Journal of Innovations in Computer Science and Engineering
Shahid Beheshti Unversity

June 2024, Special Volume 1, Issue 2

Enhancing Automated Skin Cancer Detection Through Ensemble Learning and Multi-Head Attention Mechanisms

Maryam Nazari, ✉, Orcid: 0009-0004-6309-4697
***Cyberspace research institute, Shahid Beheshti University,* Tehran, Iran**
Fatemeh Fadaei Ardestani, Orcid: 0009-0005-5865-4407
***Dept.of Computer Engineering, AI and Robotics, University of Isfahan,* Isfahan, Iran**

***Abstract*— Skin cancer is one of the deadliest but most prevalent types of cancer; as such, early diagnosis is urgently required to improve patient outcomes. This work presents a collaborative deep learning model that classifies skin cancer with respect to three different networks: EfficientnetB1, EfficientnetB2, and EfficientnetV2s on dermoscopic images. The proposed collaborative model has a multi-head attention mechanism, ensuring that this model has a better attention capability for improving its accuracy in the task of classification. The HAM10k dataset provided the proposed model with a platform for fine tuning with transfer learning, along with some augmentation techniques to handle class imbalance challenges and feature variations of lesions. The results for the ensemble model combined with Multi-Head Attention were very high: an accuracy of 97.11%, and precision, recall, and F1-score are also very high. These findings prove that our approach can dramatically improve automation in skin cancer detection. Therefore, it will be helpful in clinical dermatology for early diagnosis in medicine.**

***Keywords*— Skin lesion classification, Multi-head attention, Ensemble learning, Medical imaging**

I. Introduction

Skin cancer is one of the most prevalent cancers around the world, claiming over a thousand lives every year. Early detection significantly increases the chances of successful therapy and the percentage of survival. Skin cancer arises from skin cells whose growth becomes uncontrollable under continuous exposure to UV radiation. By the classification, skin cancer mainly could be malignant and benign, out of which malignant variants are the most uncontrollable and dangerous variants. However, all these traditional methods, including biopsy and dermoscopy, are primarily time-consuming and prone to misdiagnosis. Recent medical imaging, coupled with advances in computational techniques, has opened new prospects for the diagnosis of skin cancer, particularly with deep learning. Convolutional Neural Networks have been found fundamental in carrying out most of the image analytical applications, such as skin lesion classification. It is on record that pioneering work in the area has proved the ability of CNNs to achieve very high accuracy in distinguishing malignant lesions from benign ones. The well-known deep and simple VGG16 architecture finds applications for skin cancer classification and has obtained very impressive results in several studies [1]. In the same way, DenseNet has shown the best performance in extracting features with an architecture based on densely connected layers [2]. EfficientNet, a recent advancement in the field, enhances both accuracy and computational efficiency, thereby establishing itself as a favored option for extensive medical image analysis[3].

These models, in combination with methodologies such as transfer learning and data augmentation, have fully enhanced the accuracy and reliability of automatic skin cancer detection systems. The further development of the herein reviewed architectures, if applied to dermoscopic imaging, is foreseen to give even more intensive development of diagnostic skills and become a necessary tool for dermatologists in everyday clinical work.

II. RELATED WORK

Skin cancer is a serious public health problem and thus demands early detection to reduce mortality. Automated detection systems are very helpful in this regard for early detection. Several deep learning and transfer learning architectures are explored to improve diagnosis with high precision in cases of skin cancers. Five pre-trained models were compared for binary classification in previous work; amongst these, ResNet-50 achieved an accuracy of 93.5% using augmentation techniques to improve model robustness [4]. In another independent study, BCC, SCC, and melanoma were classified using EfficientNet models, among which EfficientNet-B4 achieved an accuracy of 79.69%, a precision rate of 81.67%, and a recall measure of 76.56% [5].

While Anand et al.[6] fine-tuned VGG16 for skin cancer classification and achieved a maximum accuracy of 89.09% in conjunction with data augmentation and hyperparameter tuning, an ensemble of the EfficientNet models ensembled with patient metadata and post-processing techniques by Ha et al.[7] obtained an AUC score of 0.9600 on cross-validation in the SIIM-ISIC Melanoma Classification Challenge, demonstrating the potential of EfficientNet ensembles in the improvement of melanoma detection.

Submit Date: 2025-03-11
Accept Date: 2025-04-12
✉ Corresponding author

Sharma et al. [8] presented a cascaded ensemble model combining ConvNet with handcrafted features to attain 98.3% accuracy, outperforming the conventional approaches like the ABCD rule. Sunarya et al. [9] introduced the use of a Graph Convolutional Network (GCN) by presenting 97.3% accuracy. It is based on processing graph-structured data for improved lesion classification.

Deep learning approaches to mobile applications include an impressive mobile-optimized algorithm that has been able to classify skin lesions from the HAM10000 dataset
with an accuracy of 98.5%[10]. Similarly, a hybrid CNN model utilizing transfer learning coupled with a random forest classifier was proposed by Mahalle et al.[11], which attained an accuracy of 90.11%.

More recently, optimization has resulted in models like that of Falcon Finch, a deep CNN classifier, which claimed an accuracy of 96.52% for detection of skin cancer [12]. Kumar et al. [13] discussed in detail the usage of AI in the healthcare sector, considering the hazards of deep fake medical data. They were able to distinguish between original and fake images using CNNs with almost 100% accuracy.

Fig. 1. Sample of each class in melanoma skin cancer dataset

It had been considered one frontier balancing accuracy and computational efficiency effectively. Very recently, works involving EfficientNet on skin lesion segmentation reported an accuracy as high as 99.78%, establishing the capability of EfficientNet concerning both classification and segmentation tasks [14, 15]. Last but not least, progress with image augmentation, including geometric transformation combined with GANs, improves the accuracy up to 96.90% [16].

III. METHOD

IV. Data Acquisition

Melanoma Skin Cancer Dataset consists of 10000 images for training set[17]. Melanoma skin cancer is a fatal cancer, and hence early detection and curing can save many precious lives. The dataset shall help in the development of deep learning models for the accurate classification of melanoma. It consists of 10682 images for training the model and 3561 images for testing model.

As such, in this work, only 2 classes will be used: benign and malignant. Samples of each selected class are shown below in Figure 1.

V. Data preprocessing

With this dataset, the pre-processing step would come before model training. The major objectives of the pre-processing of data involve organization, transformation, and cleaning of the data so that it may be suitable and usable in further steps. Quantities of the two categories can be seen in Table 1.

Therefore, this is where an appropriate data augmentation pipeline improves the generalization of deep learning models by reducing the risks of overfitting. Augmentation has been performed by a sequential Keras model, which consists of several steps that can simulate most of the variations that might actually occur to images. Then, each image was rescaled by the factor 1/255 to normalize the dataset. After that came horizontal flipping-the dataset was extended by adding horizontally mirrored samples. Rotation was also perfectly random: images may be rotated up to 40°. Images also included random zooming, scaling up to a maximum of 20%, and random translation-movement horizontally and vertically up to 20%. This augmentation strategy is required to make the model more robust regarding several geometrical transformations and to replicate conditions that include most probable variations for images of skin lesions.

VI. Modeling

This paper now proposes an overview of the deep learning architecture developed with an ensemble learning approach in extracting significant representations from medical images, further utilizing three well-known deep learning architectures: EfficientNetV2S, EfficientNetB1, and EfficientNetB0. The input image here is fed to all these three models simultaneously, each performing like a different layer comprising its architectural characteristics.

To reduce the dimensions, outputs at each functional layer are fed into a global average pooling layer; that is, the pooling step will result in flattened outputs which get jointly combined into one coherent feature vector for every input image.

TABLE I. CLASS DISTRIBUTION OF MELANOMA SKIN CANCER DATASET

Class	Number of Images
Melanoma	5341
Not Melanoma	5341
Total	**10682**

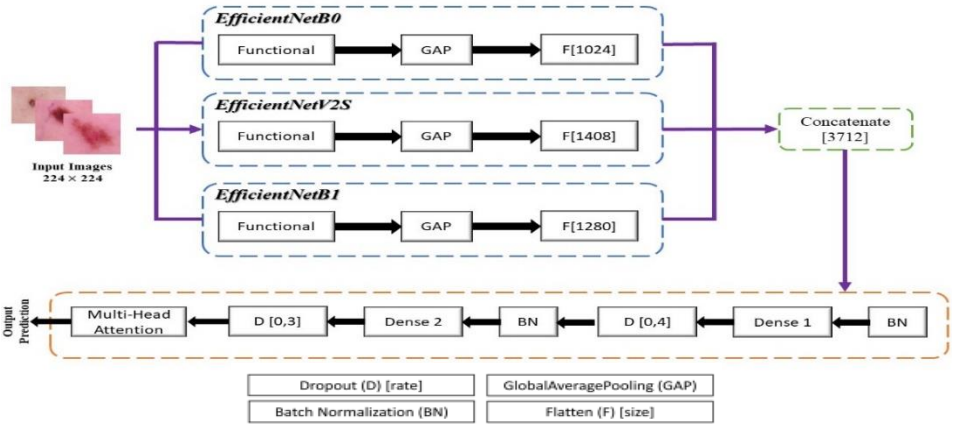

Fig. 2. overview of proposed ensemble model for skin cancer classification

Further layers are built step by step on top of this concatenated output to increase network effectiveness, reducing overfitting for better accuracy of classification. The Batch Normalization, Fully Connected, or Dense Layers, and Dropout Layers are included. Figure 3 provides an overview of these. This is finally followed by the final result of classification through a fully connected layer which merges the features processed into the output class.

In optimizing this ensemble technique, much care has been taken to obtain feature vector extraction and identification from resized skin cancer images of dimensions 224 × 224 pixels. This study will implement three pre-trained models: EfficientNetB1, EfficientNetB2, and EfficientNetV2s, each originally trained on the ImageNet dataset [16]. These flatten into feature vectors of size 1408, 2048, and 768 dimensions, respectively, in our experiments. A feature vector obtained by their concatenation, therefore, is
of 422 dimensions. To save on computation during fine-tuning, the weights of all three models are kept fixed.

A Multi-Head Attention mechanism is then added to the model after the concatenation of the outputs of EfficientNetB1, EfficientNetB2, and EfficientNetV2s [18]. That allows further abilities to attend simultaneously to lots of parts of the concatenated feature vector to get a richer contextual understanding and truly learn complex relations between data. The multi-head attention layer takes in these combined features and allows the model to assess which of the multiple features are most important dynamically.

The mathematical representation of the Multi-Head Attention mechanism is given by (1):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

In the multi-head attention mechanism, there are Q (query), K (key) and V (value), with d_k representing the key vectors dimensionality. Afterwards, the results from the head Attention layer are passed through additional dense layers, for final processing and classification purposes.

We utilized an ensemble model with multi-head attention mechanisms to improve classification performance. The multi-head attention mechanism allows the model to focus on different parts of the input image, capturing various features that contribute to accurate classification. This approach enables the model to learn complex patterns associated with different skin lesion types, enhancing its discriminative capabilities.

VII. Results

VIII. Evaluation Metrics

The effectiveness of the model is thoroughly evaluated by examining its predictions in comparison to the labels using important metrics specific to the classification task at hand. These metrics include precision and recall rates along with F1 score and accuracy measurements alongside area under the curve (AUC). The initial metrics are calculated based on classification criteria. Make use of a confusion matrix that takes into account: True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN).

The AUC metric assesses how well the model can differentiate between the two classes at threshold levels. It offers an evaluation of its capabilities [19]. Precision is determined by (2) :

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (2)$$

Measures the proportion of correctly identified positive cases out of all cases predicted as positive by the model. Recall, also known as sensitivity, is computed as (3) :

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (3)$$

It quantifies the proportion of correctly identified positive cases out of all actual positive cases in the dataset. The F1 Score, a harmonic mean of precision and recall, is calculated as (4) :

$$\text{F1 Score} = \frac{\text{Precision} .\text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

F1 score offers a measure to assess the effectiveness of the model.

IX. Analysis of Training Results

In this section, we delve into the results and accuracies of the models across various scenarios. As previously noted, we will compare the performance of the base networks with the proposed networks, highlighting the ensemble of the three best models utilizing Multi-Head Attention. Figure 3 illustrates the training and validation loss, Figure 4 highlights the training and validation accuracy.

X. Comparison of Model Performance and result

In this section, we present the results of the classification model for skin cancer using various state-of-the-art deep learning models. Results obtained from some of the existing literature on skin cancer classification are compared with our proposed models in Table II. This table points out some of the most outstanding models that have been used in relevant previous works based on different datasets with diverse classification methods.

We test our skin cancer dataset with various state-of-the-art deep learning architectures: ResNet, DenseNet, MobileNet, EfficientNet, and Xception. Our experiment results are shown in Table III. Each of the mentioned architecture was trained and tested extensively by utilizing various performance metrics for the selection of most proficient network regarding the classification of skin cancer. Further fine-tuning of the top three models has been done with the fusion of ensemble learning regarding their results on performance metrics. The entire composition of the ensemble would involve the integration of these models through Multi-Head Attention to improve classifier efficacy. This is a type of combination whereby ensemble learning combines with the mechanism of attention in order to increase the predictive capabilities of the models for higher accuracy and reliability regarding skin cancer diagnosis.

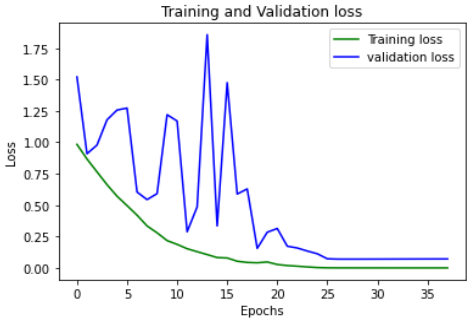

Fig. 3. Training and Validation loss of Ensemble Model with Multi-Head Attention

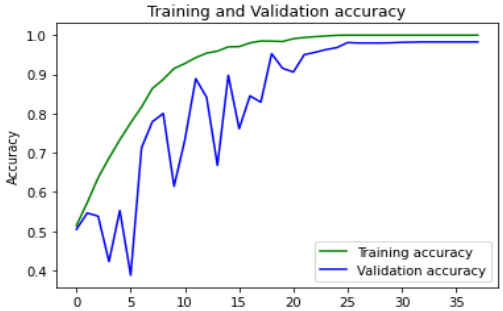

Fig. 4. Training and Validation accuracy of Ensemble Model with Multi-Head Attention

XI. Discussion

Results of this present study have also pointed out remarkable developments related to deep learning algorithms for skin cancer identification, notably those implementing convolutional neural networks. The advanced architectural features in EfficientNet, ResNet, and VGG16, added to transfer learning methodologies, have shown promise in yielding high accuracies with respect to distinguishing between benign and malignant skin lesions. That confirms the experimental results, where the ensemble models, which further combine the strengths of various pre-trained architectures, perform even better: the best performance is given by the combination of EfficientNet models.

Most useful is the component involving the use of data augmentation methodologies within the research for efficient dealing with the issue of the imbalance in the dataset. While there is some incorporation of variability within the training images for the models, much better generalization resulted in better classification performances. Application of multi-head attention considerably improved the models' ability to pay attention to the salient features in the images; this also promoted enhanced feature extraction and increased precision of the classifications.

Meanwhile, there are a couple of limitations: high accuracy was recorded, though natural variation in skin lesions seems to persist in causing misclassifications; performance is also affected by lighting conditions, the resolution of the image, and the site of lesions. Besides, the imbalance within the dataset-especially the small number of malignant images against benign ones-presents a big challenge. While augmentation is good, future work should collect even more diverse and fair data in order to further improve model robustness.

Another enormous challenge is how to integrate all these machine learning models into clinical settings. While the models do great in the lab, their translation to the field requires further validation. Apart from being accurate, models will have to be more interpretable to gain the trust of dermatologists and clinicians if they are to see wide adoption. Improvement in efforts toward explainability and transparency, such as heat map visualizations of critical -

TABLE II. RELATED WORK REPORTED RESULTS

Model	Accuracy
ResNet-50[4]	90%
EfficientNet-B4[5]	79.69%
VGG16 [6]	89.09%
Cascaded Ensemble Model [8]	85.3%%
Mobile-Optimized Algorithm[10]	98.5%
Hybrid CNN [11]	90.11%
Falcon Finch Deep CNN[13]	96.52%

-features that contribute most to the predictions, will enable translation into the wild.
Our model demonstrates high accuracy on the HAM10000 dataset, indicating its potential for skin lesion classification. However, we acknowledge that the dataset is small, which could affect the model's performance across different classes. To mitigate this, we employed data augmentation techniques. Future work could explore methods like Synthetic Minority Over-sampling Technique (SMOTE) and focal loss to further address class imbalance.
Additionally, while our model performs well on the HAM10000 dataset, its clinical applicability is limited by variability in real-world images, lighting conditions, and lesion heterogeneity. Validating the model on diverse datasets, such as the ISIC Challenge datasets, or through clinical trials would strengthen our claims.
We also recognize the importance of model interpretability in clinical settings. Incorporating visualization techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) can highlight critical areas of lesions, making the model's decision-making process more transparent and trustworthy for practitioners.
Taken all together, the results of our study demonstrate that ensemble deep learning models are promising ways of increasing the precision and reliability of diagnosis regarding skin cancers. With further refinement of algorithms on various datasets, in the near future, such systems may bring early detection into the realm of clinical dermatology and save lives due to timely interventions. Limitations These include, apart from the dataset imbalance discussed above, establishment of clinical utility in real-world clinical settings, which will be the focus of work in future studies.

XII. Conclusion

The work, therefore, presented the contribution of the deep learning approaches toward the detection of skin cancers, classifying them further into either benign or malignant lesions. In this regard, it would, therefore, be reiterated from the results that more complex models, including EfficientNet, ResNet, and DenseNet, are bound to further improve the accuracy and reliability of automatic skin

TABLE III. EXPERIMENTS RESULT ON DIFFERENT BACKBONES AND OUR PROPOSED METHOD RESULT

Model	Precision	Recall	F1Score	Accuracy
ResNet50[20]	0.94	0.94	0.94	93.91
DenseNet121[21]	0.94	0.94	0.94	94.10
DenseNet201	0.94	0.94	0.94	94.10
MobileNetV2[22]	0.94	0.94	0.94	94.15
EfficientnetV2B3[23]	0.95	0.95	0.95	95.14
Xception[24]	0.96	0.95	0.95	95.17
InceptionResNetV2[25]	0.96	0.95	0.96	95.28
InceptionV3	0.96	0.96	0.96	95.63
EfficientnetB2	0.96	0.96	0.96	96.25
Ensemble Model (EM)	0.97	0.97	0.97	96.78
EM+MultiHeadAttention	**0.97**	**0.97**	**0.97**	**97.11**

cancer detection systems when integrated with other strategies such as data augmentation and transfer learning. An ensemble coupled with multi-head attentions resulted in improved scores; hence, precision, recall, and F1 score are increased.

While this will sound quite good concerning performance on public benchmark datasets, challenges of class imbalance, variability within captured images, and interpretability within the clinical environment raise the bar toward translating performance across the real world for further research. But aside from those setbacks, deep learning in dermatology does hold great potential concerning early skin cancer diagnosis-in greatly improving patient outcomes with early and appropriate treatment.

Therefore, further research should be oriented toward enrichment of the dataset, in an effort to better capture skin cancer heterogeneity and improve model generalization. Of note, once AI-driven tools are ever implemented into clinical practice, they will need to be interpretable and actionable for health professionals. The belief that these AI-based skin cancer diagnosis systems may prove instrumental in fighting against this pandemic and deadly disease is due to continuous development in the fields of deep learning and medical image analysis.

1.1.1.1.1 *Acknowledgment*

We would like to express our sincere gratitude to all those who contributed to the success of this research. We are grateful to our colleagues and collaborators for their valuable insights and constructive feedback, which greatly enriched the study.

A special thanks to the teams and researchers behind the publicly available datasets, such as the HAM10000 dataset, which played a crucial role in the development and validation of our models. Without their dedication to open access data, this research would not have been possible.

1.1.1.1.2 References

[1] K. Simonyan, A. Zisserman, Very deep convolutional networks for largescale image recognition (2015). arXiv:1409.1556. URL https://arxiv.org/abs/1409.1556.

[2] G. Huang, Z. Liu, K. Weinberger, Densely connected convolutional networks (2016) 12.

[3] M. Tan, Q. V. Le, Efficientnet: Rethinking model scaling for convolutional neural networks (2020). arXiv:1905.11946. URL https://arxiv.org/abs/1905.11946.

[4] H. Kondaveeti, P. Edupuganti, Skin cancer classification using transfer learning, 2020. doi:10.1109/ICATMRI51801.2020.9398388.

[5] M. Harahap, A. Husein, S. Kwok, V. Wizley, J. Leonardi, D. Ong, D. Ginting, B. Silitonga, Skin cancer classification using efficientnet architecture, Bulletin of Electrical Engineering and Informatics 13 (2024) 2716–2728. doi:10.11591/eei.v13i4.7159.

[6] V. Anand, S. Gupta, A. Altameem, S. Nayak, R. Poonia, A. Saudagar, An enhanced transfer learning based classification for diagnosis of skin cancer, Diagnostics 12 (2022) 1628. doi:10.3390/diagnostics12071628.

[7] Q. Ha, B. Liu, F. Liu, Identifying melanoma images using efficientnet ensemble: Winning solution to the siim-isic melanoma classification challenge (2020). arXiv:2010.05351. URL https://arxiv.org/abs/2010.05351.

[8] A. Sharma, S. Tiwari, G. Aggarwal, N. Goenka, A. Kumar, P. Chakrabarti, T. Chakrabarti, R. Gono, Z. Leonowicz, M. Jasi´nski, Dermatologist-level classification of skin cancer using cascaded ensembling of convolutional neural network and handcrafted features based deep neural network, IEEE Access 10 (2022) 1–1. doi:10.1109/ACCESS.2022.3149824.

[9] C. Sunarya, J. Siswanto, G. Cam, F. Kurniadi, Skin cancer classification using delaunay triangulation and graph convolutional network, International Journal of Advanced Computer Science and Applications 14. doi:10.14569/IJACSA.2023.0140685.

[10] S. Osman, H. Maher, B. Sayed, Skin cancer detection using deep learning, Journal of the ACS Advances in Computer Science. URL https://api.semanticscholar.org/CorpusID:266258418.

[11] P. N. Mahalle, S. K. Shinde, P. Raka, K. Lodha, S. Mane, M. Malbhage, Skin cancer detection using cnn, 2023 International Conference on Computer Science and Emerging Technologies (CSET) (2023) 1–5. URL https://api.semanticscholar.org/CorpusID:266398345.

[12] M. M. Shukla, B. Tripathi, T. Dwivedi, A. Tripathi, B. K. Chaurasia, A hybrid cnn with transfer learning for skin cancer disease detection, Medical amp; biological engineering amp; computing 62 (10) (2024) 3057—3071. doi:10.1007/s11517-024-03115-x. URL https://doi.org/10.1007/s11517-024-03115-x.

[13] A. Kumar, M. Kumar, V. Bhardwaj, S. Kumar, S. Selvarajan, A novel skin cancer detection model using modified finch deep cnn classifier model, Scientific Reports 14. doi:10.1038/s41598-024-60954-2.

[14] M. Rajagopal, S. Ghate, R. P, E. Ganesh, Multi-class segmentation skin diseases using improved tuna swarm-based u-efficientnet, Journal of Engineering and Applied Science 71. doi:10.1186/s44147-024-00399-6.

[15] C. Supriyanto, A. Salam, J. Zeniarja, A. Wijaya, Two-stage input-space image augmentation and interpretable technique for accurate and explainable skin cancer diagnosis, Computation 11 (12). doi:10.3390/ computation11120246. URL https://www.mdpi.com/2079-3197/11/12/246.

[16] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

[17] https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images.

[18] F. Lin, C. Zhang, S. Liu, H. Ma, A hierarchical structured multi-head attention network for multi-turn response generation, IEEE Access PP (2020)1–1. doi:10.1109/ACCESS.2020.2977471.

[19] K. Ting, Confusion Matrix, 2017, pp. 260–260. doi:10.1007/ 978-1-4899-7687-1_50.

[20] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

[21] G. Huang, Z. Liu, K. Weinberger, Densely connected convolutional networks (2016) 12.

[22] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications.

[23] M. Tan, Q. V. Le, Efficientnetv2: Smaller models and faster training (2021). arXiv:2104.00298. URL https://arxiv.org/abs/2104.00298.

[24] F. Chollet, Xception: Deep learning with depthwise separable convolutions, 2017, pp. 1800–1807. doi:10.1109/CVPR.2017.195.

[25] F. Lin, C. Zhang, S. Liu, H. Ma, A hierarchical structured multi-head attention network for multi-turn response generation, IEEE Access PP (2020)1–1. doi:10.1109/ACCESS.2020.2977471.

Maryam Nazari, Cyberspace research institute, Shahid Beheshti University, Tehran, Iran

Orcid: 0009-0004-6309-4697

Fatemeh Fadaie Ardestani, Dept.of Computer Engineering, AI and Robotics, University of Isfahan, Isfahan, Iran

Orcid: 0009-0005-5865-4407