



PP-CNN: probabilistic pooling CNN for enhanced image classification

Narendra Kumar Mishra¹ · Pushpendra Singh² · Anubha Gupta³ · Shiv Dutt Joshi¹

Received: 17 June 2024 / Accepted: 29 November 2024 / Published online: 20 December 2024
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract

This study introduces a novel probabilistic pooling convolutional neural network (PP-CNN) classifier designed to enhance image classification. The PP-CNN integrates probabilistic outputs from three distinct architectures utilizing MaxPooling, MinPooling, and MaxMinPooling layers. By averaging these probabilities, the model achieves improved accuracy compared to individual models. Notably, the proposed CNN model employs Positive ReLU activation with MaxPooling, Negative ReLU activation with MinPooling, and both Positive and Negative ReLU activations with MaxMinPooling. This strategy ensures the retention of both positive and negative relevant features, enhancing the classification performance by capturing a broader range of critical information. The proposed model has been comprehensively evaluated for its generalizability on four diverse datasets: CIFAR-10, CIFAR-100, CT scan, and X-ray images. Experimental results demonstrate a consistent improvement in classification accuracy across all datasets, highlighting the versatility and effectiveness of the proposed model. The proposed model applies to various image classification tasks, specifically illustrating its utility by detecting COVID-19 from medical images. This work presents the design, implementation, and performance evaluation of the proposed model, underscoring its potential to significantly improve image classification and diagnostic accuracy in medical imaging applications.

Keywords Probabilistic Pooling Convolutional Neural Network (PP-CNN) · MaxPooling · MinPooling · MaxMinPooling · COVID-19 detection

1 Introduction

The convolutional neural networks (CNNs) represent a significant innovation in deep learning, designed to address the complexities of image classification [1, 2]. Inspired by the structure of the human visual cortex, CNNs mimic the hierarchical processing of visual information, allowing

them to manage spatial hierarchies in data effectively. This architecture is particularly suited for analyzing and interpreting visual imagery across various applications, making CNNs exceptionally effective for these tasks [3].

CNNs have become foundational in computer vision because they can automate feature extraction. Unlike traditional image processing techniques that require manual feature selection and careful engineering, CNNs learn to identify relevant features directly from the data. This capability has revolutionized image classification, enabling substantial advancements in object recognition, video, and medical image analysis. By learning to detect features from training data, CNNs have significantly improved the accuracy and efficiency of image classification tasks [4].

Recent advancements in medical image classification have demonstrated the effectiveness of CNN-based architectures for diagnosing various diseases. Akyol [5] introduced a two-stage voting framework for COVID-19 detection, while Kurt et al. [6] applied EfficientNet models for analyzing lung parenchyma to detect COVID-19, achieving promising results. Kibriya and Amin [7] further

✉ Pushpendra Singh
spushp@jnu.ac.in
Narendra Kumar Mishra
eez188568@ee.iitd.ac.in
Anubha Gupta
anubha@iitd.ac.in
Shiv Dutt Joshi
sdjoshi@iitd.ac.in

¹ Department of Electrical Engineering, IIT Delhi, Delhi, India

² School of Engineering, Jawaharlal Nehru University, Delhi, India

³ SBILab, Department of ECE, IIT-Delhi, Delhi, India

enhanced COVID-19 detection with a residual network-based framework using chest X-ray images. Additionally, CNN models have also been widely applied for cancer detection. Sahu et al. [8] proposed a hybrid CNN classifier for breast cancer detection using mammogram and ultrasound datasets, achieving high classification accuracy. Mridha et al. [9] tackled skin cancer classification with explainable AI models, employing optimized CNN architectures for interpretability. Diabetic retinopathy detection using deep learning ensemble approaches has also shown significant promise, as demonstrated by Qummar et al. [10]. Furthermore, Oguz et al. [11] developed a CNN-based hybrid model to detect glaucoma disease from medical images effectively, showcasing the broad applicability of CNNs across various medical image classification tasks.

CNNs are distinguished by their deep architecture, which consists of several layers of fundamental building blocks arranged in various patterns. The core components of a CNN, illustrated in Fig. 1, are each designed to process different aspects of input data [12]. These layers work in tandem to extract and refine features, enabling the network to perform complex image analysis tasks effectively. The convolutional layers apply filters to the input, creating feature maps highlighting key patterns. The activation layers like ReLU introduce nonlinearity, enabling the model to learn complex data relationships. The pooling layers reduce spatial dimensions, improving efficiency while retaining important features. In the final stages, the dense layers interpret high-level features and produce predictions, with the softmax layer converting outputs into class probabilities. The output layer delivers the final classification, such as distinguishing COVID-19, normal, and pneumonia.

In traditional CNNs, MaxPooling is a widely utilized technique that processes the input by extracting the maximum value from each segment of an image covered by the kernel [13]. While MaxPooling effectively highlights the most prominent features, it can result in the loss of potentially important details [14]. This loss is particularly significant in fields like medical imaging, where every pixel may carry crucial diagnostic information [15].

Recognizing the limitations of MaxPooling, especially its tendency to overlook less dominant features, there is a strong impetus to explore novel pooling strategies that preserve a more comprehensive set of information, thus potentially enhancing the overall performance of the network. Recent research underscores the potential benefits of utilizing diverse pooling strategies, including better generalization capabilities of networks from training datasets to unseen data [16–18], eventually enhancing their performance and applicability in real-world applications. For example, MinPooling captures the minimum values within the receptive field, which can be critical in scenarios where these lesser values contain important contrasts or details, such as images with subtle variations in shading or lighting. An innovative solution to leverage the strengths of both MaxPooling and MinPooling is the development of MaxMinPooling. This approach integrates outputs from both pooling strategies, preserving the highest and lowest values in the feature map. Such integration ensures that no critical details are overlooked, providing a balanced and detailed representation of the image data.

Jun Park et al. [19] introduced conditional MinPooling (CMP) to enhance CNN feature representation by preserving critical features through a tolerance mechanism. Their approach restructures CNNs by combining CMP with MaxPooling and applying a 1×1 convolution to reduce channel dimensions before pooling. This technique captures diverse features and improves robustness. Evaluated on Caltech 101 and custom datasets, CMP increased accuracy by 0.16–0.52% and decreased loss by 19.98–28.71% compared to traditional methods. Vienken, in his master thesis [20], proposed to enhance the efficiency of CNNs through multi-scale convolution filters and MinMaxPooling. This approach captures key features while being noise-resistant, addressing the computational complexity of deeper CNNs. Experiments showed that the MinMax method improved accuracy by 0.71% on CIFAR-10 and 4.9% on the Places dataset compared to the traditional methods.

One innovative approach, presented by Ozdemir, is the Avg-TopK pooling method that computes the weighted average of the most significant features by focusing on the

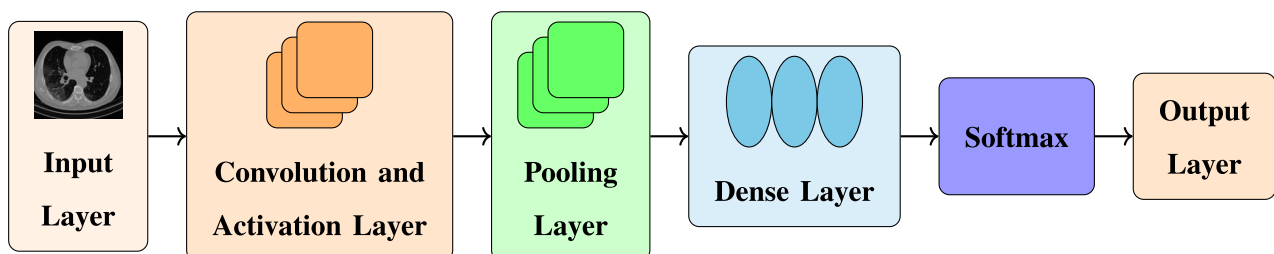


Fig. 1 Architecture of a typical CNN model

top K amplitude pixels [21]. Zhao et al. further modified this pooling strategy and introduced T-Max-Avg pool layer, which allows flexibility in feature extraction by using a threshold parameter to select the most representational/interactive pixels, enabling the maximum value or the weighted average of the top pixels to be pooled via threshold T based on data characteristics [18].

Further studies include Yu et al.'s mixed pooling method, which introduces variability by alternating between MaxPooling and average pooling in a stochastic manner, thereby enhancing regularization [22]. Er et al. have developed an attention pooling scheme that uses outputs from a bidirectional LSTM as a reference to weigh features extracted by convolutional layers, ensuring that critical information is preserved during the pooling process [23]. Cui et al. have proposed a general kernel pooling framework that leverages higher-order interactions of features through kernel approximations like the Gaussian RBF, thus capturing more complex patterns without extensive parameterization [24]. Williams et al. introduced wavelet pooling, which employs a multi-level decomposition strategy to refine feature processing by discarding less critical subbands, thus focusing on more relevant data [25].

Jie et al. explored dynamic pooling with their RunPool layer, which adapts the pooling operation dynamically to suit the training data's demands better [26]. Lastly, Sharma et al. introduced a novel method using fuzzy logic to perform dimension reduction, effectively shrinking the spatial size of convolved features while retaining essential information [27]. In a comprehensive review, Nirthika et al. discuss various pooling techniques across computer vision and medical image analysis, highlighting the breadth of research and the potential for these methods to improve CNN performance [28].

Piyush Satti et al. [29] designed a MinMax average pooling-based filter for effectively removing salt and pepper noise from images. This method combines Max- and MinPooling to better preserve edges and enhance the peak signal-to-noise ratio (PSNR), particularly in medical images corrupted by medium to high noise densities. Another study by Liyanage et al. [30] focused on hyperspectral image band selection and utilized various pooling methods, including Max- and average pooling, to reduce spectral dimensionality while retaining crucial spectral information. This method enhances classification accuracy and efficiency by addressing the limitations of traditional MaxPooling, ensuring the preservation of essential spectral details. Prior literature work is summarized in Table 1.

Motivated by the above discussion, this study aims to significantly enhance the performance of CNN-based models for image classification through innovative architectural enhancements. We aim to address the limitations of traditional CNN models by introducing refined pooling

strategies that improve the accuracy and robustness of classification in various imaging contexts. The significant contributions of the study are as follows:

- 1) **Novel MaxMinPooling Architecture with Dual-Activation Function Approach:** We have developed an innovative pooling architecture within the traditional CNN framework, termed MaxMinPooling. This architecture enhances feature extraction by uniquely combining the outputs from MaxPooling and MinPooling layers through interleaving, effectively merging the most and least activated pixels from the feature map. This method enriches the input for subsequent layers, improving the model's ability to capture detailed and nuanced features essential for accurate classification.
- 2) **To enhance the effectiveness of the MaxMinPooling layer,** we incorporate both positive ReLU $\max(0, x)$ and Negative ReLU $\min(0, x)$ activation functions and concatenate their outputs within the CNN architecture. This inclusion allows the layer to retain significant positive and negative features, ensuring a more comprehensive and nuanced feature representation. The dual-activation approach captures a broader range of data characteristics, which is essential for high-fidelity applications in image analysis, such as precise medical diagnostics and other complex imaging tasks.
- 3) **Fusion of Probabilistic Outputs:** We have integrated the probabilistic outputs from three specialized CNN architectures—each characterized by a distinct pooling layer: MaxPooling, MinPooling, and MaxMinPooling. By averaging these probabilities, we harness the diverse strengths of each pooling method, which enhances the overall stability and accuracy of the model. This ensembling effectively mitigates the weaknesses of individual architectures, leading to superior performance and better generalizability of the architecture across various datasets.
- 4) **Comprehensive Performance Evaluation:** The efficacy of the proposed model is demonstrated through a comprehensive evaluation across on four diverse datasets. We assess the model's performance on standard benchmark datasets like CIFAR-10 and CIFAR-100, as well as on specialized medical imaging datasets, including X-rays and CT scans for the detection of COVID-19 disease [31]. The CIFAR datasets were selected due to their widespread use in benchmarking image classification models, providing a reliable measure of the performance of the model against established standards. The inclusion of CT scan and X-ray images addresses the urgent need for automated tools in managing the

Table 1 Summary of recent pooling techniques used in CNNs. Each method is evaluated on different datasets, highlighting the key features of the pooling strategy and corresponding specific benefits. The symbols \uparrow and \downarrow represent an increase and a decrease, respectively

Study	Pooling method	Key features	Dataset(s)	Remarks
Jun Park et al. [19]	Conditional MinPooling (CMP)	Combines CMP and MaxPooling with 1×1 convolution for dimensionality reduction	Caltech 101, Custom datasets	Accuracy \uparrow 0.16–0.52%, Loss \downarrow 19.98–28.71%
Vienken [20]	MinmaxPooling	Multi-scale convolution with MinMaxPooling	CIFAR-10, Places dataset	Accuracy \uparrow 0.71% (CIFAR-10), \uparrow 4.9% (Places)
Ozdemir [21]	Avg-topk pooling	Weighted average of significant features	Image datasets	Accuracy \uparrow 16.62% (CIFAR-10), \uparrow 25% (CIFAR-100)
Zhao et al. [18]	T-max-avg pooling	Threshold-based selective feature extraction	ImageNet, Custom datasets	Adaptive pooling flexibility
Yu et al. [22]	Mixed pooling	Alternates between Max- and average pooling	CIFAR-100, ImageNet	Enhanced regularization
Er et al. [23]	Attention pooling	LSTM-weighted feature extraction	Medical image datasets	Preserves critical information
Cui et al. [24]	Kernel pooling	Higher-order interactions using kernel approximations	CIFAR-10, ImageNet	Captures complex patterns
Williams et al. [25]	Wavelet pooling	Multi-level decomposition, discards less critical subbands	ImageNet, Medical images	Refines feature selection
Jie et al. [26]	RunPool	Dynamic pooling adapts to data demands	CIFAR-10, MNIST	Data-adaptive pooling
Sharma et al. [27]	Fuzzy logic pooling	Dimension reduction using fuzzy logic	Medical image datasets	Retains essential spatial information
Piyush Satti et al. [29]	MinMax average pooling	Removes salt and pepper noise	Medical images	PSNR enhancement, noise reduction
Liyanage et al. [30]	Max- and average pooling	Reduces spectral dimensionality, preserves spectral details	Hyperspectral images	Classification accuracy \uparrow by 2.53%

COVID-19 pandemic, demonstrating the applicability of the model in critical real-world scenarios. This diverse selection of datasets effectively showcases the generalizability and robustness of the proposed method, proving its efficacy across both standard and specialized image classification tasks.

Next, we provide a comprehensive overview of the research objectives addressed in this study. Section II delves into the detailed methodology and theoretical concepts that foster the ideas explored in this work, setting the stage for experimental design, data selection, and architectural development. The implementation of the proposed architectures is discussed in Section III. Section IV presents the findings of our experiments along with an in-depth analysis. Finally, Section V concludes the study by summarizing the main results and discussing the limitations, and offers insights into potential future directions for research.

2 Methodology

The decision to incorporate MaxPooling, MinPooling, and MaxMinPooling layers within a single CNN architecture stems from the goal of harnessing their complementary advantages. While MaxPooling focuses on the most salient features, potentially missing finer details, MinPooling ensures these subtleties are preserved. MaxMinPooling further enriches the feature representation of the model by preserving a broad spectrum of feature activations. The proposed approach aims to balance the focus on dominant and subtle features by averaging the probabilities derived from models based on each pooling strategy, thereby achieving superior classification performance across a diverse set of image datasets. This integrated pooling strategy highlights the novelty of the model and its potential to set a benchmark in image classification tasks, particularly those requiring nuanced interpretation of visual data, such as in the detection of COVID-19 from medical imaging [32–37].

2.1 Theoretical framework

The CNN architecture has been pivotal in advancing the field of image recognition due to its ability to learn hierarchical feature representations. At the heart of these advancements are pooling layers, which reduce the spatial dimensions of the feature maps, thus decreasing computational complexity and enhancing the network's ability to capture dominant features. In this study, we explore the integration of three distinct pooling strategies, namely MaxPooling, MinPooling, and MaxMinPooling, each contributing uniquely to the robustness and accuracy of image classification tasks. This study elucidates the rationale behind the choice of these specific pooling layers and their collective impact on the proposed CNN model's performance.

- **MaxPooling:** MaxPooling is a standard pooling operation used in CNNs to down-sample the input's spatial dimensions by taking the maximum value over a specified window for each channel independently. This operation is instrumental in achieving translation invariance, reducing the sensitivity of the output to minor changes and shifts in the input image. By focusing on the most prominent features within each window, MaxPooling, along with positive ReLU, helps in accentuating features that are crucial for discrimination, thus improving the model's generalization capability [38, 39]. However, while effective in highlighting the strongest features, MaxPooling may overlook subtler, yet equally important, features present in the input.
- **MinPooling:** In contrast to MaxPooling, MinPooling operates by selecting the minimum value within each pooling window. This approach is less common in traditional CNN architectures but offers unique advantages, particularly in emphasizing low-intensity features that may be overshadowed by more dominant features in the MaxPooling process. By capturing these minimal values, MinPooling along with negative ReLU function can enhance the model's sensitivity to finer and more subtle patterns in the input data, which are often crucial in complex classification tasks, such as distinguishing between similar classes or identifying pathologies in medical images.
- **MaxMinPooling:** MaxMinPooling is a novel pooling strategy that aims to combine the strengths of MaxPooling and MinPooling by concurrently capturing both the maximum and minimum values within each pooling window. This dual approach allows the model to retain both the most and least activated features within the input feature map, offering a more comprehensive representation of the input's characteristics. The

integration of MaxMinPooling is motivated by the hypothesis that a richer feature set, encompassing both extremes of the feature activation spectrum, can significantly enhance the model's discrimination power, especially in tasks requiring fine-grained differentiation between classes.

The interleaved MaxPooling and MinPooling operations are structured such that each MaxPooling output is immediately followed by its corresponding MinPooling output. This configuration ensures an intricate balance and detailed preservation of feature information, allowing the network to capture both high- and low-intensity features of the input. Theoretical analysis suggests that this interleaving can improve gradient flow in deep networks by providing additional gradients during backpropagation [40–42]. We empirically validate this hypothesis through a series of experiments on benchmark datasets, demonstrating improved classification accuracy, particularly in scenarios where feature contrasts are critical to performance.

2.2 Activation function strategy

In our novel approach to activation functions, we propose a dual-path strategy for managing activations within neural networks, specifically designed to preserve and independently process both positive and negative information from feature maps. This approach employs a parallel combination of positive and negative ReLU (Rectified Linear Unit) functions, coupled with MaxMinPooling layers, to enhance the network's capacity for comprehensive feature interpretation. Additionally, block-wise concatenation of positive and negative ReLU activations is employed to effectively merge different feature representations, which is crucial for nuanced learning and improving the overall expressiveness of the model.

Moreover, integrating these concatenated activation functions with MaxMinPooling layers facilitates a deeper understanding and more robust feature extraction by maintaining critical spatial hierarchies. This approach not only amplifies the network's sensitivity to diverse input features, but also stabilizes the learning process by balancing the positive and negative influences within the learned representations. The combination has shown promising results in preliminary tests, particularly in complex recognition tasks where traditional models fail to capture subtle but crucial anomalies. The empirical validation of this architecture reveals significant improvements in classification accuracy, confirming its efficacy in practical applications.

2.3 Probabilistic output fusion

In this research, the integration of the probabilistic outputs of three distinct CNN models—each employing MaxPooling, MinPooling, and MaxMinPooling, respectively—forms the core of our strategy to improve image classification accuracy. By averaging these outputs before applying a softmax function to derive final scores, this fusion technique effectively captures a comprehensive spectrum of features from the images, thereby enhancing classification accuracy. This ensemble method not only mitigates individual model biases, but also leverages the diverse strengths of each model, leading to improved generalization and robustness across varied datasets.

2.4 Dataset selection rationale

Toward a comprehensive assessment of CNN models in both basic and complex image recognition tasks, we selected CIFAR-10, CIFAR-100, X-ray and CT scan datasets to evaluate the performance of our proposed CNN models, each chosen for their unique attributes and relevance to the model's application. The CIFAR-10 and CIFAR-100 datasets, known for their color images of everyday objects divided into 10 and 100 classes, respectively, are widely utilized in the machine learning community to benchmark image recognition algorithms [43]. These datasets allow us to test the CNN's ability to learn and categorize varied visual features, providing a baseline measure of performance. On the other hand, the X-ray and CT scan datasets are critical in the medical field, especially relevant under current global health conditions, such as the COVID-19 pandemic. These medical datasets contain grayscale images that present complex patterns crucial for disease diagnosis, challenging our models to capture subtle nuances that are vital for accurate medical analysis. Sample images from each dataset are shown in Fig. 2. For performance evaluation of the proposed method on various datasets, the model used an 80:20 ratio for training and testing of the model performance. The details of X-ray [44–48] and CT scan [49] dataset are tabulated in Table 2.

This broad evaluation of diverse datasets helps demonstrate the generalizability and efficacy of the models across different domains and conditions, confirming their potential for real-world applications, particularly in enhancing automated diagnostic processes. This approach underscores the importance of using diverse datasets in developing models that are robust, versatile, and capable of functioning effectively in various real-world scenarios.

2.5 Architectural design

To effectively evaluate the performance of our proposed convolutional neural network (CNN) models, we designed four custom CNN architectures tailored to each of the four distinct datasets: CIFAR-10, CIFAR-100, X-ray, and CT scans. For each dataset, the input is passed through three parallel channels of CNN models, each incorporating a different pooling strategy, MaxPooling, MinPooling, and MaxMinPooling, as illustrated in Fig. 3. This design allows us to leverage the unique advantages of each pooling method. An important aspect of our approach is the use of positive ReLU activation for MaxPooling, negative ReLU activation for MinPooling, and both positive and negative ReLU activations for MaxMinPooling. This configuration helps preserve both positive and negative relevant features, enhancing ability of the model to capture a broad spectrum of data characteristics. The outputs from these three channels are then averaged and processed through a softmax function to calculate the final classification scores. These scores are compared with the true labels to evaluate the overall performance of the model.

2.6 Structural block diagram

To clarify the proposed architecture, we have included a structural block diagram visualizing the sequence of operations, focusing on the activation and pooling layers. The diagram (refer Fig. 4) shows the flow of data starting with convolution, which extracts features from the input. These features are processed through positive ReLU (PosReLU) and negative ReLU (NegReLU) functions, resulting in distinct feature maps for each type of activation. The feature maps are then concatenated block-wise, preserving both positive and negative activations for further processing, thus providing a richer feature representation. Subsequently, MaxPooling and MinPooling layers are applied separately to emphasize high- and low-intensity features, respectively. The outputs are concatenated in an interleaved manner, ensuring balanced feature preservation and enabling the model to capture a comprehensive range of patterns, thereby improving nuanced learning. The MaxMinPooling output is then further passed through the subsequent blocks of the model.

2.7 Evaluation metrics

In our study, we chose accuracy, precision, recall, F1 score, and the confusion matrix as the primary metrics for assessing our CNN models due to their crucial role in evaluating classification performance [14, 50]. These metrics provide a comprehensive understanding of model



Fig. 2 Sample images from CIFAR-10, CIFAR-100, X-ray, and CT scan datasets

Table 2 Details of medical images dataset used for the development of COVID-19 detection model using the proposed MaxMinPooling method

Category	X-ray dataset			CT scan dataset		
	Train	Test	Validation	Train	Test	Validation
Normal	6,849	2,140	1,712	4,412	1,378	1,103
COVID-19	7,649	2,395	1,912	4,860	1,518	1,215
Pneumonia	7,208	2,253	1,802	1,676	523	419
Total	21,706	67,88	5,426	10,948	3,419	2,737
Grand Total	33,920			17,104		

accuracy and effectiveness, particularly in handling false positives and negatives-vital for applications in medical diagnostics. Additionally, we employed performance learning curves to visually track the models' progress over training epochs, aiding in the identification of potential overfitting or underfitting issues. This approach allows us to fine-tune our models effectively, ensuring robust

validation and optimization of the architectures for reliable real-world applications. This methodical evaluation not only demonstrates the models' capabilities, but also enhances their applicability across diverse scenarios, particularly in critical tasks such as disease detection. The computation of performance measures adhered to the following mathematical expressions:

Fig. 3 Block diagram of the proposed method incorporating different pooling strategies, activation functions and probabilistic averaging for enhanced image classification

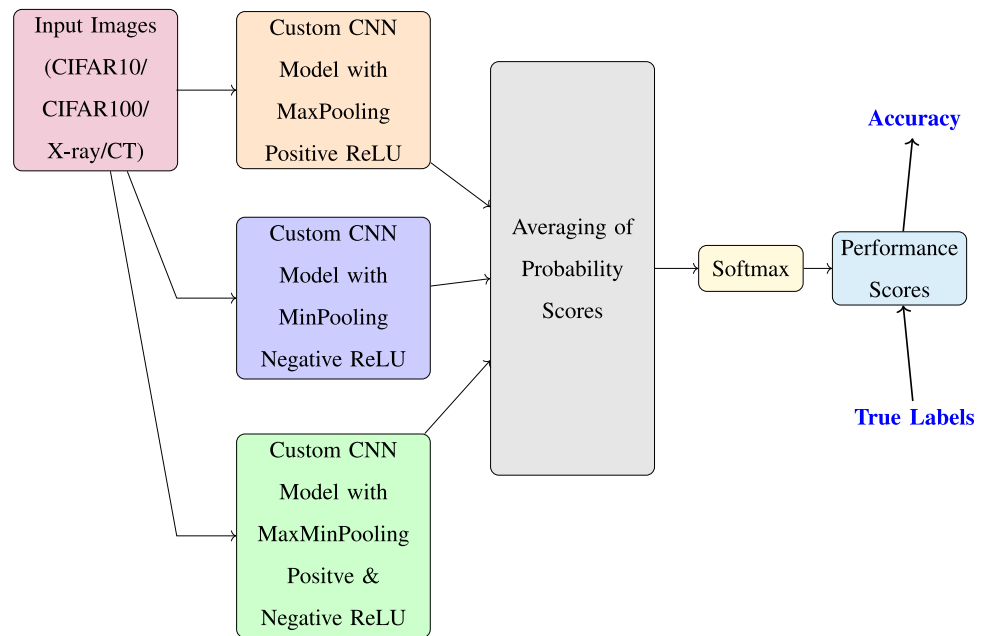
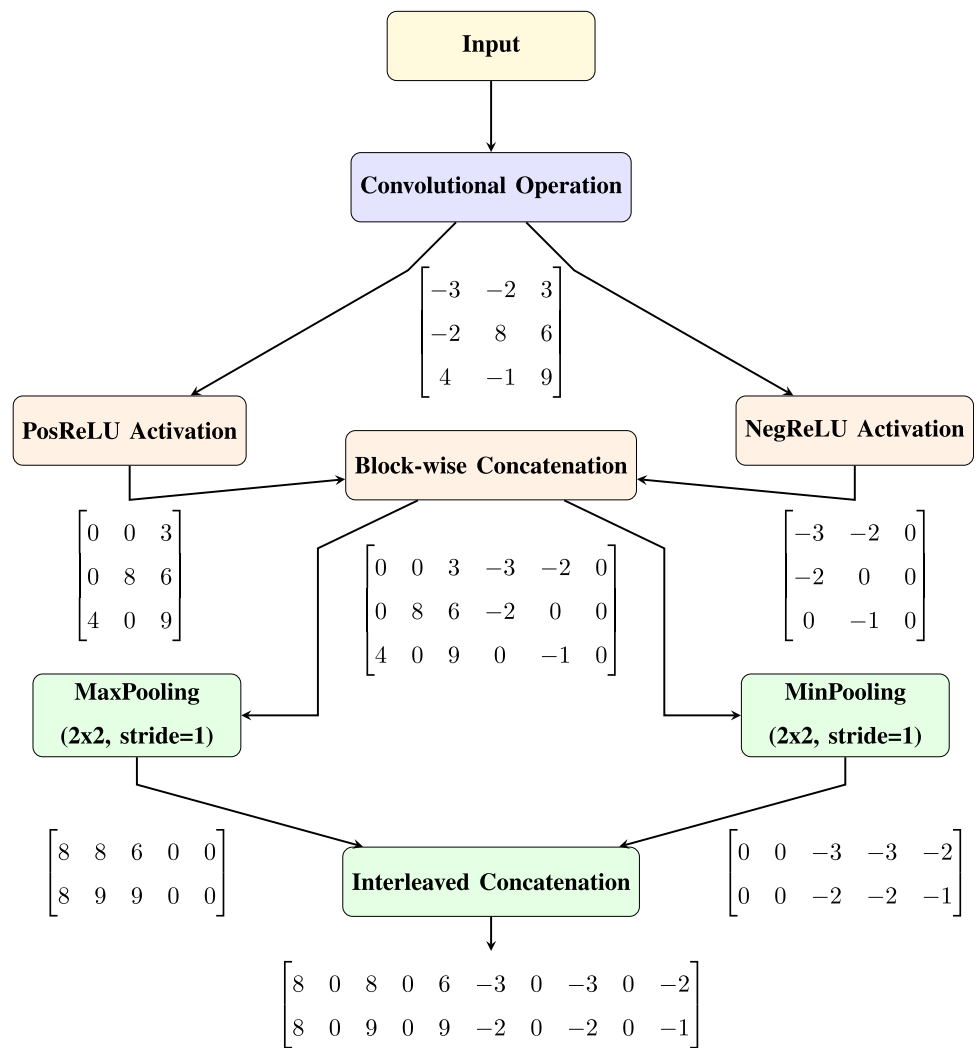


Fig. 4 Structural block diagram of the proposed architecture, detailing the sequence of operations. The diagram shows the flow from the convolutional operation, through the positive and negative ReLU activation functions, followed by block-wise concatenation, and finally the MaxPooling and MinPooling layers with interleaved concatenation



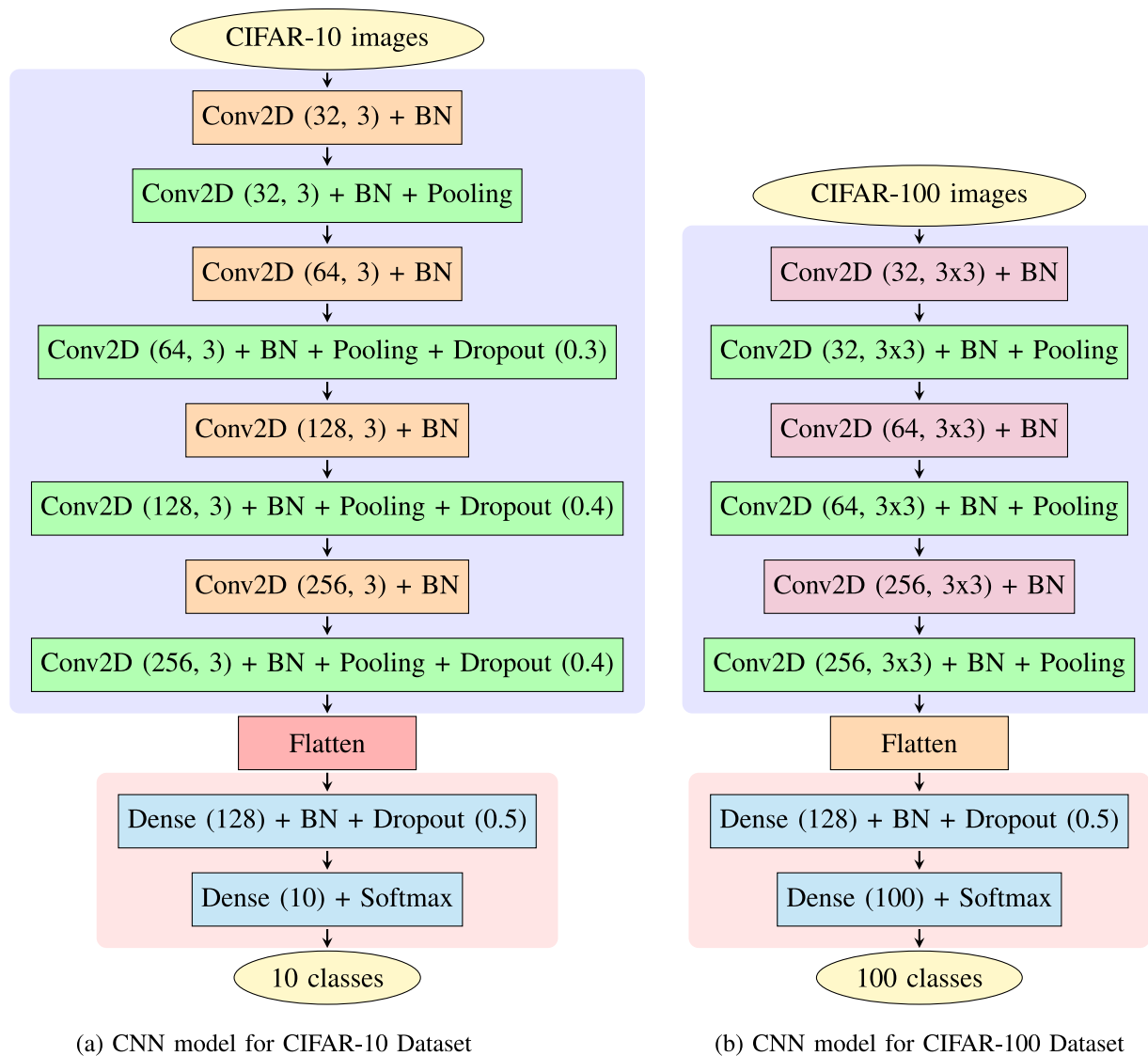


Fig. 5 A block diagram of the custom CNN model for classifying images from (a) the CIFAR-10 dataset and (b) the CIFAR-100 dataset. Conv2D refers to convolutional layer 2D followed by activation function, and BN stands for batch normalization

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

$$Precision = \frac{TP}{TP + FP},$$

$$Recall (Sensitivity) = \frac{TP}{TP + FN},$$

$$F_1 - score = 2 \times \frac{Precision}{TP + FP}$$

The above formulae utilize True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) for computing performance metrics.

3 Implementation

A custom CNN model was selected for each dataset and trained using our proposed strategy. Different pooling and activation strategies were employed to generate three distinct probability scores. The input was simultaneously passed through three parallel channels of the model, each utilizing a different pooling strategy, MaxPooling, MinPooling, and our novel MaxMinPooling, as illustrated in Fig. 3.

The MaxPooling model works by taking the maximum value over an input window, using a positive ReLU activation function at the output. In contrast, the MinPooling model emphasizes the minimum values in the feature maps and uses a negative ReLU activation function to retain the

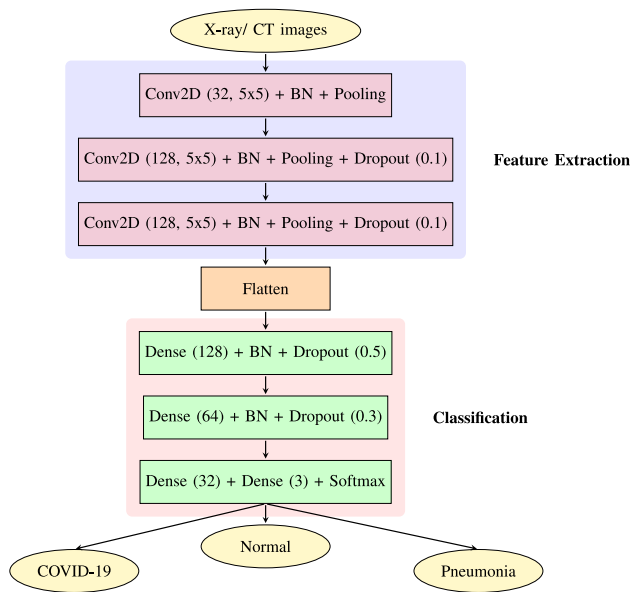


Fig. 6 A block diagram of the custom CNN model for classifying X-ray and CT scan images into three categories: COVID-19, normal, and pneumonia. Conv2D refers to convolutional layer 2D followed by activation function, and BN stands for batch normalization

negative coefficients. Our innovative MaxMinPooling model combines both MaxPooling and MinPooling outputs using an interleaving approach, incorporating both positive and negative ReLU activation functions to retain both sets of features through concatenation.

Figures 5 and 6 showcase block diagrams of each custom CNN model tailored to the CIFAR-10, CIFAR-100, and medical (X-ray and CT scan) datasets, respectively. These diagrams clarify the intricate architecture of the pooling and activation strategies and the operational flow

of our proposed models, developed empirically to achieve optimal classification performance. The training process for each model is meticulously documented to ensure a clear understanding of the approach and its effectiveness in fitting well with the training curves:

- **Data Preparation:** All images were resized to uniform 32x32 pixels and normalized to have pixel values between 0 and 1 for CIFAR datasets. The input sizes for X-ray and CT scan images were 160x160 and 224x224, respectively. Various data augmentation techniques were applied to increase dataset robustness, as detailed in Table 3.
- **Model Training:** Training was conducted on an i9 processor with 16 GB NVIDIA RTX3080 Ti GPGPU to expedite the computation. The models were evaluated at the end of each epoch using a hold-out validation set, comprising 20% of the training data, to monitor performance and adjust hyperparameters if necessary.
- **Optimizations and Monitoring:** Real-time monitoring of the training process was enabled through TensorFlow callbacks, allowing for immediate adjustments based on performance metrics like loss and accuracy.
- **Software Tools:** The implementation was carried out using CUDA 11.8, Python 3.8, TensorFlow 2.4 and Keras as the primary libraries for building and training the neural networks.

The hyperparameters for each model were selected based on an empirical intuition derived from prior studies and multiple iterations. The learning rate was optimized using an adaptive learning rate method (Adam optimizer) to adjust as training progresses. Batch size was carefully chosen to balance the trade-off between training speed and

Table 3 Experimental hyperparameter specifications for the proposed MaxMinPooling model

Hyperparameters	CIFAR-10	CIFAR-100	X-ray	CT Scan
Batchsize	64	64	128	128
Learning rate	0.001	0.001	0.00001	0.00001
Decay rate	0.9	0.91	0.89	0.827
Optimizer	Adam	Adam	Adam	Adam
Kernel_initializer	He_normal	He_Uniform	He_normal	He_normal
Kernel_regularizer	L2(0.001)	-	L2(0.001)	L2(0.001)
Activation	Softmax	Softmax	Softmax	Softmax
Loss	categorical	categorical	sparse_categorical	sparse_categorical
Early stopping	Yes (patience =20)	Yes (patience =10)	Yes (patience =10)	Yes (patience =10)
Library	Tensorflow, keras	Tensorflow, keras	Tensorflow, keras	Tensorflow, keras
Epochs	200	200	100	100
Input size	(32,32)	(32,32)	(160,160)	(224,224)
Augmentation	rotation_range(20), horizontal_flip, width_shift(0.2), height_shift(0.2)	horizontal_flip, width_shift(0.2), height_shift(0.2)	RandomContrast (0.8)	-

Table 4 Ablation study of performance metrics for various approaches on CIFAR-10 and CIFAR-100 datasets

Pooling method	Activation function	CIFAR-10 dataset				CIFAR-100 dataset			
		Accuracy	Precision	Recall	F1_Score	Accuracy	Precision	Recall	F1_Score
MaxPooling	Pos Relu	88.72	88.76	88.72	88.59	59.07	59.40	59.07	58.71
MinPooling	Neg Relu	89.08	89.05	89.08	88.94	58.66	58.78	58.66	58.27
MaxMinPooling	PosNeg Relu	88.80	88.81	88.80	88.69	63.96	65.04	64.72	64.37
MinMaxPooling	NegPos Relu	88.72	88.76	88.72	88.58	63.14	63.33	63.14	62.80
Probabilistic Average	–	90.08	90.10	90.10	89.96	64.72	65.04	64.72	64.37

Table 5 Ablation study of performance metrics of the experimented methods on X-ray and CT scan dataset

Methods	Activation function	X-ray dataset				CT scan dataset			
		Accuracy	Precision	Recall	F1_Score	Accuracy	Precision	Recall	F1_Score
MaxPooling	Pos Relu	88.60	88.60	88.55	88.50	95.61	96.34	96.54	96.40
MinPooling	Neg Relu	88.52	88.42	88.43	88.40	95.32	96.18	96.37	96.21
MaxMinPooling	PosNeg Relu	90.67	90.71	90.66	90.60	96.72	97.37	97.44	97.38
MinMaxPooling	NegPos Relu	90.0	90.0	90.0	89.95	96.57	97.20	97.30	97.24
Probabilistic Average	-	90.26	90.22	90.20	90.15	96.34	97.08	97.16	97.08

model performance stability. Each model was trained for up to 200 epochs, with early stopping implemented to prevent overfitting if the validation loss ceased to decrease for ten consecutive epochs. A detailed list of optimal hyperparameters for the MaxMinPooling model for the four different datasets is tabulated in Table 3, providing a comprehensive overview of the configuration settings employed in the experiments that ensures the reliability and reproducibility of the results.

4 Results analysis and discussion

This section details the classification accuracies obtained by the three distinct models—each employing MaxPooling, MinPooling, and our novel MaxMinPooling layers—across four different datasets, namely CIFAR-10, CIFAR-100, X-ray, and CT scan images. Performance metrics for each model, including accuracy, precision, recall, and F1 scores, are systematically presented in Tables 4 and 5. The tables provide a clear and concise comparison of the performance of each pooling strategy across the datasets, supporting the ablation study. Additionally, these comparisons are visually represented in a bar chart (Fig. 7), enhancing the interpretability of the data and providing clear insights into the strengths and weaknesses of each pooling method under different conditions.

The model training involved a rigorous process of selecting and iterating through various combinations of hyperparameters. The optimally tuned hyperparameters are detailed in Table 3. Additionally, learning curves for the datasets utilizing the proposed MaxMinPooling method are depicted in Fig. 8. These figures illustrate that both the training and validation curves for accuracy and loss converge as the number of epochs increases, signaling a well-fitting model that generalizes effectively across different data scenarios.

Notably, enhancements from the probabilistic average method over traditional MaxPooling show an increment of 1.36% for CIFAR-10, whereas a more substantial improvement of 5.65% is observed for CIFAR-100. These results underscore the effectiveness of combining MaxMinPooling with probabilistic averaging to boost the performance of MaxPooling-based neural networks. The method's efficacy was further tested on medical images such as X-rays and CT scans, specifically for distinguishing COVID-19 cases from healthy and pneumonia instances. While the proposed pooling method showed significant improvements in MaxMinPooling over MaxPooling, the gains were less pronounced when utilizing probabilistic fusion methods in medical image datasets. This could have been because probabilistic averaging may introduce noise or dilute critical features, which are essential for the precision required in medical image classification. However, it

Fig. 7 Grad-CAM analysis of the MaxMinPooling model on six CT scan images. Each row contains two pairs of images: the original CT scan (left) and the corresponding heatmap (right), highlighting the decision-making process of the model for effective pattern interpretation

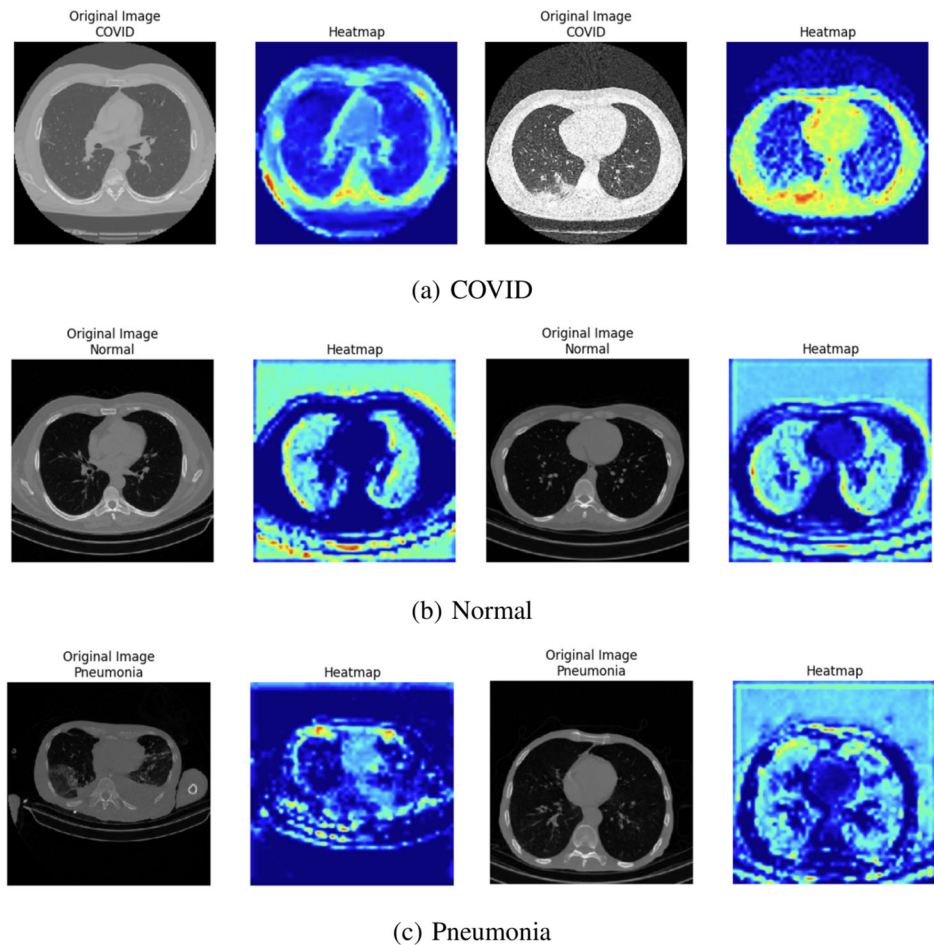
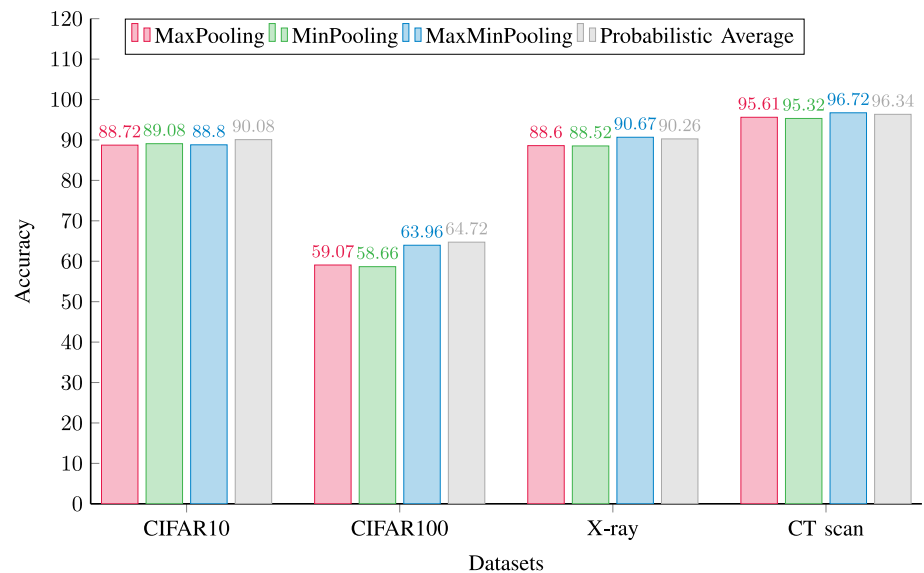
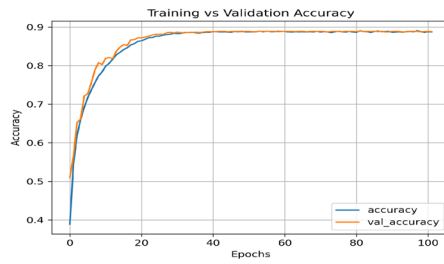


Fig. 8 Comparison of model accuracies for CIFAR-10, CIFAR-100, X-ray and CT scan datasets



is important to emphasize that the probabilistic fusion approach consistently showed improved results in non-medical datasets such as CIFAR-10 and CIFAR-100, where the fusion method performed better compared to the

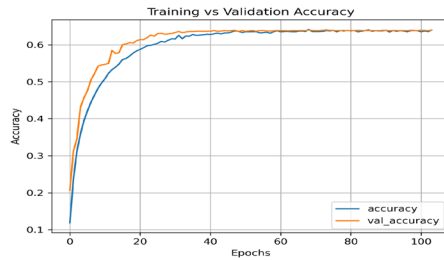
individual pooling techniques. The model's classification accuracy for each category is also detailed through confusion matrices, as illustrated in Fig. 9 and 10.



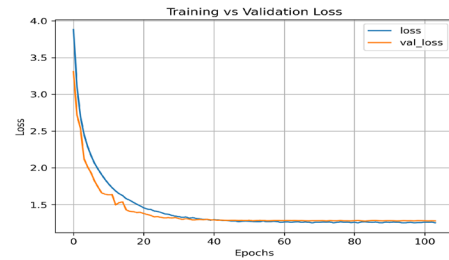
(a) Accuracy on CIFAR-10



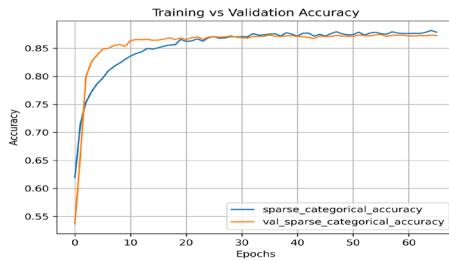
(b) Loss on CIFAR-10



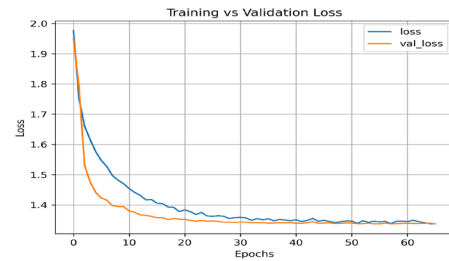
(c) Accuracy on CIFAR-100



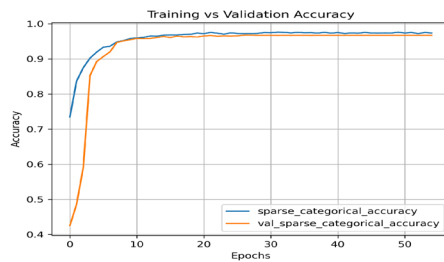
(d) Loss on CIFAR-100



(e) Accuracy on X-ray



(f) Loss on X-ray



(g) Accuracy on CT scan



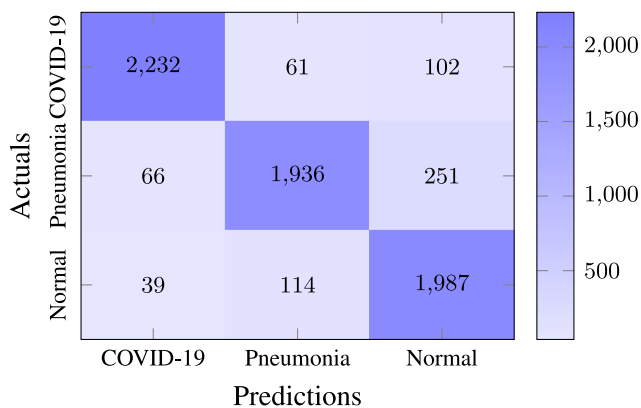
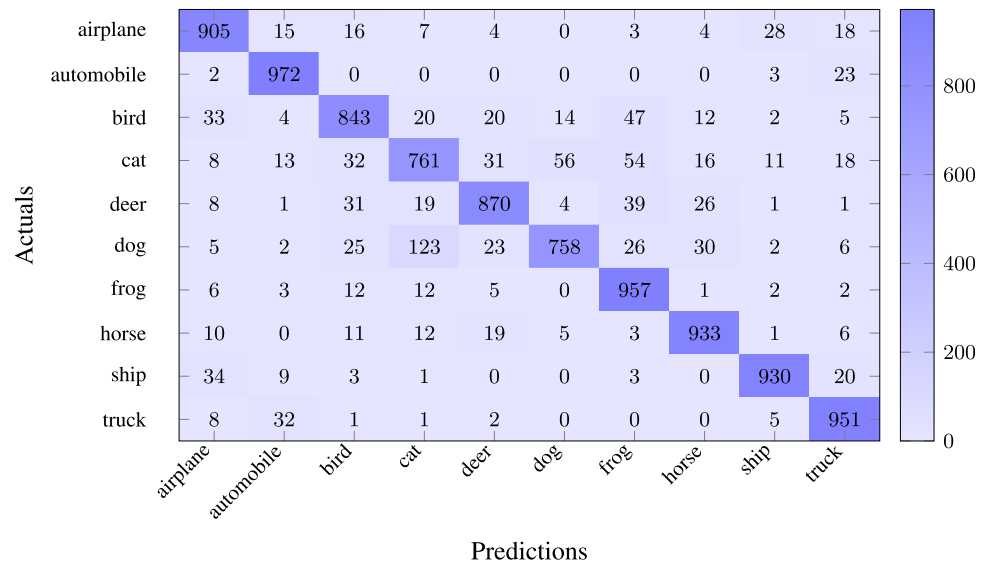
(h) Loss on CT scan

Fig. 9 Performance learning curves of the proposed MaxMinPooling method for CIFAR-10, CIFAR-100, X-ray and CT scan datasets

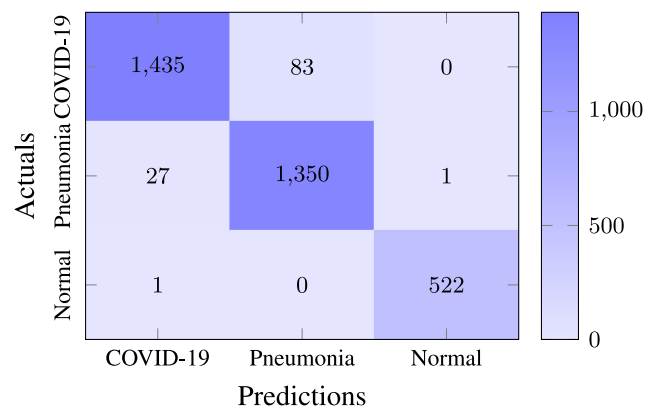
An important observation during our study was the impact of integrating both MaxPooling and MinPooling into the MaxMinPooling strategy, which effectively doubles the size of the feature maps. This increase in feature map size elevates the computational complexity of the model. However, in many AI applications, such as COVID-19 image classification, the precision of detection outweighs the need for real-time processing. Such enhancements are crucial as they lead to improved generalization of the AI solutions in scenarios where accuracy is

more critical than computational speed. Through experimental study, it has been observed that the change of order of MaxPooling and MinPooling, while concatenating the outputs in MaxMinPooling, yields almost similar classification performance. The detailed results are included in the ablation study of Tables 4 and 5 under “MinMaxPooling,” which further validate that the sequence of pooling operations has a negligible effect on the overall model accuracy.

Fig. 10 Confusion matrix for CIFAR-10 image classification using the proposed MaxMinPooling method



(a) X-ray dataset



(b) CT scan dataset

Fig. 11 Confusion matrices for detection of COVID-19 using the proposed MaxMinPooling method

Our extensive performance analysis in various datasets illustrates the benefits of integrating MaxPooling, MinPooling, and MaxMinPooling strategies. This integration enables the model to capture a broader and more nuanced spectrum of features from the input data. The study details how the model leverages high activation features typically accentuated by MaxPooling along with the subtler features often highlighted by MinPooling. This synergistic approach in the MaxMinPooling2D model enhances the overall feature representation and improves classification performance. These findings underscore the potential of the model in applications that require high levels of accuracy, providing valuable insight into its utility to improve automated AI solutions.

We employed Grad-CAM (gradient-weighted class activation mapping) to gain insights into the decision-making process of our CNN model when classifying CT scan images. Grad-CAM highlights the regions of an image

that strongly influence the predictions of the model. In Fig. 11, we present six CT scan images processed by the MaxMinPooling model, with two examples from each category. The heatmaps reveal the critical areas that contribute to the classification, where warmer colors correspond to higher activations. Upon closer examination, distinct heatmap patterns emerge for each category, providing valuable information for medical practitioners to further analyze the reasoning behind the classification. This method enhances the interpretability of AI models, offering a transparent visualization of the model's reasoning, which is crucial for building trust in AI-driven medical diagnostics.

In our study, we have drawn comparisons with several notable advancements in CNN pooling techniques from the literature, particularly focusing on their impact on classification accuracy. For instance, Ozdemir achieved

significant improvements, enhancing CIFAR-10 and CIFAR-100 accuracies by 6.28% and 7.76%, respectively, compared to traditional MaxPooling [21]. Similarly, Yu et al. reported an improvement in the CIFAR-10 test accuracy to 89.20% using a mixed pooling method, outperforming the 88.64% accuracy from MaxPooling [22]. These studies underscore the potential of alternative pooling methods in enhancing classification performance.

Our research integrates MaxPooling, MinPooling, and our novel MaxMinPooling strategies within a unified model framework. This integrated approach not only leverages the strengths of each pooling method, but also addresses the gaps noted in the single-strategy pooling models. For example, Williams et al. and Zhao et al. also reported marginal improvements using new pooling methods over MaxPooling, with Williams et al. reaching an accuracy of 81.15% on CIFAR-10 [25], and Zhao et al. enhancing CIFAR-100 and CIFAR-10 accuracies by 4.11% and 4.32% [18]. In contrast, our integrated model consistently demonstrates superior performance, showcasing more robust and versatile capabilities across different datasets, including medical imaging datasets critical for COVID-19 detection.

These comparative analyses highlight the efficacy of our novel integrated pooling approach, indicating an improvement over traditional methods and setting a new benchmark in the utilization of CNNs for complex image classification tasks. This juxtaposition with previous studies establishes the significant strides our methodology offers beyond the current state-of-the-art, particularly in applications demanding high accuracy and reliability.

5 Conclusion, limitations and future work

5.1 Conclusion

This study introduced a novel convolutional neural network (CNN) architecture that integrates three distinct pooling strategies: MaxPooling, MinPooling, and MaxMinPooling. This approach aimed to leverage the unique advantages of each pooling method to enhance the ability of the model to classify images accurately across a diverse set of datasets, including CIFAR-10, CIFAR-100, CT scans, and X-ray images. Our findings demonstrate that by averaging the probabilistic outputs from models utilizing these varied pooling layers, we achieved superior classification performance compared to models that rely on a single type of pooling layer. This improvement is consistently observed across all tested datasets, underscoring the robustness and versatility of the proposed architecture. The significance of the study lies not only in the enhanced accuracy achieved in image classification tasks but also in

the demonstration of how pooling layer diversity can contribute to the development of more sophisticated and capable CNN models. By incorporating MaxPooling, MinPooling, and MaxMinPooling layers, the proposed model effectively captures both the most salient and the subtlest features within the images, offering a more comprehensive understanding of the visual data. This balance between feature detection capabilities is particularly crucial in challenging classification scenarios, such as the detection of COVID-19 in medical imaging, where both prominent and minute patterns can be indicative of the diagnosis.

5.2 Limitations

The present study introduces several innovative enhancements, but it also faces limitations that merit consideration. Potential challenges include generalization to unseen data due to specific tuning to training datasets, increased computational complexity from integrating multiple pooling strategies, and risks of overfitting with the complexity of the model. Moreover, the efficacy of the proposed method has only been validated on a limited image dataset. While the fundamental pooling layer of CNNs suggests potential applicability across different modalities, this study restricts its evaluation to image data, which may limit the broader application of the findings. Other concerns include interpretability issues due to novel pooling layers, scalability challenges for more extensive and more varied datasets, and the intricate parameter tuning required for optimal performance across diverse scenarios, all of which could impact the practical deployment of the model in varied real-world settings.

5.3 Future work

The promising results of the study open several avenues for future research, as follows:

- 1) Further exploration into combining different pooling strategies within CNN architectures could yield even more efficient and accurate models. Investigating the impact of varying the proportion of each pooling method's contribution to the final model output may provide insights into optimal configurations for specific image classification tasks.
- 2) Extending the application of our proposed architecture to other domains, such as video analysis, natural language processing, or complex pattern recognition tasks, could demonstrate the versatility and adaptability of the model beyond static image classification. The ability of our model to capture a wide range

of feature activations makes it a promising candidate for these complex applications.

- 3) Integrating emerging deep learning techniques, such as attention mechanisms or generative adversarial networks (GANs), with our pooling strategy could enhance the model's performance. These techniques could provide additional context or augment the training data, potentially improving the ability of the model to generalize from limited or noisy datasets.
- 4) The deployment of the proposed model in real-world applications, particularly in health care for disease detection and progression monitoring tasks, could have significant implications. Further validation and adaptation of the model to meet clinical requirements would be essential steps toward this goal.

Acknowledgements The authors sincerely thank the editor and anonymous reviewers for their valuable time, insightful feedback, and suggestions, which significantly improved this work. Authors would like to acknowledge the support from the Infosys Centre for AI, IIIT Delhi, India.

Author Contributions All authors contributed to the conception and design of the study. Material preparation, data collection, and analysis were carried out collaboratively. The first draft of the manuscript was collectively written, and all authors reviewed and approved the final version. *Corresponding author

Funding This research was conducted without any financial support or external funding.

Data availability All data supporting the findings of this study are available from the authors upon request. After acceptance, all code and data will be made publicly available in an open repository, and its link will be provided here.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. LeCun, Y., Bengio, Y. (1998) Convolutional networks for images, speech, and time series, [Online]. Available: <https://api.semanticscholar.org/CorpusID:6916627>
2. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks, in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. Red Hook, NY, USA: Curran Associates Inc., p. 1097–1105
3. Karpathy A (2018) Cs231n convolutional neural networks for visual recognition, Stanford University, accessed: 2024-06-02. [Online]. Available: <https://cs231n.github.io/convolutional-networks/>
4. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization, In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921–2929
5. Akyol K (2024) Etsvf-covid19: efficient two-stage voting framework for covid-19 detection, *Neural Computing and Applications*, [Online]. Available: <https://doi.org/10.1007/s00521-024-10150-0>
6. Kurt Z, Işık Ş, Kaya Z, Anagün Y, Koca N, Çiçek S (2023) Evaluation of efficientnet models for covid-19 detection using lung parenchyma. *Neural Comput Appl* 35(16):12 121–12 132
7. Kibriya H, Amin R (2023) A residual network-based framework for covid-19 detection from cxr images,” *Neural Computing and Applications*, vol. 35, pp. 8505–8516, [Online]. Available: <https://doi.org/10.1007/s00521-022-08127-y>
8. Sahu A, Das PK, Meher S (2023) High accuracy hybrid cnn classifiers for breast cancer detection using mammogram and ultrasound datasets,” *Biomedical Signal Processing and Control*, vol. 80, p. 104292, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809422007467>
9. Mridha K, Uddin MM, Shin J, Khadka S, Mridha MF (2023) An interpretable skin cancer classification using optimized convolutional neural network for a smart healthcare system. *IEEE Access* 11:41 003–41 018
10. Qummar S, Khan FG, Shah S, Khan A, Shamshirband S, Rehman ZU, Ahmed Khan I, Jadoon W (2019) A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access* 7:150 530–150 539
11. Oguz C, Aydin T, Yaganoglu M (2024) A cnn-based hybrid model to detect glaucoma disease. *Multimedia Tools and Applications* 83:17921–17939, [Online]. Available: <https://doi.org/10.1007/s11042-023-16129-8>
12. Taye MM (2023) Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation* 11(3):52
13. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
14. Goodfellow I, Bengio Y, Courville A (2016) *Deep Learning*. MIT Press, [Online]. Available: <http://www.deeplearningbook.org>
15. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JA, van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Analys* 42:60–88
16. Gholamalnejad H, Khosravi H (2020) Pooling methods in deep neural networks, a review,” 09
17. Zafar A, Aamir M, Nawi N, Arshad A, Riaz S, Alruban A, Dutta A, Alaybani S (2022) A comparison of pooling methods for convolutional neural networks, *Applied Sciences*, 12, 8643, 08
18. Zhao L, Zhang Z (2024) A improved pooling method for convolutional neural networks. *Scientific Reports* 14:01
19. Park J, Kim J-Y, Huh J-H, Lee H-S, Jung S-H, Sim C-B (2021) A novel on conditional min pooling and restructured convolutional neural network,” *Electronics*, 10(19), [Online]. Available: <https://www.mdpi.com/2079-9292/10/19/2407>
20. Vienken G (2016) Scale selection in convolutional neural networks with dimensional min-pooling and scaling filters, [Online]. Available: <https://api.semanticscholar.org/CorpusID:44940905>
21. Özdemir C (2023) Avg-topk: A new pooling method for convolutional neural networks, *Expert Systems with Applications*, 223, 119892, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417423003937>
22. Yu D, Wang H, Chen P, Wei Z (2014) Mixed pooling for convolutional neural networks, 10 2014, pp. 364–375
23. Zhang Y, Er J, Wang N, Pratama M (2016) Attention pooling-based convolutional neural network for sentence modelling. *Information Sciences* 373:08
24. Cui Y, Zhou F, Wang J, Liu X, Lin Y, Belongie S (2017) Kernel pooling for convolutional neural networks, In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2017:3049–3058

25. Williams T, Li R (2018) Wavelet pooling for convolutional neural networks, 02 2018
26. Wanda P, Jie H (2020) Runpool: A dynamic pooling layer for convolution neural network. *Int J Comput Intell Syst* 13:01
27. Sharma T, Singh V, Sudhakaran S, Verma NK (2019) Fuzzy based pooling in convolutional neural network for image classification, In: *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–6
28. Nirthika R, Manivannan S, Ramanan A, Wang R (2022) Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study, *Neural Computing and Applications*, 34, 1–27, 04 2022
29. Satti P, Sharma N, Garg B (2020) Min-max average pooling based filter for impulse noise removal. *IEEE Signal Processing Letters* 27:1475–1479
30. Liyanage DC, Hudjakov R, Tamre M (2020) Hyperspectral image band selection using pooling, In: *International Conference Mechatronic Systems and Materials (MSM) 2020*:1–6
31. World Health Organization (2020) Coronavirus disease (covid-19) pandemic, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>, accessed on: 2024-06-02
32. Aggarwal P, Mishra NK, Fatimah B, Singh P, Gupta A, Joshi SD (2022) Covid-19 image classification using deep learning: Advances, challenges and opportunities, *Computers in Biology and Medicine*, vol. 144, p. 105350, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482522001421>
33. Mishra NK, Singh P, Joshi SD (2021) Automated detection of covid-19 from ct scan using convolutional neural network, *Bio-cybernetics and Biomedical Engineering*, 41(2), 572–588. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0208521621000437>
34. Wang L, Lin ZQ, Wong A (2020) Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Sci Reports* 10(1):19549
35. Apostolopoulos ID, Mpesiana TA (2020) Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 43(2):635–640
36. Li L, Qin B, Xu Z, Yin Y, Wang X, Kong B, Chen Y, Liu Z, Wang Q, Zhang J, Xia B (2020) Artificial intelligence distinguishes covid-19 from community-acquired pneumonia on chest ct. *Radiology* 296(2):E65–E71
37. Dwivedi D, Kushwaha SK, Kumar S (2023) Lmnet: A light-weight multi-scale cnn architecture for covid-19 detection. *BMC Med Imag* 23(1):1–16
38. Maas AL, Hannun AY, Ng AY (2013) Rectifier nonlinearities improve neural network acoustic models, In: *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pp. 3–6
39. Xu B, Wang N, Chen T, Li M (2015) Empirical evaluation of rectified activations in convolutional network, *arXiv preprint[SPACE]arXiv:1505.00853*
40. Subramanian B, Jeyaraj R, Ugli RAA, Kim J (2024) Apalu: A trainable, adaptive activation function for deep learning networks, *arXiv preprint[SPACE]arXiv:2402.08244*, [Online]. Available: <https://ar5iv.org/abs/2402.08244>
41. Kunc V, Kléma J (2024) Three decades of activations: A comprehensive survey of 400 activation functions for neural networks, *arXiv preprint[SPACE]arXiv:2402.09092*, [Online]. Available: <https://ar5iv.org/abs/2402.09092>
42. Pareto.ai, “Understanding activation functions in neural networks,” 2023. [Online]. Available: <https://www.pareto.ai/blog/understanding-activation-functions-in-neural-networks>
43. Krizhevsky A (2009) Learning multiple layers of features from tiny images, University of Toronto, Tech. Rep., cIFAR-10 and CIFAR-100 datasets
44. Tahir AM, Chowdhury ME, Khandakar A, Rahman T, Qiblawey Y, Khurshid U, Kiranyaz S, Ibtehaz N, Rahman MS, Al-Maadeed S, Mahmud S, Ezeddin M, Hameed K, Hamid T (2021) Covid-19 infection localization and severity grading from chest x-ray images, *Computers in Biology and Medicine*, 139, p. 105002, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482521007964>
45. Tahir AM, Chowdhury MEH, Qiblawey Y, Khandakar A, Rahman T, Kiranyaz S, Khurshid U, Ibtehaz N, Mahmud S, Ezeddin M (2021) COVID-QU-Ex, <https://doi.org/10.34740/kaggle/dsv/3122958>
46. Rahman T, Khandakar A, Qiblawey Y, Tahir A, Kiranyaz S, Abul Kashem SB, Islam MT, Maadeed S Al, Zughaier SM, Khan MS, Chowdhury ME (2021) Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images, *Computers in Biology and Medicine*, 132, 104319, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S001048252100113X>
47. Degerli A, Ahishali M, yamaç M, Kiranyaz S, Chowdhury M, Hameed K, Hamid T, Mazhar R, Gabbouj M (2021) Covid-19 infection map generation and detection from chest x-ray images,” *Health Information Science and Systems*, 9, 04 2021
48. Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR, Khan MS, Iqbal A, Emadi NA, Reaz MBI, Islam MT (2020) Can ai help in screening viral and covid-19 pneumonia? *IEEE Access* 8:132 665–132 676
49. Soares E, Angelov P, Biaso S, Froes MH, Abe DK (2020) Sars-cov-2 ct-scan dataset: A large dataset of real patients ct scans for sars-cov-2 identification, *medRxiv*, [Online]. Available: <https://www.medrxiv.org/content/early/2020/05/14/2020.04.24.20078584>
50. Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 45(4):427–437

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.