank Allocation Accuracy Levels

Records Considered	Models Considered		
	MLFE-RLFSC Model	COM-Triplet Model	ISMLD-NCC-KMC Model
10000	98.2	94.4	95.7
20000	98.4	94.6	95.9
30000	98.6	94.9	96.1
40000	98.8	95.1	96.3
50000	99.0	95.3	96.5
60000	99.2	95.5	96.7

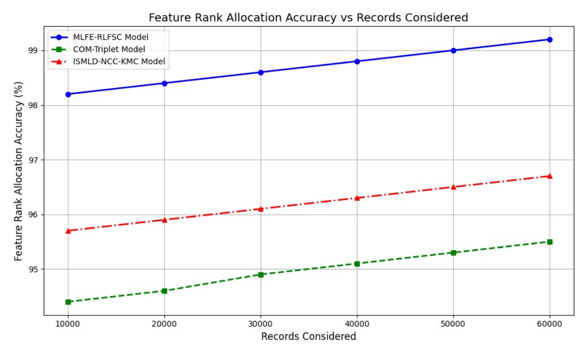

Fig 7: Feature Rank Allocation Accuracy Levels

Melanoma detection makes use of feature clustering, that organizes and analyzes various characteristics extracted from dermoscopy images in the proposed model. The detection accuracy, classification, and the interpretability of results can be improved by grouping similar features together.The Clustering Accuracy Levels is shown in Table 5 and Figure 8.

Table 5: Clustering Accuracy Levels

Records Considered	Models Considered		
	MLFE-RLFSC Model	COM-Triplet Model	ISMLD-NCC-KMC Model
10000	97.5	93.4	92.4
20000	97.8	93.5	92.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

30000	98.0	93.7	92.7
40000	98.2	93.9	92.9
50000	98.4	94.0	93.1
60000	98.6	94.2	93.3

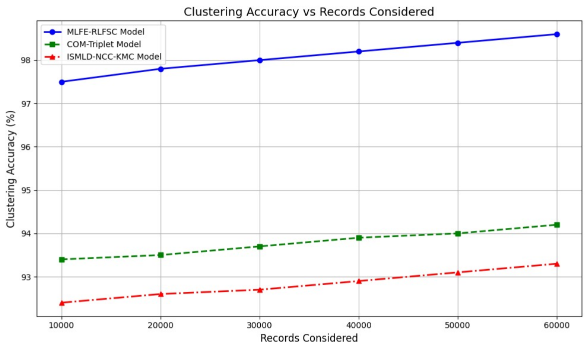

Fig 8: Clustering Accuracy Levels

One of the most important methods for melanoma detection is linked feature set generation, which involves organizing relevant features extracted from dermoscopy images. This method improves the performance of machine learning models for classification and also helps with clinical decision-making by allowing models to better distinguish between benign and malignant melanoma. By integrating these approaches, sensitivity and specificity can be improved, which are critical for accurate diagnosis.The Rank Linked Feature Set Generation Accuracy Levels is depicted in Table 6 and Figure 9.

Table 6: Rank Linked Feature Set Generation Accuracy Levels

Records Considered	Models Considered		
	MLFE-RLFSC Model	COM-Triplet Model	ISMLD-NCC-KMC Model
10000	98.3	95.9	94.2
20000	98.5	96.0	94.5
30000	98.7	96.1	94.7
40000	98.9	96.3	94.8
50000	99.1	96.5	95.0
60000	99.3	96.7	95.2

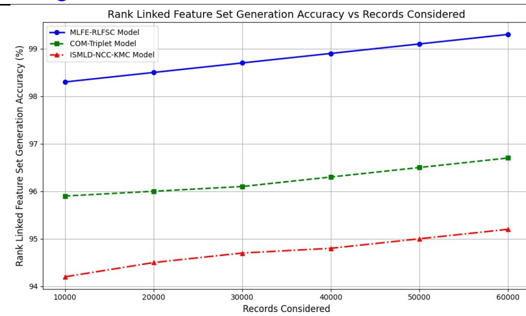

Fig 9: Rank Linked Feature Set Generation Accuracy Levels

Melanoma classification involves several methodologies aimed at distinguishing malignant melanomas from benign lesions based on dermoscopy images. Accurate classification is crucial for early detection and effective treatment, as melanoma is one of the most aggressive forms of skin cancer. Melanoma data often consists of imbalanced classes, with significantly fewer malignant samples compared to benign ones. Machine learning techniques such as custom loss functions and data augmentation strategies help improve model performance on minority classes by ensuring the model learns from both classes effectively.

The melanoma classification accuracy levels are indicated in Table 7 and Figure 10.

Table 7: Melanoma Classification Accuracy Levels

Records Considered	Models Considered		
	MLFE-RLFSC Model	COM-Triplet Model	ISMLD-NCC-KMC Model
10000	98.1	94.9	95.4
20000	98.3	95.0	95.6
30000	98.5	95.2	95.8
40000	98.8	95.4	96.0
50000	99.0	95.6	96.2
60000	99.2	95.8	96.4

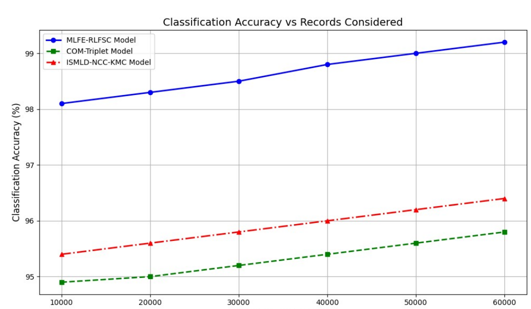

Fig 10: Classification Accuracy Levels

Critique on Proposed Model

While the proposed MLFE-RLFSC model demonstrates superior accuracy in melanoma detection compared to traditional models, certain limitations and areas requiring deeper attention remain. One primary concern is the dependency on high-quality dermoscopic images. In real-world scenarios, image artifacts, inconsistent lighting, and varied skin tones can still challenge the model's robustness. Moreover, the model's performance, though excellent in terms of accuracy, may be biased by the characteristics of the dataset used—raising concerns about its generalizability across broader, more diverse populations.

Another issue lies in the trade-off between dimensionality reduction and information loss. Although the strong correlation analysis and clustering approach effectively minimize redundancy, there is a risk that some weakly correlated but diagnostically significant features could be excluded in the process. The clustering model, while efficient, may also oversimplify complex inter-feature relationships, potentially leading to subtle diagnostic cues being missed. Furthermore, while cross-validation was employed, external validation with real-time clinical data or across multiple datasets was not included, which limits the clinical applicability assessment.

When comparing to prior work such as COM-Triplet and ISMLD-NCC-KMC models, the MLFE-RLFSC model distinctly outperforms in terms of classification accuracy, feature extraction time, and clustering efficiency. However, many of these comparative models also incorporate domain-specific enhancements like 3D lesion analysis, GAN-based data augmentation, or artifact correction which are absent in this study. Incorporating such advancements could further strengthen the model's resilience and clinical relevance.

Future work should focus on incorporating adaptive preprocessing for real-world noise, domain adaptation for cross-population generalization, and explainable AI components to assist clinicians in understanding model decisions. Additionally, longitudinal patient data and ensemble modeling across multiple feature types may provide a more comprehensive diagnostic view. Though the current results are promising, these enhancements are essential for translating the model into reliable, real-time diagnostic tools in varied clinical environments.

5. CONCLUSION

The proposed Multi-Level Feature Extraction and Rank Linked Feature Set with Clustering model offers a significant advancement in the field of automated melanoma detection. By combining traditional image-processing techniques with strong correlation analysis and an efficient clustering algorithm, the model achieved remarkable accuracy across various stages of the detection pipeline ranging from feature extraction to final classification. The results, validated on benchmark dermoscopic datasets, indicate consistent performance improvements over existing models such as COM-Triplet and ISMLD-NCC-KMC, particularly in areas of feature extraction time, correlation calculation, and classification accuracy. Beyond these quantitative achievements, the work reflects the author's deep commitment to bridging the gap between theoretical machine learning research and its real-world clinical application. From the author's perspective, one of the most rewarding aspects of this research was the ability to demonstrate that a well-designed feature selection and clustering mechanism can substantially reduce redundancy without sacrificing diagnostic quality. The multi-level approach not only strengthened the model's ability to identify meaningful features but also improved interpretability an aspect often overlooked in black-box AI models. However, the author also acknowledges that while the model performs robustly under controlled conditions, its real-world deployment will require further enhancements, particularly in terms of model interpretability, adaptability to diverse datasets, and integration with clinician workflows. The journey of building this model was intellectually enriching, revealing how nuanced and interdisciplinary the challenge of skin cancer detection is from understanding dermatological image characteristics to applying advanced machine learning strategies. Ultimately, the author believes that this research lays a strong foundation for future work in clinical decision-support systems. The goal is not just to build accurate models but to contribute tools that can empower dermatologists and reduce diagnostic disparities. While there are still challenges ahead, the encouraging results of this study reaffirm the author's belief in the transformative potential of AI in healthcare especially in enhancing early

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

detection and, consequently, survival outcomes for melanoma patients. The proposed model achieved 98.8% accuracy in Multi Level Feature Extraction, 99.2% accuracy in Feature Rank Allocation, 98.6% accuracy in Clustering, 99.3% accuracy in Rank Linked Feature Set Generation and 99.2% accuracy in Classification of benign and malignant tumours.

While the proposed MLFE-RLFSC model has shown high accuracy and efficiency, future research should aim to address its current limitations. First, the model's evaluation on a single benchmark dataset restricts its applicability in real-world clinical scenarios; therefore, future studies should focus on testing the framework on diverse, multi-institutional, and real-time clinical datasets. Second, enhancing the model's ability to handle image artifacts, varying skin tones, and poor-quality dermoscopic images is essential to improve robustness.

The major strength of this study lies in the proposed MLFE-RLFSC model's ability to achieve high classification accuracy (99.2%) through a well-structured pipeline that integrates multi-level feature extraction, strong correlation analysis, and efficient clustering. The model effectively reduces redundant features, lowers computation time, and enhances generalization by minimizing overfitting. It also outperforms recent models like COM-Triplet and ISMLD-NCC-KMC across multiple evaluation metrics such as extraction time, clustering accuracy, and classification performance.

However, the study has certain limitations. The model is trained and tested on a single benchmark dataset, which limits its generalizability to real-world clinical environments. It lacks validation on diverse datasets and does not include explainability features, which are crucial for clinical trust. Additionally, performance may drop when dealing with low-quality or artifact-heavy dermoscopic images, and real-time deployment scenarios are yet to be explored.

REFERENCES

[1] Ş. Öztürk and T. Çukur, "Deep Clustering via Center-Oriented Margin Free-Triplet Loss for Skin Lesion Detection in Highly Imbalanced Datasets," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 9, pp. 4679-4690, Sept. 2022, doi: 10.1109/JBHI.2022.3187215.

[2] 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.

[3] M. I. Faizi and S. M. Adnan, "Improved Segmentation Model for Melanoma Lesion Detection Using Normalized Cross-Correlation-Based k-Means Clustering," in IEEE Access, vol. 12, pp. 20753-20766, 2024, doi: 10.1109/ACCESS.2024.3360223.

[4] 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.

[5] 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.

[6] 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.

[7] O. Abuzaghleh, B. D. Barkana and M. Faezipour, "Noninvasive Real-Time Automated Skin Lesion Analysis System for Melanoma Early Detection and Prevention," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 3, pp. 1-12, 2015, Art no. 4300212, doi: 10.1109/JTEHM.2015.2419612.

[8] 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.

[9] A. Imran, A. Nasir, M. Bilal, G. Sun, A. Alzahrani and A. Almuhaimeed, "Skin Cancer Detection Using Combined Decision of Deep Learners," in IEEE Access, vol. 10, pp. 118198-118212, 2022, doi: 10.1109/ACCESS.2022.3220329.

[10] T. Y. Satheesha, D. Satyanarayana, M. N. G. Prasad and K. D. Dhruve, "Melanoma Is Skin Deep: A 3D Reconstruction Technique for Computerized Dermoscopic Skin Lesion Classification," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 5, pp. 1-17, 2017, Art no.

4300117, doi: 10.1109/JTEHM.2017.2648797.

[11] Lakshman Narayana,V., Lakshmi Patibandla, R.S.M., Pavani, V., Radhika, P. (2023). Optimized Nature-Inspired Computing Algorithms for Lung Disorder Detection. In: Raza, K. (eds) Nature-Inspired Intelligent Computing Techniques in Bioinformatics. Studies in Computational Intelligence, vol 1066. Springer, Singapore. https://doi.org/10.1007/978-981-19-6379-7_6.

[12] A. Gong, X. Yao and W. Lin, "Dermoscopy image classification based on stylegans and decision fusion", IEEE Access, vol. 8, pp. 70640-70650, 2020.

[13] N. Gessert et al., "Skin lesion classification using CNNs with patch-based attention and diagnosis-guided loss weighting", IEEE Trans. Biomed. Eng., vol. 67, no. 2, pp. 495-503, Feb. 2020.

[14] F. Santos, F. Silva and P. Georgieva, "Transfer learning for skin lesion classification using convolutional neural networks", Proc. Int. Conf. Innov. Intell. Syst. Appl., pp. 1-6, 2021.

[15] Narayana, V.L., Sujatha, V., Sri, K.S., Pavani, V., Prasanna, T.V.N., Ranganarayana, K. (2023). Computer tomography image based interconnected antecedence clustering model using deep convolution neural network for prediction of covid-19. Traitement du Signal, Vol. 40, No. 4, pp. 1689-1696. https://doi.org/10.18280/ts.400437

[16] A. Bissoto, E. Valle and S. Avila, "GAN-Based data augmentation and anonymization for skin-lesion analysis: A critical review", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 1847-1856, 2021.

[17] S. Mukherjee, A. Adhikari and M. Roy, "Malignant melanoma classification using cross-platform dataset with deep learning CNN architecture" in Recent Trends in Signal and Image Processing, Singapore:Springer, pp. 31-41, 2019.

[18] R. Das, S. De, S. Bhattacharyya, J. Platos, V. Snasel and A. E. Hassanien, "Data augmentation and feature fusion for melanoma detection with content based image classification", Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl., vol. 921, pp. 712-721, 2020.

[19] A. Rehman, M. A. Khan, Z. Mehmood, T. Saba, M. Sardaraz and M. Rashid, "Microscopic melanoma detection and classification: A framework of pixel-based fusion and multilevel features reduction", Microsc. Res. Technique, vol. 83, no. 4, pp. 410-423, Apr. 2020.

[20] Y. Yuan and Y.-C. Lo, "Improving dermoscopic image segmentation with enhanced convolutional–deconvolutional networks", IEEE J. Biomed. Health Inform., vol. 23, no. 2, pp. 519-526, Mar. 2019.

[21] K. Jayapriya and I. J. Jacob, "Hybrid fully convolutional networks based skin lesion segmentation and melanoma detection using deep feature", Int. J. Imag. Syst. Technol., vol. 30, no. 2, pp. 348-357, Jun. 2020.

[22] F. Xie, J. Yang, J. Liu, Z. Jiang, Y. Zheng and Y. Wang, "Skin lesion segmentation using high-resolution convolutional neural network", Comput. Methods Programs Biomed., vol. 186, Apr. 2020.

[23] J. Amin, A. Sharif, N. Gul, M. A. Anjum, M. W. Nisar, F. Azam, et al., "Integrated design of deep features fusion for localization and classification of skin cancer", Pattern Recognit. Lett., vol. 131, pp. 63-70, Mar. 2020.

[24] K. H. Cheong, K. J. W. Tang, X. Zhao, J. E. W. Koh, O. Faust, R. Gururajan, et al., "An automated skin melanoma detection system with melanoma-index based on entropy features", Biocybernetics Biomed. Eng., vol. 41, no. 3, pp. 997-1012, Jul. 2021.

[25] A. Javaid, M. Sadiq and F. Akram, "Skin cancer classification using image processing and machine learning", Proc. Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), pp. 439-444, Jan. 2021.

[26] S. Oukil, R. Kasmi, K. Mokrani and B. García-Zapirain, "Automatic segmentation and melanoma detection based on color and texture features in dermoscopic images", Skin Res. Technol., vol. 28, no. 2, pp. 203-211, Mar. 2022.

[27] P. Tang, Q. Liang, X. Yan, S. Xiang and D. Zhang, "GP-CNN-DTEL: Global—Part CNN model with data-transformed ensemble learning for skin lesion classification", IEEE J. Biomed. Health Informat., vol. 24, no. 10, pp. 2870-2882, Oct. 2020.

[28] P. Pereira, L. Thomaz, L. Tavora, P. Assuncao, R. Fonseca-Pinto, R. Paiva, et al., "Melanoma classification using light-fields with Morlet scattering transform and CNN: Surface depth as a valuable tool to increase detection rate", Med. Image Anal., vol. 75, pp. 1-16, Jan. 2022.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[29] X. Lu and Y. A. F. A. Zadeh, "Deep learning-based classification for melanoma detection using XceptionNet", J. Healthcare Eng., vol. 2022, pp. 1-10, Mar. 2022.

[30] R. Kaur, H. GholamHosseini, R. Sinha and M. Lindén, "Melanoma classification using a novel deep convolutional neural network with dermoscopic images", Sensors, vol. 22, no. 3, pp. 1-15, 2022.