International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 22 August 2023 **Revised:** 10 October 2023 **Accepted:** 26 October 2023*

Dermatological Detection and Classification using Machine Learning Techniques

Dr.Sarangam Kodati[1,2], Dr. Nethravathi P S[2] Dr.M.Dhasaratham[3]
[1]Department of Information Technology, CVR College of Engineering, Hyderabad, Telangana, India-501510.
[*]Corresponding author email id: k.sarangam@gmail.com
[2]Institute of Computer and Information Sciences, Srinivas University, Mangalore, Karnataka, India-574146.
Email id: nethrakumar590@gmail.com
[3]Department of Information Technology, TKR College of Engineering and Technology, Hyderabad, Telangana, India-500097.
Email id: dasarath.m@gmail.com

Abstract: Dermatology is the medical field that focuses on the study and treatment of skin conditions. It is a specialized branch of medicine that encompasses both diagnostic and surgical procedures related to the skin. It is a widespread disease among them. The researchers have shows a lot of attention to the early detection of lesions. Because of their proliferation ability to other parts of the body, death rates are quite high. A system that can distinguish between benign and malignant lesions is essential because melanoma can be cured with an early and accurate diagnosis. Dermoscopic skin lesion images are first segmented using data mining techniques, to identify the area of interest of the lesion part. When compared to individual classifier algorithms, dermatology datasets benefit from the various data mining techniques and feature selection methods. The SVM provides more accurate and effective skin disease prediction in terms of accuracy, precision, and Specificity.

Keywords- Support Vector Machine (SVM), Dermoscopic, Melanoma.

I. INTRODUCTION

The World Health Organization (WHO) reports that melanoma is highly prevalent globally and ranks among the most common forms of cancer. It has become increasingly prevalent in recent years and represents approximately 33% of all diagnosed cancer cases [4]. The ozone layer is responsible for shielding individuals from the significant contribution of UltraViolet (UV) radiation in preventing the abnormal proliferation of skin cells by serving as a key element and being arranged widely. The most prevalent forms of skin cancer include Melanoma (Mel), Basal Cell Carcinoma (BCC), and Squamous Cell Carcinoma (SCC). In particular, melanoma is highly aggressive and responsible for 75% of skin cancer-related fatalities. Several factors increase the likelihood of developing skin cancer, including fair skin, a history of sunburns, prolonged exposure to UltraViolet (UV) light, and the utilization of tanning beds [6].

The Skin Cancer Foundation (SCF) says cancer is considered a severe kind of cancer as it spreads to different parts of the body. Skin cancer's treatment becomes significantly more challenging. Early detection of melanoma is important to save lives. Melanoma is 100% curable when detected and treated in its early stages, according to research. Cancer can progress, spread, and even be fatal if not treated early. The presence of melanocytes in any part of the body is strongly associated with the origin of melanoma cells. As a result, the possibility that melanocytes are melanoma's precursors may appear in the body. The risk of melanoma is strongly influenced by skin exposure to UV radiation. Prolonged exposure and a higher intensity of UV radiation greatly increase this risk. Prolonged or excessive sun exposure is widely recognized as a major risk factor for melanoma.

Dermoscopy, also known as Epiluminescence Microscopy (ELM) or dermatoscopy, is a noninvasive clinical examination technique utilized in vivo to complement conventional clinical diagnosis [15]. This method involves the application of oil immersions as well as optical magnifications to enhance the transparency of the outer layer, facilitating visual inspection of the underlying structures of the skin. However, when diagnosing melanoma in pigmented skin lesions, experienced users believe that ELM is more accurate than clinical examination. Dermoscopy's diagnostic accuracy is heavily influenced by the dermatologist's training, in spite of dermoscopy's significantly higher melanoma detection rate than separate observation. Differentiating melanoma from melanocytic is difficult and frequently requires clinically determined demands for melanoma, particularly in the early phases. Expert dermatologists' melanoma diagnosis accuracy is only about 75%–84% when dermoscopy is used for diagnosis [18], which is viewed as far from good for symptomatic purposes. Due to this, there is an increasing interest in diagnostic strategies that incorporate computer-assisted analysis of lesion images. This approach offers a more effective and efficient way of analysing and interpreting the images for diagnostic purposes.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 22 August 2023 **Revised:** 10 October 2023 **Accepted:** 26 October 2023*

The epidermis has three types of cells: melanocytes, squamous cells, and pigmented cells. Abnormalities in melanin and melanocytes at the basal level can lead to various skin diseases. Among them, melanoma is the deadliest skin disease. Classifying melanoma can be challenging because of changes in colour as well as the difficulty in identifying early lesions. However, timely detection and treatment can lead to a cure. Therefore, approximate classifications of pigmentation-related skin diseases are crucial for effective subsequent medication. In clinical settings, cutting-edge skin imaging technology and expert knowledge are frequently utilized for treatment and diagnosis. The automatic classification of malignant skin diseases from dermoscopic images presents challenges, primarily due to the laborious and time-consuming nature of the image analysis involved in the process. Moreover, the factors that may not be fully valid can add more complexity to the task.

Advancements in image processing and computer-Aided Design (CAD) systems have made it possible for automatic classification of skin lesions based on their scans. These systems utilize computer algorithms to analyze and interpret the characteristics of skin lesions, allowing for automated classification. Pigmented skin diseases exhibit both similarities and differences between classes, but their early- and late-stage pathological characteristics are distinct [5]. Although there is small variation at initial manifestations of benign and malignant disease, once lesions form, melanoma (malignant) and melanoma nevus (benign) differ greatly in appearance. Because they share similar clinical manifestations and are both melanocyte-derived tumours, classifying them is very difficult. Initial and subsequent indications of seven skin injuries. The seven skin lesions in the principal segment are in their beginning phases, while the excess side effects are irrelevant. The beginning-phase lesions are basically the same in variety, size, and strange appearance, yet the late-stage sores are totally different. Hence, it is not possible to determine, which makes classification difficult.

Skin cancer is detected in different ways by doctors. An experienced dermatologist typically follows a series of benchmarks, which typically include a biopsy, dermoscopy, and naked-eye detection of suspicious tumours. The accuracy of dermoscopic image detection has improved by 50%, ranging from 75% to 84%, indicating that it may take a significant amount of time to achieve precise results and allowing the condition to individually progress to a later step. Furthermore, accurate diagnosis is one-of-a-kind and heavily reliant on the expertise of the clinician. Patients find it extremely challenging to manually identify skin conditions. Because medical professionals can use computer-assisted diagnosis to analyzed ermoscopy procedures, when they lack diagnostic expertise, they are unable to find a professional. Diminish inter- and intra-variability by using a computer-based classification. The latest computer-assisted dermatological image categorization systems face two significant challenges. Firstly, the imaging procedure itself posed difficulties as skin images were captured using dermoscopy, which differed from other medical imaging techniques such as histology and biopsy. This led to a lack of information and posed a challenge for integrating the imaging systems. Secondly, the advanced methods for classifying skin images required extensive pre-processing, segmentation, and feature extraction processes.

II. LITERATURE SURVEY

Sharma, D. K., Hota, H. S., et al. [13] employed data mining techniques such as Artificial Neural Network (ANN) and SVM to classify various erythema squamous diseases. They combined these two technologies using a secret weighted voting scheme, which results in highly accurate predictions. Throughout the training phase, their model achieved an impressive accuracy of 99.25%. Subsequently, during the testing phase, the accuracy slightly decreased to 98.99%.

Deepanker, W., Rambhajani, M., and Pathak, N. et al. [12]. The erythemato-squamous disease dataset was identified by using Bayesian classification. By employing Bayesian technology and the Best First Search feature selection method, the researchers achieved an impressive accuracy of 99.31%. They successfully identified and removed 20 features from the dataset collection available in the University of California, Irvine repository.

E.Tuba, I.Ribic, R.Capor-Hrosik, M.Tuba, et al.[9] an optimized SVM-based automatic erythemato-squamous disease detection method. The multitude advancement calculation with elephant crowding improvement was used to adjust the SVM's boundaries. They conducted their testing on a standard dataset consisting of 366 patients diagnosed with one of the six erythematosquamous diseases. Their approach achieved an effective accuracy of 99.07 percent on this dataset. C.Tao, X.Zhang, S.Wang, J.Liu, &X.Tao, et.al [7] used a hybrid strategy that combined SVM and Granular Computing (GrC). The authors investigated, evaluated, and arranged the characterization results of previous man-made reasoning frameworks for determining erythematosquamous infection. The average sensitivity and specificity of the findings were 99.71 percent and 98.43 percent, respectively.

S.Ioffe, C.Szegedy, V.Vanhoucke, J.Shlens, Z.Wojna, et. al. [11] employed Convolutional Neural Network (CNN) features extracted from both layers to classify lesions. The network could explore pathological regions in greater depth if the various convolution layers learned distinct weights in response to distinct receptive fields. Multiple convolution layers were used to build the GoogLeNet receptive field module, each with

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 22 August 2023 **Revised:** 10 October 2023 **Accepted:** 26 October 2023*

its own kernel sampled, at the center. Multireceptive fields are employed at different spatial positions to enhance the focus on the object and its background. This helps to get features of high quality and makes features easier.

M. Hasan, R. A. Hridhee, H. Sabir, I. Rabiul, et al. [3] described weighted learning system for lesions classification. This model utilized a weighted average approach by combining the prediction models: ResNeXt, SeResNeXt, ResNet, Xception, and Dense-Net. The experimental results demonstrated that this ensemble classified combination achieved an accuracy of 88%.

Nawaz, T. Nazir, M. Masood et al., [1] proposed a UNET model based on DenseNet77. DenseNet77 is used in UNET's encode to evaluate huge typical images for feature sets. Key points for computations are separated by the decoder model.

M.Attia, M.Hossny, S.Nahavandi, A.Yazdabadi, et.al. [10] conducted research using a Google Inception- Inception-v3 Convolutional Neural Network (V3CNN) model that had already been trained to classify skin cancer in a better way. The study utilized a dataset comprising 3374 dermoscopic images that were carefully selected from a larger collection of 129,450 clinical skin cancer images. When the proposed model was applied to classify skin cancer using the 2016 challenge dataset and the results showed an accuracy of 72.1 +/- 0.9. For the specific classification of Mel skin cancer, a CNN with over 50 layers was developed in 2016 and achieved the highest accuracy of 85.5% in the competition.

S.S.Han, M.S.Kim, W.Lim, G.H.Park, I.Park, S.E.Chang, et al. [8] presented a skin lesion classification system that integrated a Deep Convolutional Neural Network (DCNN) and an Optimized Color Feature (OCF) to facilitate lesion segmentation. To enhance the quality of the images and improve lesion contrast, a hybrid strategy involving artifact removal techniques was employed. After that, OCFs, a technique for separating colours, were available. By combining an existing saliency strategy with a novel pixel-based method, the OCF approach demonstrated a notable enhancement. The proposed models exhibited excellent performance, achieving accuracy rates of 92.1%, 96.5%, and 85.1% on the three datasets, respectively.

Santos et al. [2] presented a hybrid approach for integrating binary images generated by a 16-layer Convolutional Neural Network (CNN). The objective of this method was to enhance high-dimensional data and the saliency segmentation of network models. The approach involved utilizing the High-Dimensional Cosine Transform (HDCT) to extract valuable information from the binary images. The Red, green, and Blue (RGB) lesion image with segments was obtained through the maximal mutual information approach. In the classification module, a pre-trained DenseNet201 model was utilized and further retrained through transfer learning using the segmented lesion images.

Stolz, W. Riemann, A. Cognetta, et al. [20] proposed a methodological approach for classifying dermoscopy images. In their approach, the dermoscopy image was partitioned into multiple clinically relevant regions using a segmentation technique based on the Euclidean distance transform. This segmentation method facilitated the extraction of color and texture features from each region, which were subsequently employed for classification purposes. The most common method for determining melanoma's physical characteristics is the Asymmetry, Border irregularity, Color variegation, and diameter >6 mm (ABCD) skin cancer rule. Because of its simplicity and effectiveness, the majority of computer-assisted diagnosis systems use the ABCD rule to classify melanomas. For clinical observation, skin lesions are identified with significant physical characteristics of melanoma.

Chang, C. L., Chen, C. H., et al. [14] discussed how the best dermatology predictive model could be designed by utilizing a decision tree and neural network classification techniques. Six common skin conditions were predicted and analysed by the learning process. Among all the classification methods utilized, the neural network model demonstrates the highest accuracy in disease prediction, achieving a remarkable accuracy rate of 92.62 percent.

III. DERMATOLOGICAL DETECTION AND CLASSIFICATION USING DATA MINING TECHNIQUES

In this section, a block diagram of Dermatological Detection and classification using data mining techniques is described. The dataset was categorized by skin lesions in this analysis. The dataset used in this paper consists of 11,788 dermoscopic images, encompassing diverse populations and varying conditions. The original dataset includes a wide range of pigmented lesions with different characteristics.

Here are 2000 images from both benign and melanoma categories. Utilizing refined data analysis tools, The objective of data mining is to extract meaningful patterns and relationships from large datasets. To achieve this goal, data mining tools utilize various techniques, such as machine learning algorithms, mathematical approaches like neural networks or decision trees, and statistical models. Prediction and analysis are incorporated into data mining. Data and metadata that are important and relevant can be obtained using this method. Data can be categorized using this data mining technique.

In order to segment the input images, the proposed method utilizes an enhanced skin lesion image that employs frequent thresholding as well as screening operations. The initial step involves applying the standard Otsu method to automatically threshold the R, G, and B channels of the input image. This process generates the structure of a binary mask, which is detected in each colour channel, as well as a preliminary classified injury mark. To further improve the segmentation quality and achieve the three-channel masking approach, a logic operation is implemented for each channel. This technique improves the accuracy and reliability of the process.

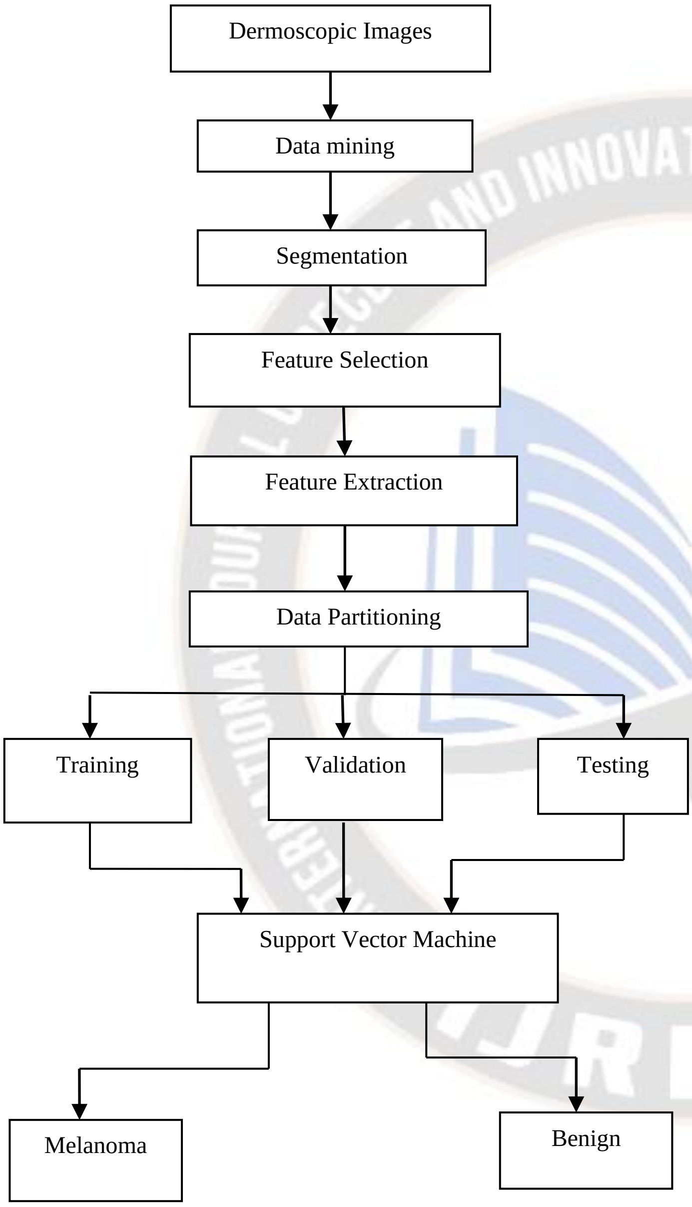

Fig.1: Block Diagram of Dermatological Detection and Classification Using Data Mining Techniques.

To address the issue of over segmentation, which can lead to an image containing numerous small blobs that are not genuine injuries, a common approach is to apply morphological area opening. This step helps in selecting markers that prevent further over segmentation and refine the segmentation results. Furthermore, In order to obtain the final classified area comprising solely lesions, a smoothing process can be employed. This helps to gradually reduce the filter size and enhance the clarity of the segmented lesions. This involves gradually decreasing the filter sizes to achieve a more refined segmentation result.

Feature selection is an automated process that aims to identify and select the most relevant features from a given dataset that significantly contribute to the prediction of disease. Numerous models, particularly those that are based on linear and logistic regression calculations, can decrease from inaccurate results if the dataset contains irrelevant attributes. The skin disease dataset underwent feature selection utilizing a range of techniques, including Univariate Selection, Recursive Feature Elimination, Principal Component Analysis, and so on.

Namely cells and blocks, which helps in understanding the underlying features. For each HOG feature, each block has 2 x 2 cells, or 16 x 16 pixels, and the typical cell size is 8 x 8. The default behavior of the HOG detector leads to overlapping HOG blocks, resulting in multiple contributions from each cell. In data mining, data partitioning is the division of all available data into two or three sets that do not overlap: the test set, the validation set, and the training set all at once. When the data set is extremely large, only a small portion is frequently chosen for the partitions. The three parts of the International Skin Imaging Collaboration (ISIC) 2020 dataset are as follows: testing, validation, and training. The training set was used to train the MobileNetV2 model. The dataset was partitioned into three subsets for different purposes: 70% for training the model, 15% for testing its performance, and 15% for validation purposes.

For training the MobileNetV2 model, the ISIC-2020 dataset was utilized, comprising 16,350 images for training, 3,500 images for validation, and an additional 3,500 images for testing. Using the feature vectors that had been extracted in the previous stages, the skin lesion images were classified using the SVM classification method. SVM was first suggested for binary problems. SVM develops a hyperplane that performs partitioning between parallel classes using the maximum margin principle. Support vectors, or data points close to the decision boundary, are all that are used to create the maximum margin. The most discriminative information about two classes is provided by these support vectors. SVM effectively handles complex classification tasks by converting input feature data to higher dimensional space using kernel trick nonlinear functions.

IV. RESULT ANALYSIS

In this section, performance analysis of Dermatological Detection and classification using data mining techniques is observed in terms of accuracy, precision, and Specificity. The performance evaluation of the classification is based on

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 22 August 2023 **Revised:** 10 October 2023 **Accepted:** 26 October 2023*

categorizing instances into True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) categories, which are defined as follows:

True Positive (TP): If a sample is correctly predicted as positive, it falls into the True Positive (TP) category.

True Negative (TN): If a sample is exactly predicted as negative, it is referred to as a True Negative (TN).

False Positive (FP): when a sample, despite being incorrectly predicted to be negative, then it is positive.

False Negative (FN): if a sample is actually positive despite being incorrectly predicted to be positive.

The performance of the provided system has various metrics which employed to assess its effectiveness. These metrics include sensitivity, precision and accuracy.

Accuracy: Accuracy is a performance parameter that shows the system's ability to make accurate predictions.

Accuracy=$\frac{TP+TN}{TP+FP+TN+FN}$x 100 ----- (1)

Precision: Precision is a metric that evaluates the system's ability to generate relevant and useful outcomes.

Precision=$\frac{TP}{(TP+FP)}$ x 100 ------ (2)

Sensitivity: True Positive Rate (TPR), also known as recall or sensitivity and is a performance parameter that assesses the positive prediction capacity of a system.

Sensitivity=$\frac{TP}{TP+FN}$ x 100 ---- (3)

Table 1 shows an evaluation of Dermatological Detection and classification using data mining techniques for different ML classifiers

Table.1: Performance Analysis

Parameters	Accuracy	Sensitivity	Precision
Decision Tree (DT)	75.8	79.8	83.5
Logistic Regression (LR)	82.9	86.1	87.9
K-Nearest Neighbors (KNN)	84	80.3	81.9
Support Vector Machine (SVM)	97.2	94.3	95.2

Accuracy

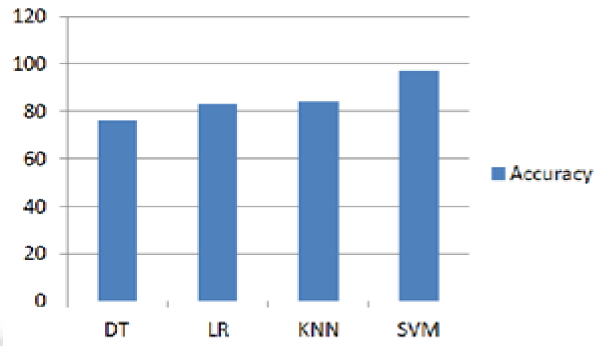

Fig. 2: Comparative Graph of The Accuracy

The accuracy comparison graph for DT, LR, KNN, and SVM is shown in Fig. 2. The SVM shows higher accuracy.

Sensitivity

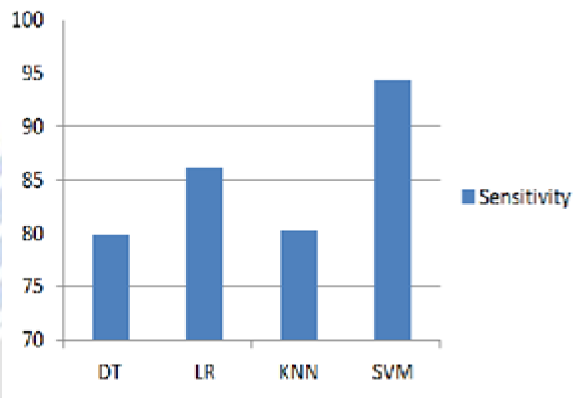

Fig.3: Comparative Graph of The Sensitivity

Figure.3, shows a comparison graph of the DT, LR, and KNN's sensitivity to skin disease prediction.

Precision

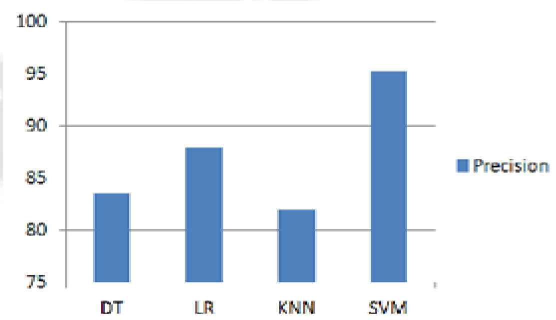

Fig. 4: Comparative Graph of The Precision

Comparison graph of DT, LR, KNN, and the SVM for prediction of skin disease is shown in Fig. 4.

V. CONCLUSION

Dermoscopic image detection of skin cancer is discussed in this analysis. The method extracts the region of interest from the image by employing data mining techniques for image segmentation. Includes, extraction was performed over a range of element descriptors. Utilizing optimized HOG-based descriptors of skin lesions, an efficient SVM model for melanoma detection has been demonstrated. The presented framework outperforms two recently developed cutting-edge alternatives, according to extensive evaluations of a large dataset of dermoscopic images. The feature's performance was evaluated using different classification methods. When compared to the other types of classifiers, the SVM approach achieves the highest performance with the highest accuracy.

REFERENCES

[1] M. Nawaz, T. Nazir, M. Masood., “Melanoma segmentation: a framework of improved DenseNet77 and UNET convolutional neural network,” International Journal of Imaging Systems and Technology, vol. 32, no. 6, pp. 2137–2153, 2022.

[2] Santos, D. Melanoma Skin Cancer Detection using Deep Learning. medRxiv 2022, doi https://doi.org/10.1101/2022.02.02.22269505

[3] R. Zillur, H. Sabir, I. Rabiul, M. M. Hasan, and R. A. Hridhee, “An approach for multiclass skin lesion classification based on ensemble learning,” Informatics in Medicine Unlocked, vol. 25, article 100659, 2021.

[4] Ashraf, R.; Afzal, S.; Rehman, A.U.; Gul, S.; Baber, J.; Bakhtyar, M.; Mehmood, I.; Song, O.Y.; Maqsood, M. Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection. IEEE Access 2020, 8, 147858–147871.

[5] I. Bakkouri and K. Afde, “Correction to: Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images,” Multimed Tools and Applications, vol. 79, no. 29-30, pp. 20519–20519, 2020.

[6] Albahar, M.A. Skin Lesion Classification Using Convolutional Neural Network with Novel Regularizer. IEEE Access 2019, 7, 38306–38313.

[7] C.Tao, X.Zhang, S.Wang, J.Liu, & X.Tao,. (2018). Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. BMC medical informatics and decision making, 18(2), 59.

[8] S.S.Han, M.S.Kim, W.Lim, G.H.Park, I.Park, S.E.Chang Classifica- tion of the clinical images for benign and malignant cutaneous tumors us- ing a deep learning algorithm. J. Investig. Dermatol. 2018, 138, 1529–1538.

[9] E.Tuba, I.Ribic, R.Capor-Hrosik, M.Tuba (2017). Support vector machine optimized by elephant herding algorithm for erythemato-squamous diseases detection. Procedia computer science, 122, 916- 923.

[10] M.Attia, M.Hossny, S.Nahavandi, A.Yazdabadi, Skin melanoma segmentation using recurrent and convolutional neural networks. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017; pp. 292–296.

[11] S.Ioffe, C.Szegedy, V.Vanhoucke, J.Shlens, Z.Wojna, “Rethinking the inception architecture for computer vision,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826, Las Vegas, NV, USA, 2016.

[12] Rambhajani, M., Deepanker, W., & Pathak, N. (2015). Classification of Dermatology Diseases through Bayes net and Best First Search. International Journal of Advanced Research in Computer and Communication Engineering, 4(5).

[13] Sharma, D. K., & Hota, H. S. (2013). Data mining techniques for prediction of different categories of dermatology diseases. Journal of Management Information and Decision Sciences, 16(2), 103.

[14] Chang, C. L., & Chen, C. H. (2009). Applying decision tree and neural network to increase quality of dermatologic diagnosis. Expert Systems with Applications, 36(2), 4035-4041.

[15] Vestergaard, M.; Macaskill, P.; Holt, P.; Menzies, S. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting. Br. J. Dermatol. 2008, 159, 669–676.

[16] Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. IEEE Conf. Comput. Vis. Pattern Recognit. 2005, 1, 886–893.

[17] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in European conference on computer vision, 2004, pp. 469-481: Springer.

[18] Argenziano, G.; Soyer, H.P.; Chimenti, S.; Talamini, R.; Corona, R.; Sera, F.; Binder, M.; Cerroni, L.; De Rosa, G.; Ferrara, G.; Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet. J. Acad. Dermatol. 2003, 48, 679–693.

[19] Menzies, S.W. A Method for the diagnosis of primary cutaneous melanoma using surface microscopy. Dermatol. Clin. 2001, 19, 299–305

[20] Stolz, W. Riemann, A, Cognetta, A. ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma. Eur. J. Dermatol. 1994, 4, 521–527.