Received: November 26, 2023. Revised: January 20, 2024.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Convolutional Neural Network with Coordinate Attention Module for the Detection of Skin Cancer

Usha Thirugnanam[1]* Nalini Joseph[1] Umarani Srikanth[2] R. Anand[3]

[1]*Department of Computer Science and Engineering,*
Bharath Institute of Higher Education and Research, Chennai, India
[2]*Department of Computer Science and Engineering, B Panimalar Engineering College, Chennai, India*
[2]*Department of Computer Science and Engineering, Prathyusha Engineering College, Thiruvallur, India*
* Corresponding author's Email: t.ushausha@yahoo.com

Abstract: Skin cancer is one of the most common types of cancer globally, by increasing occurrence rates each year. It is a predominant kind of cancer which rises from the uncontrolled growth of abnormal skin cells due to genetic mutations including various factors such as UV radiation, genetics and other factors. The death rate is decreased when skin cancer is detected at early stages. Therefore, this paper proposed a Convolutional Neural Network (CNN) with Coordinate Attention Module (CAM) for early detection of skin cancer. The data augmentation is utilized in this experiment for data pre-processing and fed into the Gray Level Co-occurrence Matrix (GLCM) based feature extraction technique. The Harris Hawk Optimization (HHO) is utilized for selecting features that have faster convergence and strong capability in local optima. The selected features are given as input to CNN with the CAM approach. This model is estimated on ISIC-2019 and ISIC-2020 datasets and attains better results using accuracy, precision, recall, specificity, and F1-score. The obtained result shows that the proposed model achieves better accuracy of 98.77% on the ISIC-2019 dataset and 99.51% on the ISIC-2020 dataset which ensures accurate detection compared to other existing methods like Inception-ResNet and Residual Deep Convolutional Neural Network (RDCNN).

Keywords: Convolutional neural network, Coordinate attention module, Dermoscopy images, Harris hawk optimization, Skin cancer detection.

1. Introduction

The skin is a major tissue of the human body that covers the whole body and the thickness varies throughout all parts and differs among men and women [1]. Skin cancer is caused by the uncontrollable growth of abnormal skin cells and UV radiation from the sun and artificial sources like tanning beds that result in malignant tumors [2]. The major kind of cancer are Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC) and Melanoma [3]. Melanoma is a most aggressive posing a significant threat to a patient's life [4]. The SCC arises from the squamous cells that are thin and flat found in the outer layer of the skin. The BCC is typically originating in the basal cells that are found in the deepest layer of the skin epidermis [5]. Early and accurate detection is essential to significantly enhance the chances of successful treatment for cancer patients minimizing the death rate [6]. However, manual detection is subjective and can vary based on the expertise of the healthcare professionals potentially leading to misdiagnosis [7]. In the past few years, Machine Learning (ML) and Deep Learning (DL) have gained the most accurate and effective for detecting skin cancers at early stages [8].

The dermoscopy is a high-resolution image captured through specialized equipment that allows dermatologists to visualize subsurface skin structures which shows a prominent role in early and accurate detection of skin cancer [9]. Numerous technologies and advancements are there, particularly in computer vision and image analysis enabling the development of automated systems that can assist dermatologists in identifying skin cancer [10, 11]. The DL based

techniques are utilized to reduce human interaction for predicting data due to the development in programming and technology. The recognition and localization of skin cancer are required to compute the image feature for detecting cancer [12]. In which the CNN shows remarkable performance for processing and analyzing medical images [13]. This technique enhances the accessibility in underserved areas which ensures timely assessment and consultation [14]. Additionally, an Optical Coherence Tomography (OCT) is explored for skin cancer detection which provides exhaustive cross-sectional images of skin and helps to distinguish between malignant and benign cancers [15]. In this research, CNN with CAM is proposed for skin cancer detection. The benefits of CNN are it can learn and extract features automatically from training images, however, in traditional method requires a manual extraction of features from images. The CAM is an attention module that is inserted into deep neural networks. It can support maximizing extraction of feature capacity of convolutional neural networks without updating many computational overheads. It utilizes an effective model to acquire channel and position of data for the best extraction of features. The integration of CNN with CAM is utilized to improve the model's capability to learn spatial relationships and focus on relevant features in an image.

- The preprocessing is done by using a data augmentation technique which improves the model performance and feeds to the feature extraction.
- The GLCM is used for feature extraction and HHO is utilized for feature selection which has faster convergence and strong capability in local optima.
- The CNN with CAM is proposed for skin cancer detection which is utilized to improve model's capability to learn spatial relationships and focus on relevant features in an image.

The rest part of research is follows: Section 2 defines literature review. Section 3 defines details of proposed methodology. Section 4 defines results and discussion and Section 5 defines conclusion and lastly this paper finish with the references.

2. Literature review

Singh [16] introduced an Inception-ResNet technique for the classification of Skin Cancer. This paper introduced a Convolutional Neural Network (CNN) with eight thick layers for classifying melanoma cancer. The Inception and Residual blocks are utilized for extracting global and local features. This classifier has a 40-layer depth enables it to obtain the maximum value of accuracy. This model utilized a bottleneck SoftMax layer with ResNet and Inception blocks that assist in training the parameter efficiently. However, this model is not adapted to handle the long-term dependencies and sequential information in input data.

Alsahafi [17] developed a deep residual network for classifying skin cancer through multi-layer feature extraction and cross-channel correlation with detection. This developed model utilized a Residual Deep Convolutional Neural Network (RDCNN) for classifying skin cancer. Additionally, it employed various conv filters for multilevel feature extraction and cross-channel correlation through gliding dot product filters. This model tackles the data imbalance issues by converting the dataset from label and image into a vector of weight and image. However, this model was not lightweight and had a high running time, so it could not run on limited memory and microdevices.

Naeem [18] presented a DL-based approach for multi-classification of skin cancer through dermoscopy images. The developed model named SCDNet integrates VGG-16 with CNN for classifying various types of skin cancer. This model was trained, tested, and validated in the ratio of 70:20:10 and is composed of the ISIC-2019 dataset. The grid search technique was utilized in various hyperparameters of SCDNet were finetuned to attain optimum performance. However, this model cannot use dark-skin people image datasets for skin cancer diagnosis.

Tahir [19] implemented a deep learning model for skin cancer diagnosis using multi-classification techniques. The developed model can extract dominant features from dermoscopy images which helps to identify the accurate disease. To overcome the issues of class imbalance in clinical data, the up-sampling and SMOTE Tomek techniques were utilized which attain concoction image samples at every class to enhance the accuracy. This model minimized the training parameters which reduced the model complexity. However, this model has a limited lightweight network selection and hyperparameters for evaluation.

Jaisakthi [20] developed a deep neural network for classification of skin cancer from dermoscopy images. The EfficientNet based on transfer learning was developed that automatically defines hyperparameters and learns more difficult and fine-grained patterns from dermoscopy images. Additionally, the ensemble technique was developed by integrating image features and metadata. The

ranger optimizer was employed to enhance the EfficientNet effectiveness which minimize the hyperparameter tuning values. However, it has less computational complexity and it attained high relevancy of features and less redundancy.

Adla [21] suggested a Full-resolution Convolutional Network with a Dynamic Graph Cut Algorithm (FrCN-DGCA) for skin cancer detection and classification. The action bundle was utilized as a hyperparameter through the segmentation initiator of skin image that was based on DGCA. This model helps to address the over and under-segmentation problems and wrongly segmented small regions. However, it cannot be utilized for over compressed images and also the comparison is limited to the classification stage.

Imran [22] introduced a convolution-based deep neural network for detecting skin cancer through the ISIC dataset. It integrates an ensemble learning technique such as VGG, ResNet, and CapsNet which exploits learner's diversity to yield a better decision. The developed model performs better than individual learners with various quality measures like accuracy, recall, f1-score, precision, and sensitivity. This model was utilized to focus on identifying features in skin mages that affect the decision of the network. However, this model is incapable of handling sequential information and long-term dependencies which affects the model's performance.

Qureshi and Roos [23] implemented transfer learning deep neural network ensembles for detecting skin cancers through imbalanced datasets. The ensemble-based CNN architecture was developed which contains few pre-trained models and few are trained on data only. This developed model handles imbalanced and limited data which enhances the capability of the model. However, it required high time to detect the cancer from the noisy images.

3. Proposed methodology

In this section, the CNN with Coordinate Attention Mechanism is proposed for effective detection of skin cancer. The ISIC-2019 and ISIC-2020 datasets are utilized for proposed method. The preprocessing is done by using a data augmentation technique which improves the model performance and fed to the feature extraction. The GLCM is used for feature extraction which is a numerical method that takes structural relationships among pixels. The extracted features are selected through Harris Hawk Optimization (HHO). Then the selected features are classified through CNN with the Coordinate Attention Module (CAM). Integrating CNN with CAM is utilized to improve the model's capability to learn spatial relationships and focus on relevant features in an image. Fig. 1 represents the block diagram of the proposed methodology.

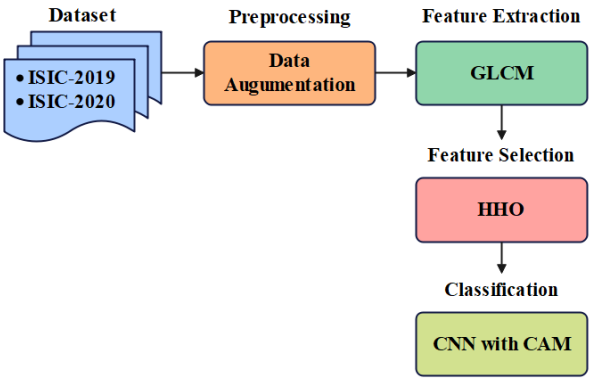

Figure. 1 Block diagram of proposed methodology

3.1 Dataset

The datasets utilized for skin cancer detection are ISIC-2019 and ISIC-2020 dataset. The ISIC-2019 dataset comprises a huge number of dermoscopy skin lesion images which are collected through numerous sources. It comprises 4522 Melanoma, 12,875 Melanocytic Nevi, 3323 Basal cell carcinoma, and 2624 Benign keratosis images. This dataset is split into 70:20:10 for training, testing, and validation tests. The ISIC-2020 dataset was derived from a variety of sources because numerous institutes contributed patient information at different ages. It comprises 33,126 dermoscopies, 579 malignant, 32,542 benign skin lesions. These two datasets are preprocessed by utilizing the data augmentation technique.

3.2 Preprocessing

Data augmentation is a technique that artificially increases the size of dataset images by making small changes or modifications in existing data. Various modifications are utilized to maximize the size of the dataset. By utilizing this technique, chances of over-fitting are reduced and the performance of the model is maximized. This model performed a range rotation, translation of height and weight, range of brightness and vertical and horizontal flips. After the data augmentation size of the dataset is increased.

3.3 Feature extraction

After preprocessing, the features are extracted through GLCM which is a numerical method that takes structural relationships among pixels. This GLCM feature is considered according to the dimension in the Region of Interest (ROI) with gray

levels in square matrix. GLCM attained 22 text-based features like entropy, contrast, correlation, energy, homogeneity, sum variance, autocorrelation, maximum probability, dissimilarity, inverse different movement (IDM), IDM normalization, and more. Among them some features are given in Eqs. (1, 2) and (3) where, $p(i,j)$ is the element of the normalized GLCM matrix and N is the number of gray levels.

$$Contrast = \sum_i \sum_j |i-j|^2 p(i,j) \quad (1)$$

The contrast is a pixel intensity and the adjacent throughout an image. Moreover, it also considers local variations established in an image. This is measured through variance in intensity and colour of each object with similar region.

$$Entropy = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \log\big(p(i,j)\big) \quad (2)$$

The entropy estimates a degree of pixels and randomness in an image and it is utilized to characterize the texture of an image. It is inversely correlated to energy when a huge grey level entropy is large.

$$Energy = \sum_{i,j} p(i,j)^2 \quad (3)$$

The energy computes a pixel pair in degrees. It is an image disorder estimation and an equivalent pixel energy value is high. After feature extraction, the extracted features are selected using an optimization process.

3.4 Feature selection

In this section, the extracted features are selected through HHO algorithm which is a nature-inspired optimization algorithm according to the hunting behavior of Harris Hawk. The HHO contains three phases exploration, transition from exploration to exploitation and exploitation. These three phases are explained in the following sections.

3.4.1. Exploration

In the exploration stage of HHO, the search agent searches different locations to explore their best prey. At every instant of tth time, the best HH is handled as a predictable target. It is based on either perch based or prey location that randomly perch on some positions in the conditions of $q < 0.5$ and $q \geq 0.5$ correspondingly [24-25]. The location of HH at the next instant $t+1$th time depends on two perching approach which is shown in Eq. (4).

$$X^{t+1} = \begin{cases} X^t_{rand} - r_1 |X^t_{rand} - 2r_2 X^t| & q \geq 0.5 \\ \left(X^t_{prey} - X^t_m\right) - r_3\left(LB + r_4(UB - LB)\right) & q < 0.5 \end{cases} \quad (5)$$

Where, X^{t+1} is the hawk position at $t+1$th time, X^t_{rand} is the randomly selected position in the population, X^t_{prey} is the prey position, X^t_m is the present position average of every hawk in the population, r_1, r_2, r_3 and r_4 and q are random numbers within a range of $[0,1]$, UB and LB are upper and lower bound variables.

3.4.2. Exploration to exploitation transition

The transitions depend on prey's escaping energy (E) that is measured through Eq. (5).

$$E = 2E_0\left(1 - \frac{t}{T}\right) \quad (5)$$

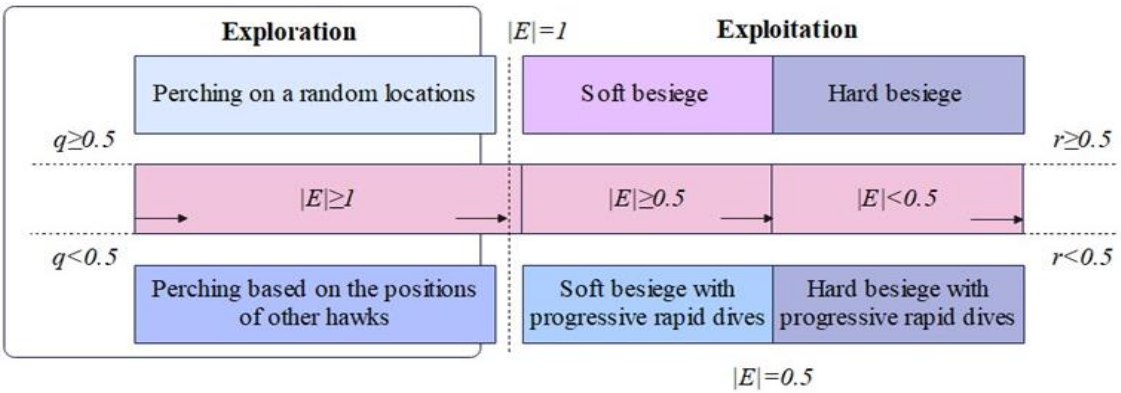

Figure. 2 Six phases of HHO

Where, E_0 is the initial prey energy which is fixed randomly within a range of $[-1,1]$ at start of every iteration, t and T are present and the highest iterations correspondingly. When, $|E| \geq 1$, the exploration takes place that usually occurs at initial iterations. When, $|E| < 1$, the exploration takes place that usually occurs at further iterations, hence preserving the balance between exploration and exploitations. Fig. 2 presents the six phases of HHO.

3.4.3. Exploitation

In the exploitation stage of HHO, the search agent maintains to exploit the solutions around obtained optimal solutions. Based on hawks' various hunting approaches, the escaping prey energy, patterns, and various exploitation methods like soft or hard besiege are utilized. The prey escape changes are calculated through parameter r , if $r < 0.5$ means there are changes that the prey will escape effectively not possible. Thus, based on E and r values, four exploitations are exhibited which are presented as follows.

A. Soft Besiege

In soft besiege, where $r \geq 0.5$ and $|E| \geq 1$ is the left prey with appropriate energy, during the time, it tries to escape by considering some random hops. During this situation, the hawks encircle preys delicately and tries to finish energy at last it creates a pounce. The Hawks' positions are estimated through Eqs. (6) and (7).

$$X^{t+1} = \Delta X^t - E\left|JX_{prey}^t - X^t\right| \tag{6}$$

$$\Delta X^t = X_{prey}^t - X^t \tag{7}$$

Here, ΔX^t is a distance between hawks' location and prey at tth time, $J = 2 \times (1 - r_5)$ is a random hop created through prey at the escaping procedure $r_5 \in (0,1)$.

B. Hard Besieg

In hard besiege, the $r \geq 0.5$ and $|E| < 0.5$ prey is exhausted that drop by small escape energy. During this time, the hawks' positions are updated through Eq. (8),

$$\Delta X^{t+1} = X_{prey}^t - E|\Delta X^t| \tag{8}$$

C. Soft besiege with progressive rapid dives

Here, $r < 0.5$ and $|E| \geq 0.5$ prey has appropriate energy to escape successfully that is finished before the pounce. Where, the further hawk's movement can be decided on the basic rules which is formulated in Eq. (9),

$$Y = X_{prey}^t - E\left|JX_{prey}^t - X^t\right| \tag{9}$$

This may perform asymmetrical, unpredicted, and fast dives depending on levy flight patterns that are shown in Eq. (10),

$$Z = Y + S \times LF(D) \tag{10}$$

Here, D is the problem dimensions, S is the random vector dimensions and the LF is attained through Eq. (11),

$$Lf(x) = 0.01 \times \frac{u \times \sigma}{|v|^{\frac{1}{\beta}}},$$
$$\sigma = \left(\frac{\Gamma(1+\beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}}\right. \tag{11}$$

Where, u and v are random numbers in the range of $(0,1)$ and β is the constant that value is fixed as 1.5. Lastly, hawks' positions are updated through Eq. (12),

$$X^{t+1} = \begin{cases} Y\ F(Y) < F(X^t) \\ Z\ F(Z) < F(X^t) \end{cases} \tag{12}$$

D. Hard besiege with progressive rapid dives

Here, $r < 0.5$ and $|E| < 0.5$ prey has an inappropriate energy that is utilized before the pounce. The hawks' positions are updated through Eq. (13),

$$X^{t+1} = \begin{cases} Y\ F(Y) < F(X^t) \\ Z\ F(Z) < F(X^t) \end{cases} \tag{13}$$

Here, Y and Z are measured through Eqs. (14) and (15),

$$Y = X_{prey}^t - E\left|JX_{prey}^t - X^t\right| \tag{14}$$

$$Z = Y + S \times LF(D) \tag{15}$$

Where, X_{prey}^t is the prey position, X_m^t is the present position average of every hawk in population, D is the problem dimensions, S is the random vector dimension. The selected features are given as input to the CNN based CAM.

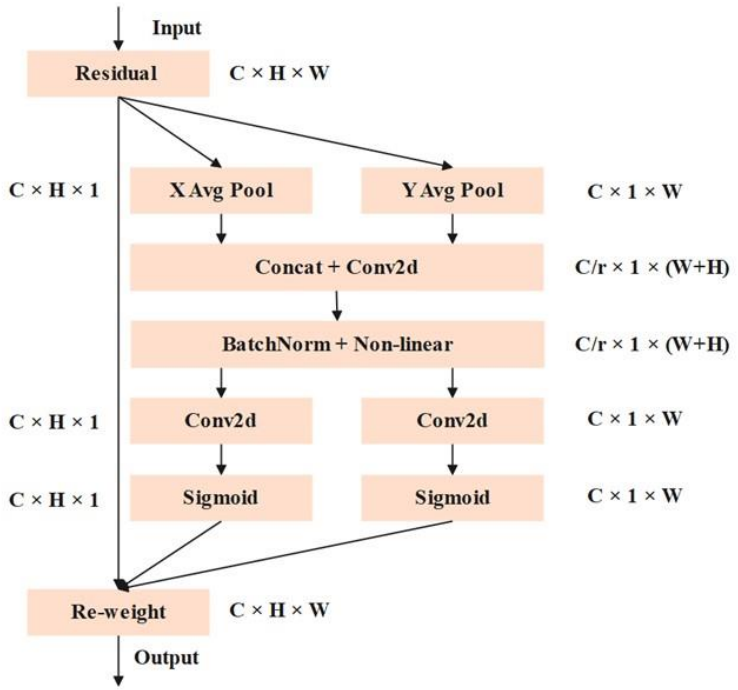

Figure. 3 Structure of Coordinate Attention Module

3.5 Convolutional Neural Network for Classification

CNN is utilized for classification which is a relatable algorithm for the identification of huge amounts of data. The benefits of CNN are it can learn and extract features automatically from training images, however, in traditional method requires a manual extraction of features from images. CNN has three various layers Convolutional (Conv), Pooling, and Fully Connected (FC) layer. The conv layer is the first and major critical element of CNN that extracts the features from the input image and it has a small array of numbers known as kernels applied over the input that generates the result known as feature maps. Various convolutional kernels are utilized to extract various features. Many conv layers are present that depend on the size of input images. Next, the pooling layer is processed which is responsible for minimizing the convolutional feature map dimension. The pooling layer performs down-sampling operations by minimizing the feature map dimension that helps in minimizing the needed complexity of computation to data process. There are various pooling operations like max, min and average pooling. The resulting feature maps of conv or pooling layer are changed to a 1D vector where each input is merged with every output by weight and this layer is known as the dense layer. Lastly, there are one or more FC layers and the final FC layer consists of a similar number of outputs as number of classes.

3.5.1. Coordinate Attention Module

The attention mechanism concentrates on certain relevant elements and ignores other elements in deep neural networks. The CAM is an effective attention module that is simply inserted into deep neural networks. It can support maximizing extraction of feature capacity of convolutional neural networks without updating many computational overheads. It utilizes an effective model to acquire channel and position of data for the best extraction of features. Fig. 3 represents the structure of the Coordinate Attention module.

The input of CAM is illustrated as $C \times H \times W$, where C represents the input channel, H and W a feature map height and width. The process of CAM is illustrated below:

- The coordinate Attention module decomposes 2D into two 1D global pooling operations through the directions of perpendicular and horizontal. The result of the 1D pooling operation is $C \times 1 \times W$ and $C \times H \times 1$ correspondingly.
- Results of two 1D pooling are merged and managed to utilize 1×1 conv, batch norm, and non-linear layer. The result is $C/r \times 1 \times (W + H)$, here r is a compression ratio.
- The result of the above stage is divided into two sets ($C \times 1 \times W \; and \; C \times H \times 1$), these are computed to obtain weight through the direction of perpendicular and horizontal.

- At last, the input feature $C \times H \times W$ is re-weighted by utilizing results from the third stage.

Let's consider that, the input of the Coordinate Attention module is a vector which is denoted as Eq. (16),

$$X = [x_1, x_2, \dots, x_c] \in R^{C \times H \times W} \quad (16)$$

The input is computed by two 1D pooling operation through vertical and horizontal paths. The result of channel c at height h is formulated as Eq. (17),

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i \le W} x_c(h, i) \quad (17)$$

The result of channel c at width w is presented in Eq. (18),

$$z_c^w(w) = \frac{1}{H} \sum_{0 \le i \le H} x_c(j, w) \quad (18)$$

The above two results of 1D global pooling operation, utilizing a couple of feature maps. The two feature maps are merged and fed into 1×1 convolutional transformation function F_1 illustrated in Eq. (19),

$$f = \delta\left(F_1\left(z^h, z^w\right)\right) \quad (19)$$

Where, $[.,.]$ is the addition operation with geometric dimension, δ denotes activation function for non-linear, $f \in R^{C/r \times (H+W)}$ presents an in-between feature map that encodes geometric data in the horizontal and vertical direction, r is the ratio of reduction for managing the size of the block. The f divides geometric dimension into two various tensors $f \in R^{C/r \times H}$ and $f \in R^{C/r \times W}$. Other two 1×1 conv transformation function $F_h \; and \; F_w$ are used for divided transform f^h and f^w into tensors, by a similar channel to input, X can be presented in Eqs. (20) and (21),

$$g^h = \delta\left(F_h\left(f^h\right)\right) \quad (20)$$

$$g^w = \delta\left(F_w\left(f^w\right)\right) \quad (21)$$

The results of g^h and g^w in 2D are extended and utilized as attention weights. The result of CAM is represented as Eq. (22),

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (22)$$

Where, y_c is the output of CAM and the weights of g^h and g^w in 2D are fused with the input X to obtain the coordinate attention module output. Integrating CNN with CAM is utilized to improve the model's capability to learn spatial relationships and focus on relevant features in an image.

4. Experimental result

In this paper, the proposed model is stimulated by utilizing a Python environment with the system configuration: RAM:16GB, processor: intel core i7 and operating system: windows 10. The accuracy, precision, recall, specificity and f1-score are utilized to estimate a model performance which are given in Eq. (23), (24), (25), (26) and (27),

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (23)$$

$$Precision = \frac{TP}{TP + FP} \quad (24)$$

$$Recall = \frac{TP}{TP + FN} \quad (25)$$

$$Specificity = \frac{TN}{TN + FP} \quad (26)$$

$$F1 - score = 2 \times \frac{Precision \times recall}{Precision + recall} \quad (27)$$

Where, TP, TN, FP and FN illustrate the True Positive, True Negative, False Positive, and False Negatives respectively.

4.1 Quantitative and qualitative analysis

This section shows the quantitative and qualitative analysis of the proposed CNN with CAM using accuracy, precision, recall, specificity and f1-score are shown in Tables 1 and 2. Table 1 illustrates the performance of the proposed model on ISIC-2019 dataset and Table 2 illustrates the proposed model performance on the ISIC-2020 dataset. The Artificial Neural Network (ANN) requires maximum training data and it is slow to train on large datasets. The Recurrent Neural Network (RNN) is incapable of handling lengthy sentences due to its vanishing gradient problems. The LSTM requires maximum training data and it is slow to train on large datasets. The CNN is not adopting to handle the long-term dependencies and sequential information in input data. Integrating CNN with CAM is utilized to improve the model's capability to learn spatial

relationships and focus on relevant features in an image.

Table 1 and Fig. 4 shows the performance of CNN with CAM on ISIC-2019 dataset. The performance of ANN, RNN, LSTM and CNN are measured and matched with the proposed CNN with CAM model. The obtained result shows that the CNN with CAM model achieves better results by utilizing accuracy, precision, recall, specificity, and f1-score values about 98.77%, 98.54%, 97.61%, 97.48% and 97.36% correspondingly while comparing other classifiers.

Table 1. The performance of CNN with CAM on the ISIC-2019 dataset

Method	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
ANN	91.89	91.42	90.65	90.52	90.19
RNN	93.84	93.56	92.17	92.73	92.61
LSTM	94.65	94.31	93.64	93.22	93.48
CNN	96.81	96.62	95.46	95.39	95.24
CNN with CAM	98.77	98.54	97.61	97.48	97.36

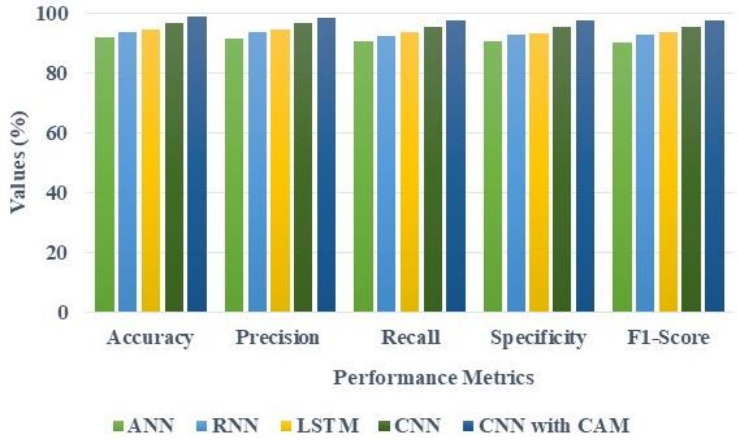

Figure. 4 Performance of CNN with CAM on ISIC-2019 dataset

Table 2. The performance of CNN with CAM on ISIC-2020 dataset

Method	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
ANN	92.77	92.69	92.61	91.72	91.61
RNN	94.86	94.72	93.54	93.31	93.29
LSTM	96.73	96.64	95.69	95.47	95.36
CNN	98.64	98.51	97.83	97.78	97.57
CNN with CAM	99.51	99.48	98.72	98.63	98.25

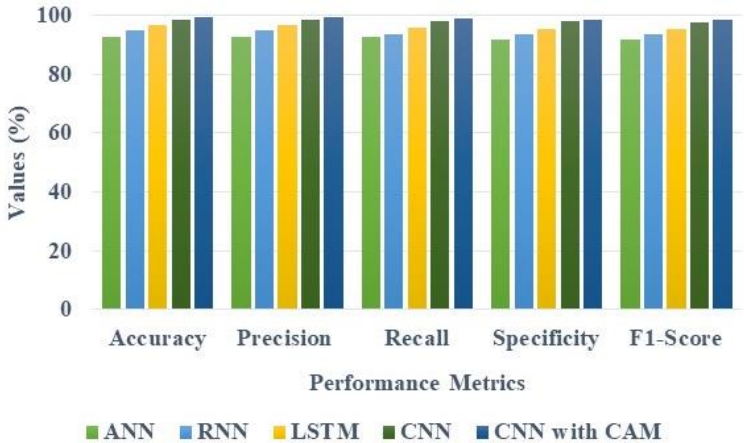

Figure. 5 Performance of CNN with CAM on ISIC-2019 dataset

Table 3. Comparative analysis of proposed model on ISIC-2019 dataset

Method	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
Inception-ResNet [16]	96.13	N/A	94.67	96.64	N/A
RDCNN [17]	94.65	72.56	70.78	96.78	71.33
SCDNet [18]	96.91	92.19	92.18	N/A	92.18
Proposed CNN with CAM	98.77	98.54	97.61	97.48	97.36

Table 4. Comparative analysis of proposed model on ISIC-2020 dataset

Method	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
Inception-ResNet [16]	96.28	N/A	95.21	96.60	N/A
RDCNN [17]	99.05	99.42	96.57	96.57	96.57
DSCC_Net with SMOTE Tomek [19]	94.17	93.76	94.28	N/A	93.93
Proposed CNN with CAM	99.51	99.48	98.72	98.63	98. 25

Table 1 and Fig. 5 shows the performance of CNN with CAM on ISIC-2020 dataset. The performance of ANN, RNN, LSTM and CNN are measured and matched with the proposed CNN with CAM model. The obtained result shows that the CNN with CAM model achieves better results by utilizing accuracy, precision, recall, specificity and f1-score values about 99.51%, 99.48%, 98.72%, 98.63% and 98.25% correspondingly while comparing other classifiers.

4.2 Comparative analysis

This section demonstrates a comparative analysis of the proposed model with accuracy, precision, recall, specificity and f1-score are illustrated in Tables 3 and 4. The existing result such as [16], [17], [18], and [19] are utilized to evaluate the ability of the classifier. The CNN with CAM is trained, tested and validated using ISIC-2019 and 2020 dataset. The result obtained from Table 3 and 4 shows that the proposed model attains better performance when compared with the existing methods. In ISIC-2019 dataset the accuracy was improved to 98.77%, precision of 98.54%, recall of 97.61%, specificity of 97.48% and f1-score of 97.36%. In ISIC-2020 dataset the accuracy was improved to 99.51%, precision of 99.48%, recall of 98.72%, specificity of 98.63% and f1-score of 98.25%.

4.2.1. Discussion

In this section, the advantages of the proposed method and the limitations of existing methods are discussed. The existing method has some limitations, such as the Inception-ResNet [16] model was not adopted to handle long-term dependencies and consecutive information in input data. The RDCNN [17] model was not lightweight and high running time so it cannot run on limited memory and microdevices. The SCDNet [18] model cannot use dark-skin people image datasets for skin cancer diagnosis. The DSCC_Net with SMOTE Tomek [19] model has a limited lightweight network selection and hyperparameters for evaluation. The proposed CNN with CAM model overcomes these existing model limitations. The integration of CNN with CAM is utilized to improve the model's capability to learn spatial relationships and focus on relevant features in an image.

5. Conclusion

Early and accurate detection is essential to significantly enhance the chances of successful treatment for cancer patients which minimizes the death rate. The Convolutional Neural Network (CNN) with Coordinate Attention Module (CAM) is proposed in this research for the effective detection of skin cancer. The ISIC-2019 and ISIC-2020 dataset is utilized in this research for efficient skin cancer detection. The data augmentation is utilized in this experiment for data pre-processing and fed into Gray Level Co-occurrence Matrix (GLCM) based feature extraction technique. The Harris Hawk Optimization (HHO) is utilized for selecting features that have faster convergence and strong capability in local optima. The selected features are given as input to CNN with CAM approach. By integrating CNN with CAM is utilized to improve the model's capability to learn spatial relationships and focus on relevant features in an image. The obtained result shows that the CNN with CAM attains better accuracy of 98.77% on ISIC-2019 dataset and 99.51% on ISIC-2020 dataset which ensures accurate detection compared to other existing methods. In the future, hyperparameter tuning can be applied to improve the classification performance.

Notations

Notation	Description
$p(i,j)$	Element of the normalized GLCM matrix
N	Number of gray levels
X^{t+1}	Hawk position at $t+1$th time
X_{rand}^{t}	Randomly selected position in the population
X_{prey}^{t}	Prey position
X_{m}^{t}	Present position average of every hawk in the population
r_1 , r_2 , r_3 , r_4 and q	Random numbers in the range of [0,1]
UB	Upper bound variable
LB	Lower bound variable
E_0	Initial prey energy
t and T	Present and the highest iterations
ΔX^t	Distance between hawks' location and prey at tth time
$\mathrm{J} = 2 \times (1 - r_5)$	Random hop
D	Problem dimension
S	Random vector dimension
u and v	Random numbers in the range of $(0,1)$
β	Constant
C	Input channel
H and W	Feature map height and width
r	Compression ratio
δ	Activation function
y_c	Output
TP	True Positive
TN	True Negative
FP	True Positive
FN	False Negative

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision and project administration, have been done by 2nd, 3rd and 4th author.

References

[1] M. Shorfuzzaman, "An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection", *Multimedia Systems*, Vol. 28, No. 4, pp. 1309-1323, 2022.

[2] B. Mazoure, A. Mazoure, J. Bédard, and V. Makarenkov, "DUNEScan: a web server for uncertainty estimation in skin cancer detection with deep neural networks", *Scientific Reports*, Vol. 12, pp. 179, 2022.

[3] O.K. Pal, D. Paul, E. Hasan, M. Mohammad, M.A.H. Bhuiyan, and F. Ahammed, "Advanced Convolutional Neural Network Model to Identify Melanoma Skin Cancer", In: *Proc. of 2023 IEEE International Conference on Contemporary Computing and Communications (InC4)*, Bangalore, India, pp. 1-5, 2023.

[4] H.M. Balaha, A.E.-S. Hassan, E.M. El-Gendy, H. ZainEldin, and M.M. Saafan, "An aseptic approach towards skin lesion localization and grading using deep learning and harris hawks optimization", *Multimedia Tools and Applications*, 2023.

[5] A.K. Sharma, S. Tiwari, G. Aggarwal, N. Goenka, A. Kumar, P. Chakrabarti, T. Chakrabarti, R. Gono, Z. Leonowicz, and M. Jasiński, "Dermatologist-level classification of skin cancer using cascaded ensembling of convolutional neural network and handcrafted features based deep neural network", *IEEE Access*, Vol. 10, pp. 17920-17932, 2022.

[6] M. Nawaz, Z. Mehmood, T. Nazir, R.A. Naqvi, A. Rehman, M. Iqbal, and T. Saba, "Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering", *Microscopy research and technique*, Vol. 85, No. 1, pp. 339-351, 2022.

[7] H.M. Balaha, and A.E.-S. Hassan, "Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm", *Neural Computing and Applications*, Vol. 35, No. 1, pp. 815-853, 2023.

[8] V. Venugopal, N.I. Raj, M.K. Nath, and N. Stephen, "A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images", *Decision Analytics Journal*, Vol. 8, pp. 100278, 2023.

[9] S.P. Maniraj, and P.S. Maran, "A hybrid deep learning approach for skin cancer diagnosis using subband fusion of 3D wavelets", *The Journal of Supercomputing*, Vol. 78, No. 10, pp. 12394-12409, 2022.

[10] H.C. Reis, V. Turk, K. Khoshelham, and S. Kaya, "InSiNet: a deep convolutional approach to skin cancer detection and segmentation", *Medical & Biological Engineering & Computing*, Vol. 60, No. 3, pp. 643-662, 2022.

[11] L. Liu, M. Qi, Y. Li, Y. Liu, X. Liu, Z. Zhang, and J. Qu, "Staging of skin cancer based on hyperspectral microscopic imaging and machine learning", *Biosensors*, Vol. 12, No. 10, pp. 790, 2022.

[12] K.M. Hosny, and M.A. Kassem, "Refined residual deep convolutional network for skin lesion classification", *Journal of Digital Imaging*, Vol. 35, No. 2, pp. 258-280, 2022.

[13] S. Barın, and G.E. Güraksın, "An automatic skin lesion segmentation system with hybrid FCN-ResAlexNet", *Engineering Science and Technology, an International Journal*, Vol. 34, pp. 101174, 2022.

[14] S.P.G. Jasil, and V. Ulagamuthalvi, "A hybrid CNN architecture for skin lesion classification using deep learning", *Soft Computing*, 2023.

[15] M.A. Rasel, U.H. Obaidellah, and S.A. Kareem, "Convolutional Neural Network-Based Skin Lesion Classification With Variable Nonlinear Activation Functions", *IEEE Access*, Vol. 10, pp. 83398-83414, 2022.

[16] S.K. Singh, S. Banerjee, A. Chakraborty, and A. Bandyopadhyay, "Classification of Melanoma Skin Cancer Using Inception-ResNet", In: *Proc. of Frontiers of ICT in Healthcare: Proceedings of EAIT 2022, Part of the Lecture Notes in Networks and Systems book series (LNNS)*, Singapore: Springer Nature Singapore, pp. 65-74, 2022.

[17] Y.S. Alsahafi, M.A. Kassem, and K.M. Hosny, "Skin-Net: a novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier", *Journal of Big Data*, Vol. 10, pp. 105, 2023.

[18] A. Naeem, T. Anees, M. Fiza, R.A. Naqvi, and S.-W. Lee, "SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images", *Sensors*, Vol. 22, No. 15, pp. 5652, 2022.

[19] M. Tahir, A. Naeem, H. Malik, J. Tanveer, R.A. Naqvi, and S.-W. Lee, "DSCC_Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopic Images", *Cancers*, Vol. 15, No. 7, pp. 2179, 2023.

[20] S.M. Jaisakthi, P. Mirunalini, C. Aravindan, and R. Appavu, "Classification of skin cancer from dermoscopic images using deep neural network architectures", *Multimedia Tools and Applications*, Vol. 82, No. 10, pp. 15763-15778, 2023.

[21] D. Adla, G.V.R. Reddy, P. Nayak, and G. Karuna, "A full-resolution convolutional network with a dynamic graph cut algorithm for skin cancer classification and detection", *Healthcare Analytics*, Vol. 3, pp. 100154, 2023.

[22] A. Imran, A. Nasir, M. Bilal, G. Sun, A. Alzahrani, and A. Almuhaimeed, "Skin Cancer Detection Using Combined Decision of Deep Learners", *IEEE Access*, Vol. 10, pp. 118198-118212, 2022.

[23] A.S. Qureshi, and T. Roos, "Transfer learning with ensembles of deep neural networks for skin cancer detection in imbalanced data sets", *Neural Processing Letters*, Vol. 55, No. 4, pp. 4461-4479, 2023.

[24] Dataset link: https://www.kaggle.com/datasets/qikangdeng/isic-2019-and-2020-melanoma-dataset (accessed on 26 October 2023).

[25] P. Bansal, A. Vanjani, A. Mehta, J.C. Kavitha, and S. Kumar, "Improving the classification accuracy of melanoma detection by performing feature selection using binary Harris hawks optimization algorithm", *Soft Computing*, Vol. 26, No. 17, pp. 8163-8181, 2022.