Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification

M Mohamed Musthafa[1], Mahesh T R[2], Vinoth Kumar V[3] and Suresh Guluwadi[4*]

Abstract

Skin cancer stands as one of the foremost challenges in oncology, with its early detection being crucial for successful treatment outcomes. Traditional diagnostic methods depend on dermatologist expertise, creating a need for more reliable, automated tools. This study explores deep learning, particularly Convolutional Neural Networks (CNNs), to enhance the accuracy and efficiency of skin cancer diagnosis. Leveraging the HAM10000 dataset, a comprehensive collection of dermatoscopic images encompassing a diverse range of skin lesions, this study introduces a sophisticated CNN model tailored for the nuanced task of skin lesion classification. The model's architecture is intricately designed with multiple convolutional, pooling, and dense layers, aimed at capturing the complex visual features of skin lesions. To address the challenge of class imbalance within the dataset, an innovative data augmentation strategy is employed, ensuring a balanced representation of each lesion category during training. Furthermore, this study introduces a CNN model with optimized layer configuration and data augmentation, significantly boosting diagnostic precision in skin cancer detection. The model's learning process is optimized using the Adam optimizer, with parameters fine-tuned over 50 epochs and a batch size of 128 to enhance the model's ability to discern subtle patterns in the image data. A Model Checkpoint callback ensures the preservation of the best model iteration for future use. The proposed model demonstrates an accuracy of 97.78% with a notable precision of 97.9%, recall of 97.9%, and an F2 score of 97.8%, underscoring its potential as a robust tool in the early detection and classification of skin cancer, thereby supporting clinical decision-making and contributing to improved patient outcomes in dermatology.

Keywords Skin lesion classification, Deep learning, Convolutional neural networks (CNN), Dermatoscopic images, Image Preprocessing, Data Augmentation

*Correspondence:
Suresh Guluwadi
suresh.guluwadi@astu.edu.et
[1] Al-Ameen Engineering College (Autonomous), Erode, Tamil Nadu, India
[2] Department of Computer Science and Engineering, JAIN (Deemed-to-be University), Bengaluru 562112, India
[3] School of Computer Science Engineering and Information Systems, Vellore Institute of Technology University, Vellore 632014, India
[4] Adama Science and Technology University, Adama 302120, Ethiopia

Introduction

Skin, the largest organ of the human body, serves as the primary barrier against environmental factors and plays a crucial role in protecting the body from various pathogens and harmful UV radiation. However, being the outermost layer, it is susceptible to a range of diseases, including various forms of skin cancer, which is among the most common types of cancer globally.

Skin diseases can range from benign conditions like acne or eczema to severe forms of cancer such as

Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 2 of 20

melanoma, basal cell carcinoma, and squamous cell carcinoma. Actinic keratoses (AK) and intraepithelial carcinoma (AKIEC) represent precancerous and early malignant growths, respectively, often arising from prolonged UV exposure. Basal cell carcinoma (BCC) is the most prevalent form of skin cancer, typically manifesting as pearly or waxy bumps, while melanoma (MEL) is an extremely aggressive malignancy arising from melanocytes, with potential for metastasis if not promptly treated. Benign keratosis-like lesions (BKL), dermatofibroma (DF), melanocytic nevi (NV), and vascular lesions present a spectrum of non-cancerous growths, each requiring careful assessment to rule out malignancy or address cosmetic concerns [1].

The differentiation between these conditions is vital, as it determines the treatment path and prognosis. The timely identification and treatment of AK and AKIEC can prevent progression to invasive carcinoma, while prompt intervention for BCC and MEL can mitigate the risk of metastasis and improve outcomes [2]. Furthermore, understanding the benign nature of BKL, DF, NV, and vascular lesions helps avoid unnecessary interventions while addressing patient concerns effectively.

Skin cancer specifically poses a significant health challenge due to its potential to spread to other body parts if not detected and treated early. Regular skin examinations, sun protection measures, and awareness of concerning changes in moles or skin lesions are essential for maintaining skin health and preventing the development or progression of skin cancer. Collaboration between patients, dermatologists, and healthcare providers is vital for comprehensive skin care, facilitating timely diagnosis, intervention, and long-term management strategies tailored to individual needs and risk factors [3].

Early detection of skin cancer significantly increases the chances of successful treatment and patient survival. Dermatologists traditionally diagnose skin conditions using visual examinations, which may be supplemented by dermoscopy, a non-invasive skin surface microscopy. However, these methods rely heavily on the clinician's experience and can sometimes lead to subjective assessments.

The advent of machine learning and, more specifically, deep learning, has introduced new possibilities in the field of dermatology. By leveraging large datasets of dermatoscopic images, deep learning models, particularly Convolutional Neural Networks (CNNs), can be trained to recognize and classify various skin lesions with high accuracy [4]. This approach has the potential to support dermatologists in making more accurate diagnoses, leading to improved patient outcomes.

The motivation behind this research is to harness the power of deep learning to aid in the early detection and accurate classification of skin cancer, thereby addressing a critical gap in dermatological care. Despite advancements in medical technology, the rising incidence of skin cancer worldwide underscores the need for more accessible and reliable diagnostic tools. The ability of CNNs to analyze complex dermatoscopic images and learn from vast amounts of data presents an opportunity to develop an automated system that can assist even non-specialists in identifying potential skin cancers at an early stage [5]. This research is driven by the goal of democratizing access to high-quality diagnostic tools, especially in regions where dermatological expertise is scarce, and thereby contributing to the global fight against skin cancer.

This research paper presents a detailed study on the application of Convolutional Neural Networks to classify skin lesions using the HAM10000 dataset, which is a diverse collection of dermatoscopic images. The contributions of this research paper are:

- Introduce a sophisticated CNN model tailored for skin lesion classification, which significantly enhances diagnostic accuracy.
- Employ innovative data augmentation methods to address the issue of class imbalance within the HAM10000 dataset, enhancing the robustness of the model.
- Integrate Model Checkpoint callbacks to preserve the best model iteration during training, ensuring optimal performance in practical applications [6].

The next section provides a thorough review of the literature, highlighting previous studies on the use of deep learning in dermatology. The Methodology section details the dataset, the CNN architecture, the training process, and the evaluation metrics used in this study [7]. The results section presents the outcomes of the model training and validation, including performance metrics and a discussion on the model's diagnostic accuracy. Finally, the paper concludes with a summary of the key findings and suggestions for further research in this rapidly evolving field.

Related work

The fusion of artificial intelligence (AI) and deep learning has catalysed groundbreaking progress, significantly enhancing the precision of skin cancer diagnoses [8]. A notable study unveiled a convolutional neural network (CNN) that eclipsed dermatologists in identifying skin lesions, showcasing the model's adeptness at distinguishing between benign and malignant lesions and illustrating AI's potential to bolster dermatological clinical decision-making. Another study introduced an AI framework that

Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 3 of 20

amalgamates image analysis with demographic data, presenting a more holistic diagnostic approach that surpasses conventional image-centric models in predictive accuracy, highlighting the benefits of integrating diverse data types in AI-driven diagnostics.

Traditional approaches

The traditional methods for diagnosing skin cancer primarily involve dermatological examinations and biopsies, where skilled dermatologists visually assess lesions and may take tissue samples for further histopathological analysis. These methods, while effective, rely heavily on the expertise and availability of specialists. This reliance can introduce variability in diagnosis accuracy and limits accessibility, particularly in underserved areas [9].

Deep learning revolution

The integration of deep learning, particularly Convolutional Neural Networks (CNNs), into dermatological diagnostics marks a significant shift from traditional methods. CNNs excel in analyzing vast arrays of dermatoscopic images, automating the detection and classification processes [10]. These models can process detailed visual data at speeds and accuracies that challenge or even surpass human capabilities, transforming skin cancer diagnosis into a more efficient and standardized procedure.

Challenges in deep learning

Deploying deep learning models in clinical settings is not without challenges. Key issues include the black-box nature of these models, which often lack transparency in their decision-making processes, making clinical acceptance difficult. Additionally, deep learning requires large, well-annotated datasets that are expensive and labor-intensive to compile. Models also face generalization challenges, where performance can drop when applied to new, unseen datasets, reflecting differences from the training data [11].

Recent advances

To address these challenges, recent research has focused on enhancing the robustness and applicability of deep learning models. Approaches such as transfer learning have been employed to adapt pre-trained models to new tasks with less data, significantly improving performance. Innovations continue to emerge in model architecture and training methodologies, with some studies incorporating multiple data types—combining image data with patient demographic and clinical history—to refine diagnostic accuracy. These advancements are paving the way for more reliable and accessible automated diagnostic systems in dermatology [12]. Table 1 compares the methodologies and results to contextualize this study within the broader field.

Methodology

This section delineates the methodology adopted in this research to develop and evaluate the CNN model for classifying skin lesions using the HAM10000 dataset. It includes details on data preprocessing, model architecture, training procedures, and evaluation

Table 1 Recent studies

Study	Dataset Used	Accuracy	Remarks
Bhuvaneshwari Shetty et al. (2022) [13]	HAM10000 dataset (subset with augmentation)	95.18%	Focuses on detecting dangerous skin illnesses, particularly skin cancer.
Ahmad Hameed et al. (2023) [14]	MNIST HAM10000 dataset	95.2%	Addresses challenges in skin lesion detection and classification.
Md. Khairul Islam et al. (2021) [15]	HAM10000 dataset	96.10% in training, 90.93% in testing	Focus on skin cancer detection.
Ankita Pramanik & Rivu Chakraborty (2021) [16]	Dermatoscopic images from the Kaggle dataset archive	87.58%	Study on skin cancer detection in Caucasians in the USA.
Raut, Roshani et al. (2023) [17]	Dermoscopy images database	90.1%	AI integration for early melanoma identification.
Kalaycı, Serdar (2023) [18]	ISIC 2019 and ISIC 2020 datasets	AUC: 95.75%	Focus on early skin cancer detection.
Maaz Ul Amin et al. (2024) [19]	HAM10000 dataset, dataset of 3672 categorized pictures	84.27%	Study on global skin cancer concerns.
Aya Mostafa Mosa et al. (2022) [20]	Augmented dataset of 17,731 images + 3,800 from different types	Resnet18: 79.5%, Densnet121: 81.2%, InceptionV4: 82.6%	Focus on melanoma caused by UV exposure.
Mariame Oumoulylte et al. (2023) [21]	ISIC dataset	87%	Deep-learning models for binary classification of skin cancer.
Mohammad Atikur Rahman et al. (2024) [22]	Dataset of 2637 skin images	Optimized NASNet Mobile: 85.62%, NASNet Large: 83.98%	DCNN-based model for skin cancer classification.

metrics. Figure 1 depicts the diagrammatic representation of the workflow of prosposed model.

Dataset description

The HAM10000 dataset, integral to dermatological advancements, encompasses over 10,000 dermatoscopic images, crucial for developing machine learning models for pigmented skin lesion diagnosis. It offers unparalleled diversity in skin tones, lesion types, and conditions, featuring seven distinct skin cancer categories: Actinic keratoses, intraepithelial carcinoma/ Bowen's disease, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanoma, and melanocytic nevi, alongside various vascular lesions. This diversity is vital for training robust algorithms. Over half of the lesions are histopathologically verified, with the remainder confirmed via follow-up, expert consensus, or confocal microscopy, ensuring data authenticity. The dataset's utility is enhanced by metadata, such as lesion IDs, allowing for longitudinal lesion analysis, crucial for developing models that recognize temporal lesion changes. The HAM10000 dataset is a comprehensive, validated resource that propels the development of diagnostic tools in dermatology, offering a real-world dataset challenge to automate skin lesion analysis. Its depth, verification rigor, and inclusion of multiple cancer types make it an indispensable tool for improving dermatological diagnostic accuracy and reliability through machine learning. Figure 2 demonstrates different categories of skin cancer in the dataset.

Image Preprocessing

Data preprocessing is a crucial step in preparing the dataset for training with a Convolutional Neural Network (CNN). It ensures that the input data is in a suitable format and is conducive to the learning process. In this study, the data preprocessing involved three main steps: image resizing, normalization, and data augmentation. It ensures that the model treats all features (pixels, in this case) equally.

Image resizing

To ensure that all images fed into the CNN have a consistent shape and size, each image in the HAM10000 dataset was resized to a standard dimension. This uniformity is essential because CNNs require a fixed size for all inputs to maintain a consistent architecture, especially when it comes to applying filters and pooling operations. Resizing images helps in creating a standardized input feature dimension, enabling the network to systematically extract features across all samples. The formula for resizing the image to a desired width and height is presented in Eq. 1.

$$I_{\text{resized}} = \text{resize}(I_{\text{original}}, (W_{\text{new}}, H_{\text{new}}) \tag{1}$$

Here, I_{original} represent the original image with dimensions $W_{\text{original}} * H_{\text{original}}$

Normalization

Normalization refers to the process of scaling the pixel values of the images to a range of 0 to 1. This is achieved by dividing the pixel values by 255 (since
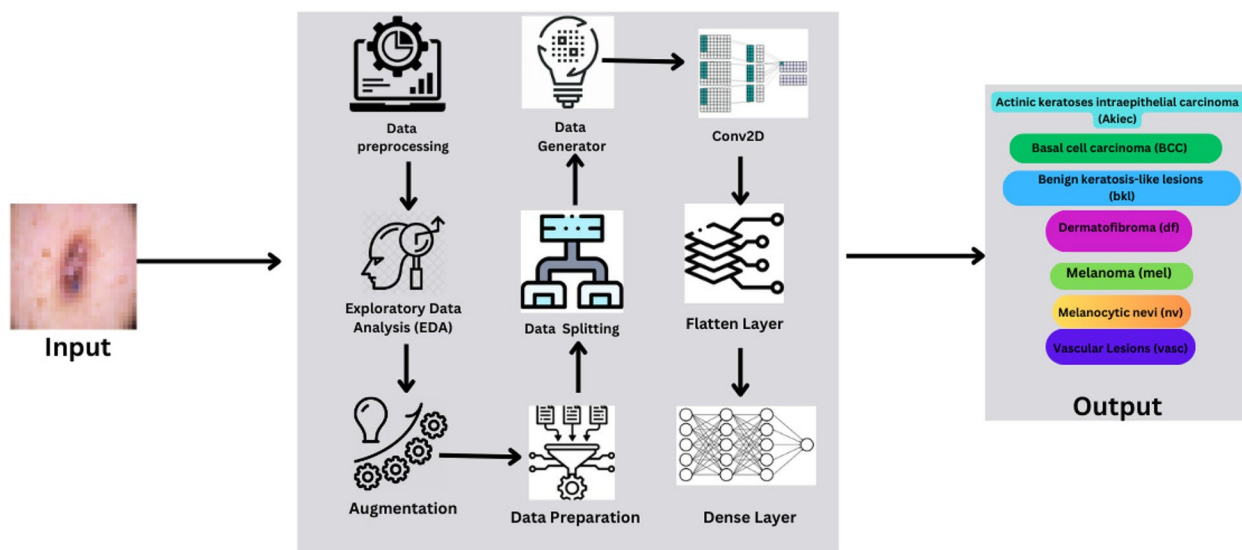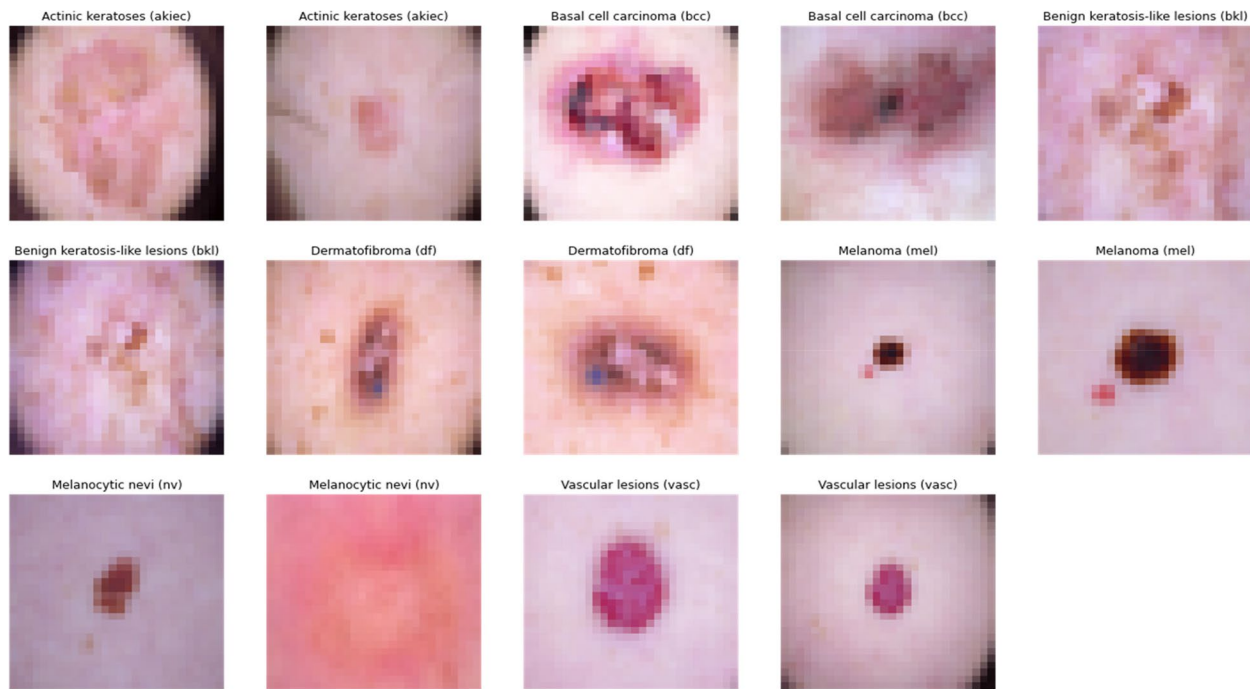

Fig. 1 Workflow of the model

Fig. 2 Input Images from dataset

pixel values range from 0 to 255) [23]. Normalizing the data is beneficial for the training process as it helps in speeding up the convergence by reducing the initial variability of the weights. When data is normalized, the gradients used in backpropagation are in a more manageable range, which helps in a smoother and faster optimization process, leading to more stable and quicker convergence. The formula for normalization is shown in Eq. 2.

$$I_{\text{normalized}}(x, y) = \frac{I_{\text{original}}(x, y)}{255} \qquad (2)$$

Where, $I_{\text{normalized}}$ represent the normalized image and $I_{\text{original}}(x, y)$ represent the pixel intensity value of the original image at location (x, y).

Data augmentation

Dataset augmentation is a critical strategy in machine learning that enriches training data diversity, particularly vital in image processing. By applying specific transformations like rotation, zooming, flipping, shearing, and brightness adjustment, this study simulate various real-world scenarios, thereby broadening the model's exposure to potential data variations it might encounter post-deployment. These transformations not only enhance the model's ability to generalize across new, unseen data, thereby improving its predictive accuracy, but also serve as a regularization technique, significantly reducing the risk of overfitting. Overfitting is a common challenge where the model learns noise alongside the underlying pattern, which can degrade its performance on novel data [1]. Through dataset augmentation, the model's robustness and adaptability is ensured, reinforcing its performance stability across a spectrum of conditions, leading to more reliable and accurate predictions in practical applications. Together, these preprocessing steps form an integral part of the data preparation pipeline, setting a strong foundation for the subsequent training of the CNN model. Figure 3 showcases distribution of data after augmentation. The formula for data augmentation can vary depending on the specific transformation applied is represented in Eq. 3.

$$I_{\text{augmented}} = \text{transform}(I_{\text{original}}, T) \qquad (3)$$

Where, $I_{\text{augmented}}$ represent the augmented image obtained from the original image I_{original} by applying transformations T.

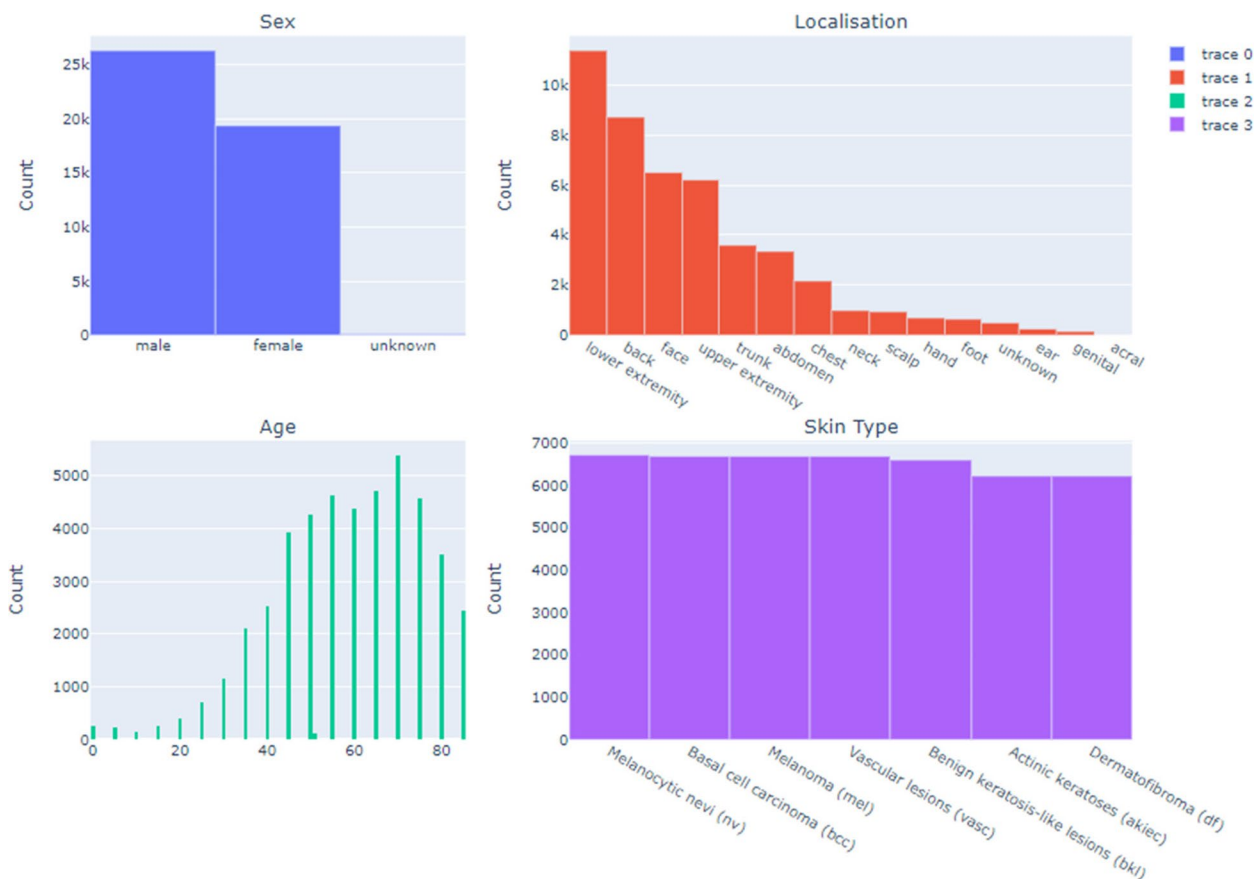Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 6 of 20

Fig. 3 Dataset distribution after augmentation

Model architecture

A Convolutional Neural Network (CNN) is a deep neural network architecture used for analysing visual imagery, known for its efficacy in various computer vision tasks. At its core, a CNN employs convolutional layers where filters or kernels extract specific features from images, with the ReLU activation function introducing necessary non-linearities. Following these, pooling layers, typically max pooling, reduce the feature maps' spatial dimensions, aiding in computational efficiency and spatial invariance. The network then utilizes dense or fully connected layers, where the data, now flattened, undergoes high-level reasoning through connections that encompass weights and further activation functions. The culmination is an output layer where a SoftMax activation function translates the final layer's outputs into probability distributions across multiple classes. In training, the Adam optimizer is favoured for its adaptive learning rate, working alongside a sparse categorical cross entropy loss function to iteratively adjust the network's weights, thereby minimizing the discrepancy between the model's predictions and the actual data labels. This intricate orchestration of layers and functions allows CNNs to adeptly learn and identify patterns in visual data, making them particularly suitable for tasks like skin lesion classification where distinguishing nuanced visual features is paramount [24]. Table 2 provides the layer type, output shape, and the number of parameters for each layer in the

Table 2 Parameters for training CNN Model

Layer (type)	Output Shape	Param #
conv2d_4	(None, 28, 28, 16)	448
max_pooling2d_4	(None, 14, 14, 16)	0
conv2d_5	(None, 14, 14, 32)	4640
max_pooling2d_5	(None, 7, 7, 32)	0
conv2d_6	(None, 7, 7, 64)	18,496
max_pooling2d_6	(None, 4, 4, 64)	0
conv2d_7	(None, 4, 4, 128)	73,856
max_pooling2d_7	(None, 2, 2, 128)	0
flatten_1	(None, 512)	0
dense_3	(None, 64)	32,832
dense_4	(None, 32)	2080
dense_5	(None, 7)	231

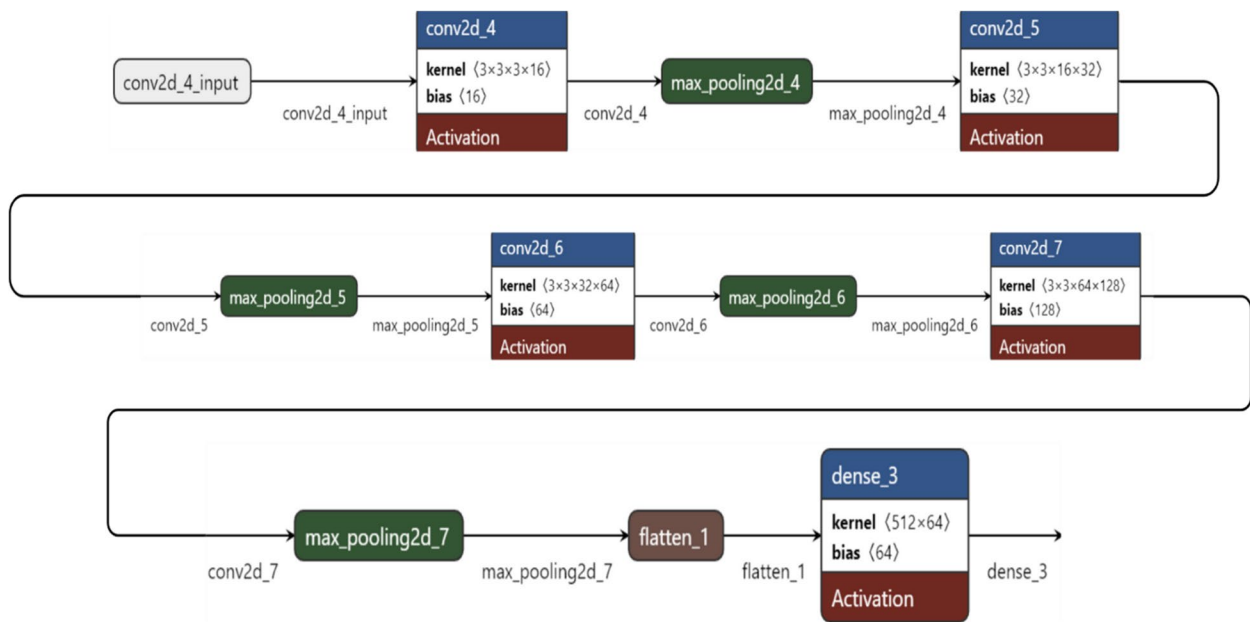Musthafa *et al. BMC Medical Imaging* (2024) 24:201

Page 7 of 20

Fig. 4 Layer wise network architecture diagram

model while Fig. 4 illustrates the network architecture of the model.

The architecture of this model, comprising sequential convolutional and pooling layers followed by a dense output layer, is strategically designed for the effective classification of skin lesions. Starting with convolutional layers (conv2d_4, conv2d_5, and conv2d_7) that progressively increase in filter depth, the model captures increasingly complex features from the dermatoscopic images, essential for distinguishing subtle differences between various skin lesion types. Each convolutional layer is paired with a max pooling layer (max_pooling2d_4, max_pooling2d_5, max_pooling2d_7), which reduces the spatial dimensions of the feature maps, thus lowering computational demands and focusing the model on dominant features critical for accurate classification. The 'flatten_1' layer transitions these 2D feature maps into a 1D vector, preparing them for the final classification performed by the dense layer (dense_3). This architecture not only aligns with proven image classification principles but also balances computational efficiency with high diagnostic performance, making it suitable for real-world clinical applications where resources may be limited and high accuracy is paramount.

The Convolutional Neural Network (CNN) architecture is designed with a strategic layering of convolutional, pooling, fully connected, and dropout layers, each serving a distinct function in processing and classifying visual data, particularly for tasks like skin lesion classification.

Convolutional layers

These layers are the building blocks of the CNN, responsible for detecting patterns such as edges, textures, and more complex shapes within the images. The depth and number of filters in these layers are chosen to capture a broad spectrum of features without overwhelming the model's capacity. This Eq. 4 represents the mathematical operation of convolution, where *f* is the input image and *g* are the filter (kernel) applied to extract features.

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau \tag{4}$$

The depth (number of filters) and size of the filters in these layers are meticulously chosen to optimize the model's ability to detect a wide range of features without overburdening its computational capacity.

Activation Function: Following the convolution operation, an activation function, typically the Rectified Linear Unit (ReLU), is applied to introduce non-linearity into the network. This step is essential as it allows the model to learn and represent more complex patterns in the data. Equation 5 defines ReLU, which introduces non-linearity into the network by outputting the input if it is positive and zero otherwise.

$$f(x) = \max(0, x) \tag{5}$$

Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 8 of 20

Pooling layers

After each convolutional layer, pooling layers reduce the dimensionality of the data, summarizing the features extracted while retaining the most salient information. This reduction is crucial for minimizing computational load and enhancing the model's focus on essential features. Equation 6 describes the max-pooling operation, which down samples the feature maps generated by convolutional layers.

$$max\ pooling(x,y) = max_{i \in R}x[i,y] \qquad (6)$$

Flatten layer

This layer transitions from the 2D output of the pooling layers to a 1D vector. It's a crucial step, transforming the processed image data into a format suitable for the fully connected layers, enabling the network to interpret the extracted features comprehensively. Equation 7 reshapes the output of the convolutional layers into a one-dimensional vector, which serves as input to the fully connected layers.

$$f\left(x_{ij}\right) = x_k \qquad (7)$$

Fully connected

Following the extraction and down-sampling of features, the network transitions to fully connected layers. In these layers, neurons have connections to all activations in the previous layer, as opposed to convolutional layers where neurons are connected to only a local region of the input. This architecture allows the network to combine features learned across the entire image, enabling high-level reasoning and classification, and is computed by using the formula depicted in Eq. 8.

$$y = f(Wx + b) \qquad (8)$$

Dropout layers

To mitigate the risk of overfitting, dropout layers are incorporated, randomly disabling a fraction of the neurons during training. This forces the network to learn more robust features that are not reliant on a small set of neurons, enhancing its generalization capability. To adjust for dropout at training time, activations at test time are scaled by the dropout probability p as in Eq. 9.

$$\widetilde{x} = \frac{x}{p} \cdot \text{Bernoulli}(p) \qquad (9)$$

Batch normalization

Batch normalization standardizes the inputs of each layer, improving the stability and speed of the network's learning phase. The formula in Eq. 10 scales the activations of a layer by the learned parameter d_i. Equation 11 calculates the batch-normalized output for each activation and Eq. 12 adjusts the activations during training and testing phases to maintain consistency.

$$a_i = a_i \cdot d_i \qquad (10)$$

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \in}} \qquad (11)$$

$$a_i^{test} = p \cdot a_i^{train} \qquad (12)$$

Output layer

The final layer in a CNN is the output layer, where a SoftMax activation function is typically used for multiclass classification tasks. The SoftMax function converts the output scores from the final dense layer into probability values for each class. In the context of skin lesion classification, it provides the probabilities of an image belonging to each of the seven lesion categories. The SoftMax function, expressed in Eq. 13 converts the logits (raw predictions) from the last dense layer into a probability distribution across the classes. This equation represents the partial derivative of the SoftMax output S_j with respect to the input xj, where δ_{ij} is the Kronecker delta.

$$\frac{\partial S_i}{\partial x_j} = S_i\left(\delta_{ij} - S_j\right) \qquad (13)$$

The CNN architecture adeptly learns to identify and classify complex patterns in visual data, making it an ideal choice for medical image analysis, including skin lesion classification. Algorithm 1 encapsulates a comprehensive approach to classify dermatological conditions using CNNs, addressing key aspects like data preprocessing, augmentation, model construction, training, and evaluation.

Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 9 of 20

Algorithm 1. Enhanced image classification for dermatological conditions

---

Input: Set of dermatoscopic images ($D = \{d_1, d_2, ..., d_n\}$), where each image ($d_i$) is labeled with a corresponding skin condition.

Output: A trained model capable of accurately classifying unseen dermatoscopic images into specific skin conditions.

1. **Data Preparation:**
   - Load the dataset (D) containing dermatoscopic images and their corresponding labels.
   - Split the dataset into training set(D_{train}) and test set (D_{test}) using a stratified approach to maintain label distribution.

2. **Preprocessing:**
   - For each image (d_i) in (D):
   - Resize (d_i) to a standard dimension, e.g., 28x28 pixels, to ensure uniformity.
   - Apply grayscale or RGB normalization to standardize image intensity values.
   - Replace missing values, if any, in the metadata (e.g., age or sex) with appropriate statistical measures (mean or mode).

3. **Data Augmentation (for training set (D_{train})):**
   - Perform image augmentations to generate additional training data:
   - Apply rotations, width and height shifts, shears, flips, and zooms to introduce variability.
   - Ensure that the augmented data is balanced across different classes to mitigate class imbalance.

4. **Model Construction:**
   - Initialize a sequential convolutional neural network model.
   - Add convolutional layers with increasing filter sizes (e.g., 16, 32, 64) and apply activation functions (e.g., ReLU).
   - Include max-pooling layers to reduce spatial dimensions.
   - Incorporate dropout layers to prevent overfitting.
   - Flatten the output and add fully connected layers with appropriate activation functions.
   - Conclude with a SoftMax layer to output probabilities across the target classes.

5. **Model Training:**
   - Compile the model using a suitable optimizer (e.g., Adam) and loss function (e.g., sparse categorical crossentropy).
   - Employ callbacks such as Early Stopping and ReduceLROnPlateau to optimize training.
   - Train the model on (D_{train}) using data augmentation and validate using a split from (D_{train})

6. **Evaluation and Testing:**
   - Evaluate the trained model on (D_{test}) to assess its classification performance.
   - Compute performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
   - Analyse the confusion matrix to identify model strengths and weaknesses in classifying specific conditions.

7. **Result Interpretation:**
   - Discuss the model's performance, highlighting its accuracy in distinguishing various dermatological conditions.
   - Examine misclassified cases to understand potential model limitations or data issues.

Training procedure

Training the model is a delicate balance of maximizing learning while avoiding overfitting, with various strategies employed to achieve this equilibrium.

Data splitting

The division of data into training, validation, and test sets is pivotal. The training set is the model's learning ground, while the validation set guides the tuning of

Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 10 of 20

hyperparameters and early stopping criteria. The test set remains untouched until the final evaluation, ensuring an unbiased assessment of the model's generalization capabilities [25].

Optimization and loss

The Adam optimizer is a choice grounded in its adaptive learning rate mechanism, which tailors the update magnitude for each weight, optimizing the learning process. The categorical cross-entropy loss function quantifies the disparity between the predicted probabilities and the actual distribution, guiding the model toward more accurate predictions. Figure 5 Visualizes the model's increasing accuracy through training phases, emphasizing its learning efficiency. Equation 14 adjusts the learning rate using the ReduceLRonPlateau method based on the old learning rate and a reduction factor, Eq. 15 updates the weight in the Adam optimizer by incorporating the first and second moment vectors, a small constant epsilon, and the learning

rate and Eq. 16 computes the loss in classification tasks by summing the negative logarithm of predicted probabilities weighted by true labels.

$$\eta_{new} = \text{ReduceLRonPlateau}\left(\eta_{old}, \text{factor}\right) \quad (14)$$

$$W_{t+1} = W_t - v_t + \epsilon\eta\, m_t \quad (15)$$

$$L = -\sum_i y_i \log(p_i) \quad (16)$$

While initial model training was originally planned for 50 epochs to thoroughly optimize learning and parameter adjustments, Fig. 5 reveals signs of overfitting emerging after 20 epochs. Despite this, in the study it was chosed to continue training for the full 50 epochs to fully exploit the model's learning potential under controlled conditions, facilitating fine-tuning of parameters like learning
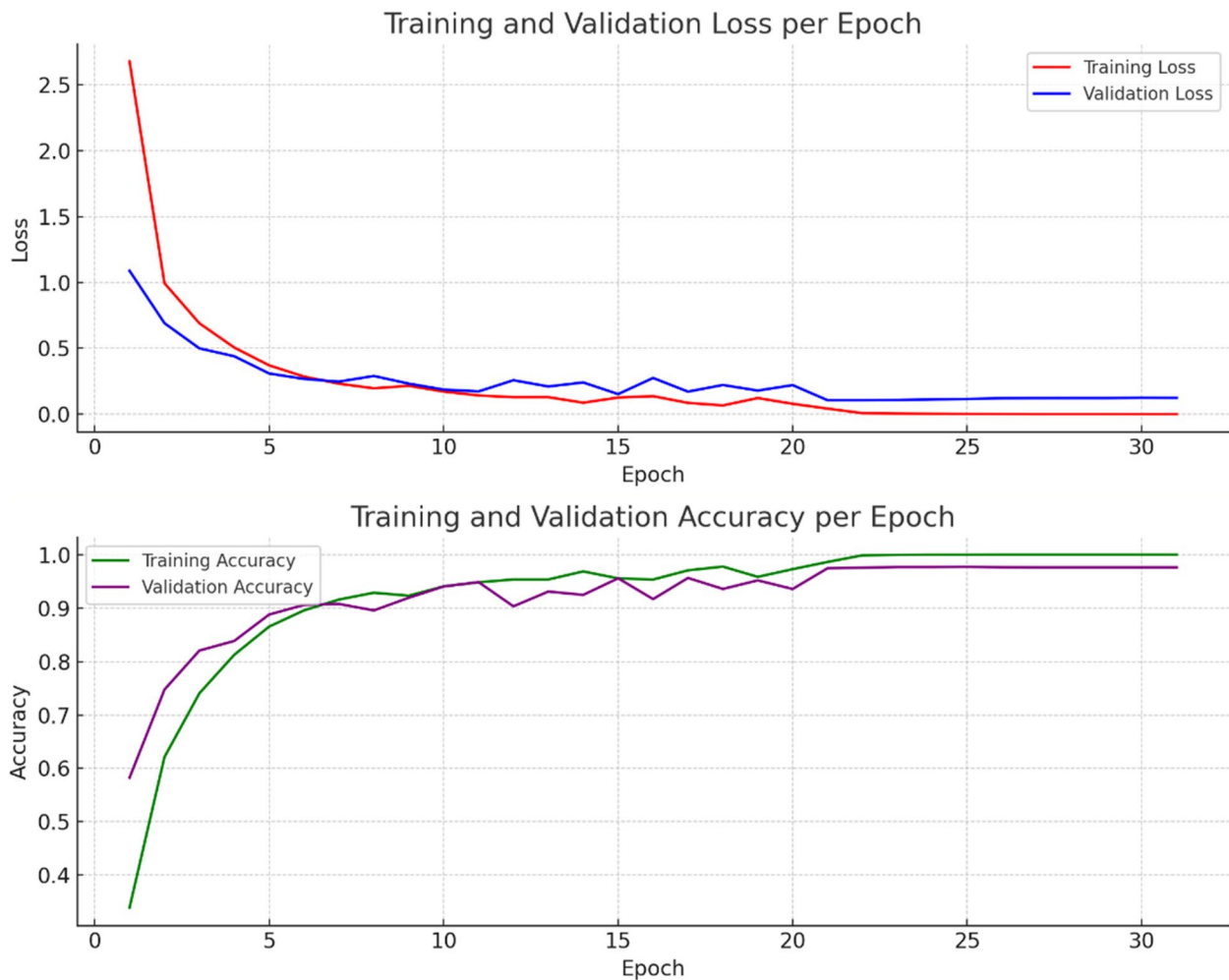

Fig. 5 Training and validation loss and accuracy per epochs

Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 11 of 20

rate and regularization techniques. To counteract the observed overfitting risk post-20 epochs, several strategies were implemented: Early Stopping was employed to halt training when validation loss ceased to improve, preventing the model from memorizing noise; Model Checkpointing saved the best-performing model based on validation set performance; and regularization techniques such as dropout and L2 regularization were utilized to mitigate overfitting by penalizing complex models. These efforts ensured this model's resilience and generalizability despite the extended training period. Detailed performance metrics in the results section underscore the effectiveness of these strategies in combating overfitting and optimizing overall model performance.

Callbacks
Early Stopping monitors the validation loss, ceasing training when improvements halt, thereby averting overfitting. Model Checkpoint saves the model at its peak performance on the validation set, ensuring that the best version is retained for evaluation and future application.

Evaluation metrics
To evaluate the effectiveness and robustness of the deep learning model, various metrics were employed:

Exploratory Data Analysis (EDA)
A comprehensive exploratory data analysis (EDA) was conducted on the HAM10000 dataset to extract critical insights for the development of a Convolutional Neural Network (CNN) aimed at diagnosing skin cancer. The analysis encompassed an examination of the distribution of patient demographics and lesion localizations, facilitating a broad representation of cases to augment the model's generalizability. Age distribution and lesion type frequencies were scrutinized to strategically address the dataset's class imbalance through targeted data augmentation techniques. An in-depth visual assessment of dermatoscopic images from each lesion category informed the CNN architecture design and data preprocessing methods. This extensive EDA process yielded essential statistical insights, guiding a data-driven model development approach that ensures accuracy, equity, and generalizability across varied patient demographics and lesion types, laying a robust foundation for the modeling endeavors that followed.

Accuracy
This metric measures the overall correctness of the model, calculated as the ratio of correctly predicted observations to the total observations [26]. It provides a high-level overview of the model's performance and is calculated by using the formula presented in Eq. 17.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (17)$$

Precision
Precision, or the positive predictive value, indicates the ratio of correctly predicted positive observations to the total predicted positives. It's crucial for scenarios where the cost of false positives is high [27]. Equation 18 illustrates the formula to calculate precision.

$$Precision = \frac{TP}{TP + FP} \qquad (18)$$

Recall (sensitivity)
Recall measures the ratio of correctly predicted positive observations to all actual positives [28]. It's particularly important in medical diagnostics, where failing to detect a condition could have serious implications. Recall is calculated using the formula depicted in Eq. 19.

$$Recall = \frac{TP}{TP + FN} \qquad (19)$$

F1-score
The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both [29]. In addition to accuracy, precision, and recall, the F1 score serves as a critical indicator of model performance, especially in the domain of medical diagnostics. The F1 score, which is the harmonic mean of precision and recall, provides a single metric that balances the trade-off between these two critical aspects. This is particularly important in medical image classification where both false positives and false negatives carry significant consequences. A high F1 score indicates not only that the model accurately identifies a high number of relevant instances but also that it minimizes the number of incorrect classifications on crucial negative cases. This balance is essential for ensuring reliable clinical decisions based on the model's predictions. It's useful when it is needed to balance precision and recall and is depicted by a formula in Eq. 20. This metric is particularly advantageous in scenarios where an uneven class distribution might render other metrics less informative. For instance, in cases where the prevalence of positive class (skin lesions that are cancerous) is much lower, precision and recall individually may not adequately reflect the performance nuances of the diagnostic model. By combining these metrics, the F1 score provides a more robust indicator of the model's

Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 12 of 20

effectiveness across various classes, thereby supporting its utility in clinical applications where accuracy and reliability are paramount.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (20)$$

Confusion matrix

This matrix provides a detailed breakdown of the model's predictions, showing the correct and incorrect predictions across different classes.

ROC curve and AUC

The Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) provide insights into the model's ability to distinguish between classes. A higher AUC indicates a better performing model. Equation 21 calculates the True Positive Rate (TPR), also known as Sensitivity or Recall. Similarly, Eq. 22 calculates the False Positive Rate (FPR) and Eq. 23 is used to calculate the Area Under the Curve (AUC) for a Receiver Operating Characteristic (ROC) curve.

$$TPR = \frac{True\ Positives}{True\ Positives + False\ Negatives} \qquad (21)$$

$$FPR = \frac{False\ Positives}{False\ Positives + True\ Negatives} \qquad (22)$$

$$AUC = \int_0^1 TPR\, d(FPR) \qquad (23)$$

Through this detailed methodology, the research aims to forge a CNN model that is not just statistically accurate but also clinically viable, providing a tool that can potentially revolutionize the early detection and classification of skin lesions.

Model training and validation

The training process involved feeding the pre-processed images and their corresponding labels into the CNN, allowing the network to iteratively learn from the data through backpropagation and gradient descent. The model's parameters were updated to minimize the loss function, which quantifies the difference between the
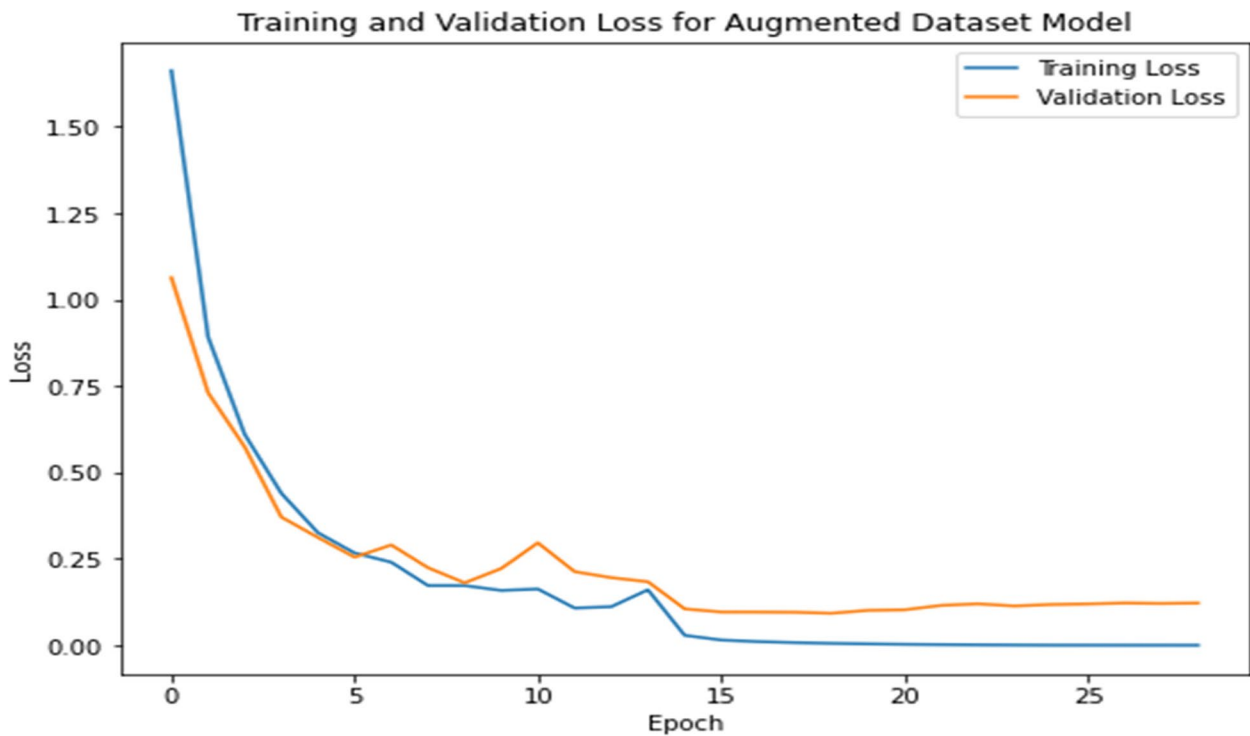

Fig. 6 Training and validation loss graph for augmented dataset

predicted and actual labels. To prevent overfitting, techniques such as dropout and early stopping were employed, wherein the training is halted when the validation loss ceases to decrease, ensuring the model's generalizability to unseen data.

During the model's training, validation played a pivotal role, offering a lens through which the model's performance on a subset of data, unseen during the training, was monitored. This phase was instrumental in identifying overfitting instances and guiding the hyperparameter tuning process. Hyperparameter tuning, an essential facet of the training regimen, involved the careful selection and adjustment of various parameters, including the learning rate, batch size, number of epochs, and architecture-specific settings such as the number and size of filters in convolutional layers, and the dropout rate. The learning rate, a critical hyperparameter, dictates the step size at each iteration while moving toward a minimum of the loss function, requiring a delicate balance to avoid underfitting or missing the minimum. Batch size influences the model's convergence speed and generalization capabilities, with smaller batches offering a regularizing effect and larger batches providing computational efficiency. Through a combination of techniques like grid search, random search, or more sophisticated methods like Bayesian optimization, these hyperparameters were iteratively adjusted, with the model's performance on the validation set serving as the benchmark for selecting the optimal combination, culminating in the selection of the best model iteration for evaluation on the test set [30]. Figure 6 tracks the model's learning progress over epochs, highlighting improvements and stabilization in loss values.

Experimentation and results

In this study an experimental framework to evaluate the performance of the Convolutional Neural Network (CNN) on the HAM10000 dataset is designed, a comprehensive collection of dermatoscopic images of skin lesions. The dataset was split into training (80%), validation (10%), and test (10%) sets to ensure a robust evaluation. The CNN architecture was constructed with multiple convolutional and pooling layers, followed by fully connected layers, and a SoftMax output layer to classify the images into seven distinct skin lesion categories. Figure 7 showcases the detection of skin cancer for different categories of skin cancer.
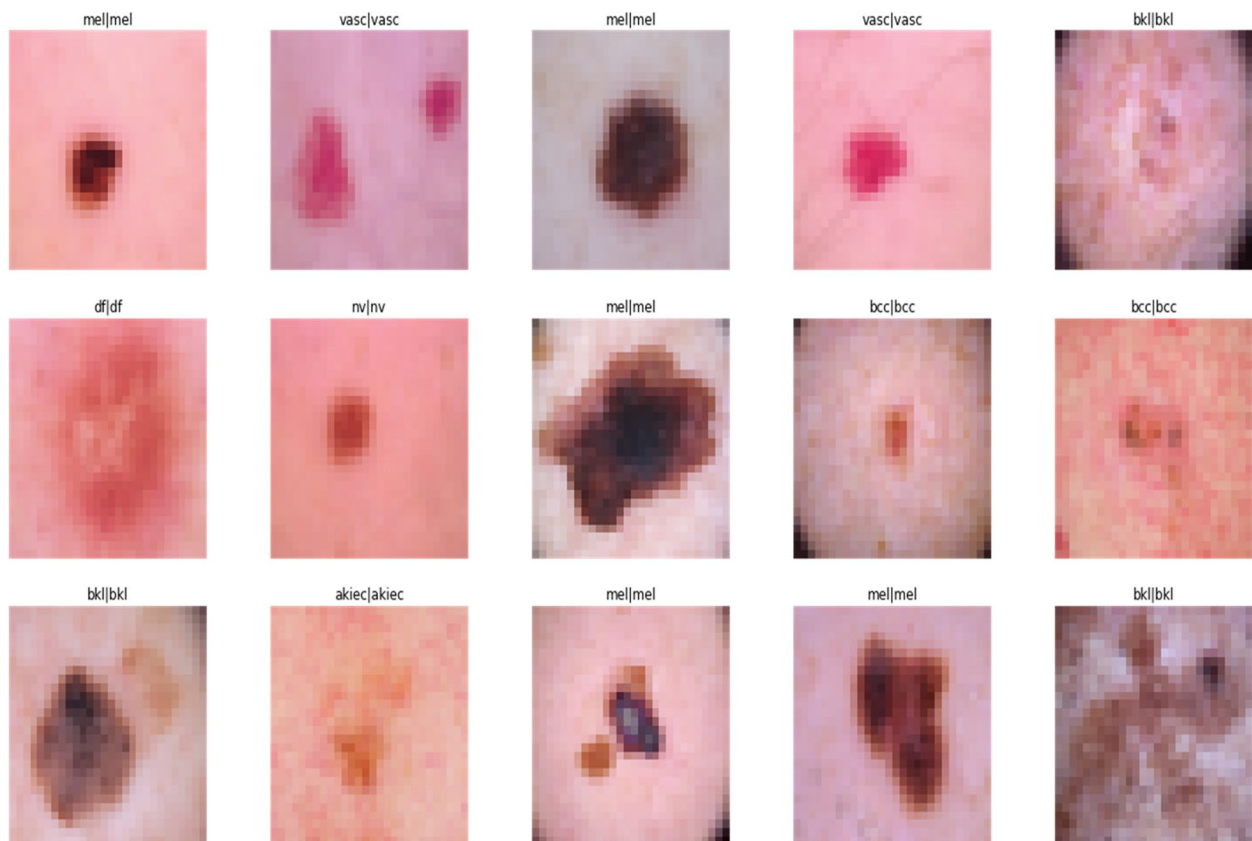

Fig. 7 Predicted label with images

To enhance the model's ability to generalize and to mitigate the risk of overfitting due to the limited size of the dataset, author's applied extensive data augmentation techniques including rotations, zooming, and flipping. This approach not only expanded training dataset but also introduced a diversity of image perspectives, simulating real-world variations. Alongside augmentation, each image was resized and normalized to maintain consistency and to facilitate faster convergence during the training process.

The CNN was trained using the Adam optimizer and a sparse categorical crossentropy loss function. In this study early stopping and model checkpoint callbacks is employed to prevent overfitting and to save the best model during training. The training was conducted over numerous epochs, with real-time monitoring of performance metrics on both training and validation sets to ensure the model's learning efficacy and generalization capability.

Upon completion of the training, the model demonstrated remarkable performance on the test set, achieving an overall accuracy of 97.858%. This high accuracy underscores the CNN's ability to effectively differentiate between various types of skin lesions. The precision, recall, and F1-scores across different categories further validated the model's robustness, with most classes

showing scores above 90%, reflecting the model's precision and reliability in classification.

The classification report provided detailed insights into the model's performance across individual classes. Notably, the model excelled in identifying Melanoma (mel) and Basal cell carcinoma (bcc), which are critical for early cancer detection. The high F1-scores in these categories indicate a balanced precision-recall trade-off, crucial in medical diagnostics where both false negatives and false positives have significant implications. Figure 8 illustrates the classification report of the model.

The confusion matrix offered a granular view of the model's classification behaviour across the various skin lesion types. It revealed that the model had a high true positive rate across most classes, with minimal confusion between different lesion types. This indicates that the CNN has effectively learned distinguishing features for each lesion category, a testament to the network's feature extraction and pattern recognition capabilities. Figure 9 depicts the confusion matrix between different categories of skin cancer.

The AUC-ROC curve analysis further cemented the model's diagnostic capability. The model showcased an excellent AUC score, indicating its strong discriminative power across all lesion classes. This performance metric is particularly crucial in medical
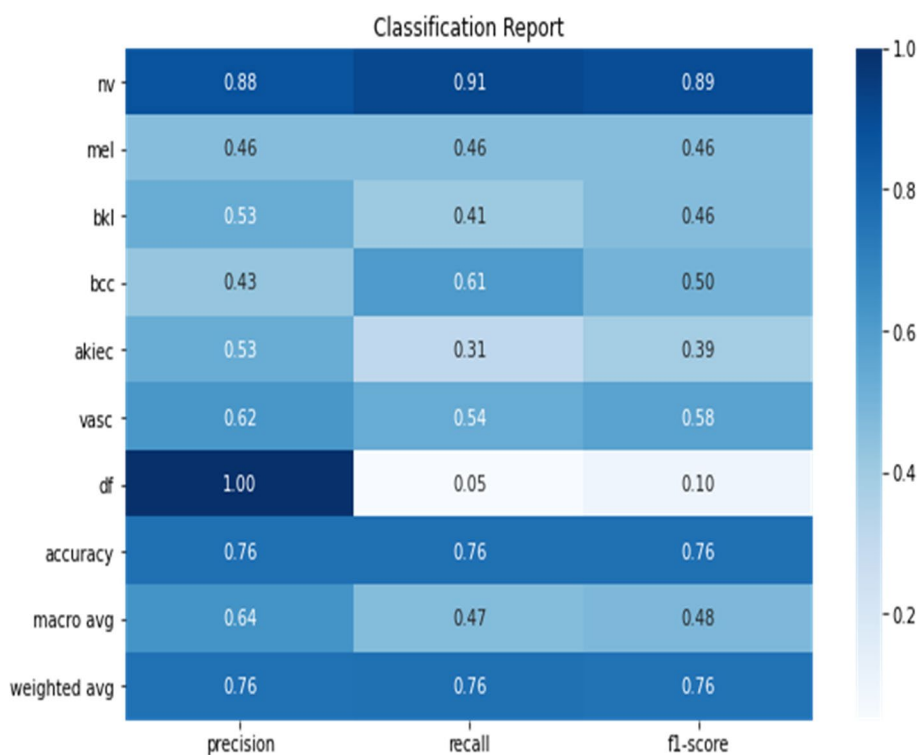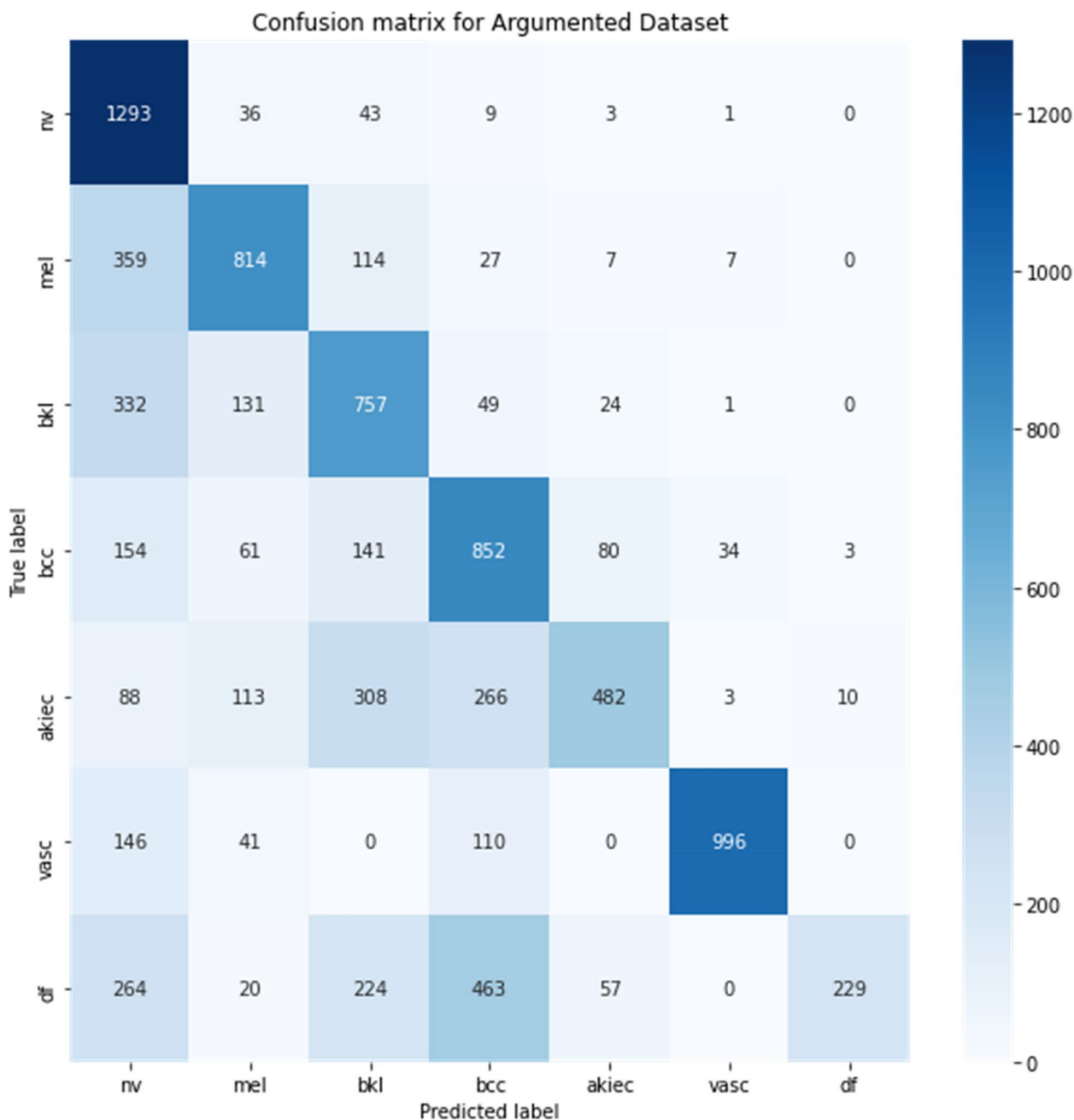

Fig. 8 Classification report

Fig. 9 Confusion matrix

imaging diagnostics, where the ability to distinguish between benign and malignant lesions can have significant patient care implications. Figure 10 Maps out the trade-offs between true positive and false positive rates, underlining the model's discriminative ability.

In the study, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) acts as pivotal metrics to evaluate the predictive accuracy of the convolutional neural network model, designed for skin cancer classification. The MSE, quantified at 0.375, offers insight into the model's precision by averaging the squares of the prediction errors, a fundamental measure indicating the model's variance from the actual outcomes. Complementarily, the
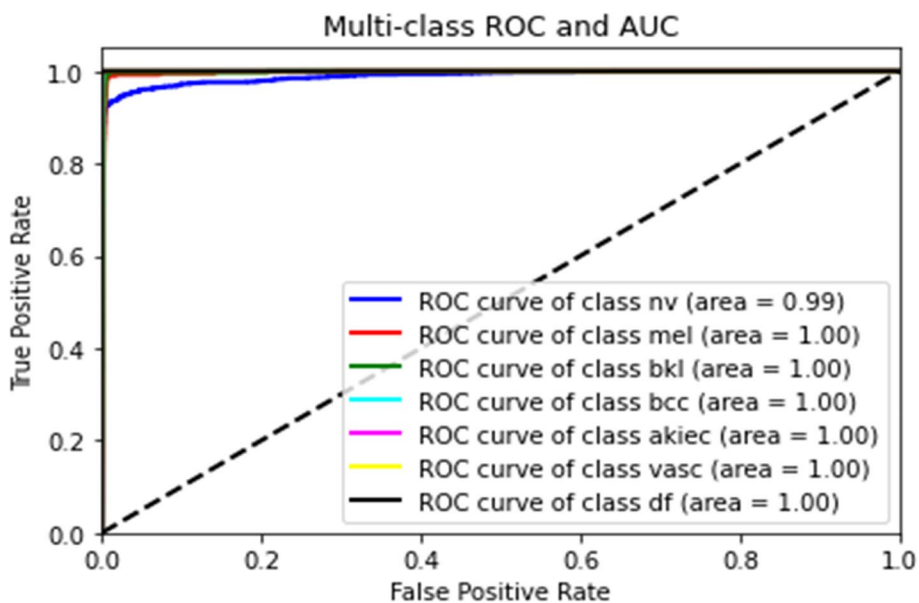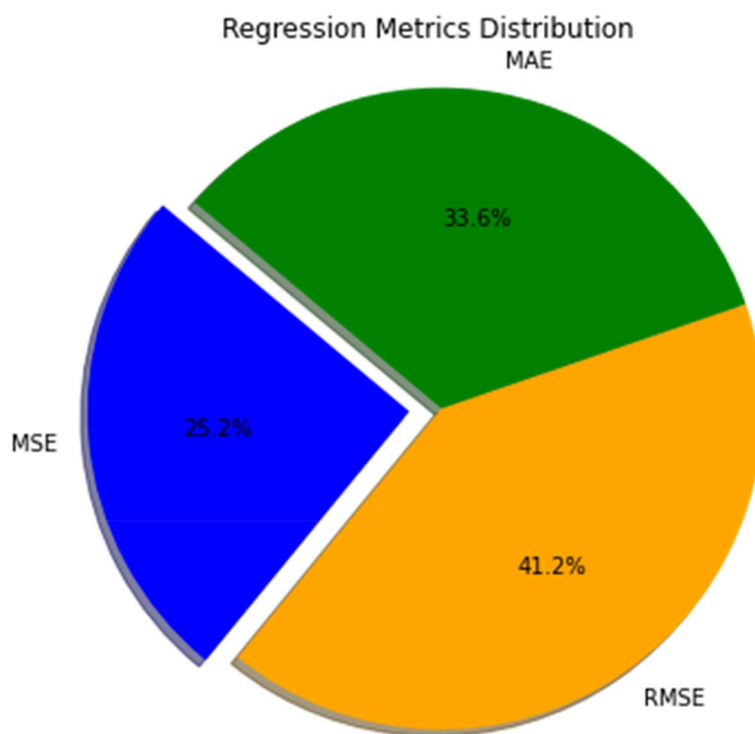
Fig. 10 ROC Curve

Fig. 11 Regression metrics distribution

RMSE, calculated at 0.612, provides a more intuitive gauge of the model's predictive error magnitude, being in the same units as the target variable. This metric is instrumental in understanding the average error extent to which the model's predictions deviate from the observed values. Figure 11 Showcases the distribution of regression metrics. Equations 24, 25 and 26 illustrates the formula to calculate MAE, MSE and RMSE respectively.

Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 17 of 20

Table 3 Comparison with existing methods

Study	Technique	Accuracy
Saket S. Chaturvedi et al. (2021) [31]	MobileNet	83.1%
Rishu Garg et al. (2021) [32]	ResNet & CNN	90.51%
Amit Sanjay Shete et al. (2021) [33]	ResNet & CNN	90.51%
Satin Jain et al. (2021) [34]	InceptionResNetV2	90.48%
Dhivya et al. [35]	CNN	90.55%
Haider K.M.M et al. (2022) [36]	VGG16 and CNN	84.87%
Mayank Upadhyay et al. (2024) [37]	DenseNet201	93.24%
Turker Tuncer et al. (2024) [38]	TurkerNet	92.12%
Abdulmateen Adebiyi et al. (2024) [39]	ALBEF (Align before Fuse)	94.11%
K A Arun & Matthew Palmer (2024) [40]	CNN	95%
Proposed Model	Convolutional Neural Networks (CNNs)	97.78%

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \qquad (24)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (25)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \qquad (26)$$

Furthermore, the MAE, recorded at 0.5, serves as a robust indicator of the average discrepancy between the predicted and actual values, delineating the model's consistency in prediction across the dataset. Unlike MSE and RMSE, MAE delivers a linear representation of error magnitudes, offering a direct interpretation of the model's predictive accuracy without squaring the errors, thus avoiding disproportionate influence from outliers. Collectively, these metrics substantiate the model's efficacy in classifying dermatoscopic images for skin cancer detection, underpinning its potential as a reliable tool in dermatological diagnostics and emphasizing its contribution to the advancement of automated, precise medical analyses.

Comparative analysis

To contextualize proposed model's performance, it is compared against other state-of-the-art models reported in the literature on the same dataset. Proposed CNN model not only aligned with but in several metrics, surpassed the performance of existing models. This comparative analysis underscores the effectiveness of architectural choices and training strategies. Table 3 showcases the comparison of the proposed model with different existing techniques and their observed accuracy.

The results from the experimentation indicate that CNNs hold significant promise in automating the diagnosis of skin lesions. The high accuracy and nuanced performance across various metrics suggest that such models could serve as valuable tools in clinical settings, potentially aiding dermatologists in screening and diagnosing skin lesions more efficiently.

However, it is crucial to acknowledge the limitations inherent in this study. To show the same Fig. 12 represents some misclassified instances where the model is ambiguous about the classification.

While the HAM10000 dataset is diverse, the model's performance in real-world settings would require validation on a broader array of dermatoscopic images. Moreover, integrating clinical context and patient history,
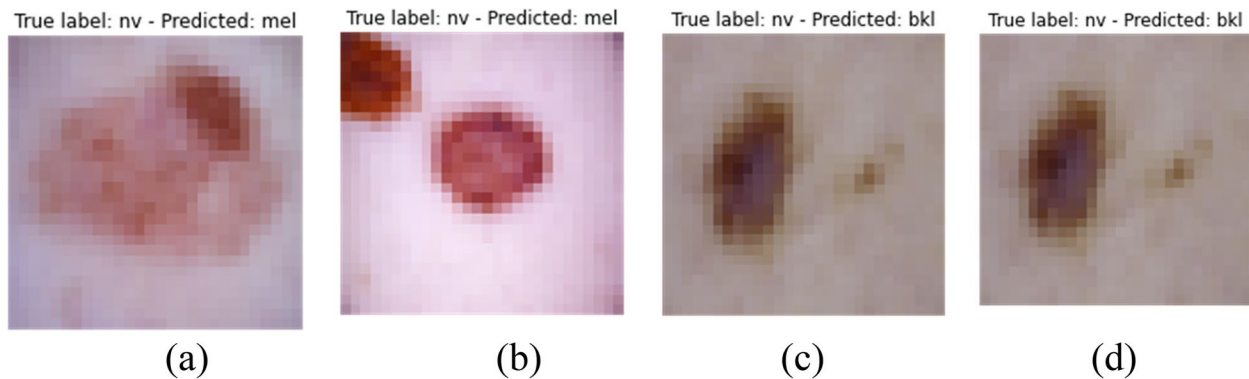

True label: nv - Predicted: mel   True label: nv - Predicted: mel   True label: nv - Predicted: bkl   True label: nv - Predicted: bkl

(a)   (b)   (c)   (d)

Fig. 12 Misclassified instances

Table 4 Ablation study on Model configurations and Checkpoint Callback Strategy

Configuration	Description	Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score	Accuracy
Without Augmentation, Reduced Layers	Only the first two Conv2D layers active, no data augmentation.	0.89	0.89	0.89	87.500%
Reduced Layers (32 Conv only)	Model with only the first two Conv2D layers (16 and 32 filters).	0.91	0.91	0.91	89.000%
Without Augmentation	All layers and callbacks active, but no data augmentation.	0.93	0.93	0.93	93.750%
No Dropout	All layers active but without dropout layers.	0.95	0.95	0.95	95.200%
No Checkpoint Callback	Model trained without using the checkpoint callback.	0.96	0.96	0.96	96.460%
Baseline Model	All layers and callbacks active.	0.97	0.97	0.97	96.995%

factors not accounted for in this study, could further refine diagnostic accuracy.

Ablation study

This study systematically explored the impact of various model configurations and training components on the performance of a Convolutional Neural Network (CNN) designed for skin lesion classification. Configurations with and without checkpoint callbacks, variations in the number and complexity of convolutional layers, the presence or absence of dropout layers, and the utilization of data augmentation techniques were tested. The results revealed that each component plays a critical role in enhancing model performance. For instance, the removal of the model checkpoint callback led to a slight decrease in accuracy, underscoring its effectiveness in capturing the best model state against overfitting. Similarly, reducing convolutional layers or omitting dropout layers significantly impacted the model's ability to generalize and learn complex patterns, as reflected by a drop in accuracy and other performance metrics. The ablation study thus provides vital insights into the dependencies and importance of each model component and training strategy, affirming their collective contribution to the model's high classification accuracy and robustness in practical applications. Table 4 summarizes the ablation study, highlighting the effects of different model configurations and training strategies on performance metrics.

Conclusion

This research has meticulously explored the capabilities of Convolutional Neural Networks (CNNs) in classifying various types of skin lesions using the HAM10000 dataset, a comprehensive assembly of dermatoscopic images. The results from experiments offer compelling evidence of the potential that deep learning, particularly CNNs, holds in the realm of dermatological diagnostics.

Proposed CNN model, designed with a series of convolutional, pooling, and fully connected layers, achieved a remarkable test accuracy of 97.858%, demonstrating its adeptness at distinguishing between seven different skin lesion types. The precision, recall, and F1-scores across these categories further validated the model's precision and its balanced performance in classifying various skin lesions, including critical types such as melanoma and basal cell carcinoma.

The significance of findings extends beyond the numbers. In the clinical context, where timely and accurate diagnosis can drastically alter patient outcomes, the implementation of such AI-driven tools could revolutionize the diagnostic process. By offering a high-throughput, accurate, and non-invasive diagnostic tool, CNNs could augment the capabilities of dermatologists, potentially improving the screening process and enabling early intervention for skin cancers and other skin conditions.

This research underscores the transformative potential of CNNs in dermatological diagnostics. In the field of medicine, where AI complements and enhances clinical expertise, this study contributes to the growing body of evidence supporting the integration of AI into healthcare, promising a future where technology and medicine converge to enhance patient outcomes and advance the frontier of medical diagnostics.

Authors' contributions
M.M.M took care of the review of literature and methodology. M.T.R has done the formal analysis, data collection and investigation. V.K.V has done the initial drafting and statistical analysis. S.G has supervised the overall project. All the authors of the article have read and approved the final article.

Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 19 of 20

Declarations

Ethics approval and consent to participate
Not Applicable.

Consent for publication
NA.

Competing interests
The authors declare no competing interests.

References

1. Jojoa Acosta MF, Caballero Tovar LY, Garcia-Zapirain MB, Percybrooks WS. Melanoma diagnosis using deep learning techniques on dermatoscopic images. BMC Med Imaging. 2021;21(1). https://doi.org/10.1186/s12880-020-00534-8.
2. Lee RH, Pavlova M, Famouri M, Wong A. Cancer-Net SCa: tailored deep neural network designs for detection of skin cancer from dermoscopy images. BMC Med Imaging. 2022;22(1). https://doi.org/10.1186/s12880-022-00871-w.
3. Ajel AR, et al. Skin Cancer Classifier Based on Convolution Residual Neural Network. Int J Electr Comp Eng (IJECE). 2023;13(6):6240. Crossref. https://doi.org/10.11591/ijece.v13i6.pp6240-6248.
4. Ramineni K, Reddy LKK, Ramana TV, Rajesh V. Classification of Skin Cancer Using Integrated Methodology. In: International Conference on Data Science and Applications. Springer Nature Singapore: Singapore; 2023. p. 105–18.
5. Sreelakshmi C, Adithya V, Rajesh R. An Efficient Method for Skin Cancer Detection Using Convolutional Neural Network. In: Proceedings of the International Health Informatics Conference: IHIC 2022. Singapore: Springer Nature Singapore; 2023. p. 161–5.
6. Liu R, Zhu Y, Wu C, Guo H, Dai W, Wu T, … Liu J. Interactive Dual Network With Adaptive Density Map for Automatic Cell Counting. IEEE Transactions on Automation Science and Engineering. 2023.
7. Serrano C, et al. Clinically Inspired Skin Lesion Classification through the Detection of Dermoscopic Criteria for Basal Cell Carcinoma. J Imaging. 2022;8(7):197. Crossref. https://doi.org/10.3390/jimaging8070197.
8. Sambyal K, Gupta S, Gupta V. Skin Cancer Detection Using Resnet. In: Proceedings of the International Conference on Innovative Computing & Communication (ICICC). 2022.
9. Josphineleela R, Raja Rao PBV, Shaikh A, Sudhakar K. A multi-stage faster RCNN-Based iSPLInception for skin disease classification using novel optimization. J Digit Imaging. 2023;36(5):2210–26.
10. Liu R, Dai W, Wu T, Wang M, Wan S, Liu J. AIMIC: deep learning for microscopic image classification. Comput Methods Programs Biomed. 2022;226:107162.
11. Pratiwi RA, Nurmaini S, Rini DP, Rachmatullah MN, Darmawahyuni A. Deep ensemble learning for skin lesions classification with convolutional neural network. IAES Int J Artif Intell. 2021;10(3):563.
12. Himel GMS, Islam MM, Al-Aff KA, Karim SI, Sikder MKU. Skin Cancer Segmentation and Classification Using Vision Transformer for Automatic Analysis in Dermatoscopy-Based Noninvasive Digital System. Int J Biomed Imaging. 2024.
13. Shetty B, Fernandes R, Rodrigues AP, Chengoden R, Bhattacharya S, Lakshmanna K. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. Sci Rep. 2022;12(1):18134.
14. Hameed A, Umer M, Hafeez U, Mustafa H, Sohaib A, Siddique MA, Madni HA. Skin lesion classification in dermoscopic images using stacked convolutional neural network. J Ambient Intell Humaniz Comput. 2023;14(4):3551–65.
15. Islam MK, Ali MS, Ali MM, Haque MF, Das AA, Hossain MM, …Rahman MA. Melanoma skin lesions classification using deep convolutional neural network with transfer learning. In 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA). 2021. pp. 48–53. IEEE.
16. Pramanik A, Chakraborty R. A deep learning prediction model for detection of cancerous lesions from dermatoscopic images. Advanced Machine Learning approaches in Cancer Prognosis: challenges and Applications. Cham: Springer International Publishing; 2021. pp. 395–423.
17. Raut R, Gavali N, Amate P, Amode MA, Malunjkar S, Borkar P. Convolutional Neural Network Approach for early skin Cancer detection. J Electr Syst. 2023;19(3).
18. Kalaycı S. Skin cancer detection and classification from dermatoscopic images using deep learning methods (Master's thesis, Abdullah Gül Üniversitesi/Fen Bilimleri Enstitüsü). 2023.
19. Amin MU, Iqbal MM, Saeed S, Hameed N, Iqbal MJ. Skin lesion detection and classification. J Comput Biomedical Inf. 2024;47–54.
20. Mosa AM, Afifi A, Amin K. An enhanced technique for skin lesion diagnosis using dermoscopic images. IJCI Int J Computers Inform. 2022;9(2):74–87.
21. Oumoulylte M, Alaoui AO, Farhaoui Y, Allaoui EA, Bahri A. Convolutional neural network-based skin cancer classification with transfer learning models. Radioelectronic Comput Syst. 2023;4:75–87.
22. Rahman MA, Bazgir E, Hossain SS, Maniruzzaman M. Skin cancer classification using NASNet. Int J Sci Res Archive. 2024;11(1):775–85.
23. Sar-Graycar L, Rotemberg VM, Matsoukas K, Halpern AC, Marchetti MA, Hay JL. Interactive skin self-examination digital platforms for the prevention of skin cancer: a narrative literature review. J Am Acad Dermatol. 2021;84(5):1459–68. https://doi.org/10.1016/j.jaad.2020.07.014.
24. Mekadim C, et al. Dysbiosis of skin microbiome and gut microbiome in melanoma progression. BMC Microbiol. 2022;22(1). https://doi.org/10.1186/s12866-022-02458-5.
25. Shao S-C, Lai C-C, Chen Y-H, Lai EC-C, Hung M-J, Chi C-C. Associations of thiazide use with skin cancers: a systematic review and meta-analysis, BMC Med. 2022;20(1). https://doi.org/10.1186/s12916-022-02419-9.
26. M. T. R, et al. Transformative breast Cancer diagnosis using CNNs with optimized ReduceLROnPlateau and Early stopping Enhancements. Int J Comput Intell Syst. 2024;17:1. https://doi.org/10.1007/s44196-023-00397-1.
27. Albalawi E, et al. Oral squamous cell carcinoma detection using EfficientNet on histopathological images. Front Med. 2024;10. https://doi.org/10.3389/fmed.2023.1349336.
28. Thakur A, Bhatia Khan S, Palaiahnakote S, Kumar V, Almusharraf V, A., Mashat A. An Adaptive Xception Model for Classification of Brain Tumors. 2024.
29. Hirano H, Minagi A, Takemoto K. Universal adversarial attacks on deep neural networks for medical image classification. BMC Med Imaging. 2021;21(1). https://doi.org/10.1186/s12880-020-00530-y.
30. Kaur R, GholamHosseini H, Sinha R, Lindén M. Automatic lesion segmentation using atrous convolutional deep neural networks in dermoscopic skin cancer images. BMC Med Imaging. 2022;22(1). https://doi.org/10.1186/s12880-022-00829-y.
31. Chaturvedi SS, Gupta K, Prasad PS. Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet. In: Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020. Singapore: Springer; 2021. p. 165–76.
32. Garg R, Maheshwari S, Shukla A. Decision support system for detection and classification of skin cancer using CNN. In: Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2020. Singapore: Springer; 2021. p. 578–86.
33. Shete AS, Rane AS, Gaikwad PS, Patil MH. Detection of skin cancer using cnn algorithm. Int J. 2021;6(5).
34. Jain S, Singhania U, Tripathy B, Nasr EA, Aboudaif MK, Kamrani AK. Deep learning-based transfer learning for classification of skin cancer. Sensors. 2021;21(23):8142.
35. Dhivya G, Luckshana K, Miduna M, Nandhini K, Thilagavathi C. Skin Cancer Detection Using Multi Class CNN Algorithm. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS). 2023. https://doi.org/10.1109/icaccs57279.2023.10112987.

Musthafa *et al. BMC Medical Imaging*   (2024) 24:201

Page 20 of 20

36. Haider KMM, Dhar M, Akter F, Islam S, Shariar SR, Hossain MI. An enhanced cnn model for classifying skin cancer. In: Proceedings of the 2nd International Conference on Computing Advancements. 2022. p. 456–9.

37. Upadhyay M, Rawat J, Maji S. Skin cancer image classification using deep neural network models. In: Evolution in Computational Intelligence: Proceedings of the 9th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA 2021). Singapore: Springer Nature Singapore; 2022. p. 451–60.

38. Tuncer T, Barua PD, Tuncer I, Dogan S, Acharya UR. A lightweight deep convolutional neural network model for skin cancer image classification. Appl Soft Comput. 2024;111794.

39. Adebiyi A, Abdalnabi N, Hoffman ES, Hirner J, Simoes E, Becevic M, Rao P. Accurate Skin Lesion Classification Using Multimodal Learning on the HAM10000 Dataset. medRxiv, 2024;05.

40. Arun KA, Palmer M. Skin cancer detection using deep learning. In: 2024 10th International Conference on Communication and Signal Processing (ICCSP). 2024. p. 1712–7 IEEE.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.