



Attention-Based Hybrid RGB-Edge-Histogram Network for Melanoma Classification

Navya Sri Dodda

Master's Student, Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India – 522502
Mobile : 9491728010

Email: navyasridodda@gmail.com

Dr. MALLIKARJUNAMALLU K,

Assistant Professor,

Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, A.P, 522302, India.
Mobile: +91 9010349521

E-mail: mallikarjunamalluk@gmail.com

Abstract— Even with the photos of skin lesions, the identification of the melanoma remains challenging, because lesions that appear to be benign could have similar colour, texture, and structural features to melanoma. In this paper, we suggest a hybrid DL system to automatically classify melanomas with the help of complementary picture representations. For the implementation, a data set of 1001 labelled images of the benign and malignant cases is used. The dataset is randomly undersampled to get a balanced set of 1000 images. The photographs are scaled to 50×50 pixels and rendered in three different forms: a normalised RGB image, a Canny edge map and an intensity histogram of the image with 256 bins. These representations are then fed into special convolutional and dense branches that learn features, which are then concatenated and further enhanced by layer norm and four-head multi-head self-attention before being classified using binary softmax. The framework is trained using the Adam optimiser with categorical cross-entropy, and is evaluated against a CNN baseline using only images. The proposed hybrid model is able to achieve best validation accuracy of 86.50%, while image-only baseline model can achieve the test accuracy of 88.06% in the same experimental setting. The results show that it is possible to incorporate appearance, structural and intensity-distribution information into an attention-based classification pipeline, but also emphasise the necessity for bigger and externally validated datasets to provide robust generalisation. Furthermore, the solution also allows batch and real-time inference, suggesting the potential use as a computational screening aid.

“Keywords— Melanoma Classification, Skin Lesion Analysis, Convolutional Neural Network, Canny Edge Detection, Feature Fusion, Multi-Head Attention, Deep Learning, Medical Image Analysis.”

I. INTRODUCTION

Melanoma is a dangerous type of skin cancer, in which early detection and correct separation from benign lesions is critical for timely clinical evaluation. Image based automated analysis is a landmark study area because of the variety of skin lesions that are quite different in appearance. Medical image analysis is an area, where DL has been quite successful, by which models learn hierarchical representations from images without having the task of designing descriptors by hand [1]. In particular, DNN have exhibited powerful skills for automatic skin-cancer classification and have demonstrated potential for aiding expert assessment [2].

The availability of dermoscopic image collections to the public has further boosted research on automated skin-lesion analysis. The HAM10000 dataset consists of a large and heterogeneous collection of dermatoscopic images of different and common types of pigmented skin lesions [3]. In the same way, the ISIC has developed benchmark datasets and challenges, which have enabled the development and comparative assessment of automated melanoma detection systems [4]. These resources offer an important foundation for computer techniques to discriminate among lesions that have similar visual appearance.

Although there have been advances in DL, it is still difficult to differentiate between benign and malignant tumours due to differences in lesion appearance including pigmentation, form, and structure characteristics. RGB methods are efficient at capturing picture information (colour and spatial patterns), but may be less effective in providing explicit information of boundaries and global intensity features. Structural information can complement the information of the lesion morphology and intensity distributions can carry global properties that are not necessarily represented in localised convolutional features. Edge detection has been a good way to find important intensity transitions and structural boundaries in images for a long time [5].

CNN based technique has demonstrated an effective performance for image identification with hierarchical spatial feature extraction. Just increasing the depth of a model will not ensure that all complementary properties of an image are captured, however. Hierarchical feature learning has been proven to be powerful for complex recognition problems in deep residual architectures [6]. The ability to gain complementary representations using parallel processing has been demonstrated in multi-path convolutional architectures [7]. The improvements indicate that several feature

*Corresponding Author:

Received: 25th May, 2026; Revised: 6th July, 2026; Accepted: 28th July, 2026; Available Online: 29th August, 2026
(DOI): 10.70102/AEJ.2026.18.8s.9

representations can be combined to provide a richer description of visually complex things.

It is further confirmed by deeper CNN architectures that discriminative representation learning from visual data is important [8]. Current methods for melanoma classification, however, tend to concentrate on spatial and colour features and additional structural and intensity-distribution information could be used. Automated melanoma detection systems work well in classification tasks [9, 10] but due to the properties of lesions and different sets of images, generalisation becomes a challenge. This promotes the study of feature representations other than the basic learning from RGB.

This paper deals with a hybrid representation combining RGB appearance with complementary structural and intensity-distribution information, for melanoma classification. Another novel approach investigated in this paper is attention-based fusion, which allows for interaction between the representations generated, and it is compared to an image-only CNN baseline approach. The resulting framework aims to provide a wider representation of skin-lesion features while keeping a feasible computational process for classification. The analysis also enables us to understand whether the characterisation of benign and malignant lesions can be advanced by jointly exploiting visual information with complementing RGB image features, or just relying on the RGB image features.

II. RELATED WORK

Automated skin-lesion classification has been the mainstream approach and the DL approach has been the current trend because CNNs learn discriminative representations directly from dermoscopic images. To gain from the complementary deep representation, Mahbod et al. presented a hybrid method which extracts features from various pretrained CNN architectures and abstraction levels before the classification stage and score fusion [11]. Yu et al. suggested aggregated deep convolutional features for melanoma detection, and showed that learnt features can be combined for improved discriminating of lesion categories [12]. The efficiency of CNN derived feature representations is demonstrated in these studies, which are mostly confined to deep spatial-feature domain.

Later, feature fusion has been studied to develop more complete representations of lesion. To improve the capability of the CNNs to focus on discriminative lesion sites, Zhang et al. introduced attention residual learning [13]. Kwasigroch et al. studied neural architecture search for skin-lesion categorisation, demonstrating automatically optimised network topologies [14]. Using regularly spaced picture shifting, Thurnhofer-Hemsi et al. showed that for classification, the advantages of data and model diversity in the generation of an ensemble of deep convolutional networks should be leveraged [15]. Recently, feature-fusion techniques have used multiple pretrained CNNs to input multiple feature maps into a fusion layer prior to feature selection and classification [16]. This shows that using multiple backbone features is better than a single backbone feature. However, these methods typically use a plethora of deep models to extract their free information, rather than directly

incorporating RGB appearance that includes structural and intensity-distribution properties.

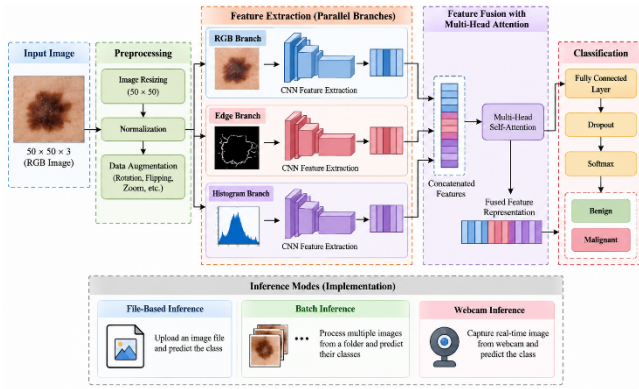
One possibility to enhance the representation of lesions is to make use of attention mechanisms to attend to the channels or regions of features that are relevant to the lesion. Wu et al. investigated scale-aware transformer representations for melanocytic lesion identification, with a focus on the importance of attention-driven modelling to capture information at various scales [17]. For skin lesions classification, attention and self-attention methods have also demonstrated their effectiveness, and, in this context, self-attention is shown to be particularly effective in enhancing feature interactions in convolutional architecture [18]. Recent studies have introduced attention mechanisms in conjunction with sophisticated feature-fusion modules to account for colour, texture and multi-scale representations on skin-lesion datasets [19]. Recently, attention based ensemble architectures have demonstrated that a dynamic attention to emphasize relevant features could be beneficial for CNN feature extraction [20].

While this advancement has taken place, there is a particular need to combine multiple low-level and learnt representations within a compact classification system for melanoma. Most of the existing studies focus on deeper CNN features, multiple pretrained networks, ensembles or attention modules. Explicit integration of RGB appearance, edge-based structural information and greyscale intensity distributions as a single representation is seldom explored. The proposed technique fills this gap by exploring complementary RGB, edge and histogram representations followed by an attention-based feature fusion while retaining a direct comparison with an image-only CNN baseline.

III. PROPOSED METHODOLOGY

A. System Overview

Visual, structural and intensity distribution information are provided complementarily to the proposed system which classifies a given skin-lesion image as benign or malignant. The total method starts with an input lesion image which is first pre-processed in a standardised way and then converted into three parallel representations: a normalised RGB image, a Canny-derived edge map, and a gray-scale intensity histogram. The dedicated branches of the network, called feature extraction branches, process these representations to detect additional properties of relation to lesion morphology, borders and intensity distribution. The retrieved features are then merged into a single representation and further enhanced by a multi-head self-attention technique. This feature representation is then passed on to a classification layer that classifies the lesion as benign or malignant based on the probability. This sequential organization allows for the complementary information in each of several representations to add to the final categorisation decision.

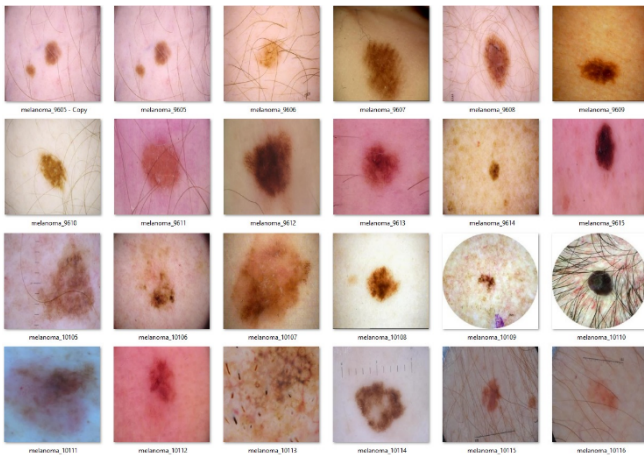


“Fig. 1. Overall System Architecture”

Figure 1 shows the entire working of the proposed system. The preprocessed image is mapped to three complementary feature representations after the pre-processing. Two feature extraction branches are learnt from the RGB image, edge map and histogram are stitched together to output. Afterwards, multi-head self-attention is used to further enhance the fused representation based on the relationships between the extracted characteristics. Finally, the classification layer is responsible for the classification on benign or malignant.

B. Dataset Description and Preprocessing

The collection consists of 1,001 dermoscopic photos of skin-lesions in two class directories, benign and malignant. The original distribution has 501 benign and 500 cancerous pictures. We address the small class imbalance by random undersampling of the benign class, to create a balanced set of 1000 photos distributed equally across the two classes of 500. Fig. depicts a typical image of a lesion from the data set. 2.



“Fig. 2. Representative Dataset Image”

Each image used in the preprocessing stage is resized to 50×50 pixels to normalise the spatial dimension for the next processing step, which is the feature extraction. The integer pixel values in the RGB format are converted to floating point values and normalised to the $[0, 1]$ range by dividing pixel intensities by 255. The balanced data set is then divided into 80% train and 20% test (random state 42). The obtained experimental partition is therefore composed of 800 training images and 200 testing images for a balanced dataset.

C. RGB Feature Extraction

The $50 \times 50 \times 3$ RGB image is preprocessed and passed through the RGB branch to extract the spatial and appearance relevant characteristics by utilising a convolutional neural network. It is composed of two convolutional blocks with 32 and 64 filters each with 3×3 kernels and same padding. After every convolution, there is a batch normalisation layer followed by a ReLu activation layer to ensure the features are trained more stably and to introduce some nonlinear mappings. After each block, max-pooling is used to progressively lower the spatial dimensions, while maintaining important feature responses. The feature maps generated are flattened to give a 1-D feature representation. At the end, a dropout rate of 0.3 is employed to prevent overfitting before the RGB representation that was derived is passed over to the subsequent feature-fusion stage. The main features that are collected in this branch are colour, texture and spatial features of the original lesion image.

D. Edge Feature Extraction

Edge branch is intended to capture structures and edge information which could be incorporated into the appearance attributes learned from RGB representation. The images taken are scaled and converted to greyscale representation from RGB. Next, lower and upper thresholds of 100 and 200 are used to obtain a canny edge detection to detect salient intensity changes corresponding to the border of the lesion and structural patterns. The resulting edge map is a single channel image of size 50×50 and is normalised prior to being passed into the feature-extraction network. A CNN branch is dedicated to this representation, which it uses to learn spatial features of the identified edges. The recovered edge features are further passed on to feature fusion operation where they are combined with RGB and histogram feature to provide additional structural information for the classification of melanoma.

E. Histogram Feature Extraction

In addition to the CNN branches, which recover the spatial characteristics of each lesion image, the histogram branch is used to capture the global intensity distribution of each lesion image. The resized input image is then converted to grey scale, and a 256-bin intensity histogram is created to indicate the frequency distribution of the grey scale values of the input image. This histogram is then flattened to a feature vector of size 256 and each number in the range is normalised using Min-Max scaling. This normalised vector is then sent to a dense feature extraction layer with 128 units, batch normalisation and ReLU activation. Then, the regularisation is done with a dropout rate of 0.3. This provides a dense representation of the intensity characteristics of the image across the entire image, which is then used at the feature-fusion stage, along with the RGB and edge features.

F. Multimodal Feature Fusion and Attention

The representations taken from the RGB, edge and histogram branches are concatenated to provide a 19,344-dimensional unified feature representation. This fusion fuses appearance, structural and global intensity information in a common feature space. Next, the concatenated representation is layer-normalized to make the feature distribution more

stable prior to the attention-based refinement. The attention mechanism is formed as a sequence-wise shape, where the characteristics are normalised. Then, a multi-head self-attention module with 4 attention heads and a key dimension of 32 is used to model the relations between the fused features and find the informative feature interactions. The resulting representation of attention is pooled globally, yielding a short vector of features for next classification stage. The complimentary information from the three branches can contribute jointly to the ultimate choice with this fusion technique.

G. Classification and Training

A fully connected classification head, with a 256 unit dense layer, batch normalisation and ReLU activation is used to ingest this attention refined feature representation. A dropout rate of 0.4 is used to prevent overfitting. The class probabilities for benign and malignant lesions are obtained by a two-unit softmax layer. The hybrid model is trained with the Adam optimiser and the loss function is category cross entropy. A batch size of 8 is used, and the number of epochs is set to 30, with a random state set to 42 for the reliable data partitioning. As a baseline, we also apply the image-only CNN baseline model without the RGB modality branches (edge, histogram) and multimodal fusion based on attention. This baseline provide a reference to investigate the contribution of the proposed complimentary feature representations.

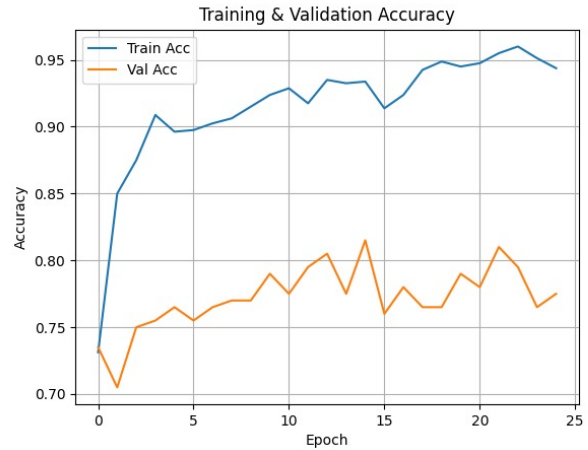
IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Configuration

The proposed framework is implemented in Python with TensorFlow/Keras and leverages OpenCV, NumPy, Scikit-learn, and Imbalanced-learn for image processing, numerical computations, dataset preparation, and assessment. The trials are conducted on a carefully balanced set of 1,000 images, with 80% of the images in the training set and 20% in the testing set, and a random state of 42 for the split. The hybrid model was trained for 30 epochs on 8-batch using Adam optimisation with categorical cross-entropy loss. The performance of the model is assessed via classification accuracy and class-level precision, recall and F1-score and prediction behaviour is examined using confusion-matrix analysis, across benign and malignant classes.

B. Training Performance

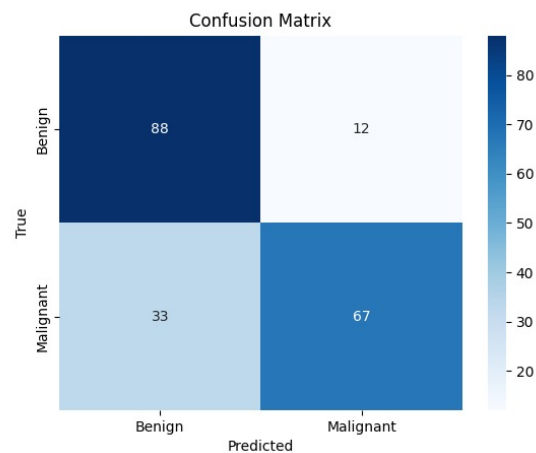
The training/validation accuracy evolution of the proposed hybrid model is depicted in Fig. 3. Training accuracy is significantly improved in the first few epochs, and improves along the training and is at about 94-96% in the latter part of the training. The accuracy of the validation set also increases in the initial training process, but it fluctuates over the epochs. The model obtains a best validation accuracy of 86.50% and validation accuracy at the last epoch is 86.00%. The distance between the training and validation curves increases with the number of epochs as a measurable generalisation gap, meaning the model was better able to learn the training data than the validation data during later epochs. This behaviour suggests that there exists other forms of regularisation or more varied and extensive training data that can be used to enhance generalisation.



“Fig. 3. Training and Validation Accuracy”

C. Confusion Matrix and Class-Level Performance

The classification behaviour of the model for each class is presented in the confusion matrix as shown in Fig. 4. There are 88 instances where the benign lesion is correctly labelled as a benign lesion, and 12 instances where the benign lesion is incorrectly labelled as a malignant tumour. In the case of malignant class, a total of 67 samples are correctly classified as malignant, whereas 33 samples that are truly malignant are projected as benign. The study reveals that the model is more consistent in the recognition of benign lesions than of malignant lesions. The classification error made in particular is more apparent with the higher prevalence of malignant samples being classified as benign, which is crucial because it would impact the reliability of automated screening if malignant samples were mistakenly classified as benign. The 12 samples considered as benign, however, classified as malignant indicate that the rate of false positive predictions in the case of malignancy is relatively low.

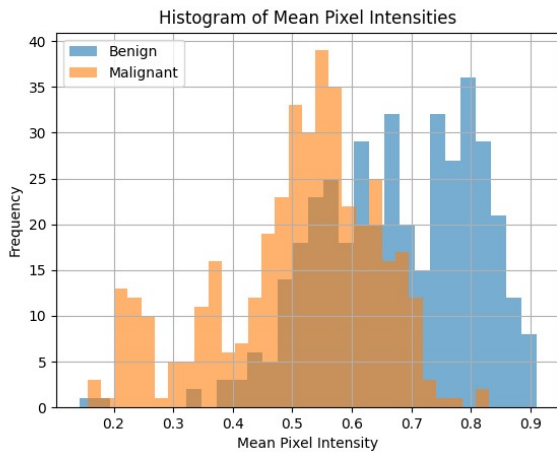


“Fig. 4. Confusion Matrix of the Classification Model”

D. Feature Distribution Analysis

The distribution of the mean pixel intensity of normal and malignant samples is shown in . 5. The two classifications overlap in the intensity range and mean pixel intensity alone is not a good differentiator between benign and malignant tumours. The distributions do have differences, however, with

the benign samples being more concentrated at the higher values of intensity, and the malignant samples being more concentrated at the lower and intermediate values of intensity. The differences suggest that intensity feature information can be used for global intensity features that will provide useful information about the appearance of the lesions. Thus, histogram-based intensity information is not considered as a separate diagnostic criterion, but as a feature representation to be used as a complementary feature representation together with RGB and edge features.



“Fig. 5. Distribution of Mean Pixel Intensities”

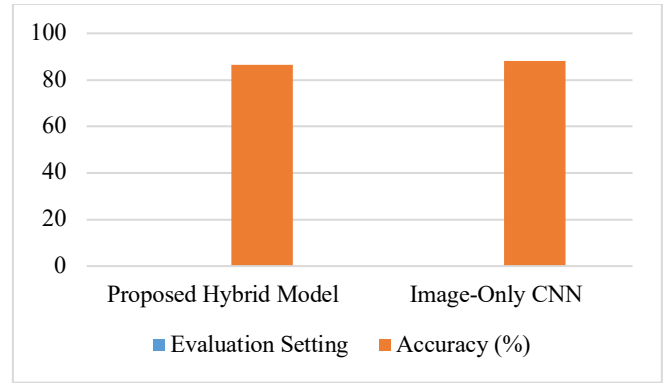
E. Comparative Model Evaluation

The results of the classification models analyzed are presented in Table II. The proposed hybrid model achieves the best validation accuracy of 86.50%, while the CNN baseline using only images achieves the test accuracy of 88.06%. The results show that the models could well classify benign and malignant skin lesions, with the image-only CNN model being able to acquire a higher accuracy in the test conducted. As they are generated at different levels of assessment, they are not considered to be the same assessment level, but rather accompanied by their respective validation and test conditions.

“Table II. Performance Comparison of the Evaluated Models”

Model	Evaluation Setting	Accuracy (%)
Proposed Hybrid Model	Best Validation Accuracy	86.50
Image-Only CNN	Test Accuracy	88.06

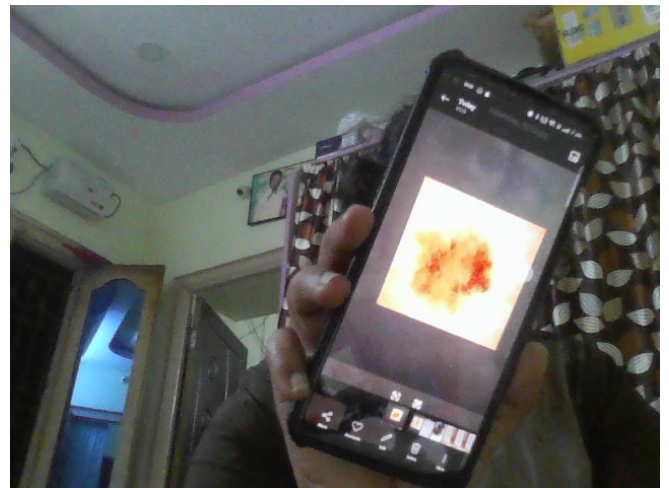
The results of the model accuracies are graphically compared in Fig. 6. The graph illustrates the accuracy results recorded and highlights the gap between the best accuracy achieved during the validation phase with the hybrid framework and the test accuracy of the image-only CNN. The comparison results give a brief overview of the classification results observed for the respective model, while preserving the assessment setting of the model.



“Fig. 6. Graphical Comparison of Model Performance”

F. Real-Time and Batch Inference Discussion

The proposed system introduces multiple inference strategies for utilization of the trained classifier in the images of lesions. The file based prediction enables the loading of an image for categorisation, and offers an immediate assessment of a selected lesion image. Batch classification takes this a step further by classifying a bunch of images and providing a prediction for a batch of images. This enables more efficient processing of batches of images. The method also features webcam based inference which allows the system to capture an image from the webcam, and classify the image as it is happening. The prediction interface generates a class label and confidence score for each prediction. The prediction for the displayed webcam example in Fig. 7 is 76.61% with a confidence value. This value is not an actual performance statistic from the model, but rather the level of confidence on this particular inference. The inference modes point out the potential of the framework for real world uses in interactive and automatic image categorisation processes, however, further validation is needed before clinical use can be considered.



“Fig. 7. Real-Time Webcam-Based Melanoma Classification”

G. Discussion

The performance of the proposed framework shows that the framework is advantageous by utilizing the complementarity of free representation of the images of skin lesions. The RGB branch is used to extract the visual and colour information, while the edge representation accentuates structural edges and lesion boundaries. In addition to spatial

features, the histogram branch provides the global intensity distribution information as a histogram. The multi-head attention technique, on the other hand, enhances the fused representation by allowing the different components of the features to interact prior to classification.

The accuracy achieved in the experiments shows that there is a big gap between the proposed hybrid model and the baseline image-only CNN model. The hybrid model achieves 86.50% best validation accuracy while the image-only CNN achieves the test accuracy of 88.06%. Comparing these numbers as a direct test comparison is not appropriate; rather they should be considered in the context of the circumstances of their evaluation. There is also a generalisation gap in the training-validation curves: the validation accuracy is decreasing while training accuracy is increasing in later epochs.

Installation-wise, the system offers file-based, batch and webcam-based inferences, visualizing the practical application of predictive inference. The study, however, has some limitations in the scope of the relatively small data set, 50×50 image resolution, limited external validation and absence of clinical validation. Hence, the results shown here are experimental classification results which are not meant to be clinical diagnosis.

V. CONCLUSION AND FUTURE SCOPE

A. Conclusion

In this paper, a hybrid DL system for automated classification of skin lesions into benign and malignant class is proposed. In the proposed architecture, for each lesion image, three complementary representations are used: RGB features for visual and colour properties, Canny-based edge features for structural and border information and greyscale histogram features for global intensity distribution. These scattered representations are retrieved by dedicated branches and combined into a cohesive feature space. Then, multi-head self-attention is used to optimise interactions among fused features for classification. The best validation result for the proposed hybrid model is shown to be 86.50% in the experimental evaluation. Under the evaluation condition, the CNN baseline model only achieves a test accuracy of 88.06%. Results indicate that picture representations can be incorporated in a unified categorisation system, but external evaluations and generalisation are required to be improved. This study provides a systematic basis for the multi-modal feature extraction and attention-based fusion for the classification of skin-lesion pictures. In addition, the system offers a practical file-based, batch and webcam inference that can take the paradigm further than the offline model evaluation.

B. Future Scope

Larger and more diverse skin lesion datasets, higher resolution image inputs and transfer-learning inspired architectures could be used in future studies to boost feature representation and generalisation. Validation of the model with independent data sets is possible, and might be more convincing if performed outside the model. Explainable AI algorithms can enhance the understanding of classification judgements, and clinical deployment studies can evaluate the

practical utility of the algorithm in real world scenarios. Further optimisation enables resources-constrained inference contexts, such as mobile and edge-devices.

REFERENCES

- [1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, doi: 10.1038/nature21056.
- [2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [3] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, Art. no. 180161, 2018, doi: 10.1038/sdata.2018.161.
- [4] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC)," in *Proc. 2018 IEEE 15th Int. Symp. Biomed. Imaging (ISBI)*, 2018, pp. 168–172, doi: 10.1109/ISBI.2018.8363547.
- [5] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC)," in *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2019.
- [6] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, 1986, doi: 10.1109/TPAMI.1986.4767851.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
- [9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015, pp. 1–14.
- [10] A. A. Adegun and S. Viriri, "Deep learning-based system for automatic melanoma detection," *IEEE Access*, vol. 8, pp. 7160–7172, 2020, doi: 10.1109/ACCESS.2019.2962812.
- [11] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinger, "Skin lesion classification using hybrid deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2019, pp. 1229–1233, doi: 10.1109/ICASSP.2019.8683352.
- [12] Z. Yu, X. Jiang, F. Zhou, J. Qin, D. Ni, S. Chen, B. Lei, and T. Wang, "Melanoma recognition in dermoscopy images via aggregated deep convolutional features," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 1006–1016, 2019, doi: 10.1109/TBME.2018.2866166.
- [13] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention residual learning for skin lesion classification," *IEEE Trans. Med. Imaging*, vol. 38, no. 9, pp. 2092–2103, 2019, doi: 10.1109/TMI.2019.2893944.
- [14] A. Kwasigroch, M. Grochowski, and A. Mikołajczyk, "Neural architecture search for skin lesion classification," *IEEE Access*, vol. 8, pp. 9061–9071, 2020, doi: 10.1109/ACCESS.2020.2964424.
- [15] K. Thurnhofer-Hemsi, E. López-Rubio, E. Domínguez, and D. A. Elizondo, "Skin lesion classification by ensembles of deep convolutional networks and regularly spaced shifting," *IEEE Access*, vol. 9, pp. 112193–112205, 2021, doi: 10.1109/ACCESS.2021.3103410.
- [16] R. Maurya, A. K. Bais, T. Gopalakrishnan, M. K. Dutta, N. N. Pandey, and Y. V. Srinivasa Murthy, "Skin lesion classification using deep feature fusion and selection using XGBoost classifier," in *Proc. 2024 IEEE Int. Students' Conf. Electr., Electron. Comput. Sci. (SCEECS)*, 2024, doi: 10.1109/SCEECS61402.2024.10481955.
- [17] W. Wu, S. Mehta, S. Nofallah, S. Knezevich, C. J. May, O. H. Chang, J. G. Elmore, and L. G. Shapiro, "Scale-aware transformers for

- diagnosing melanocytic lesions,” *IEEE Access*, vol. 9, pp. 163526–163541, 2021, doi: 10.1109/ACCESS.2021.3132958.
- [18] R. Pedro and A. L. Oliveira, “Assessing the impact of attention and self-attention mechanisms on the classification of skin lesions,” in *Proc. 2022 Int. Joint Conf. Neural Networks (IJCNN)*, 2022, pp. 1–8, doi: 10.1109/IJCNN55064.2022.9892274.
- [19] Z. Ji, X. Wang, C. Liu, Z. Wang, N. Yuan, and I. Ganchev, “EFAM-Net: A multi-class skin lesion classification model utilizing enhanced feature fusion and attention mechanisms,” *IEEE Access*, vol. 12, pp. 143029–143041, 2024, doi: 10.1109/ACCESS.2024.3468612.
- [20] S. T. Tabibi, A. Nikravanshalmani, and H. Saboohi, “An ensemble classifier based on diverse convolutional neural networks for skin lesions classification,” *IEEE Access*, vol. 13, pp. 195673–195686, 2024, doi: 10.1109/ACCESS.2024.3442827.