IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 30 October 2023, accepted 10 November 2023, date of publication 14 November 2023,
date of current version 27 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3332628

RESEARCH ARTICLE

The Power of Generative AI to Augment for Enhanced Skin Cancer Classification: A Deep Learning Approach

MUDASSIR SAEED[1], **ASMA NASEER**[1], **HASSAN MASOOD**[1], **SHAFIQ UR REHMAN**[2], **AND VOLKER GRUHN**[3]

[1]Department of Computer Science, National University of Computer and Emerging Sciences, Lahore 44000, Pakistan
[2]College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia
[3]Department of Software Engineering, University of Duisburg–Essen, 45141 Essen, Germany

Corresponding author: Shafiq Ur Rehman (srehman@imamu.edu.sa)

This work was supported by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) under Grant IMSIU-RP23042.

ABSTRACT Skin cancer, particularly the malignant melanoma subtype, is widely recognized as a highly lethal form of cancer characterized by abnormal melanocyte cell growth. However, diagnosing and classifying skin lesions, as well as automatically recognizing malignant tumors from dermoscopy images, present significant challenges. To address this challenge, our study employs variants of Convolutional Neural Networks (CNNs) to effectively diagnose and classify various skin lesion types using the latest benchmark datasets ISIC 2019 and 2020. The dataset underwent rigorous preprocessing, which involves employing advanced Generative Artificial Intelligence (AI) techniques i.e., Generative Adversarial Networks (GANs) and Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN), for augmentation. These generative techniques are carefully evaluated and compared for their effectiveness. Our CNN-based approach involves aggregating results from multiple transfer learning models, including VGG16, VGG19, SVM along with a hybrid model in combination of VGG19 and SVM. On ISIC 2019, we have achieved promising accuracies of 92% for VGG16 and 93% for VGG19. Notably, the hybrid VGG19+SVM model exhibits the highest accuracy of 96%. On ISIC 2020, VGG16, VGG19, and SVM achieves accuracies of 90%, 92%, and 92%, respectively. Our findings underscore the potential of generative AI for augmentation, and the efficacy of CNN-based transfer learning models in improving skin cancer classification accuracy.

INDEX TERMS Skin cancer, CNNs, VGG16, VGG19, GAN, ESRGAN, classification.

I. INTRODUCTION

Skin tumors develop when melanocytic cells in the skin increase abnormally. Melanocytes are cells that give color to the skin, and they are often involved in the more aggressive types of skin cancer [1]. In other words, skin cancer occurs when certain cells in the skin start to grow and divide uncontrollably, which can be harmful and potentially life-threatening if left untreated [2]. Specifically by the year 2030, it's projected that around 13.1 million individuals diagnosed with skin cancer, based on data from the (WHO), are predicted to witness a substantial upsurge in the upcoming decade [3]. This indicates a concerning trend of increasing incidence rates of this disease and underscores the importance of taking preventative measures such as protecting oneself from excessive sun exposure and getting regular check-ups with a dermatologist. Non-Hispanic white individuals, particularly men, and women, are more commonly affected by melanoma [4]. This highlights the urgent need for early detection and effective treatment of melanoma, as it is a particularly aggressive form of skin cancer that can be deadly if not properly managed. Melanoma, a type of skin tumor, is recognized as a perilous type of skin cancer [5]. The main reason is melanoma

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

has a tendency to spread quickly to various regions of the human body and might be challenging to treat if not caught early. It is crucial for individuals to take steps to prevent the development of melanoma, such as limiting sun exposure, wearing protective clothing, and regularly checking the skin for any unusual moles or growths [6]. Skin cancers include Dermato-fibroma, squamous cell-carcinoma, benign-keratosis, melanocytic-nevus, actinic-keratosis, basal cell-carcinoma, and vascular-lesion are among the different types of skin conditions. Each type of skin cancer has its own unique characteristics, and some are more aggressive or difficult to treat than others [7]. It is important to be aware of the different types of skin cancer and to regularly check the skin for any changes or abnormalities in order to catch any potential issues early on. Fortunately, early diagnosis and treatment might greatly lower the alarming rates of this disease. Research suggests that malignant skin lesions can be successfully treated if diagnosed at the beginning [8]. It is consequently critical to differentiate among lesions that are malignant and benign to identify detecting whether a skin lesion is cancerous or not. Necessitates a careful evaluation performed by a qualified healthcare professional in order to ensure prompt and appropriate medical intervention [9]. Early diagnosis is key to effective treatment, so it is important for individuals to regularly check their skin and seek medical attention if they notice any changes or abnormalities. However, despite this knowledge, identifying malignant skin lesions can be challenging [10]. As a result, accurate detection is crucial in providing effective treatment options for patients, and recent research indicates that data models may play an important role in improving illness diagnosis. Additionally, numerous researchers encounter obstacles when obtaining vast and accurate related datasets to thoroughly experiment and estimate suggested approaches, which can pose a major obstacle to the advancement of skin cancer research [11]. As a consequence, many studies seem to have access to datasets containing neural networks with less than 5000 entries. To address this challenge, we used publicly available ISIC 2019 and ISIC 2020 classification datasets and applied different preprocessing techniques like labeling, noise removal, data augmentation resizing These datasets are widely used in skin cancer research and are valuable resources for developing and testing new approaches to skin lesion diagnosis and classification using intelligent technology techniques [12]. Using these datasets, we classified skin lesions using a variety of machine and deep learning machine-driven procedures. These approaches allow us to compare the performance of different algorithms and identify which ones are most effective for accurately identifying malignant skin lesions [13]. By utilizing a variety of models, we can also gain insights into the underlying mechanisms that drive accurate skin cancer diagnosis and develop more sophisticated and effective approaches for diagnosing and treating this disease [14]. In this research, we encounter two main challenges related to image data. One is that some types of skin lesions are represented much more frequently in the data than others, which makes it difficult to train our machine and deep learning models to accurately recognize all skin lesions. The second challenge is ensuring that all the skin lesion images are correctly labeled, which is critical for the accuracy of our models. Despite these challenges, we are able to develop effective approaches to address them and obtain promising results in skin lesion classification research.

FIGURE 1. Melanoma and Non-Melanoma.

A. MAJOR RESEARCH CONTRIBUTIONS

This research focuses on the diagnose of skin lesions and skin cancer classification via deep learners and Generative AI by employing latest benchmark datasets and achieves promising accuracy. The major research contribution of the current study campuses:

- A novel approach to image augmentation is applied involving generative AI through a combination of ESR-GAN and standard GAN models. Initially, ESRGAN is employed for image augmentation for up-scaling and up-sampling, followed by GAN to enhance the sharpness of the images to addresses the challenges in diagnosing and classifying skin lesions and recognizing malignant tumors from dermoscopic images.
- Investigated the effectiveness of the fusion of deep networks and machine learning by utilizing CNN architecture in the context of feature extraction and machine learning classifier for categorization of skin lesions.
- An extensive comparative analysis is undertaken to evaluate the outcomes achieved through the implementation of two disparate Data Augmentation Techniques.
- Convolutional Neural Networks VGG-16 and VGG-19 along with a hybrid model (VGG19 + SVM) are trained on benchmark and latest datasets, ISIC-2019 and ISIC-2020, and promising results are achieved by using the fusion of VGG19 and SVM.

The organization of the paper in the subsequent sections is as follows: Section II introduces a comprehensive overview of the background and presents a detailed examination of previous research, focusing on various classification techniques. In Section III, a detailed description is provided regarding the methodology which provides a comprehensive account of the experimental procedures. Section IV presents an in-depth analysis of the results obtained, while Section V provides the concluding remarks and implications of the research.

II. RELATED WORK

SkinCancer one of the most frequently encountered forms of cancer, affecting a significant population throughout the globe. Skin cancer identification and diagnosis may improve treatment outcomes and lower death rates. Abbasi et al. oriented InSiNet, a deep-convolutional technique capable of high-accuracy skin cancer detection and segmentation [1]. Skin classification has turned into a major topic of study as a result of the high incidence and risk of skin cancer as well as the possibility of early identification and therapy. The paper "DermoExpert: proposes a novel technique for improving classification accuracy with a hybrid (CNNs). Hasan et al. obtained a 92.2 percent accuracy rating on the ISIC 2018 dataset, outperforming advanced methods [4]. Deep learner and machine learner algorithms shown the encouraging results of identification and categorization of skin lesions, with various studies proposing novel approaches proposed by Shorfuzzaman et al. using Convolutional Neural Networks(CNNs), Transfer-learning with augmentation techniques to improve accuracy.They evaluated their approach on the ISIC 2019 dataset with an outstanding accuracy rate of 94.87% [2]. Benyahia et al. sought to help enhance the categorization of skin lesions by utilizing a fusion of features extracted from the VGG16, ResNet50, and Inception-v3 models. The suggested strategy is tested using the ISIC 2018 dataset, and the authors report a high accuracy rate of 88.1%, surpassing several modern techniques [3]. Agyenta et al. Convolutional neural networks (CNNs) are used to create a novel technique for skin lesion categorization. The suggested approach achieves a high accuracy rate of 94.52% on the ISIC 2019 dataset, outperforming various modern methods, demonstrating the effectiveness of (CNNs) for categorization [5]. In this research, comparisons with traditional models such as (SV) classification, (knn), naive-Bayes, random-forest, and gradient-boosting algorithms. Alkhushayni et al. observed a significant performance advantage of CNN-based algorithms [6]. By utilizing a well-established melanoma dataset, a two-stage learning platform is meticulously constructed and thoroughly assessed. Allugunti focused on the development and evaluation of (CNNs) model specifically designed for precise skin cancer diagnosis [7]. In this research, Soham Bhosale et al. provided a professional assessment of three skin cancer classification models: MobileNet TensorFlow, MobileNet on PyTorch, and Random Forest. While overall accuracy favored the MobileNet TensorFlow model at 86.5%, followed by PyTorch MobileNet at 78.3%, and Random Forest at 74.9% [9]. Solene et al. presented a detailed assessment of extensively examined various machine learning (ML) algorithms, including (CNNs) and popular qualified models, which underwent thorough evaluation in the research. Data for the evaluation is obtained from the reputable source dermnetnz.org. The findings repeatedly revealed DL models' improved performance, with accuracies reaching 0.88 [10]. The use of neural networks (also called AI) in dermatology this topic has garnered substantial interest and has been the subject of extensive attention, especially in terms of enhancing patient care via improved skin lesion detection. Das et al. Collaborative efforts aimed at gathering diverse clinical data play a pivotal role in advancing AI research within this field [11]. A practical solution is proposed in this study by Shri et al. combining image processing(IP) and Machine-learning(ML) methods. Impressive results are achieved, with an accuracy rate of 89.5% and a training accuracy of 93.7%, using publicly available datasets [8]. In this research, Walaa et al. Using the ISIC2018 dataset, techniques based on deep learning, notably (CNNs), are used to identify malignant and benign skin tumours. For the CNN models, the suggested method attained an accuracy percentage of 83.2, demonstrating its effectiveness in classifying skin tumors using the ISIC2018 dataset [12]. Furthermore, Shuwei et al. outperformed previous research in the Derm7pt dataset, with the greatest BACC of 0.735. Grad-CAM++'s model-based heatmaps confirmed the correct selection of lesion characteristics. This strategy reduces computational costs and holds promise for clinically intelligent diagnosis and mobile-based skin cancer screening [13]. Afza et al. used a mix of deep neural network feature fusion and an extreme learning machine to present a unique strategy for multi-class skin lesion categorization. The advised method's effectiveness examined using two publicly free datasets, HAM10000 and ISIC2018. Achieving accuracies of 93.40% and 94.36% respectively [15]. In this research, Dan Popescu et al. used multiple (CNNs) with HAM10000 dataset trained to effectively detect different forms of lesions. They identified the best (CNNs) models after intensive study and assessment, including, GoogLeNet, ResNet-50, DenseNet201, InceptionResNet-V2, Xception, MobileNet-V2, ResNet-101, and AlexNet [16]. Tom McIlwain aimed to achieve high performance in skin lesion classification by combining image segmentation and image + metadata classification. The author obtained competitive AUC scores, comparable to other state-of-the-art approaches [17]. Effective melanoma diagnosis plays a crucial role in patient care. In this study, Yiming Zhang et al. employed the DenseNet model to accurately recognize melanomas in skin lesion images. Their method is trained and evaluated on the ISIC2020 dataset, demonstrating superior performance compared to other deep-learning approaches. With an AUC score of 0.925, this research showcased the potential of their approach in improving melanoma analysis and supporting dermatologists in delivering efficient and precise diagnoses [18]. Wang Jiahao et al. established a n-etwork on the basis of Efficient-B5 to classify melanoma in dermoscopic images. Their network's advanced architecture allowed for accurate feature extraction and outperformed existing methods [19]. In this study, Subroto Singha et al. conducted an analysis using the public ISIC 2020 database to categorize lesions. They benchmarked the performance of three top pretrained models (ResNet, VGG16, and MobileNetV2) demonstrated excellent training accuracy, ranging from 98.73% to 99.76%, and 98.39% [20]. In this research, Panja et al. implemented a (CNNs) model to recognise and categorise the cancerous state of skin into malignant (melanoma) and benign (non-malignant) categories [14]. Duggani Keerthana et al. suggested a framework for action significantly improved the accuracy of dermoscopy image classification. When compared against the most advanced CNN techniques on the ISBI 2016 dataset, their models outperformed with accuracy rates of 88.02% and 87.43% [21]. Ahmad Naeem et al. Proposed a model, known as SCDNet, that merges Vgg16 with convolutional neural networks (CNN) for skin cancer classification citenaeem2022scdnet. Evaluation with the ISIC 2019 dataset demonstrated SCDNet's superior performance over other classifiers. A comprehensive close analysis of six distinct networks for the purpose of multi-class classification. The evaluation is carried out using the HAM10000 dataset. Among the studied models, the Xception Net performed the best, with an accuracy of 90.48% [22], [23]. Due to the success of deep learners and generative AI in multiple domains of medical imaging [24], [25], [26], [27], [28], in this research we have focused on GANs and different flavours of CNN. The details of all hyperparameters are listed in Table 4.

TABLE 1. Strengths and weaknesses of previous methods.

Paper	Strengths	Weaknesses
[1] [3] [4] [7] [12] [13] [19]	Combined deep CNN approaches, Multi-feature extraction, segmentation, and CNN to investigate skin cancer detection.	Insufficient dataset, methodology details, and performance evaluation; more analysis and model comparisons needed.
[2] [5] [9] [14] [15] [16] [20]	Utilized ensemble models, CNN, Keras, TensorFlow; compared deep learning and machine learning; and explored collective intelligence.	Limited explanations of certain approaches, preprint status, and potential gaps in comparative analysis.
[6] [8] [10] [17] [21] [23] [24]	Explored hybrid approaches combining various techniques, including machine learning and deep learning, for skin cancer detection.	Dataset and model evaluation details need expansion; limited SVM integration information.
[11] [18] [22] [25] [26] [27] [28]	Utilized adaptive Otsu-based algorithm for computer-aided COVID-19 and melanoma diagnosis, and investigated glaucoma detection.	Limited algorithm implementation details, need clarification on applicability to skin cancer detection.

The state-of-the-art has certain strengths and weaknesses as illustrated in table 1.The current research has used the strengths to overcome the research gap and address the weaknesses.

TABLE 2. Review of related studies.

Paper	Year	Dataset	Technique	Results
[1]–[4]	2021-22	HAM10000, ISIC 2016-20	CNN	90-94%
[5]–[8]	2022	HAM10000, ISIC 2017-19	CNN	86-93%
[9]–[12] [16]	2021-22	HAM10000, ISIC 2017	Resnet, InceptionV3,VGG16, MobileNet	84-88%
[13]–[15], [17]	2021-22	HAM10000, ISIC 2017-18	CNN	90-94%
[18]–[21]	2021-23	ISIC 2016-20	CNN SVM	86-92%
[22]–[28]	2021-23	HAM10000, ISIC 2016-20	VGG16-19, EfficientNet-B0, ResNet-152	86-98%

III. METHODOLOGY

The current research addresses the challenges in diagnosing skin lesions and once malignant tumors are recognized from dermoscopic images, classifying them into different cancer types. To achieve the objective of accurately classifying instances of skin cancer, various preprocessing steps are applied to the dataset followed by augmentation and training various models.

A. DATASET

When it comes to determining if a skin spot is cancerous or not, having a marked amount of reliable and diverse data is important for training computer models to make accurate predictions. However, finding datasets with a sufficient number of examples for effective model training can be a challenge. In this study, we make use of two well-known datasets called ISIC 2019 and ISIC 2020. The ISIC 2019 dataset consists of over 25,000 images showcasing different types of skin spots, including moles, basal cell carcinoma, and others. These images serve as valuable resources for training models to improve prediction accuracy. Additionally, the ISIC 2020 dataset includes over 33,000 dermoscopic images captured using a specialized device called a dermoscope. This dataset offers a wide range of images showing both benign and potentially cancerous skin spots, collected from a diverse group of individuals. By utilizing these datasets, we aim to enhance our understanding of skin cancer and develop better methods for identifying cancerous spots accurately. Table 3 depicts the metadata of both the datasets.

The dataset's clarity is paramount for the credibility of our methodology and results. We have ensured dataset integrity by using verified data from the International Skin Imaging Collaboration (ISIC), known for its reliability in dermatological data. This approach upholds scientific rigor and strengthens our methodology.

TABLE 3. Metadata of ISIC-2019 and ISIC-2020 datasets.

Datasets	Year	Classes	Total Images
ISIC 2019	2019	8	25331
ISIC 2020	2020	2	33126

Each image in the dataset is associated with a unique patient identifier, ensuring its distinctiveness and privacy. Malignant diagnoses are verified through detailed histopathology examinations, while benign cases are confirmed using a combination of expert consensus, long-term follow-up data, and additional histopathology analysis. The dataset is meticulously compiled by the (ISIC) "International Skin Imaging Collaboration" and comprises a diverse collection of images sourced from renowned institutions worldwide, including the Hospital Clinic de Barcelona, Medical University of Vienna, Memorial Sloan Kettering Cancer Center, Melanoma Institute Australia, University of Queensland, and University of Athens Medical School.

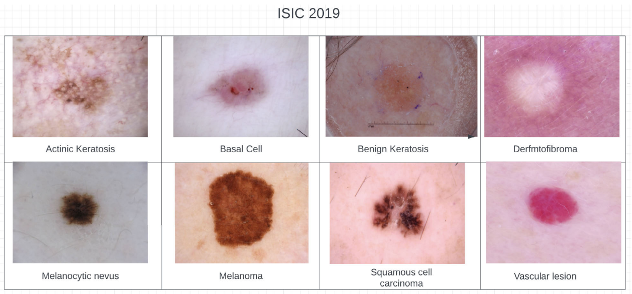

FIGURE 2. ISIC 2019 dataset images of Eight Classes.

FIGURE 3. ISIC 2020 dataset images of benign and malignant tumor.

B. DATA PREPROCESSING

To optimize the quality and practicality of the ISIC Data Set, a diverse set of preprocessing techniques are applied. One of the primary measures involved resizing the images to ensure consistent dimensions across the dataset. Additionally, data labeling is utilized to accurately categorize each image within the dataset. Furthermore, data-augmentation is implemented to boost the variety and size of the dataset by applying transformations such as rotations and flips to the images. These preprocessing steps improved the dataset's accuracy and reliability, making it a crucial resource for researchers in the field in Figure 4.

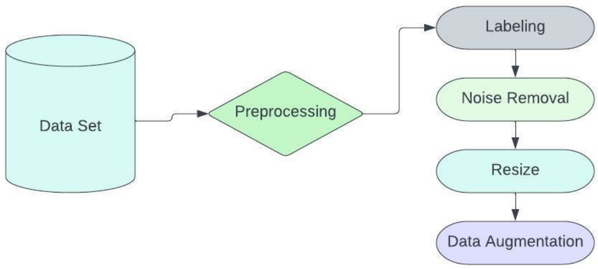

FIGURE 4. Major Steps of Preprocessing.

1) DATA LABELING

Most Deep learning models today rely on supervised learning. It involves using an algorithm to correlate an input with its output. To achieve this, a labeled dataset is essential. Human judgment is often used to label data, and the accuracy of this process affects the model's quality. Ground truth datasets serve as valuable resources for training and evaluating deep learning models. These datasets provide reliable and verified data that serve as benchmarks for assessing the performance and effectiveness of the models. By using these datasets, researchers can train and test their deep learning models to ensure accurate and reliable results. In this research, the ISIC 2019 and ISIC 2020 datasets are initially unlabeled. However, with the help of ground truth, we labeled all images with class name tags using different techniques Figure 5.

In our study, we made sure to be very thorough and accurate when labeling the classes in our dataset. We wanted to use the most trustworthy sources, so we relied on a well-respected organization called the International Skin Imaging Collaboration (ISIC) for our class labels. The folks at ISIC are like the experts of experts when it comes to dermatology. They've put together a really top-notch collection of skin lesion images that have been carefully checked by dermatologists and medical pros who really know their stuff. ISIC is known worldwide and is supported by the International Society for Digital Imaging of the Skin (ISDIS). So, we used ISIC's expertise to make sure our dataset was spot on. Dermatologists and medical pros associated with ISIC double-checked and verified all the class labels for each image. These experts have years of experience and deep medical knowledge, so they're about as reliable as it gets when it comes to deciding if an image shows melanoma or something else. We tagged our dataset with these already checked class labels before we got started with our experiments. It was like the rock-solid foundation of our research, giving us the confidence to move forward with our training and testing. Our goal was to make sure our study had top-notch data quality and followed all the right rules in dermatology and medical research, just like ISIC recommends.

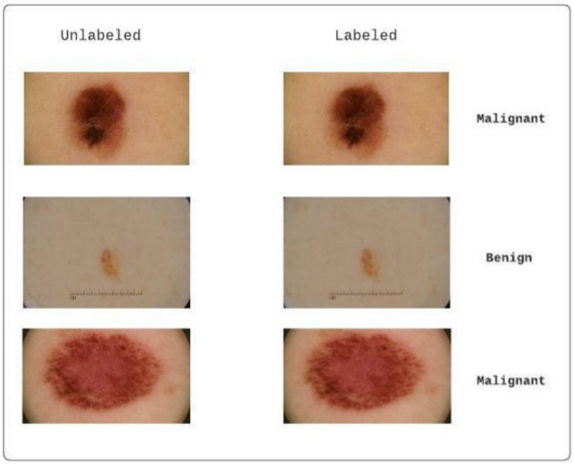

FIGURE 5. **Example of Labeled Classes.**

2) NOISE REMOVAL

To enhance the quality of an image, five steps are involved in hair noise removal. Firstly, converted the image to grayscale, which is essential for many image processing tasks. Secondly, created a kernel using morphology to extract features. Applied MORPH BLACKHAT to the grayscale image to highlight dark regions. Then, the blackhat image is converted into a binary format using a threshold value. Lastly, used inpainting to fill the original image with the thresholded image to restore missing or damaged regions as illustrated in Figure 6.

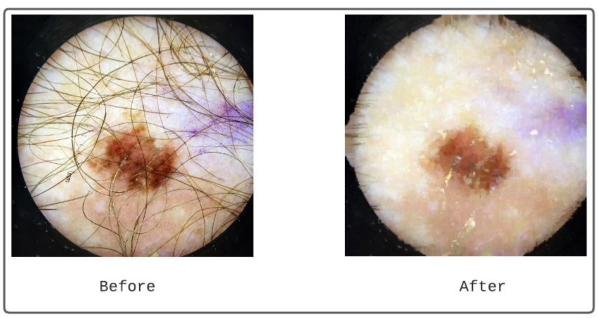

FIGURE 6. **Hair Noise Removal Effects Before and After.**

3) RESIZING

Data rescaling is vital in optimizing classification rates in imaging datasets, particularly when dealing with numerous photos of varying sizes. We addressed this challenge by standardizing all images to a resolution of 224 × 224 pixels using various methods. To overcome memory issues caused by the original image size, we developed a resizing algorithm in Python that reduced the images' size without compromising quality. By applying these techniques, we improved the quality of our data and achieved more accurate results Figure 7.

FIGURE 7. **Example of Resized Skin Lesion Image.**

4) DATA AUGMENTATION

In our research, we utilized augmentation techniques to address the challenge of overfitting in our training dataset. To this end, we applied various image transformations, such as random rotations, cropping, and splitting, with the aim of artificially expanding the dataset and extracting more information from the available images. We leveraged powerful tools like ImageDataGenerator and Generative Adversarial Networks (GANs) to generate new images from the existing ones. This approach helped us to enhance the quality and quantity of the training dataset, resulting in improved performance of our deep learning models. Moreover, data augmentation played a critical role in preventing overfitting, which is a common problem in deep learning. Figure 8 illustrates the results of augmentation while Figure 9 portrayes the results of ESRGAN and the results of augmentation with GAN are illustrated in Figure 10.

C. CLASSIFICATION ARCHITECTURE FOR ISIC 2019, ISIC 2020

The core architecture of the system utilized three distinct learning models, namely SVM, VGG 16, VGG 19, and VGG-SVM. This integration enabled the system to achieve highly accurate skin cancer detection. The VGG models are specialized (CNNs) that are already trained on large image datasets, making them adept at feature extraction and successful in a variety of image recognition tasks. The VGG 16 model comprises 16 layers, while VGG 19 has 19 layers, and the VGG-SVM model combines VGG 19 with Support-Vector-Machines (SVM), An extensively utilized algorithm in machine learning for classification purposes is the Support Vector Machine (SVM). This algorithm has gained popularity due to its ability to effectively categorize data into distinct classes with high accuracy. The system leverages the strengths of each of these models to accurately identify different types of skin cancer, even in cases where the

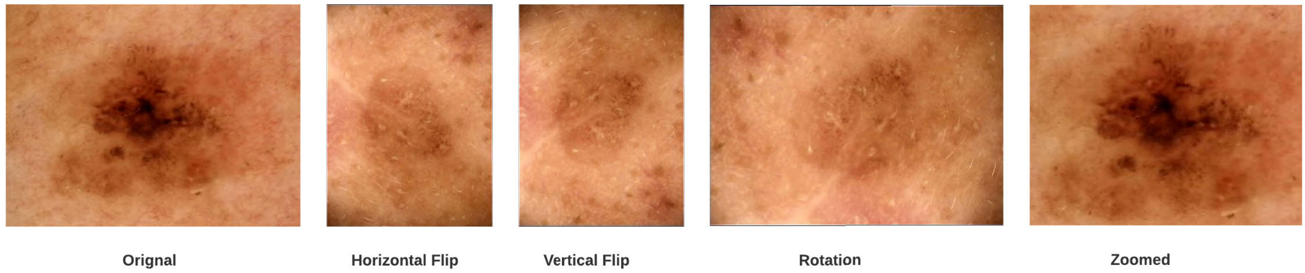

FIGURE 8. Effects of augmentation on Skin Lesion image.

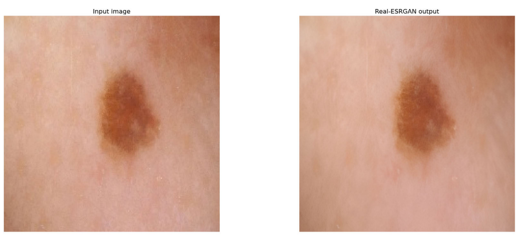

FIGURE 9. Effects of Upscalling image with ESRGAN.

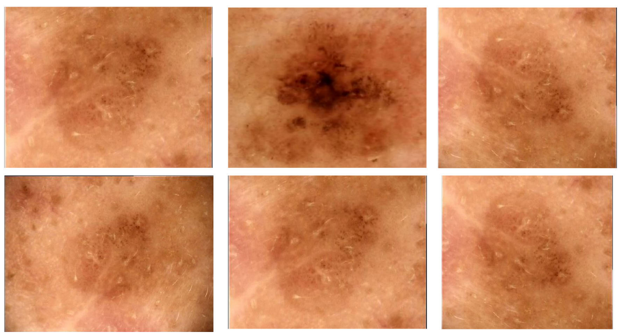

FIGURE 10. Effects of augmentation on Skin Lesion image with GAN.

cancer is challenging to detect. This comprehensive approach ensures a reliable and precise diagnosis, crucial for effective treatment and patient care.

We prioritize the use of skin images because they provide a non-invasive way to start the process. In contrast, biopsies, although very accurate, involve invasive procedures and might not be the best choice for routine checks, early detection, or ongoing monitoring. Skin images, on the other hand, give us a quick and easy way to look at unusual spots on the skin. They're useful for finding areas of concern, especially for those who can't or prefer not to undergo invasive procedures. Early detection is really important in managing cancer, and skin images help us spot these unusual spots early when treatment works best. Dermatologists can keep an eye on how these spots change over time, potentially catching problems before they become serious. This approach not only reduces the need for more invasive tests but also makes things easier for patients and lowers healthcare costs. Our confidence in using skin images is also backed by the knowledge of dermatologists. These experts are highly trained and really understand skin issues. They take into account factors like gender, race, and economic situations when making diagnoses, ensuring they're accurate, even among diverse groups of patients. Plus, advances in imaging technology, including dermoscopy, have made skin images much better in quality. These technologies create high-resolution images that show tiny details we can't see with the naked eye. Teledermatology and computer systems help dermatologists make precise diagnoses based on these images. To make sure our research is spot on, we use well-verified datasets from the International Skin Imaging Collaboration (ISIC). ISIC is known worldwide for its quality and trustworthy dermatological data. The vast collection of dermatoscopic images checked and approved by medical professionals at ISIC guarantees that the data we rely on is solid. We follow a scientific approach built on these reliable datasets, keeping the highest standards of scientific accuracy in our classification models and findings.

The details of all hyperparameters are listed in Table 4.

TABLE 4. Hyperparameters of the models.

HYPERPARAMETERS	
Activation Function	ReLU
Epochs	50 - 800
Batch Size	64
Optimizer	Adam
Loss Function	Categorical Cross Entropy
Drop out	0.5
Output Activation Function	softmax

1) SVM

In this study, we evaluated the effectiveness of the Support Vector Machines (SVMs) algorithm on the ISIC 2020 Data Set using a machine learning approach. SVMs are a commonly used algorithm for classification problems, and we specifically focused on binary classification. To establish a well-optimized dataset, we implemented a meticulous

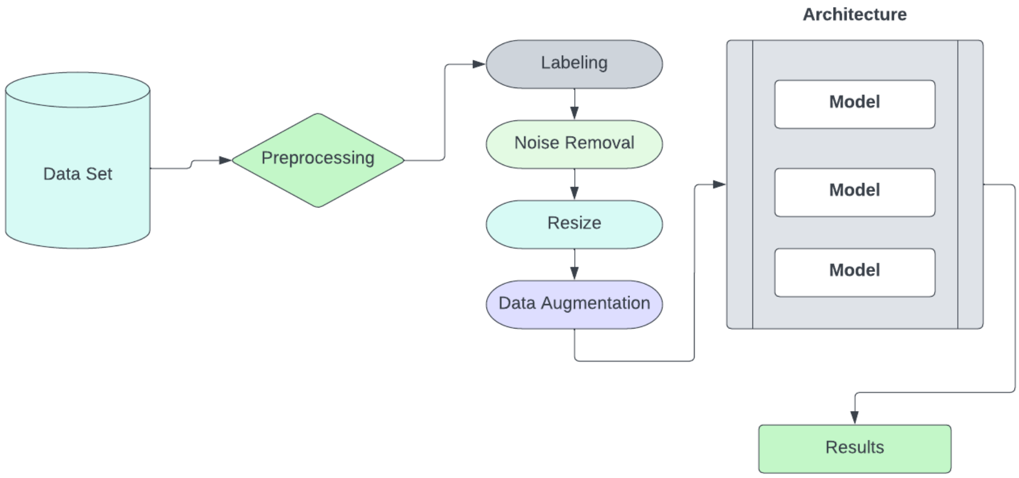

FIGURE 11. **Architecture of the proposed models for skin cancer detection and classification.**

splitting technique, allocating 80% of the data for comprehensive training while reserving a unique and distinct 20% portion solely for rigorous testing and validation purposes. To build our model, we utilized the (RBF) kernel, which is a popular kernel used in SVMs. Additionally, we explored the effect of different gamma values, which is a hyperparameter that controls the shape of the decision boundary. Specifically, we tested gamma values of 0.001 and 0.0001. Our results suggest that SVMs are effective for image classification in the ISIC 2020 Data Set using the RBF kernel and a specific gamma value.

2) VGG 16

In our research, we utilized the VGG16 model, a widely recognized convolutional neural network (CNN) known for its exceptional accuracy in classifying images. We employed a technique called transfer learning with cross-validation, which involved dividing the dataset into subsets for training and testing purposes. To enhance the learning process, we used the Rectified Linear-Unit (ReLU) activation function within the CNN and applied the sigmoid function in the output layer for binary classification. With a batch size of 32, we initialized the model's weights using pre-trained Imagenet weights. During the training phase, we employed the cross-entropy loss function for 100 epochs, using a learning rate of 0.001. Our study revealed that the VGG16 model demonstrated effective image classification capabilities, solidifying its position as a reliable and powerful tool in deep learning.

3) VGG 19

To address the challenge of classifying large-scale images, our study utilized an advanced deep neural network architecture consisting of 19 layers. We employed a transfer learning approach combined with cross-validation, where the dataset is divided into subsets for both training and testing. The activation function used in the network is Rectified Linear Unit (ReLU), a widely adopted choice known for its effectiveness in deep learning models. For the final layer, we applied the Softmax activation function to handle multi-class classification tasks. The model is trained using a batch size of 32, and the initial weights are pre-trained on the ImageNet dataset. To optimize performance, we employed the cross-entropy loss function, commonly used in classification tasks. Training is performed for 100 epochs with a learning rate of 0.001. Our experimental findings revealed the exceptional accuracy of our deep neural network, suggesting its potential as a robust solution for large-scale image classification tasks.

4) VGG-SVM

In our study, we proposed a novel approach for classifying skin cancer images as Malignant or Benign using a hybrid model that combines VGG and SVM. We utilized the ISIC 2020 dataset and trained the model to perform binary classification. To achieve high accuracy, we trained VGG models using the PyTorch Python library. Our approach consisted of three main steps: image normalization and resizing, feature extraction using variants of the VGG model, and classification using non-linear SVM. Our experimental results demonstrate that the hybrid VGG-SVM model, specifically VGG 19 with SVM, outperforms existing techniques and achieves high accuracy in classifying brain tumors. This research has the potential to improve the accuracy of tumor classification and contribute to better diagnosis and treatment options for patients.

D. EVALUATION METRICS

Precision is a measurement that focuses on the accuracy and relevance of the top-ranked results, specifically catering to the user's needs. It quantifies the extent to which these results align with the user's expectations, and its calculation involves assessing the quality and appropriateness of these

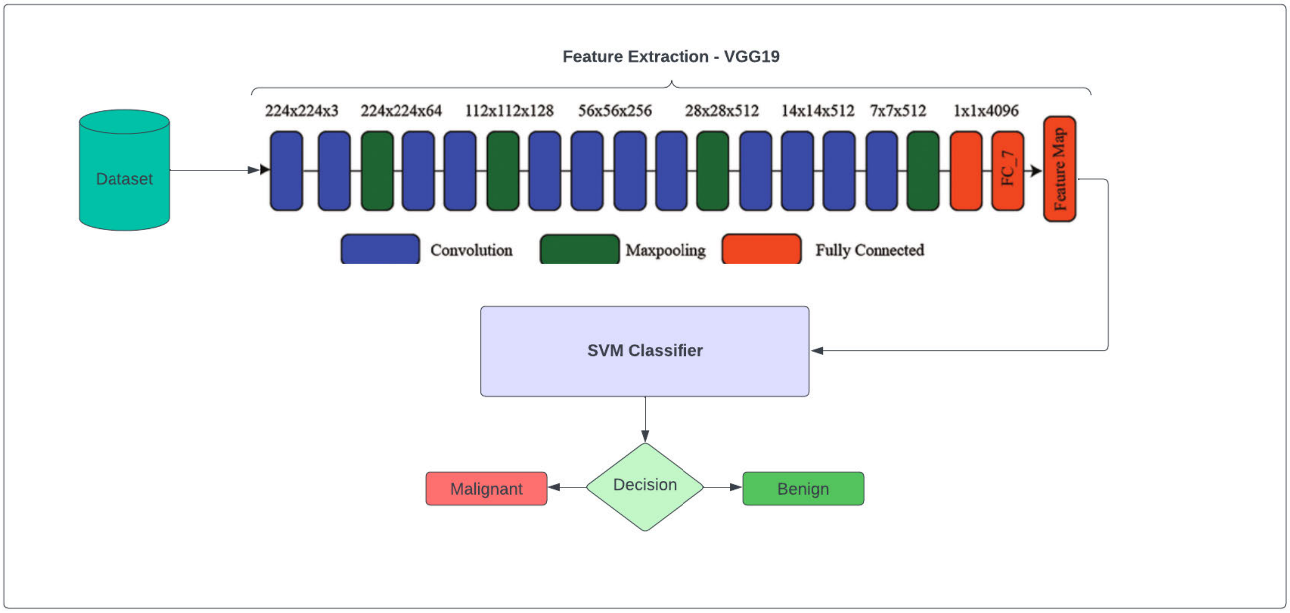

FIGURE 12. Architecture of the best performing model VGG19+SVM.

top outcomes:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

The recall pertains to the extent of retrieving all pertinent results from a corpus. It assesses the ability to capture and include all relevant information in the search outcomes. The calculation of recall involves evaluating the comprehensiveness and inclusiveness of the retrieved results, ensuring that no relevant information is missed:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The F-Score, commonly known as the F1 score, incorporates both precision and recall to assess performance. By calculating the harmonic mean of these two metrics, it offers a balanced measure of effectiveness. Higher values of both precision and recall contribute to an elevated F-Score. The calculation of the F-Score can be determined using the following formula:

$$F\text{-}Score = \frac{2 * P * R}{P + R} \tag{3}$$

IV. RESULTS

The present study involved a series of experiments aimed at evaluating the performance of different Machine Learning and Deep Learning models on the ISIC 2019 and ISIC 2020 datasets. The experiments utilized various models, including SVM, VGG16, VGG19, and VGG-SVM. To further enhance the accuracy of the models, augmentation techniques are employed, including the ImageDataGenerator and GAN. The outcomes of the experiments are summarized in Tables from 5 to 8 out of which Table 7 and Table 8 have shown the outcomes following the implementation of augmentation techniques through GAN.

As presented in Table 5, the maximum precision, recall and F1-score are achieved by VGG-SVM model on ISIC-2020 dataset. VGG19 is the 2nd best model in precision while in recall SVM is performing better after VGG-SVM. On ISIC-2019 dataset, the performance of VGG16 and VGG19 is better than the one achieved on ISIC-2020. Both the models reveal above 0.91 scores for precision, recall and Fa-score, as depicted in Table 6.

When the ISIC-202 dataset is augmented by using GAN, the precision, recall and F1-score of all the models increased significantly by the 3% to 4% which is depicted in Table 7. The same holds true for recall and F1-score. Still VGG-SVM model outperforms all the other models while the second best models turns out to be both SVM and VGG19 in all the evaluation measures. For ISIC-2019 augmented dataset, VGG16 and VGG19 have revealed better results, please refer to Table 8, than non augmented datasets, please refer to Table 6. The precision, and F1-score of VGG16 has increased by 1% while recall remains almost the same. On the other hand, VGG19 performs better in precision, recall and F1-score by the factor of 0.01, 0.02 and 0.02 respectively. Figure 14 and Figure 15 represent the visual representation of results which contain Evaluation Curves and Confusion Matrices.

V. DISCUSSION

In summary, the VGG-SVM model outperforms the other models in terms of precision, recall, and F1-score for all the four types of datasets i.e., ISIC-2019 and ISIC-2020 with and without augmentation. The overall results reveal that the SVM model reveals a tendency in precision, recall, and

TABLE 5. Performance of the models for ISIC 2020 dataset.

	Precision	Recall	F1-Score
SVM	0.88	0.89	0.89
VGG16	0.86	0.89	0.87
VGG 19	0.89	0.88	0.89
VGG-SVM	0.92	0.92	0.92

TABLE 6. Performance od the models for ISIC 2019 dataset.

	Precision	Recall	F1-Score
VGG16	0.91	0.92	0.91
VGG 19	0.92	0.91	0.91

TABLE 7. Performance of the models for ISIC 2020 dataset with GAN-based augmentation.

	Precision	Recall	F1-Score
SVM	0.92	0.92	0.92
VGG16	0.90	0.89	0.90
VGG 19	0.92	0.92	0.92
VGG-SVM	0.96	0.96	0.96

TABLE 8. Performance of the models for ISIC-2019 dataset with GAN-based augmentation.

	Precision	Recall	F1-Score
VGG16	0.92	0.92	0.92
VGG 19	0.93	0.93	0.93

an F1-score that indicate that the model has a high ability to correctly classify positive instances (precision scores) and effectively identify all relevant positive instances (recall scores) thus resulting in a balanced F1-score. On the other hand, the evaluation measures of VGG16 suggest that the model performs well in terms of recall, similar to SVM, but has a slightly lower performance for positive instances and F1-score. VGG19 is the model with the ability to maintain a balanced performance in terms of precision and recall, and attains a consistent F1-score. The VGG-SVM model demonstrates the best performance among the models. The scores indicate a high level of precision, recall, and overall classification accuracy, making it the most effective model for skin cancer classification based on the provided metrics.

A. EFFECT OF GAN-BASED AUGMENTATION

The results also reveal that data augmentation via GAN has significantly improved skin cancer classification performance by increased diversity in data by adding synthetic samples. The diversity helps the classifiers to generalize in a better way by learning robust features and reducing overfitting which is highly probable if only original dataset is used. GAN has extended the original dataset, thus resulting in a larger training data for capturing more patterns and variations. Data imbalance is also addressed by GAN by generating reasonable synthetic samples for classes with lesser samples. Finally, as data augmentation brings variations by generating new samples this helps the models to become robust to noise.

B. COMPARISON BETWEEN MODELS

Figure 13 illustrates the performance improvement of all the models following the application of data augmentation. In general, the models exhibited an enhancement in F-score, ranging from 0.3 to 0.4, for the diagnosis of cancer using the ISIC-2020 dataset.

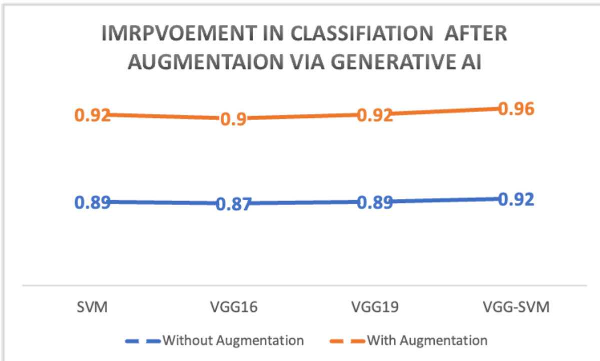

FIGURE 13. Effects of Generative AI model GAN-based Augmentation on the classification of skin cancer. The graph depicts the F1-score.

C. COMPARISON WITH THE STATE-OF-THE-ART

As shown in Table 9, the prosed implementations, trained on a combination of Generative AI-generated synthetic data samples and real data, have achieved superior results. The F1-scores is presented in the last two columns of the table, showcasing the results of the models trained in this research with and without augmentation for the ISIC-2020 dataset and is compared with the state-of-the-art in F1-scores. The SVM model has shown a 1% improvement in F1-Score, while the VGG-16 model has exhibited a remarkable 4% increase in F1-Score. Notably, the VGG19+SVM model has achieved an impressive 96% F1-Score, surpassing the state-of-the-art [21] which obtains a 94% F1-Score.

TABLE 9. Performance comparison with the State-of-the-Art. The last two columns depict the results of the models trained in this research with and without augmentation for ISCI-2020.

Models	State-of-the-Art		Augmentation NO	Augmentation YES
SVM	[11]	0.91	0.89	0.92
VGG	[10]	0.88	0.87	0.92
VGG	[23]	0.92	0.91	0.93
VGG19+SVM	[21]	0.94	0.92	0.96

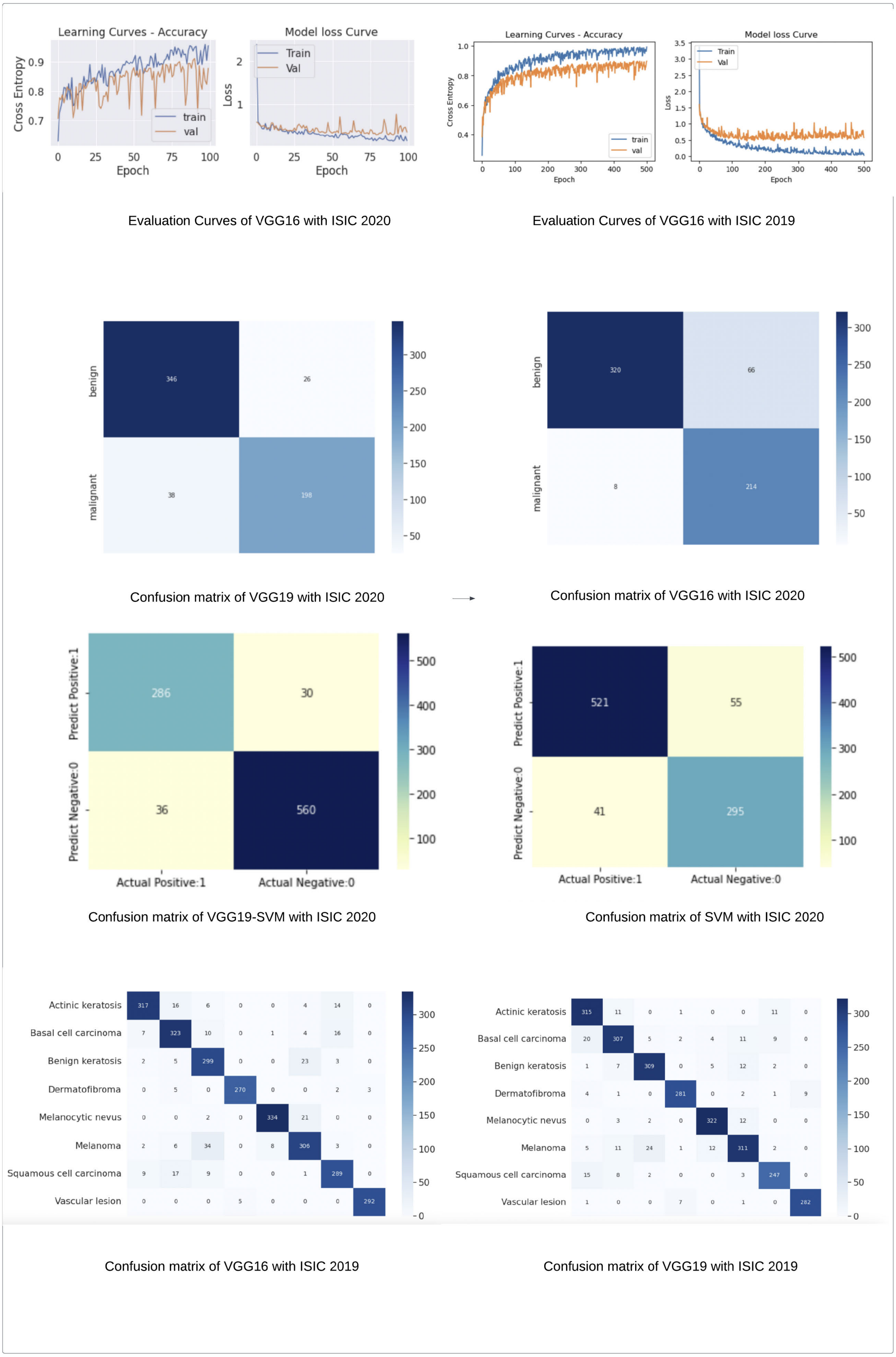

FIGURE 14. (a,b) Learning and Model Loss trends of VGG16 for ISIC-2020 dataset. (c,d) Confusion Matrices of VGG-19 and VGG-16 respectively for ISIC-2020. (e,f) Confusion Matrices of VGG19-SVM and SVM respectively for ISIC-2020. (g,h) (c,d) Confusion Matrices of VGG-19 and VGG-16 respectively for ISIC-2019.

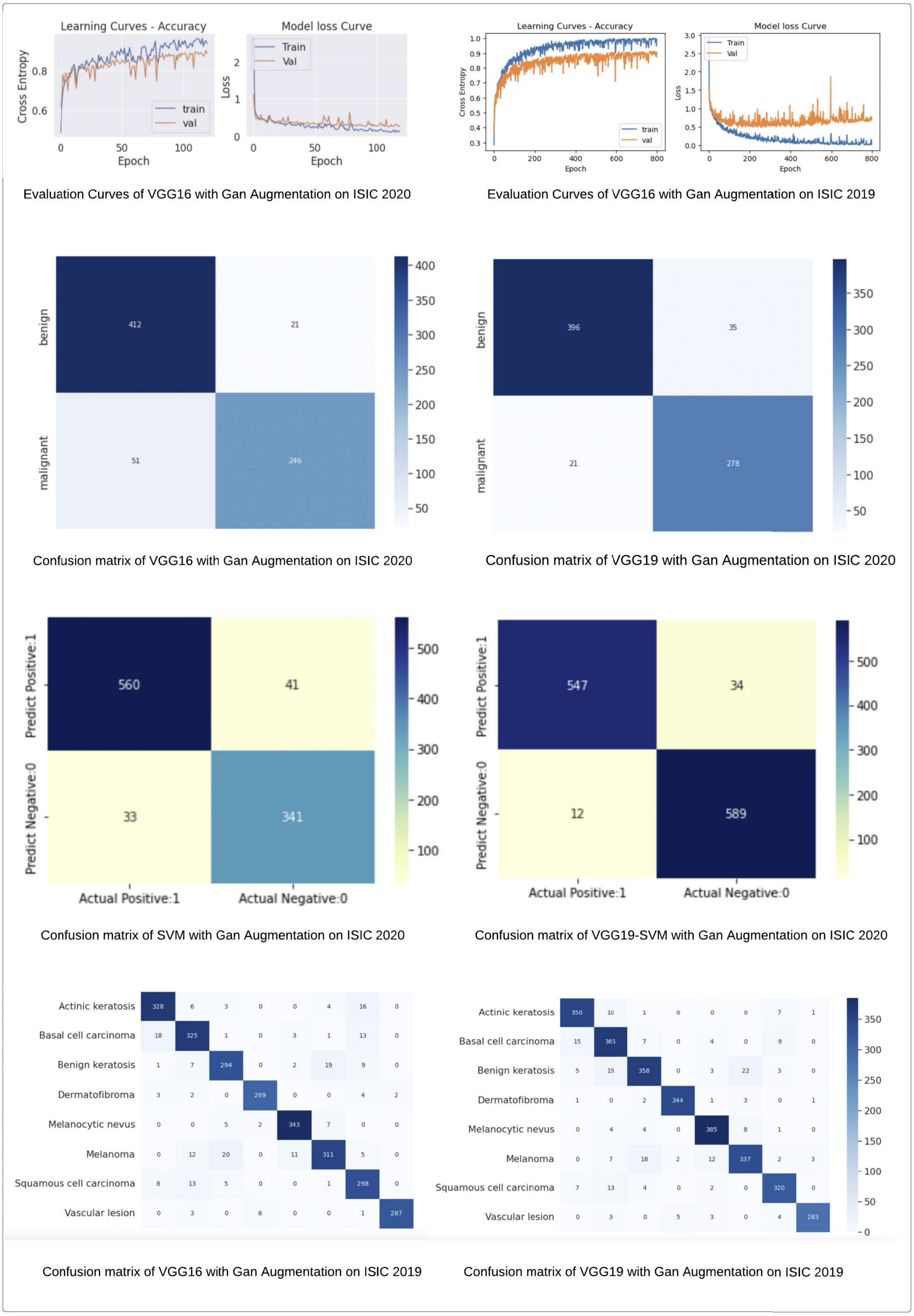

FIGURE 15. **(a,b) Learning and Model Loss trends of VGG16 for GAN Augmented ISIC-2020 dataset. (c,d) Confusion Matrices of VGG-19 and VGG-16 respectively for GAN Augmented ISIC-2020. (e,f) Confusion Matrices of VGG19-SVM and SVM respectively for GAN Augmented ISIC-2020. (g,h) (c,d) Confusion Matrices of VGG-19 and VGG-16 respectively for GAN Augmented ISIC-2019.**

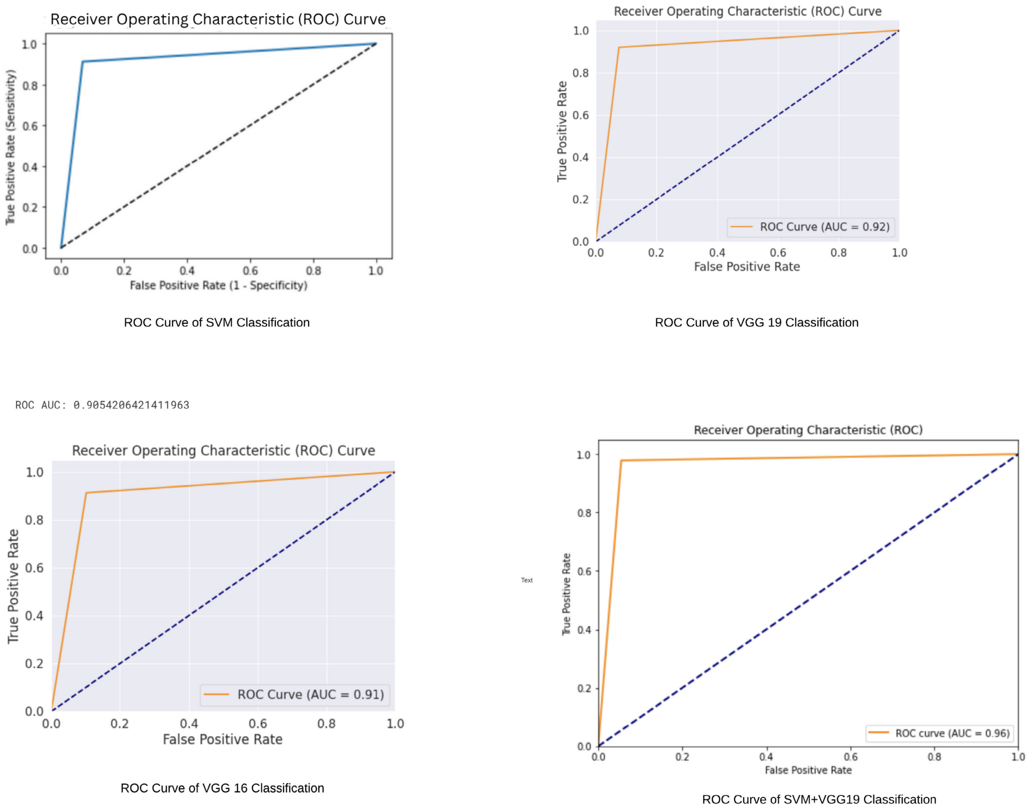

FIGURE 16. (a, b) Receiver Operating Characteristic Curve (ROC) for Support Vector Machine and VGG 19 (c,d) ROC of VGG 16 and SVM+VGG19 Classification.

TABLE 10. Strengths and weaknesses of proposed methods.

Strengths	Weaknesses
Effective use of Convolutional Neural Networks (CNNs) for skin cancer classification. - Achieved promising accuracies on ISIC 2019 and ISIC 2020 datasets, demonstrating the efficacy of CNN-based models.	Optimizing the computational complexity could enhance the practicality and scalability of the proposed methods.
Rigorous preprocessing involving Generative Adversarial Networks (GANs) and Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) for data augmentation. - Evaluation and comparison of generative techniques for augmentation.	To further enhance the clinical applicability and trustworthiness of the model, future research could focus on improving interpretability and providing insights into why the model makes specific predictions, facilitating better understanding by healthcare professionals.
Rigorous preprocessing of the dataset using generative AI techniques, enhancing dataset quality. - Use of benchmark datasets (ISIC 2019 and ISIC 2020) adds credibility to the research.	

VI. CONCLUSION

Skin cancer, recognized as one of the most deadly forms of cancer, poses a significant threat to individuals worldwide. Among its types, malignant melanoma stands out as a particularly dreadful form, characterized by the abnormal growth of melanocyte cells. While skin lesions are prevalent, accurately characterizing them and automating the identification of malignant tumors from dermoscopy images remain complex challenges in the field.

This study aimed to address these challenges by leveraging convolutional neural networks (CNNs) to diagnose and classify various types of skin lesions, using the extensive ISIC 2019 and ISIC 2020 datasets. To prepare the data for analysis, a comprehensive preprocessing pipeline is implemented, including image augmentation, normalization, and resizing. Notably, advanced techniques such as ImageDataGenerator and GAN are employed for effective data augmentation, and the resulting outcomes are thoroughly compared. The proposed CNN-based approach involved aggregating results from multiple iterations, enabling the enhancement of overall classification accuracy. Moreover, an extensive evaluation of different transfer learning models, including VGG16, VGG19, SVM, and a hybrid VGG19+SVM model, is conducted to assess their efficacy in skin lesion classification.

The experimental results on the ISIC 2019 dataset showcased promising accuracy levels, with the VGG16 and VGG19 models achieving accuracies of 92% and 93% respectively. Particularly noteworthy is the hybrid

VGG19+SVM model, which exhibited the highest accuracy of 96%. On the ISIC 2020 dataset, the VGG16, VGG19 and SVM models achieved accuracies of 90%, 92% and 92% respectively. These findings underscore the potential of CNN-based approaches in the accurate diagnosis and classification of skin lesions. Leveraging transfer learning models, such as VGG16, VGG19, SVM, and the hybrid VGG19+SVM model, proved instrumental in achieving higher classification accuracy. This study contributes valuable insights to the field of automated skin cancer detection and classification, paving the way for improved diagnosis and treatment strategies. The strengths and weaknesses of this research are summarized in table 10.

ACKNOWLEDGMENT

This work was supported by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) under Grant IMSIU-RP23042.

DECLARATIONS

CONFLICT OF INTEREST

There is no conflict of interest between the authors.

AVAILABILITY OF DATA

The datasets (ISIC-2019 and ISIC-2020) used in this research are bench mark datasets, available publicly.

CODE AVAILABILITY

Code can be provided on request.

AUTHORS' CONTRIBUTIONS

The author Mudassir Saeed worked on algorithm implementation and wrote the initial draft of the article. The author Asma Naseer floated the idea, supervised the implementation and improved the write-up. The author Hassan Masood helped in implementation and paper write-up. The author Shafiq Ur Rehman verified and analyzed the results and contributed in article write-up. The author Volker Gruhn improved the algorithm and the article write-up.

REFERENCES

[1] H. C. Reis, V. Turk, K. Khoshelham, and S. Kaya, "InSiNet: A deep convolutional approach to skin cancer detection and segmentation," *Med. Biol. Eng. Comput.*, vol. 60, no. 3, pp. 643–662, Mar. 2022.

[2] M. Shorfuzzaman, "An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection," *Multimedia Syst.*, vol. 28, no. 4, pp. 1309–1323, Aug. 2022.

[3] S. Benyahia, B. Meftah, and O. Lézoray, "Multi-features extraction based on deep learning for skin lesion classification," *Tissue Cell*, vol. 74, Feb. 2022, Art. no. 101701.

[4] M. K. Hasan, M. T. E. Elahi, M. A. Alam, M. T. Jawad, and R. Martí, "DermoExpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation," *Informat. Med. Unlocked*, vol. 28, Jan. 2022, Art. no. 100819.

[5] C. Agyenta and M. Akanzawon, "Skin lesion classification based on convolutional neural network," *J. Appl. Sci. Technol. Trends*, vol. 3, no. 1, pp. 14–19, Jul. 2022.

[6] S. Alkhushayni, D. Al-zaleq, L. Andradi, and P. Flynn, "The application of differing machine learning algorithms and their related performance in detecting skin cancers and melanomas," *J. Skin Cancer*, vol. 2022, pp. 1–11, May 2022.

[7] V. R. Allugunti, "A machine learning model for skin disease classification using convolution neural network," *Int. J. Comput., Program. Database Manag.*, vol. 3, no. 1, pp. 141–147, Jan. 2022.

[8] D. C. Malo, M. M. Rahman, J. Mahbub, and M. M. Khan, "Skin cancer detection using convolutional neural network," in *Proc. IEEE 12th Annu. Comput. Commun. Workshop Conf. (CCWC)*, 2022, pp. 0169–0176.

[9] S. Bhosale, "Comparison of deep learning and machine learning models and frameworks for skin lesion classification," 2022, *arXiv:2207.12715*.

[10] S. Bechelli and J. Delhommelle, "Machine learning and deep learning algorithms for skin cancer classification from dermoscopic images," *Bioengineering*, vol. 9, no. 3, p. 97, Feb. 2022.

[11] K. Das, C. J. Cockerell, A. Patil, P. Pietkiewicz, M. Giulini, S. Grabbe, and M. Goldust, "Machine learning and its application in skin cancer," *Int. J. Environ. Res. Public Health*, vol. 18, no. 24, p. 13409, Dec. 2021.

[12] W. Gouda, N. U. Sama, G. Al-Waakid, M. Humayun, and N. Z. Jhanjhi, "Detection of skin cancer based on skin lesion images using deep learning," *Healthcare*, vol. 10, no. 7, p. 1183, Jun. 2022.

[13] S. Shen et al., "A low-cost high-performance data augmentation for deep learning-based skin lesion classification," *BME Frontiers*, vol. 2022, 2022.

[14] A. Panja, J. J. Christy, and Q. M. Abdul, "An approach to skin cancer detection using Keras and tensorflow," *J. Phys., Conf.*, vol. 1911, no. 1, May 2021, Art. no. 012032.

[15] F. Afza, M. Sharif, M. A. Khan, U. Tariq, H.-S. Yong, and J. Cha, "Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine," *Sensors*, vol. 22, no. 3, p. 799, Jan. 2022.

[16] D. Popescu, M. El-khatib, and L. Ichim, "Skin lesion classification using collective intelligence of multiple neural networks," *Sensors*, vol. 22, no. 12, p. 4399, Jun. 2022.

[17] T. McIlwain, "A two-step, two-pronged approach to binary classification of melanoma," Dept. Bioeng., Stanford Univ., Tech. Rep. 103172581, 2021.

[18] Y. Zhang and C. Wang, "SIIM-ISIC melanoma classification with DenseNet," in *Proc. IEEE 2nd Int. Conf. Big Data, Artif. Intell. Internet Things Eng. (ICBAIE)*, Mar. 2021, pp. 14–17.

[19] W. Jiahao, J. Xingguang, W. Yuan, Z. Luo, and Z. Yu, "Deep neural network for melanoma classification in dermoscopic images," in *Proc. IEEE Int. Conf. Consum. Electron. Comput. Eng. (ICCECE)*, Jan. 2021, pp. 666–669.

[20] S. Singha and P. Roy, "Skin cancer classification and comparison of pretrained models performance using transfer learning," pp. 1-8, 2022. [Online]. Available: https://doi.org/10.20944/preprints202209.0215.v1

[21] D. Keerthana, V. Venugopal, M. K. Nath, and M. Mishra, "Hybrid convolutional neural networks with SVM classifier for classification of skin cancer," *Biomed. Eng. Adv.*, vol. 5, Jun. 2023, Art. no. 100069.

[22] S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, "Deep learning-based transfer learning for classification of skin cancer," *Sensors*, vol. 21, no. 23, p. 8142, Dec. 2021.

[23] M. Tahir, A. Naeem, H. Malik, J. Tanveer, R. A. Naqvi, and S.-W. Lee, "DSCC_Net: Multi-classification deep learning models for diagnosing of skin cancer using dermoscopic images," *Cancers*, vol. 15, no. 7, p. 2179, Apr. 2023.

[24] A. Naseer, T. Yasir, A. Azhar, T. Shakeel, and K. Zafar, "Computer-aided brain tumor diagnosis: Performance evaluation of deep learner CNN using augmented brain MRI," *Int. J. Biomed. Imag.*, vol. 2021, pp. 1–11, Jun. 2021.

[25] Y. S. Malik, M. Tamoor, A. Naseer, A. Wali, and A. Khan, "Applying an adaptive Otsu-based initialization algorithm to optimize active contour models for skin lesion segmentation," *J. X-Ray Sci. Technol.*, vol. 30, no. 6, pp. 1169–1184, Nov. 2022.

[26] A. Naseer, M. Tamoor, and A. Azhar, "Computer-aided COVID-19 diagnosis and a comparison of deep learners using augmented CXRs," *J. X-Ray Sci. Technol.*, vol. 30, no. 1, pp. 89–109, Jan. 2022.

[27] N. Raza, A. Naseer, M. Tamoor, and K. Zafar, "Alzheimer disease classification through transfer learning approach," *Diagnostics*, vol. 13, no. 4, p. 801, Feb. 2023.

[28] J. Raja, P. Shanmugam, and R. Pitchai, "An automated early detection of glaucoma using support vector machine based visual geometry group 19 (VGG-19) convolutional neural network," *Wireless Pers. Commun.*, vol. 118, no. 1, pp. 523–534, May 2021.

MUDASSIR SAEED received the B.S. degree in computer science from The Islamia University of Bahawalpur, Pakistan, and the M.S. degree in computer science from the National University of Computer and Emerging Science (NUCES), Lahore, Pakistan, in 2023. He has been a Software Engineer (remotely) in UT, USA, since 2022.

ASMA NASEER received the M.S. and Ph.D. degrees in computer science from the National University of Computer and Emerging Science (NUCES), Lahore, Pakistan, in 2008 and 2019, respectively.

She is currently with NUCES. Before joining NUCES, she was a Faculty Member with the Department of Computer Science, University of Management and Technology (UMT), from 2010 to 2021. She is a dedicated Artificial Intelligence (AI) and Machine Learning (ML) Expert, with almost 15 years of experience. She has been awarded full scholarships from the Higher Education Commission (HEC) Pakistan and other local and foreign bodies for the master's and Ph.D. studies.

HASSAN MASOOD received the B.S. degree in computer science from COMSATS University Islamabad, Abbottabad Campus, Pakistan, in 2020, and the M.S. degree in computer science from NUCES, Lahore, Pakistan, in 2023. He is currently a Data Scientist.

SHAFIQ UR REHMAN received the M.S. degree in computer science from the Dresden University of Technology, Dresden, Germany, and the Ph.D. degree in computer science from the Department of Software Engineering, University of Duisburg–Essen, Germany, in 2020. He was a Consultant (Requirements Engineer) in well-renowned international organizations in Germany. He is currently an Assistant Professor with the College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia. He is involved in different international-funded projects in the field of cyber-physical systems and cybersecurity. He has published several research papers in high-ranked international conferences and ISI-indexed journals. His research interests include AI, cyber-physical systems, cybersecurity, and requirements engineering.

VOLKER GRUHN received the M.S. and Ph.D. degrees in computer science from the University of Dortmund, Germany, in 1987 and 1991, respectively. He was with the Fraunhofer Institute for Software and System Technology. He co-founded adesso AG, in 1997, where he is currently the Supervisory Board Chairperson. He has been the Chair of software engineering with the University of Duisburg–Essen, Germany, since 2010. He supervised several Ph.D. students. He has published more than 300 research papers in ISI-indexed journals and high-ranked international conferences. His research interests include methods for industrial software engineering, as well as the effects of digital transformation on enterprises.