Skin Cancer Detection using Transfer Learning Models and Ensemble Approach to Enhanced Diagnostic Accuracy

[1] [1]
Khan[1], Hee-Cheol Kim[1,2]*

Md Ariful Islam Mozumder , Rashadul Islam Sumon , Tagne Poupi Theodore Armand , Mohammad Omair

[1] Digital Anti-Aging Health Care, Inje University, Gimhae, Republic of Korea

[2] Department of Computer Engineering, Inje University, Gimhae, Republic of Korea

arifulislamro@gmail.com, sumon39.cst@gmail.com, poupiarmand2@gmail.com, mumairkhan690@gmail.com, heeki@inje.ac.kr

***Abstract*—Skin cancer remains a critical global health concern, necessitating efficient and accurate diagnostic methods. Early diagnosis and proper identification could greatly increase skin cancer patients' survival and cure rates. This study explores the potential of advanced transfer learning models for detecting and classifying skin cancer. Each model leverages pre-trained weights, allowing fine-tuning to optimize the skin lesion dataset's performance. An ensemble approach integrates the strengths of these models, enhancing classification accuracy and robustness. The proposed methodology achieves 94% accuracy, demonstrating its effectiveness in distinguishing malignant and benign skin lesions. By combining state-of-the-art transfer learning architectures with ensemble learning, this study provides a reliable diagnostic tool, reducing reliance on traditional biopsy procedures. The results highlight the promise of artificial intelligence in dermatological applications, paving the way for improved early detection and treatment strategies.**

Keywords—skin cancer, image processing, deep learning, transfer learning, classification

I. INTRODUCTION

Skin cancer is one of the most frequent forms of cancer in the world, and melanoma is an extremely aggressive skin cancer, having one of the highest mortality rates. Traditional diagnostic methods are very subjective and depend completely on expert clinical judgment based on visual examination. Deep learning techniques, a subset of AI, have recently been developed, and these promise incredible results in the automated analysis of skin lesions.

The newest stride in deep learning technology, in cancer analysis including skin cancer detection and classification, shows considerable advances in different methodology directions, [1 - 11]. They proposed a two-tier framework that concentrated on data augmentation and the implementation of MVT. The proposed model performed well on the HAM10000 dataset while setting a new state-of-the-art by introducing an MVT-based model [12]. Integrating multi-modal data using a novel cross-modality-fusion module integrated into a transformer-based framework. Their approach achieved, on average, an increase in F1 score of 6.5% and improved the accuracy by 2.8% across five common skin tumor types on the Derm7pt dataset (1011 cases [13]. They have discussed the drawbacks of traditional CNN models. Though CNN efficiently extracts the features of small objects, it is not good at locating important parts within an image. Based on this observation, more studies on other architectures have been done. CNNs and transformers focusing on the HAM10000 dataset. Included in their study were VGGNet, ResNet, Vision Transformers, and DeepViT. In the study, they accounted for the balance in the dataset using weighted assignments. The results indicated that both techniques gave good results; however, CNN methods slightly outperformed in classifying skin cancer lesions [14]. In that work, they used two sets of pre-trained Vision Transformers with impressive results. The ViT-L32 model achieved 91.57% accuracy with 58.54% melanoma recall, while the ViT-L16 achieved 92.79% accuracy with 56.10% [15]. They proposed the modified lightweight vision transformer architecture combined with Ensemble Learning in MABSCNET [16]. Their hybrid framework reached state-of-the-art results on several datasets, such as 92.74% for ISIC 2020, 100% for ISIC 2018, and 94.24% for the Kaggle dataset. An optimized vision transformer approach was developed, where an elaborate preprocessing technique was performed. Based on the ISIC 2019 database, their approach obtained very good metrics: accuracy at 99.81%, precision at 96.65%, sensitivity at 98.21%, and Mathew's correlation coefficient at 98.89% [17].

The authors introduced SkinViT, integrating outlook, transformer, and MLP head blocks. Their architecture achieved varying accuracy rates across different datasets: 0.9109 on Dataset1 (ISIC2019), 0.8611 on Dataset2, and 0.8911 on Dataset3, demonstrating consistent performance improvements over existing methods [18]. This study proposed a hybrid model combining Residual Learning Machines, Swin Transformers, and Fast Neural Networks. Testing across ISIC-2008, PH-2, and HM007 datasets, their model achieved 98.78% classification accuracy, 98.7% precision, and an MCC of 0.9863 [19]. They compared CNNs and ViTs, finding that EfficientNet-B3 outperformed other models at 224×224 resolution. Their study utilized pre-trained models on ImageNet-21k and ImageNet 2012 datasets [20]. Evaluated multiple models, including ViTs, Swin transformers, and CNNs. Their results showed ResNet50 achieving the highest performance with 88.8% accuracy, while ViT models demonstrated competitive performance at 88.6% [21]. This study utilizing DinoV2 achieved a remarkable 96.48% test accuracy and 0.9727 F1-Score on a 31-class skin disease dataset, significantly improving benchmark results [22]. They compared various algorithms for monkeypox prediction using 1300 skin lesion images. Their Vision Transformer implementation achieved 93% accuracy, outperforming traditional machine-learning and CNN-based methods [23]. They integrated CNN, transformer, and InceptionV3 architectures, achieving 78.1% accuracy on

ISBN: 979-11-88428-13-7

the ImageNet dataset for melanoma classification [24]. They focused on histopathology-based diagnosis, employing transformer-based classification in an incremental learning setting. Their approach demonstrated superior results on histopathology datasets, emphasizing the importance of tissue-level analysis in skin cancer diagnosis. They have approached multi-instant learning for cancer analysis, which can also be used in skin cancer analysis [25].

II. METHODOLOGY

Our study's skin cancer classification process is structured around these main stages: data collection, data preprocessing, and cancer detection. This section presents the implementation of reliable methodologies, establishing an accurate and scalable framework for skin cancer classification while emphasizing innovation and precision in computational analysis.

A. Data Collection

The dataset employed in this study for skin cancer detection and classification originates from the ISIC 2018 dataset [19], publicly accessible via the Kaggle data repository. Specifically designed for distinguishing between malignant and benign skin lesions, the dataset consists of 3,297 RGB images, each with dimensions of 224x244 pixels and a total size of 340 MB. Among these, 1,800 images belong to the benign class, while 1,497 represent the malignant class. For model development and evaluation, the dataset was carefully partitioned into training and testing subsets. The training set includes 1,440 benign images and 1,197 malignant images, whereas the testing set comprises 360 benign and 300 malignant images shown in Table 1.

TABLE I. DISTRIBUTION OF DATA

Class	*Training*	*Testing*
Benign	1440	360
Malignant	1197	300
Total	2637	660

This deliberate dataset organization ensures a well-balanced and robust framework for training and rigorously evaluating deep learning models aimed at accurate skin cancer detection and classification. Figure 1 shows a sample of the dataset images.

B. Data Preprocessing

The data augmentation strategy focuses on increasing the dataset's diversity and robustness to improve model performance. The original images are retained as a baseline for comparison. Three primary augmentation techniques with specific parameter configurations are applied to address data imbalance and reduce the risk of overfitting during training. Data augmentation generates additional samples for underrepresented classes by applying random transformations to existing images. Techniques such as rotation, translation, Gaussian blur, multiplication, and flipping are used to enhance the dataset and improve the model's classification accuracy.

C. Workflow

The workflow for skin cancer classification begins with acquiring skin lesion images, resized to 224x224 pixels, followed by preprocessing and augmentation to enhance model generalizability. Preprocessing includes resizing, noise filtering, flipping, rotation, and intensity adjustments. The dataset is split into training and testing subsets, with transfer learning models like Inception, Xception, VGG19, and ResNet50 fine-tuned for classification. Fine-tuning involves adjusting final layers for malignant and benign patterns while freezing earlier layers to retain general features. An ensemble learning approach combines predictions from multiple models, leveraging their strengths to improve classification accuracy and reduce bias. Metrics such as accuracy, precision, recall, and F1 score evaluate model performance, ensuring reliable identification of malignant lesions. This workflow addresses challenges like data scarcity, class imbalance, and overfitting, providing a robust tool for skin cancer detection. Figure 2 shows the workflow for this study.

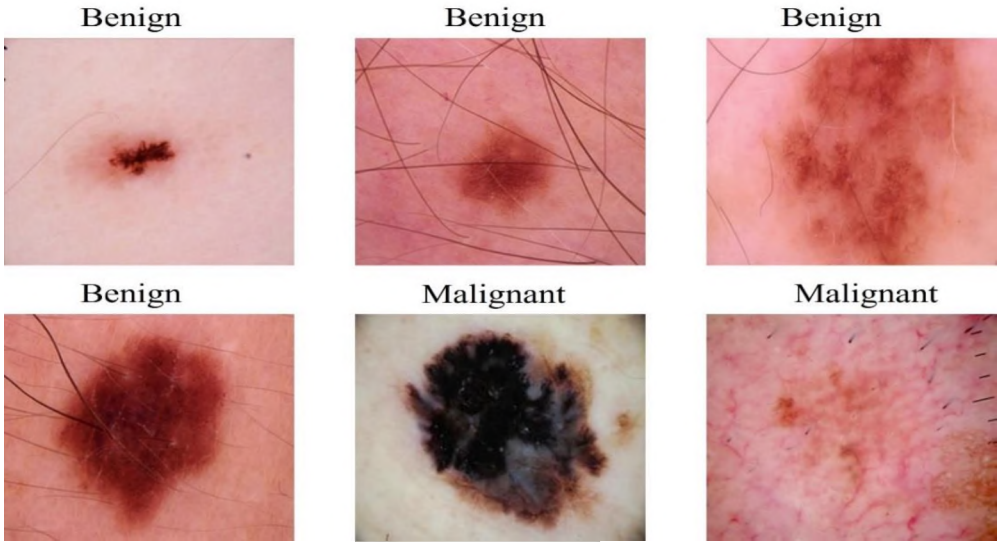

Figure 1. Sample of data from classes

ISBN: 979-11-88428-13-7

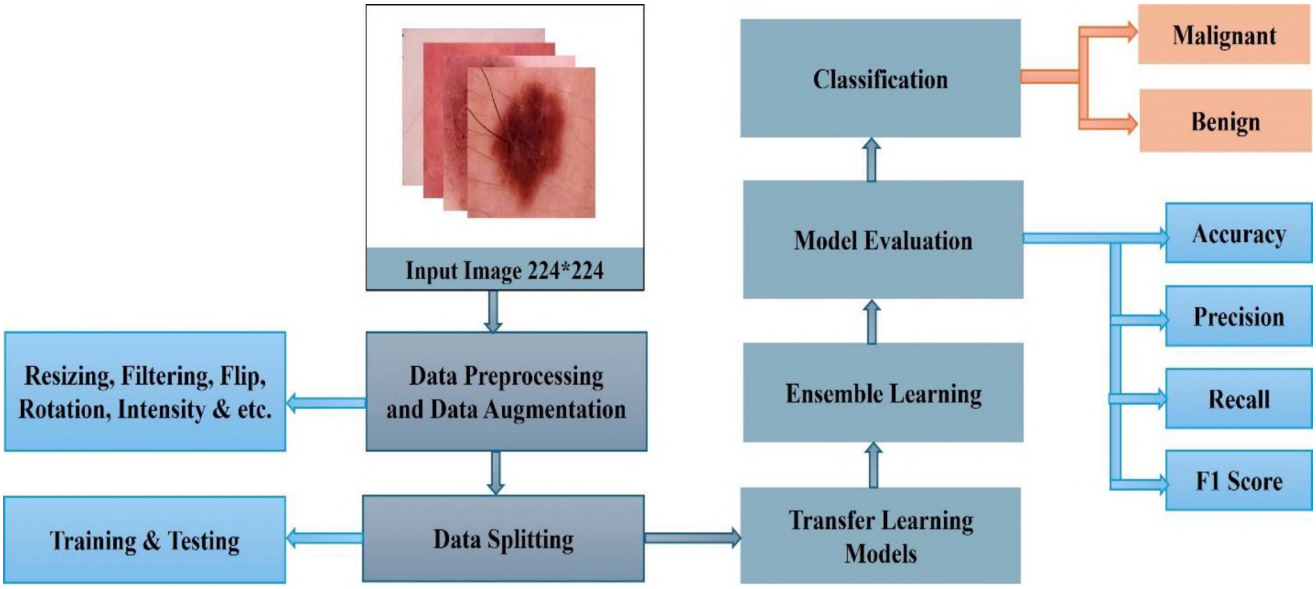

Figure 2. The study workflow

D. *Ensemble Approch*

The figure illustrates our ensemble learning approach that integrates predictions from multiple deep learning models Inception, VGG19, Xception, and ResNet50 for skin cancer classification. Input data is initially processed and trained on each model, with each model independently predicting whether the skin lesion is benign or malignant. The predictions from all models are then aggregated using a mean-based strategy, producing a final prediction representing an averaged or consensus output. This ensemble approach leverages the unique strengths of each model, enhancing the classification accuracy and robustness compared to using a single model alone. By combining these models, the ensemble approach reduces prediction variance and mitigates the limitations of individual models, resulting in a more reliable diagnostic outcome for skin cancer detection. Figure 3 shows the ensemble approach.

III. RESULTS

The details of the experimental findings are covered in this section.

A. *Performance Evaluation*

The suggested model architecture was evaluated for its performance in predicting skin cancers using many performance assessment indicators. Accuracy, recall, precision, and F1 score are the four measures.

Accuracy: The ratio of the number of classes a model successfully predicts to the total number of predictions

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

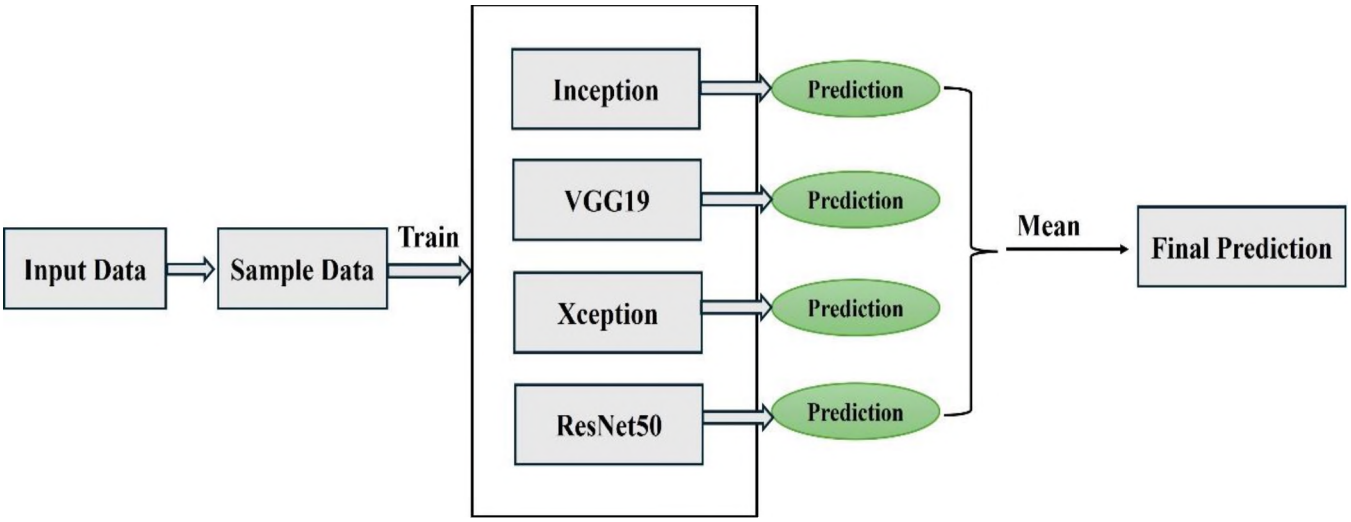

Figure 3. Transfer ensemble method for this study

ISBN: 979-11-88428-13-7

Precision: Precision is defined as the proportion of the number of correct predictions divided by the total number of positive class predictions

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

Recall: Recall is the proportion of correct predictions divided by the number of actual counts of the positive class in the dataset.

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

F1 score: The F1 score represents the balance between Precision and Recall

$$F1 - Scroe = \frac{2*(Rec*Pre)}{(Rec+Pre)} \quad (4)$$

Here, True Positives (TP) and False Negatives (FN) are the numbers of positive images accurately predicted. In contrast, the number of incorrectly anticipated negative images is known as False Positives (FP), while the number of accurately predicted negative images is known as True Negatives (TN).

The overall performance of the models can be visualized using the Area Under the Curve (AUC) of the Receiver Characteristic Operator (ROC), below in Figure 4 & the prediction outcome of the ensemble classification model is summarized in the confusion matrix shown in Figure 5.

Figure 4: Validation accuracy & loss for models.

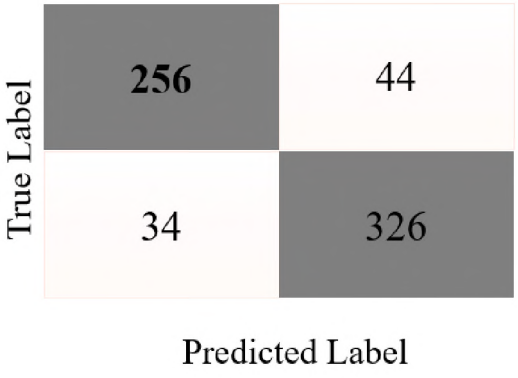

Figure 5: Confusion matrix for ensemble model.

TABLE II. COMPARISON TABLE

Models	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1-Score (%)*
VGG19	93.00	93.00	93.00	93.00
InceptionV3	90.50	91.00	91.00	91.00
Xception	89.65	89.90	89.90	89.90
ResNet50	93.50	93.50	93.50	93.50
Proposed Ensemble	94.50	94.50	94.50	94.50

IV. CONCLUSION

This study demonstrates the efficacy of an ensemble learning approach using advanced transfer learning models Inception, VGG19, Xception, and ResNet50 for accurate skin cancer classification. Leveraging pre-trained weights and fine-tuning each model, followed by a consensus-based ensemble method, the proposed system achieves high accuracy, distinguishing between malignant and benign lesions with robust performance. This technique holds promise for enhancing early skin cancer detection, offering a non-invasive, reliable diagnostic tool. The findings underscore the potential of AI-driven solutions in supporting healthcare professionals in dermatology.

ACKNOWLEDGMENT

This research was supported by a grant from Korea Health Technology R&D Project through the Korea Health Industry Development Institute validation (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (Grant No: HI21C0977) and the MSIT (Ministry of Science ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation) in 2022 (2022-0-01091, 1711175863).

REFERENCES

[1] Aladhadh, Suliman, Majed Alsanea, Mohammed Aloraini, Taimoor Khan, Shabana Habib, and Muhammad Islam. "An effective skin cancer classification mechanism via medical vision transformer." *Sensors* 22, no. 11 (2022): 4008.

[2] Xu, J., Gao, Y., Liu, W., Huang, K., Zhao, S., Lu, L., Wang, X., Hua, X.S., Wang, Y. and Chen, X., 2022, September. RemixFormer: a transformer model for precision skin tumor differential diagnosis via multi-modal imaging and non-imaging data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 624-633). Cham: Springer Nature Switzerland.

[3] Sumon, Rashadul Islam, et al. "Enhanced Nuclei Segmentation in Histopathology Image Leveraging RGB Channels through Triple-Encoder and Single-Decoder Architectures." *2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2023.

[4] Rahman, Wahidur, et al. "A Deep CNN-Based Salinity and Freshwater Fish Identification and Classification Using Deep Learning and Machine Learning." *Sustainability* 16.18 (2024): 7933.

[5] Sumon, Rashadul Islam, et al. "A Deep Learning-Based Approach for Precise Emotion Recognition in Domestic Animals Using EfficientNetB5 Architecture." *Eng* 6.1 (2025): 9.

[6] Mozumder, Md Ariful Islam, et al. "Automated Pipeline for Robust Cat Activity Detection Based on Deep Learning and Wearable Sensor Data." *Sensors* 24.23 (2024): 7436.

[7] Hussain, Ali, Md Ariful Islam Mozumdar, and Hee Cheol Kim. "Predicting The Survival and Death of Patients with Idiopathic Pulmonary Fibrosis Disease using SVM, Random Forest and Naïve Bayes."

ISBN: 979-11-88428-13-7

[8] Hosen, Md Nayeem, et al. "Prediction of Chronic Kidney Disease Using Machine Learning." *2023 The 20th World Congress of the International Fuzzy Systems Association (IFSA)*. 2023.

[9] Theodore Armand, Tagne Poupi, et al. "ELIPF: Explicit Learning Framework for Pre-Emptive Forecasting, Early Detection and Curtailment of Idiopathic Pulmonary Fibrosis Disease." *BioMedInformatics* 4.3 (2024): 1807-1821.

[10] Xin, C., Liu, Z., Zhao, K., Miao, L., Ma, Y., Zhu, X., Zhou, Q., Wang, S., Li, L., Yang, F. and Xu, S., 2022. An improved transformer network for skin cancer classification. *Computers in Biology and Medicine*, *149*, p.105939.

[11] Zhao, Z., 2022, December. Skin cancer classification based on convolutional neural networks and vision transformers. In *Journal of Physics: Conference Series* (Vol. 2405, No. 1, p. 012037). IOP Publishing.

[12] Flosdorf, C., Engelker, J., Keller, I. and Mohr, N., 2024. Skin Cancer Detection utilizing Deep Learning: Classification of Skin Lesion Images using a Vision Transformer. *arXiv preprint arXiv:2407.18554*.

[13] Reis, H.C. and Turk, V., 2024. Fusion of transformer attention and CNN features for skin cancer detection. *Applied Soft Computing*, *164*, p.112013.

[14] Desale, R.P. and Patil, P.S., 2024. An efficient multi-class classification of skin cancer using optimized vision transformer. *Medical & Biological Engineering & Computing*, *62*(3), pp.773-789.

[15] Khan, S. and Khan, A., 2023. SkinViT: A transformer-based method for Melanoma and Nonmelanoma classification. *Plos one*, *18*(12), p.e0295151.

[16] Ramkumar, K., Medeiros, E.P., Dong, A., de Albuquerque, V.H.C., Hassan, M.R. and Hassan, M.M., 2024. A novel deep learning framework based swin transformer for dermal cancer cell classification. *Engineering Applications of Artificial Intelligence*, *133*, p.108097.

[17] Ashfaq, M. and Ahmad, A., 2023. Skin Cancer Classification with Convolutional Deep Neural Networks and Vision Transformers Using Transfer Learning. In *Advances in Deep Generative Models for Medical Artificial Intelligence* (pp. 151-176). Cham: Springer Nature Switzerland.

[18] Dagnaw, G.H., El Mouhtadi, M. and Mustapha, M., 2024. Skin cancer classification using vision transformers and explainable artificial intelligence. *Journal of Medical Artificial Intelligence*.

[19] Mohan, J., Sivasubramanian, A., Sowmya, V. and Vinayakumar, R., 2024. Enhancing skin disease classification leveraging transformer-based deep learning architectures and explainable ai. *arXiv preprint arXiv:2407.14757*.

[20] Kundu, D., Siddiqi, U.R. and Rahman, M.M., 2022, December. Vision transformer based deep learning model for monkeypox detection. In *2022 25th International Conference on Computer and Information Technology (ICCIT)* (pp. 1021-1026). IEEE.

[21] Telagam, N. and Kandasamy, N., 2023. Classification of Melanoma Skin Cancer Based on Transformer Deep Learning Model. In *Ecological and Evolutionary Perspectives on Infections and Morbidity* (pp. 208-227). IGI Global.

[22] Imran, M., Akram, M.U. and Salam, A.A., 2024, July. Transformer-Based Skin Carcinoma Classification using Histopathology Images via Incremental Learning. In *2024 14th International Conference on Pattern Recognition Systems (ICPRS)* (pp. 1-7). IEEE.

[23] Codella, Noel, et al. "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)." *arXiv preprint arXiv:1902.03368* (2019).

[24] Sumon, Rashadul Islam, et al. "Exploring Deep Learning and Machine Learning Techniques for Histopathological Image Classification in Lung Cancer Diagnosis." 2024 International Conference on Electrical, Computer and Energy Technologies (ICECET. IEEE, 2024.

[25] Mozumder, Md Ariful Islam, et al. "Multiple instance learning: state-of-the-art in cancer analysis, applications & future potential." *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET*. IEEE, 2024.

Md Ariful Islam Mozumder is pursuing his Ph.D. in the Institute of Digital Anti-Aging Healthcare & Computer Science at Inje University. His research interest aligns with Deep Learning, Computer Vision, Medical Image Processing, Digital Pathology Images, Weakly Supervised Learning, Foundation Models, Sensor Data Analysis, Bio Signal Processing, NLP, and Blockchain.

Rashedul Islam Sumon is pursuing his Ph.D. in the Institute of Digital Anti-Aging Healthcare at Inje University. His research interest aligns with Computer Vision, Medical Image Processing, Metaverse, Artificial Intelligence, and Bio Signal Processing.

Tagne Poupi Theodore A. is a Ph.D. student at the Institute of Digital Anti-aging and Healthcare at Inje University. His research interests include Computer Vision, Metaverse, Image Processing, Deep Learning, and Machine Learning.

Mohammad Omair Khan is pursuing his Ph.D. in the Institute of Digital Anti-Aging Healthcare at Inje University. His research interest aligns with Computer Vision, Medical Image Processing, Metaverse, Artificial Intelligence, and Bio Signal Processing.

Hee-Cheol Kim Ph.D. at Numerical Analysis and Computing Science, at Stockholm University in Sweden. He is a professor and Head of Institute of the Digital Anti-aging Healthcare, Inje University, South Korea. His research interests include Machine learning, Computer Vision, Text mining, Bioinformatics, Blockchain, Metaverse, and XAI.