VERİ BİLİMİ DERGİSİ
www.dergipark.gov.tr/veri

SHAP ve LRP Analizlerinden Elde Edilen İçgörülerle CNN Modelleri Aracılığıyla Cilt Hastalığı Sınıflandırma Doğruluğunun Artırılması

Cem ÖZKURT[1, 2]

¹Sakarya Uygulamalı Bilimler Üniversitesi, Teknoloji Fakültesi Bilgisayar Mühendisliği Bölümü
² Sakarya Uygulamalı Bilimler Üniversitesi, Yapay Zeka ve Veri Bilimi Uygulama ve Araştırma Merkezi

Özet

Bu çalışma, HAM10000 veri kümesini ve evrişimli sinir ağı (CNN) modellerini kullanarak cilt hastalığı sınıflandırmasını derinlemesine incelemektedir. Titiz veri hazırlığı ve kapsamlı model eğitimi yoluyla, normal ve hastalıklı cilt koşulları arasındaki ayrımı yüksek doğrulukla başarıyoruz. SHAP modelini kullanmak, CNN modelinin karar verme sürecine dair değerli içgörüler sunarak tahmin yorumlanabilirliğini artırır.

Çalışmamız, bireyler arasında normal ve hasta olanlar arasında ayırma konusunda umut verici doğruluk sergilemektedir. 5000 kişiden 4274'ü doğru şekilde normal olarak sınıflandırılırken, 5015 kişiden 2962'si doğru şekilde hastalıklı olarak tanımlanmıştır. Ancak, modelimiz önemli hatalar göstermektedir, özellikle 726 normal kişiyi hasta olarak yanlış sınıflandırır ve yanlış negatifleri azaltmada iyileştirme alanları sunar. SHAP ve LRP analizlerini kullanarak, ortalama değerlerinin sırasıyla $2.11251442 \times 10^{-5}$ ve yaklaşık olarak 0.032795×10^{-5} olduğunu gözlemledik, bu da özellik önemini ve model davranışını anlamada değerli içgörüler sağlar. Bu bulgular, tıbbi uygulamalarda tanı doğruluğunu artırma ve yanlış sınıflandırmaları azaltma potansiyelini vurgular.

Karışıklık matrisleri ve SHAP ve LRP modellerinden elde edilen çıktılar da dahil olmak üzere çeşitli görsel temsiller kullanarak, CNN modelinin güçlü yanları ve sınırlamaları hakkında kapsamlı perspektifler sunuyoruz ve genel performansı artırmayı amaçlayan potansiyel iyileştirmelere rehberlik ediyoruz.

Normal ve hastalıklı bireyler arasında dengeli sınıflandırma sağlamamıza rağmen, yanlış sınıflandırmaları azaltmak ve genel doğruluğu artırmak için daha fazla geliştirme gerekmektedir. SHAP ve LRP çıktılarının derinlemesine sayısal analizi, SHAP'ın LRP'den daha detaylı bir analiz sunduğunu ortaya koymaktadır, bu da onu bu bağlamda tercih edilen metodoloji olarak konumlandırır.

Bu araştırma, yapay zeka destekli cilt hastalığı teşhisi alanındaki ilerlemeye önemli katkıda bulunmaktadır ve özellikle dermatoloji uygulamalarında CNN modellerinin sağlık uygulamalarındaki potansiyelini vurgular. Gelecekteki çalışmaların, klinik karar verme sürecini güçlendirmeye odaklanması ve böylece dermatolojik uygulamalarda hastaların sonuçlarını ilerletmesi gerekmektedir.

Anahtar Kelimeler: *SHAP, Dermatoloji, Layerwise Relevance Propagation (LRP), Açıklanabilir YZ, CNN*

Enhancing Skin Disease Classification Accuracy through CNN Models: Insights from SHAP and LRP Analyses

Abstract

This study presents an in-depth exploration of skin disease classification utilizing the HAM10000 dataset and convolutional neural network (CNN) models. Through meticulous data preparation and extensive model training, we achieve high accuracy in distinguishing between normal and diseased skin conditions. Employing the SHAP model provides valuable insights into the decision-making process of the CNN model, enhancing prediction interpretability.

Our study demonstrates promising accuracy in distinguishing between normal individuals and patients, with 4274 out of 5000 correctly classified as normal and 2962 out of 5015 accurately identified as diseased. However, our model exhibits significant errors, notably misclassifying 726 normals as patients and displaying areas for improvement in reducing false negatives. Leveraging SHAP and LRP analyses, we observed an average value of $2.11251442 \times 10^{-5}$ and approximately 0.032795×10^{-5}, respectively, suggesting valuable insights into feature importance and model behavior. These findings underscore the potential for enhancing diagnostic accuracy and mitigating misclassifications in medical applications.

Utilizing various visual representations, including confusion matrices and outputs from SHAP and LRP models, we give comprehensive perspectives on the strengths and limitations of the CNN model, guiding potential refinements aimed at enhancing overall performance.

Despite achieving balanced classification between normal and diseased individuals, further enhancements are warranted to reduce misclassifications and improve overall accuracy. In-depth numerical analysis of SHAP and LRP outputs reveals differences in interpretation capabilities, with SHAP providing a more detailed analysis than LRP, positioning it as the preferred methodology in this context.

This research significantly contributes to the advancement of AI-driven skin disease diagnosis and underscores the potential of CNN models in healthcare applications, particularly in dermatological practice. Future endeavors should focus on enhancing methodologies to bolster clinical decision-making, thereby advancing patient outcomes within dermatologic practice.

Keywords: *SHAP, Dermatology, Layerwise Relevance Propagation (LRP), Explainable AI, CNN*

* Corresponding e-mail: cemozkurt@subu.edu.tr

1 Introduction

The imperative role of interpretable machine learning models in promoting transparency and understanding complex decision-making processes is paramount. Algorithms like LRP (Layer-wise Relevance Propagation) and SHAP (SHapley Additive exPlanations) have emerged as formidable tools for unraveling the intricacies of black-box models, contributing to transparency, reliability, and improved decision-making processes in diverse domains [1,2]. To demonstrate the practical application of these algorithms, we embarked on artificial intelligence training using the HAM10000 dataset in Python, with a specific emphasis on a dermatology model. The subsequent sections of our code meticulously illustrate how SHAP and LRP values decode the decisions of a pre-trained dermatology model. This real-world Python implementation aims to bridge the theoretical understanding and practical application gap, offering a tangible perspective on the transformative benefits of SHAP and LRP in elucidating machine learning models.

The realm of Explainable AI (xAI) has garnered considerable attention across diverse domains, particularly in the medical sector, where transparency and traceability are imperative for opaque AI/ML systems in healthcare[1,2]. A notable method in xAI, Layer-wise Relevance Propagation (LRP), focuses on elucidating relevant components of inputs and representations in neural networks, thereby enhancing the interpretability of decision-making processes [2]. Explainable AI (xAI) is being used in medicine for tasks like helping doctors diagnose diseases from X-rays, like COVID-19. This helps doctors understand how the AI makes its decisions. Techniques like LRP and FLRP are useful for this. [3,4]. The advent of explainable AI addresses the black-box problem in deep learning models, with Selective Layer-wise Relevance Propagation proposed to generate clearer heatmaps for model predictions [5]. These studies underscore xAI's significance in medicine and the development of methods like LRP and FLRP to enhance transparency and interpretability in AI/ML systems.

Chlebus et al. [6] provided an approach to explain semantic segmentation networks through layer-wise relevance propagation, offering a valuable means for model validation and improvement. A study by Ahmed et al. [7] developed a system to automatically analyze medical images. They used a special kind of AI model (generative adversarial network) to identify polyps and instruments in endoscopy images. They also included a method (layer-wise relevance propagation) to explain how the model makes its decisions. Karim et al. [4] presented a transparent deep neural network approach for identifying COVID-19 symptoms in chest radiography images, utilizing layer-wise relevance propagation for human-interpretable explanations. Alam et al. [8] investigated the application of layer-wise relevance propagation in interpreting deep neural networks within the medical domain, focusing on specific pathology classes. Bassi et al. [9] employed LRP to improve the interpretability of deep neural convolutional networks for COVID-19 detection using chest X-rays. Collectively, layer-wise relevance propagation has been effectively applied in various medical image analysis tasks, highlighting its potential in providing human-interpretable explanations for predictions made by deep neural networks in the medical and healthcare domain [6, 7, 4, 8, 9].

The comprehensibility of deep neural networks has emerged as a central issue in the medical and healthcare sector, especially concerning artificial intelligence (AI) and medical imaging. This focus arises from apprehensions regarding clarity, legal and ethical implications, and the clinical significance of predictions produced by deep neural networks within clinical decision support systems [8]. Explainable Deep Learning has gained traction in AI, especially in medical imaging, where accurate and interpretable machine learning models are vital for effective diagnosis and treatment planning[10]. Explainable AI techniques elucidate a machine learning model's predictive behavior using human-interpretable features and predictions, without relying on the model's architecture, parameters, or training strategies [11]. In multi-instance learning, attention pooling helps explain the results for each individual image, but some medical uses need even

more detail-explanations for each individual pixel in the image. Current models are not yet able to do this. [12]. Magnetic resonance imaging is crucial for prostate cancer detection, and AI systems can assist in radiological evaluation by categorizing lesions as either clinically significant (csPCa) or nonclinically significant (ncsPCa). [13]. These studies underscore the potential of Explainable Deep Learning and various interpretability methods, such as SHAP, Grad-CAM, and LRP, in enhancing the accuracy and interpretability of deep learning models in medical imaging [8, 10, 12]. They seek to deliver a thorough evaluation via both quantitative and qualitative analyses, furnishing clinicians with insights to enhance their comprehension and confidence in decisions made by computer-aided diagnosis systems [12]. Furthermore, there's a strong emphasis on conducting a comprehensive evaluation that encompasses all sub-modules within the system, underscoring the importance of adopting a multimodal-based approach that integrates various sources of information beyond medical images to bolster interpretability in the medical field. [8, 13].

2 General methods

This research deals with the classification of different skin diseases. They used a large collection of images called HAM10000 to train a model. Then, techniques such as SHAP and LRP were used to understand how the model makes its decisions. The steps involved in this process were presented in Figure 1.

2.1 Dataset

We have a valuable resource for skin cancer research. This collection contains over 10,000 images of skin growths, including melanoma, a serious form of skin cancer. The dataset is like a comprehensive encyclopedia of skin conditions, meticulously organized for scientists to study. Each image captures intricate details of the growths, providing a wealth of information for analysis. These details include color variations, shape characteristics, size measurements, and other visual features.

The HAM10000 dataset, introduced by Tschandl, Rosendahl, and Kittler in 2018 [14], encompasses a vast array of dermatoscopic images portraying common pigmented skin lesions. Subsequent to the release of this dataset, scholars such as Nugroho, Slamet, and Sugiyanto in 2019 employed convolutional neural networks (CNNs) for discerning skin cancer within the HAM10000 dataset [15]. Additionally, Khan, Javed, Sharif, Saba, and Rehman proposed a methodology focused on deep neural networks for feature extraction and optimality selection in dermatological skin lesion classification, as outlined in their publication from April 2019 [16]. These endeavors underscore the ongoing commitment within the scientific community to harness advanced computational techniques for enhanced diagnosis and comprehension of skin conditions.

2.2 CNN - Convolutional neural network

Li et al. (2021) conducted a comprehensive survey of CNNs, analyzing their evolution, applications, and future prospects [17]. Their study not only
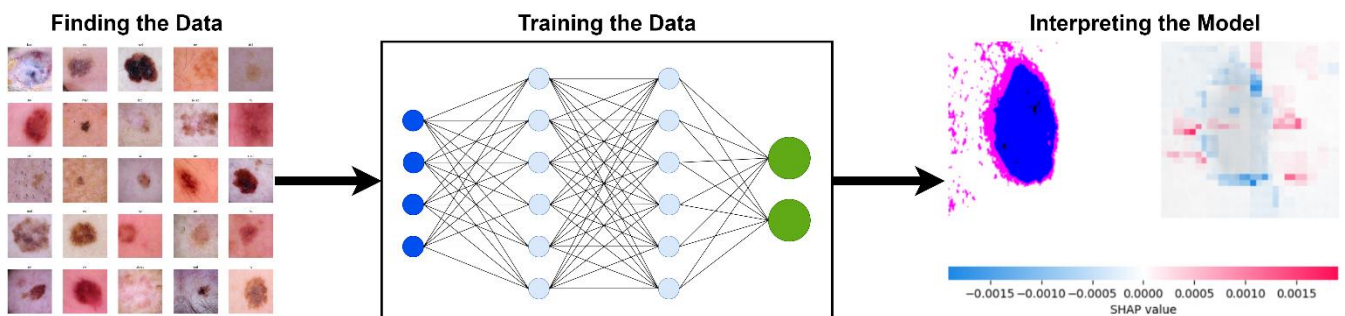

Figure 1. The advancement of the article

summarized the key findings from previous research but also identified emerging trends and potential directions for further exploration. The study highlighted how CNNs are used for many things, like recognizing images, understanding language, diagnosing diseases, and even self-driving cars.

CNNs are a powerful type of deep learning model, especially good at making sense of pictures and videos. Albawi et al. (2017) provided a foundational understanding of CNNs, outlining their architecture and operation principles [18]. They came up with three key building blocks for CNNs: fully connected layers, convolutional layers and pooling layers.

Subsequent research, as exemplified by Gu et al. (2018) [19], delved into advanced techniques and algorithms that have significantly improved the performance of CNNs. These advancements include novel convolutional operations, regularization methods, and optimization algorithms tailored specifically for CNN architectures. Gu and his team explored these developments, providing insights into the latest trends and breakthroughs in CNN research.

In sum, these studies add to our comprehensive comprehension of Convolutional Neural Networks (CNNs) and their extensive array of applications. By automatically learning hierarchical representations from raw data, CNNs excel at feature extraction and pattern recognition tasks. As CNN technology continues to evolve and integrate with other disciplines, such as reinforcement learning and generative modeling, they hold tremendous promise for addressing complex real-world challenges and driving innovation across various domains.

2.2.1 CM – The confusion matrix

The CM-confusion matrix acts as a one-stop shop for assessing how well a classification model performs. It provides a clear breakdown of the model's predictions compared to the real results. This breakdown is captured in four key categories, each revealing a different aspect of the model's strengths and weaknesses.

TP – True Positive: The model correctly guesses something is positive, and it actually is positive (like catching a spam email).

TN – True Negative: The model correctly guesses something is negative, and it actually is negative (like letting a real email through).

FP – False Positive: The model mistakenly guesses something is positive, but it's actually negative (like flagging a normal email as spam). This is also called a Type I error.

FN – False Negative: The model mistakenly guesses something is negative, but it's actually positive (like missing a spam email). This is also called a Type II error.

This table, called a confusion matrix, is a powerful tool for evaluating how well a classification algorithm performs. It helps us calculate the recall, precision, accuracy and F1 score as important measures. By breaking down the model's predictions into these categories, we can gain a clearer picture of the classification system's strengths and weaknesses. These metrics provide a more detailed understanding of how well the model makes accurate predictions, finds positives and negatives in true types, and identifies their effects of errors and types. This allows us to make better choices about how to improve and optimize the model. Figure 2 shows a confusion matrix with it's values.

Actual Values

	Positive	Negative
Positive	TP	FP
Negative	FN	TN

Predicted Values

Figure 2. An example of a confusion matrix

2.3 XAI – Explainable artificial ıntelligence

XAI research is committed to advancing the visibility and interpretive capabilities of artificial intelligence (AI) systems, enabling humans to understand the decisions and processes underlying their functioning. In this section, we offer a comprehensive overview of seminal studies that have significantly contributed to the advancement of XAI. Confalonieri et al. [20] provide an insightful historical analysis of XAI, tracing its evolutionary journey and elucidating key milestones that have shaped its trajectory over time. Their study not only provides valuable insights into the origins of XAI but also offers a roadmap for guiding future research endeavors in this domain. Angelov et al. [21] offer a comprehensive analytical review of XAI, critically examining the latest developments and emerging trends in the field. By synthesizing existing literature, their study provides a nuanced understanding of the current state of XAI and identifies potential avenues for future exploration. Furthermore, Došilović et al. [22] conduct an extensive survey on XAI, meticulously cataloging various explainability techniques and methodologies employed in the field. The meticulous scrutiny of these outputs emerges as a pivotal resource for practitioners and researchers endeavoring to navigate the intricate realm of eXplainable Artificial Intelligence (XAI) and devise effective strategies for enhancing explainability in AI systems. Additionally, Tjoa and Guan [23] focus specifically on the application of XAI in the medical domain, exploring its potential implications for healthcare delivery and patient outcomes. Through their in-depth investigation, they shed light on the unique challenges and opportunities associated with integrating XAI into clinical practice, offering insights that can inform the development of ethically sound and clinically relevant AI-driven solutions. Collectively, these seminal studies illuminate the multifaceted nature of XAI and provide valuable guidance for future research efforts aimed at advancing the field and realizing its full potential in real-world applications.

2.4 SHapley additive exPlanations (SHAP)

SHAP is a methodology employed for elucidating the predictions made by machine learning models [24]. It assesses the contribution of individual features or variables to a prediction, serving to comprehend these contributions. SHAP elucidates the rationales behind a prediction through the computation of a SHAP value corresponding to each feature [24]. This value indicates why the model's output deviates from the expected value.

SHAP employs the concept of "shapley values" to make these explanations. The shapley values measure the contribution of each player in a group to win or lose a game while representing different scenarios [24]. Similarly, SHAP determines how much each feature or variable contributed to a prediction. This is used to assess the importance of each feature or variable in a prediction and to understand the behavior of the model.

SHAP can be used with various machine learning models, but it is particularly useful in complex models, especially those referred to as black-box models [24]. SHAP functions as a vital instrument for grasping the internal operations of the model and producing trustworthy predictions.

$$SHAP_G(X) = \sum_k \lambda_k SHAP_{F_k}(X) \qquad (1)$$

As seen in equation 1:

$SHAP_G(X)$ = The SHAP value ($SHAP_G(X)$) for a specific data point (X) shows how much each feature affects the model's final prediction.

Σ_k = This symbol signifies an addition over a group of values, particularly over diverse SHapley values.

λ_k weights show how much influence each SHAP value has on the overall SHAP value ($SHAP_G(X)$). They basically act like importance scores for the features.

$SHAP_{F_k}(X)$ denotes the SHapley value allocated to a certain feature or collection of features, calculated using a designated SHapley value estimation methodology. It assesses the discrepancy in the predicted outcome of the model (E[F]) from the actual output (F(x)).

$$\phi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (2)$$

As seen in equation 2:

$\phi_i(f)$ is the i Shapley value for the relevant in model f.

$\sum_{S \subseteq N \setminus \{i\}}$ is a summation over all subsets S of the set N, excluding i. This element incorporates every possible arrangement of characteristics, barring feature i.

$\frac{|S|!(|N|-|S|-1)!}{|N|!}$ binomial coefficient, denoting the selection of |S| elements out of |N| elements, ensures a fair distribution of contributions.

[f(S ∪ {i}) – f(S)] is representing the disparity in the model's forecast when feature i is encompassed in the subset |S| versus when it is excluded. It measures the additional influence of feature i on the model's outcome.

Figure 3 is showing SHAP analysis along with the original image of a random data extracted from the dataset.
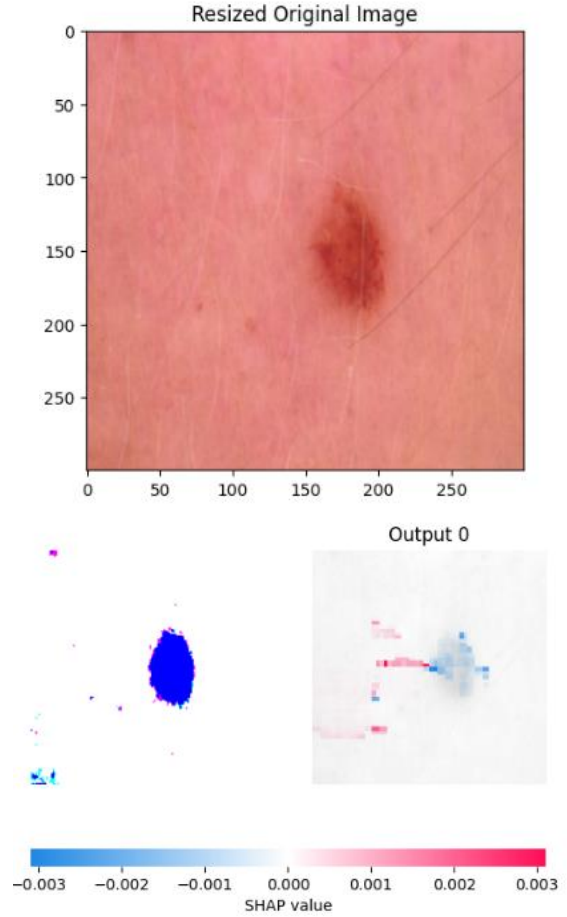

Figure 3. SHAP analysis for a randomly selected image from the dataset

2.5 Layer-wise Relevance Propagation (LRP)

In the world of Explainable Artificial Intelligence (XAI), Layer Relevance Propagation (LRP) is an important detective that sheds light on the decision-making process of complex machine learning models, especially neural networks. LRP works by assigning importance scores to individual pieces of information fed into the model, revealing how much each one influences the final outcome. It does this by meticulously backtracking through the layers of

the model, taking into account the influence of each "neuron" and its connections. This allows us to see inside the "black box" of the model, making it more reliable, especially in sensitive areas such as healthcare where transparency is crucial. LRP is a versatile tool, compatible with a variety of neural network designs, from those that recognize images (convolutional neural networks) to those that understand language (recurrent neural networks). As a result, LRP is a driving force in the exciting field of Explainable Artificial Intelligence.

Figure 4 shows the LRP analysis with the original image of a data randomly taken from the dataset.

$$R_i = \sum_j \frac{z_{ij}.R_j}{z_j} \tag{3}$$

R_i is the relevance of the input feature i.

Σ_j represents the sum over all neurons in the layer to which the input feature i is connected.

z_{ij} is the connection weight between the input feature i and the neuron j.

R_j is the relevance of neuron j in the layer above (closer to the output).

z_j is the total activation of neuron j in the layer above.

3 Results and Discussion

First, we found a dataset named HAM10000 that includes skin diseases. Subsequently, we resized the dataset to make it suitable for training. Later on, we conducted an analysis of the trained model by applying the SHapley model. Using this approach, we were able to pinpoint distinct regions within the scrutinized images that were projected to be afflicted by the condition and those categorized as non-pathological. Subsequently, we analyzed the model we trained by subjecting it to the SHapley model. This strategy facilitated the identification of the sections within the scrutinized images anticipated to be influenced by the ailment and those classified as typical.

Figure 5 illustrates the SHAP analysis findings for a randomly sampled data point extracted from the dataset.
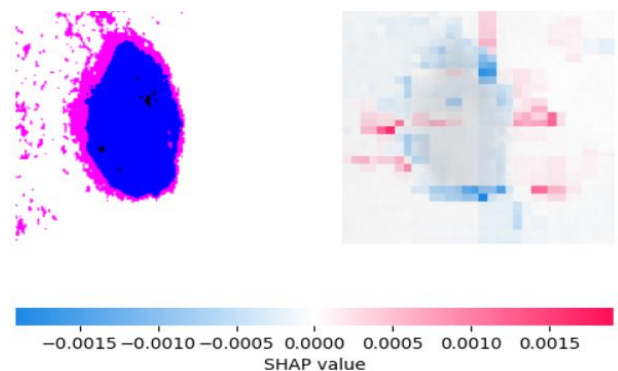

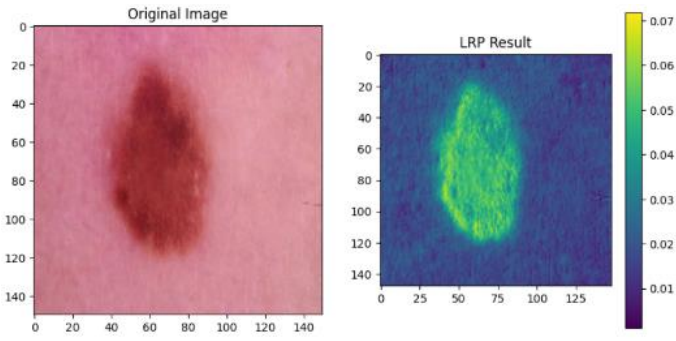Figure 4. LRP analysis for a randomly selected image from the dataset

Figure 5. Shap analysis for a randomly selected image from the dataset

Figure 6. LRP analysis of the same image

Figure 7. Confusion matrix of the trained model

Subsequently, we deployed this approach across the entire dataset, generating predictions for all the data. As seen in Figure 7, it correctly predicted 4274 individuals as normal, which is an accurate prediction. Unfortunately, it incorrectly classified 726 normal individuals as patients. However, as we proceed, we observe that it correctly predicted 2962 patients as diseased. On the flip side, it erroneously identified 2053 patients as normal despite their actual state being diseased. The model we trained classified 49.9% of the total 10,015 data as normal while 50.1% as diseased. Of the data classified as normal, 85.48% were actually normal, while the remaining 14.52% were mistakenly classified as diseased. For the data classified as diseased, 59.06% were correctly classified as diseased, while 41.94% were mistakenly classified as normal when they were actually diseased.

In Figure 8, we observe a graphical presentation of confusion matrix we generated. As previously noted, in the top and left quadrant, our artificial intelligence model accurately classified normal data as normal. Conversely, in the image on the right side, it erroneously classified normal data as diseased. Moving to the bottom-left quadrant, it misclassified diseased data as normal, while adjacent to it, it correctly identified diseased data as diseased.
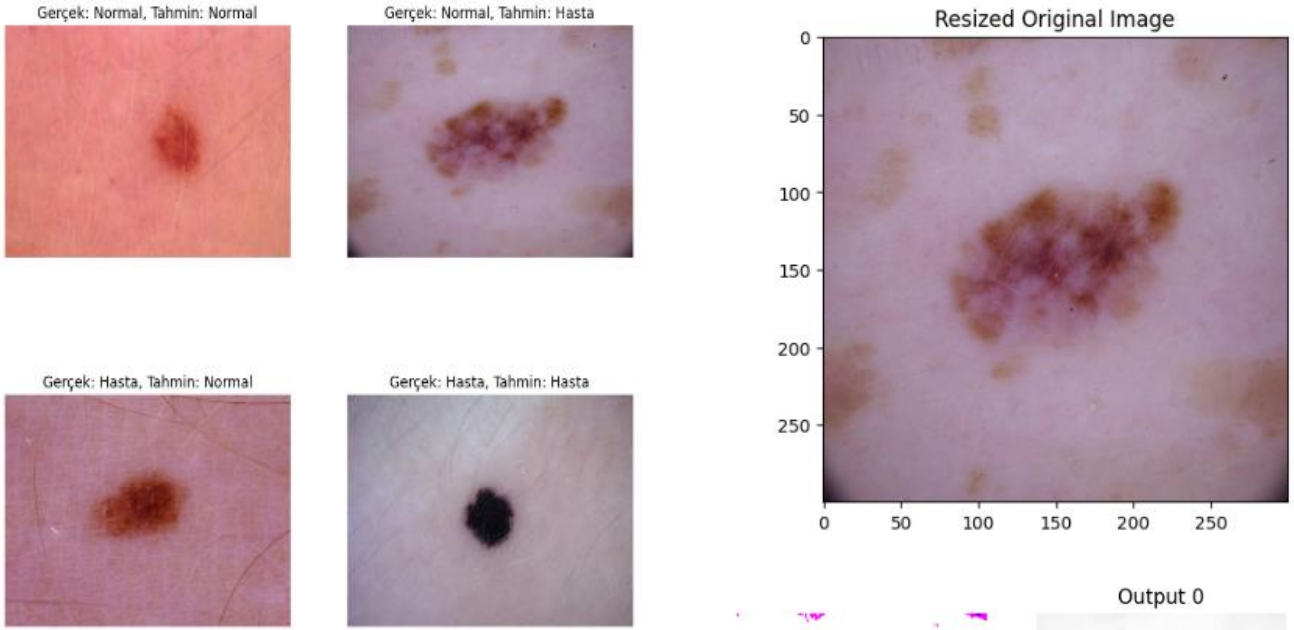
Figure 8. Confusion matrix of dataset with images

In Figure 9, we examine the output of the SHAP model, indicating potential areas of disease presence in the image. In the below, in Figure 10, the analysis of the image obtained through LRP can be seen. A comparative analysis of these two outputs enables us to derive a more precise conclusion.
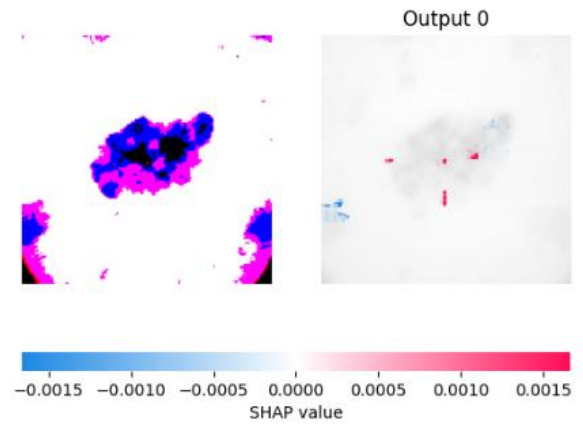

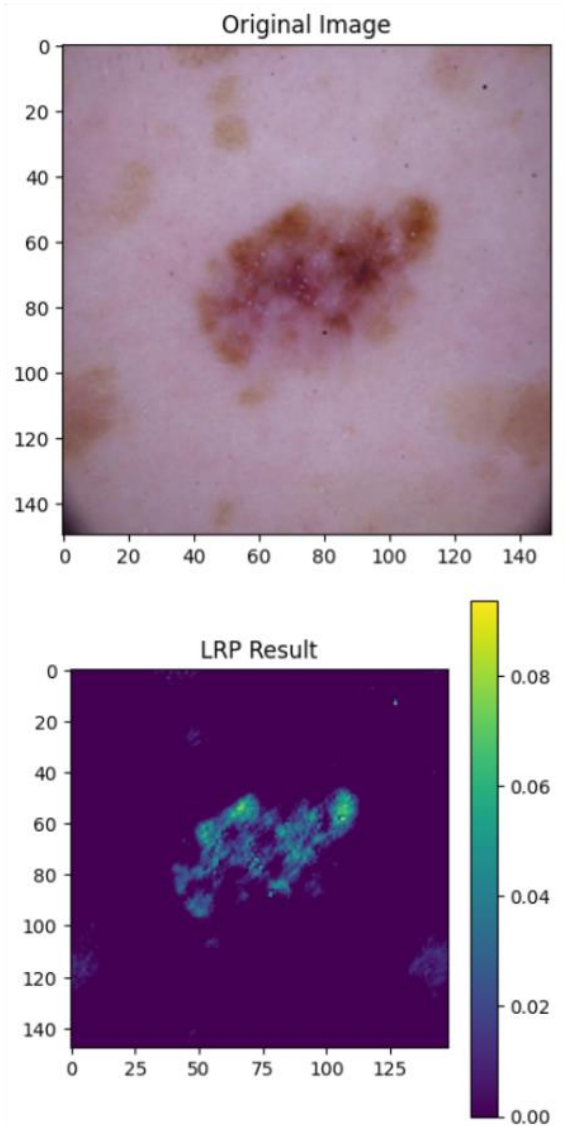Figure 9. SHAP analysis for the image
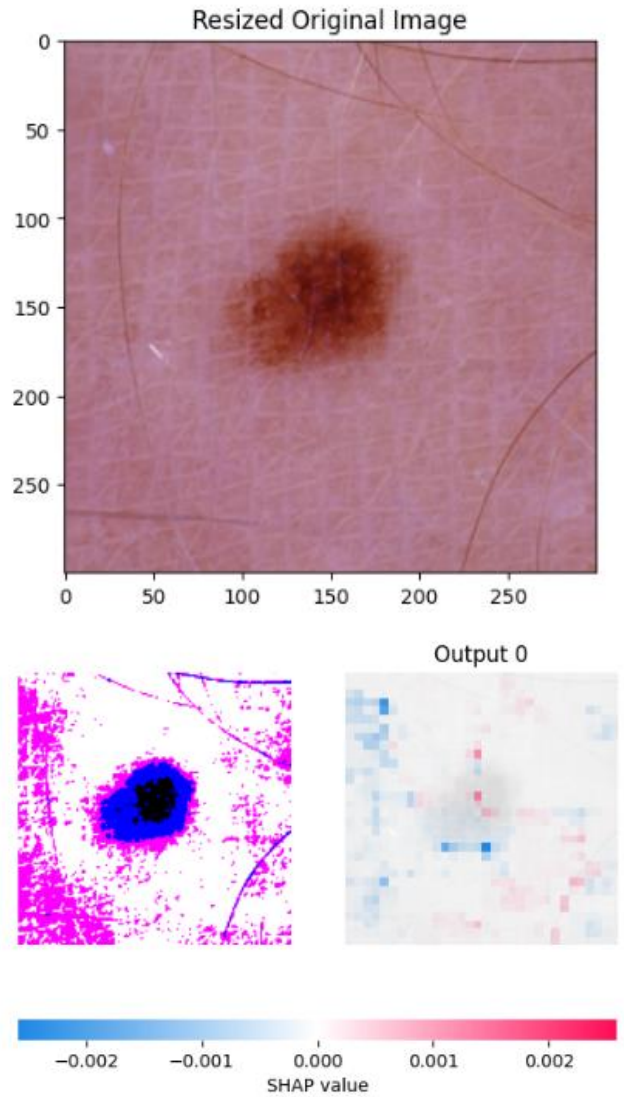
Figure 10. LRP analysis for the same image

Figure 11. The SHAP analysis of the data, which has been misinterpreted as normal when the patient is actually diseased

In Figure 11, we observe an image that the model has misinterpreted. The reason for this is the presence of different data outside the region where the disease is located in the image. As seen in the output, the model has indeed examined areas beyond where the disease is present. In Figure 12, we see the same image after undergoing Layer-wise Relevance Propagation.
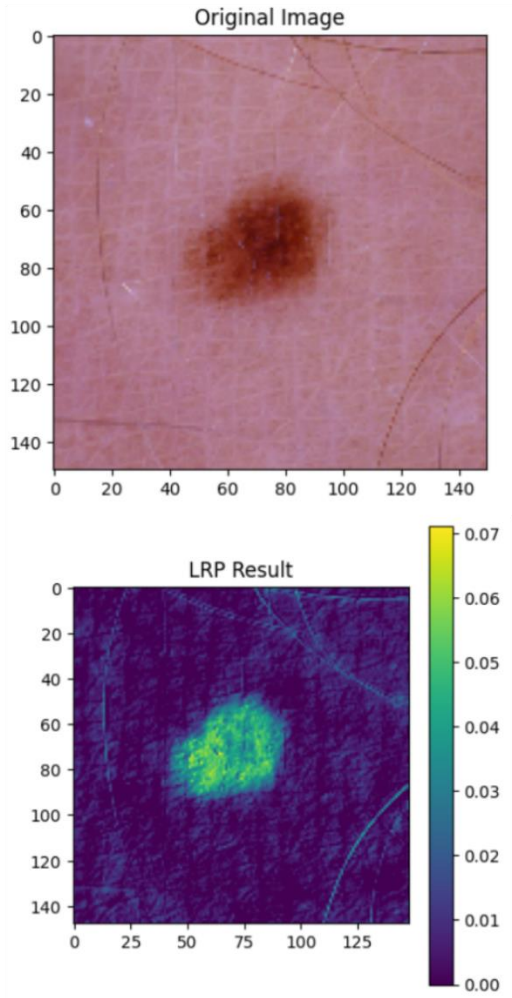
Figure 12. LRP analysis of the same image

In Figure 13, we observe that the model correctly interprets a data point as diseased and identifies the areas that are considered diseased. In Figure 14, we see the output of the same image obtained with LRP.
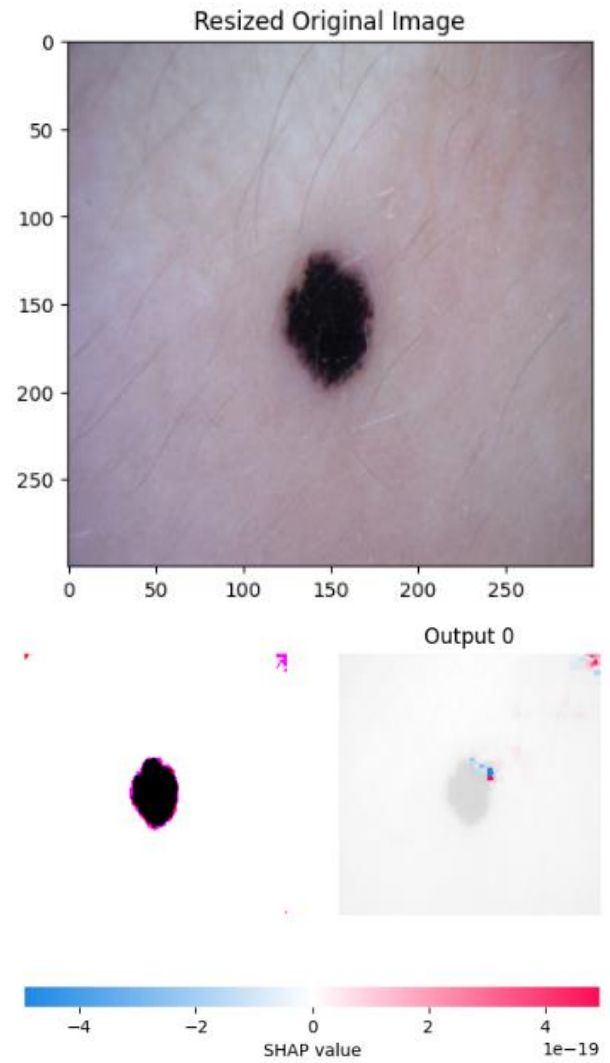

Figure 13. The SHAP analysis of the data, which has been interpreted as diseased when the patient is actually diseased
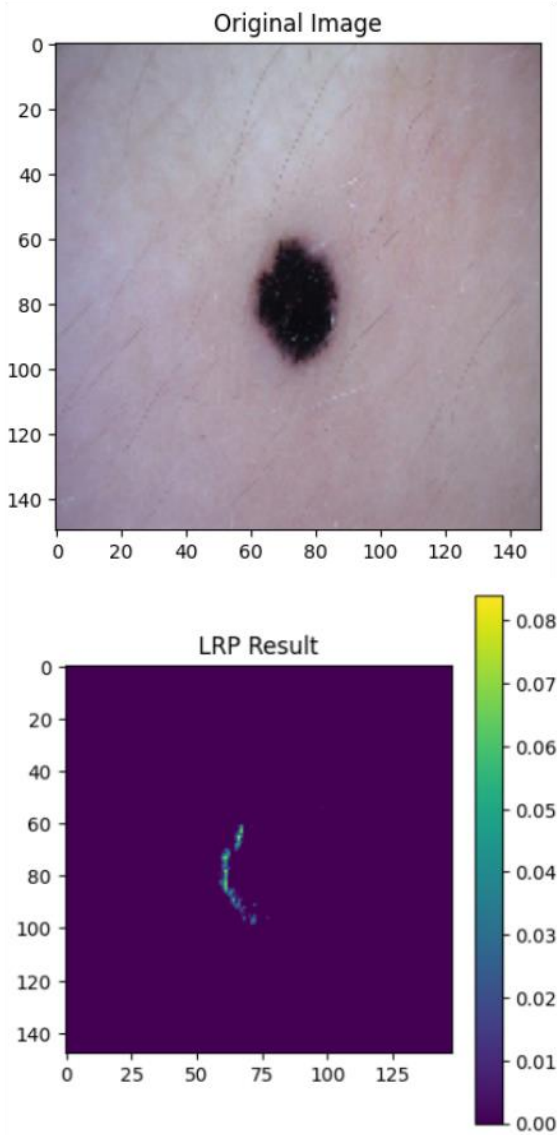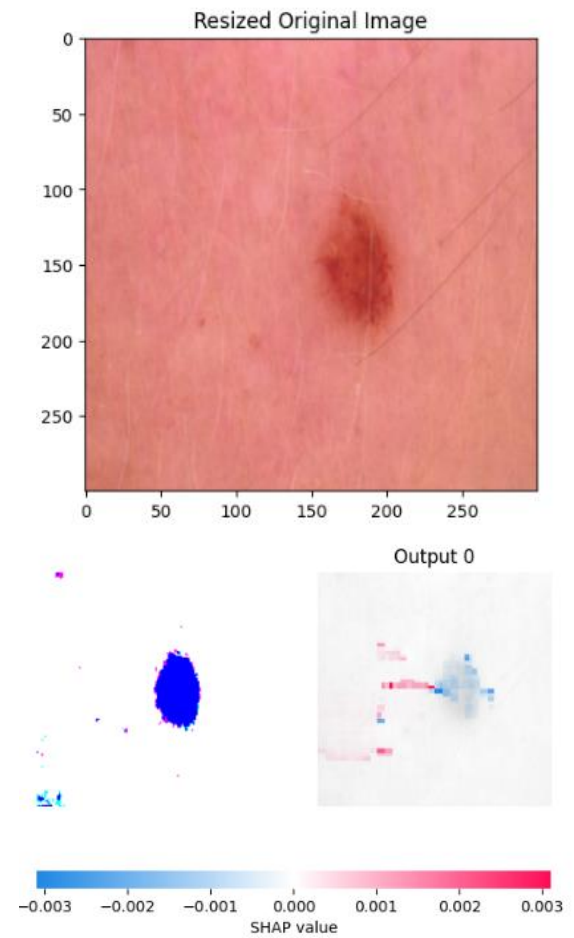
Figure 14. LRP analysis of the same image

Figure 15. SHAP analysis of the data, which has been interpreted as normal when the patient is actually normal

In Figure 15, we observe that it correctly interprets normal data and identifies the areas that are considered normal. In Figure 16, on the other hand, we see the output obtained with LRP for the same image.
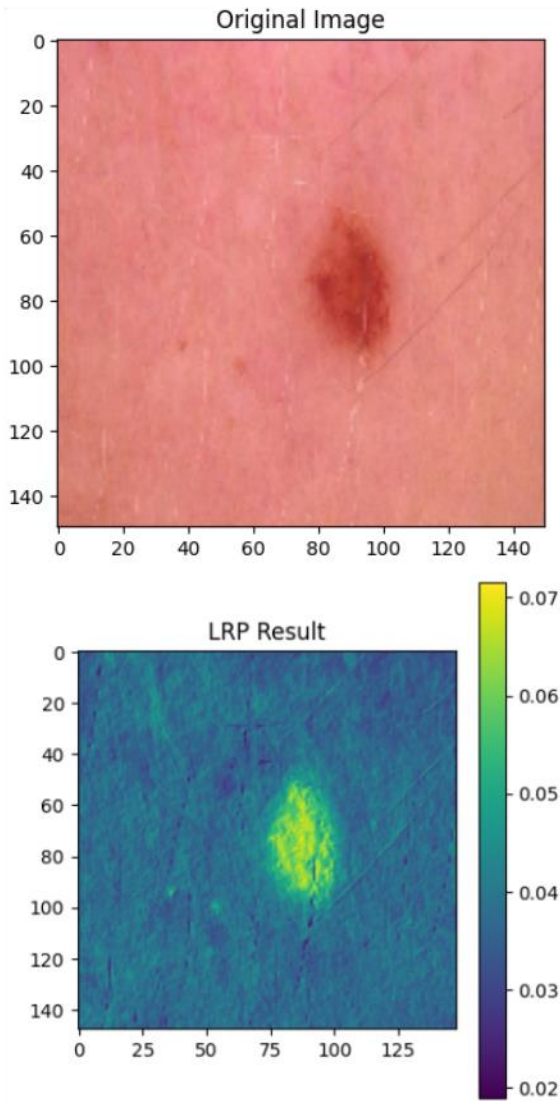
Figure 16. LRP analysis of the same image

SHAP provides a detailed description by coloring the areas of healing in blue and the areas of deterioration in red, and these colors have a numerical equivalent. LRP, on the other hand, only identifies the location of the wound by coloring it yellow and does not provide a detailed description, and LRP, like SHAP, has a numerical equivalent according to the shades of yellow color.

The equivalent of these numerical values in the outputs gives an average value of $2.11251442 \times 10^{-5}$

in SHAP. When the same process is done in LRP, the result is approximately 0.032795×10^{-5}.

4 Conclusion

Visual presentations, like the confusion matrix and results from the SHAP and LRP models, provided insightful perspectives into the model's performance and interpretability. These graphical representations highlighted the model's capabilities and limitations, aiding in potential enhancements.

Our findings revealed a commendable accuracy in identifying normal individuals, with 4274 out of 5000 correctly classified. However, the model exhibited a notable margin of error, misclassifying 726 normal individuals as patients. On the other hand, it demonstrated a promising ability to accurately predict diseased cases, correctly identifying 2962 out of 5015 individuals. Nonetheless, it erroneously labeled 2053 patients as normal, highlighting areas for improvement.

While our model achieved a balanced classification between normal and diseased individuals, further enhancements are warranted to reduce misclassifications and enhance overall accuracy. Future research could explore incorporating additional features or refining existing methodologies to bolster the model's predictive capabilities.

In essence, our study contributes to the ongoing efforts in leveraging artificial intelligence for skin disease diagnosis, underscoring the potential of CNN models in healthcare applications. Through continued refinement and validation, such models hold promise in augmenting clinical decision-making and improving patient outcomes in dermatological practice.

Looking at the numerical results of SHAP and LRP in Results, LRP has a higher numerical value because it only uses a darker shade of yellow in the wounds when analyzing. SHAP has a lower value than LRP because it uses lighter shades of blue in the healing part of the wound. It should not be misunderstood that LRP has a higher value, SHAP has a lower value

because it interprets in more detail and is more qualified than LRP's output.

As a result, it is concluded that the SHAP is better in this area as it is more detailed than the LRP in the area of health.

References

[1] A. Holzinger, A. Carrington, H. Müller, "Measuring The Quality Of Explanations: The System Causability Scale (SCS). Comparing Human And Machine Explanations", ARXIV-CS.AI, 2019.

[2] A. Holzinger, A. M. Carrington, H. Müller, "Measuring The Quality of Explanations: The System Causability Scale (SCS)", KUNSTLICHE INTELLIGENZ, 2020. (IF: 5).

[3] C. Seibold, A. Hilsmann, P. Eisert, "Focused LRP: Explainable AI for Face Morphing Attack Detection", ARXIV-CS.CV, 2021. (IF: 3).

[4] M. R. Karim, T. Döhmen, M. Cochez, O. Beyan, D. Rebholz-Schuhmann, S. Decker, "DeepCOVIDExplainer: Explainable COVID-19 Diagnosis from Chest X-ray Images", IEEE International Conference On Bioinformatics And ..., 2020. (IF:3).

[5] Y. J. Jung, S. H. Han, H. J. Choi, "Explaining CNN and RNN Using Selective Layer-Wise Relevance Propagation", IEEE ACCESS, 2021. (IF: 3).

[6] G. Chlebus, N. Abolmaali, A. Schenk, H. Meine, "Relevance Analysis Of MRI Sequences For Automatic Liver Tumor Segmentation", ARXIV-EESS.IV, 2019.

[7] A. M. A. Ahmed, L. A. M. Ali, "Explainable Medical Image Segmentation Via Generative Adversarial Networks and Layer-wise Relevance Propagation", ARXIV-EESS.IV, 2021.

[8] M. U. Alam, J. Hollmen, J. R. Baldvinsson, R. Rahmani, "SHAMSUL: Systematic Holistic Analysis to Investigate Medical Significance Utilizing Local Interpretability Methods in Deep Learning for Chest Radiography Pathology Prediction", ARXIV-EESS.IV, 2023

[9] P. R. A. S. Bassi, R. Attux, "A Deep Convolutional Neural Network for COVID-19 Detection Using Chest X-Rays", ARXIV-EESS.IV, 2020. (IF: 4).

[10] S. Suara, A. Jha, P. Sinha, A. A. Sekh, "Is GradCAM Explainable in Medical Images?", ARXIV-EESS.IV, 2023

[11] S. Suthaharan, "NEGXAI: A Negation-based Explainable AI Through Feature Learning in Fourier Domain", NANOSCIENCE + ENGINEERING, 2023.

[12] A. Sadafi, O. Adonkina, A. Khakzar, P. Lienemann, R. M. Hehr, D. Rueckert, N. Navab, C. Marr, "Pixel-Level Explanation of Multiple Instance Learning Models in Biomedical Single Cell Images", ARXIV-EESS.IV, 2023

[13] E. S. Rolfsnes, P. Thangngat, T. Eftestøl, T. Nordström, F. Jäderling, M. Eklund, A. Fernandez-Quilez, "Reconsidering Evaluation Practices in Modular Systems: On The Propagation of Errors in MRI Prostate Cancer Detection", ARXIV-EESS.IV, 2023.

[14] P. Tschandl, C. Rosendahl, H. Kittler "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions." Scientific data, 5(1), 1-9, 2018.

[15] A. A. Nugroho, I. Slamet, S. Sugiyanto, "Skins cancer identification system of HAMl0000 skin cancer dataset using convolutional neural network." In AIP conference proceedings (Vol. 2202, No. 1), December, 2019. AIP Publishing.

[16] M. A. Khan, M. Y. Javed, M. Sharif, T. Saba, A. Rehman, "Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification." In 2019 international conference on computer and information sciences (ICCIS) (pp. 1-7), April, 2019. IEEE.

[17] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects." IEEE transactions on neural networks and learning systems, 2021.

[18] S. Albawi, T. A. Mohammed, S. Al-Zawi, "Understanding of a convolutional neural network." In 2017 international conference on engineering and technology (ICET) (pp. 1-6), August, 2017. IEEE.

[19] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Chen, "Recent advances in convolutional neural networks." Pattern recognition, 77, 354-377,2018.

[20] R. Confalonieri, L. Coba, B. Wagner, T. R. Besold. "A historical perspective of explainable Artificial Intelligence." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11(1), e1391, 2021.

[21] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, P. M. Atkinson. "Explainable artificial intelligence: an analytical review." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11(5), e1424, 2021.

[22] F. K. Došilović, M. Brčić, N. Hlupić, "Explainable artificial intelligence: A survey." In 2018 41st

International convention on information and communication technology, electronics and microelectronics (MIPRO) (pp. 0210-0215), May, 2018. IEEE.

[23] E. Tjoa, C. Guan, "A survey on explainable artificial intelligence (xai): Toward medical xai." IEEE transactions on neural networks and learning systems, 32(11), 4793-4813, 2020.

[24] G. Van den Broeck, A. Lykov, M. Schleich, D. Suciu, "On the tractability of SHAP explanations." Journal of Artificial Intelligence Research, 74, 851-886, 2022.