International Journal of Advances in Applied Sciences (IJAAS)
Vol. 15, No. 3, September 2026, pp. 1072~1084
ISSN: 2252-8814, DOI: 10.11591/ijaas.v15.i3.pp1072-1084

COVID-19 infection detection using convolutional self-attention network with voting classifier

Mustafa Hussein Zwayyer[1], Ammar A. Ali[2], Rusul Hussein Hasan[3]
[1]Department of Islamic Banking and Financial Science, College of Islamic Sciences, University of Baghdad, Baghdad, Iraq
[2]Department of Islamic Educational and Psychological Sciences, College of Islamic Sciences, University of Baghdad, Baghdad, Iraq
[3]Department of Public Law, College of Law, University of Baghdad, Baghdad, Iraq

Article Info

Article history:

Received Jan 11, 2025
Revised Jul 2, 2026
Accepted Aug 13, 2026

Keywords:

Artificial intelligence
Computer vision
COVID-19
Deep learning
Gamma correction
Voting ensemble classifier

ABSTRACT

Early and accurate detection of COVID-19 from chest computed tomography (CT) scans are becoming essential for effective clinical decision-making and disease control. This study is proposing a robust deep learning framework that integrates a convolutional self-attention network (CSAN), gamma correction for image enhancement, and a voting-based ensemble classifier to improving diagnostic performance. The model is being evaluated on a dataset of 2,271 CT images and is achieving an accuracy of 95.12%, sensitivity of 97.25%, specificity of 98.11%, F1-score of 96.46%, and area under the curve (AUC) of 0.977. Experimental results are demonstrating that the proposed method significantly surpasses baseline models, including standalone CSAN, residual networks 50 (ResNet-50), densely connected convolutional network 121 (DenseNet-121), and visual geometry group (VGG) 16, with improvements of up to 9.5% in accuracy. The integration of attention mechanisms, image enhancement, and ensemble learning are proving effective in capturing both local and global features, leading to more reliable classification. These findings are suggesting that the proposed framework is supporting automated COVID-19 diagnosis in real-world clinical applications.

Corresponding Author:

Rusul Hussein Hasan
Department of Public Law, College of Law, University of Baghdad
Baghdad, Iraq
Email: russl@colaw.uobaghdad.edu.iq

1. INTRODUCTION

The novel coronavirus (SARS-CoV-2), which was first detected in December 2019, poses a significant threat to healthcare systems worldwide. Rapid and precise diagnosis is key to managing patients and preventing the disease from spreading further. Although reverse transcription (RT)-polymerase chain reaction (PCR) tests are considered the most reliable method of diagnosis, radiological imaging, particularly computed tomography (CT) scans, has proven to be informative and more sensitive than PCR in certain cases [1].

Artificial intelligence, especially deep learning methods, have showed significant capability in medical image processing. The capability of deep learning models to automatically identify features from complicated imaging information make them suitable for detecting several diseases, including the diagnosis of COVID-19 from X-ray and CT images. These approaches can improve accessibility and decrease diagnostic time, in addition to providing a second assessment for healthcare professionals [2], [3].

In recent years, several deep learning models have been considered for detection of SARS-CoV-2 via imaging. Convolutional neural networks (CNNs) have become particularly popular due to their performance in image processing tasks. Architecture, the self-attention network (SAN), has also proved effective in identifying long-range correlations in images. Group learning has also proved useful in improving classification performance. However, these approaches either work separately or use basic combination techniques. Improved strategies for combining these approaches are still needed [4]–[7].

This study introduces an innovative concept that integrates three fundamental pillars with the aim of improving the detection rate of SARS-CoV-2 by using computerized CT images.

i) Convolutional self-attention network (CSAN): these harness the benefits of feature extraction offered by CNNs, but also includes self-attention systems that can help identify and capture long-range connections within images more effectively.
ii) Voting ensemble classifier: this utilizes a number of different fundamental classifiers in an ensemble system to enhance classification performance.
iii) Gamma correction: sharpening the attributes within images using pre-processing techniques to extract information more efficiently.

2. RELATED WORK

2.1. Deep learning for medical image analysis

This research identified multiple structures, utilizing dense convolutional network (DenseNet) and residual network (ResNet) models to diagnose the presence of SARS-CoV-2 from chest X-ray scans with a precision rate of up to 91%. While this demonstrates the immense potential of deep learning technology in medical imaging, it also highlights the importance of designing specific models for particular imaging technologies [4], [8]. In this research, a review of deep learning techniques for COVID-19 diagnosis and machine learning was conducted. Different types of data were used to analyze the results, including X-rays, reverse transcriptase PCR tests, and CT, in addition to cough sounds. It was proven that there is a need for automated computer-aided techniques due to the time and cost constraints associated with manual analysis [9].

2.2. COVID-19 detection using computed tomography scans

Gürsoy and Kaya [10] proposed performance of ensemble and hybrid models compared to individual architectures, providing an overview of deep learning techniques for COVID-19 detection. In additional, the methodologies, challenges, and potential future directions are studied, indicating that combining several approaches can lead to improved disease detection rates [10]. Mijwil [11] suggest to detect COVID-19 from chest X-rays using deep CNN architecture. The results of this study achieved an accuracy rate of 93.5% using transfer learning with the modified ResNet-50 architecture, demonstrating the potential of transfer learning in situations with limited training data [11].

2.3. Self-attention networks in medical imaging

Mondal *et al.* [12] conducted a study to predict COVID-19 using different architectures using deep learning. The obtain results focusing on relevant image regions, attention mechanisms can significantly improve the model performance, especially in CT scans where COVID-19 features may be localized in specific lung regions [12]. Das *et al.* [13] proposed method to detection the COVID-19 by using deep learning and machine learning techniques was conducted. The results illustrated that hybrid approaches combining CNNs and attention mechanisms gave higher accuracy results than standard CNNs alone, indicating the potential benefit of incorporating attention into model architectures [13].

2.4. Ensemble methods for medical image classification

Akter *et al.* [14] proposed an ensemble learning method for COVID-19 detection from chest X-rays by combining voting mechanism and multiple CNN models. The accuracy results were achieved at 94.1%, demonstrating the effectiveness of ensemble methods in improving classification performance [14]. Cruz [15] developed a multi-model ensemble approach that combined features from different CNN architectures for COVID-19 detection from CT scans. Ismael [16] purpose of this paper is the improvement of quality of natural images through various enhancement techniques.

2.5. Image enhancement techniques

Kumar *et al.* [7] employed gamma correction and several image enhancement methods as pre-processing steps for the detection of a SARS-CoV-2 infection using chest X-ray images. The results showed a 2.3% increase in model performance accuracy when using enhanced images. Zhou *et al.* [8] showed the rapid detection of COVID-19. This paper proposes an ensemble deep learning model for novel COVID-19 detection from CT images.

3. COVID-19 AND ARTIFICIAL INTELLIGENCE

3.1. COVID-19 pandemic overview

The SARS-CoV-2 virus caused the novel pandemic of the disease now known as 'COVID-19', which was first identified in Wuhan, China, in December 2019 and soon began to spread worldwide. According to World Health Organization data up to September 2020, there were already more than 33 million reported cases, and over 991,000 people had died in 213 countries worldwide, resulting in a case fatality ratio of about 3.3% [17]. Early and accurate diagnosis is one of the key elements in the efficient treatment of the disease and the prevention of further spread of infection. Although RT-PCR is recognized as the most effective test for diagnosing SARS-CoV-2, research suggests that CT scans may offer supplementary information in certain situations. Such diagnostic imaging methods can sometimes provide more sensitive results than RT-PCR [1].

3.2. Artificial intelligence in COVID-19 detection

It is worth noting that artificial intelligence has become an effective means of combatting the global pandemic of COVID-19, as its application in healthcare enables more efficient diagnosis, tracking, and treatment of the disease. Deep learning has produced significant results in terms of image processing and analysis aimed at detecting the virus [18]. Different neural network architectures and learning strategies have also been investigated to improve the representation and classification of complex patterns in various applications [19]. CNNs are commonly used in many images processing operations, including identifying the presence of the virus in patients based on their radiographs. This is made possible by bypasses that can autonomously recognize hierarchies of attributes and identify details in images that people cannot consistently recognize, as shown in Figure 1.

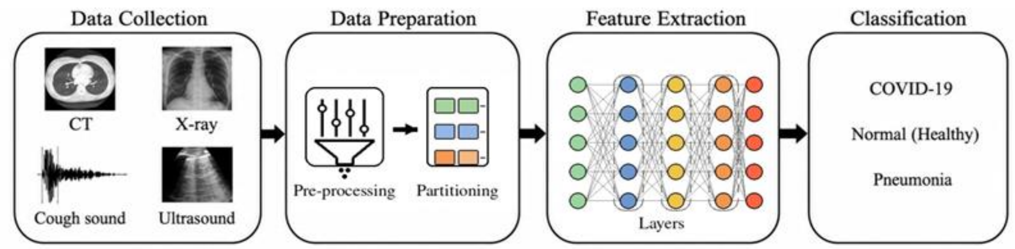


Figure 1. Overview of COVID-19 detection systems using deep learning

CNN architecture includes several types of layers, including convolutional, clustering, and fully connected layers. The former use filters to extract useful attributes from images. Clustering layers minimize spatial dimensions and extract static attributes, while fully connected layers analyze these attributes and classify or regress them Figure 2 [20], [21].

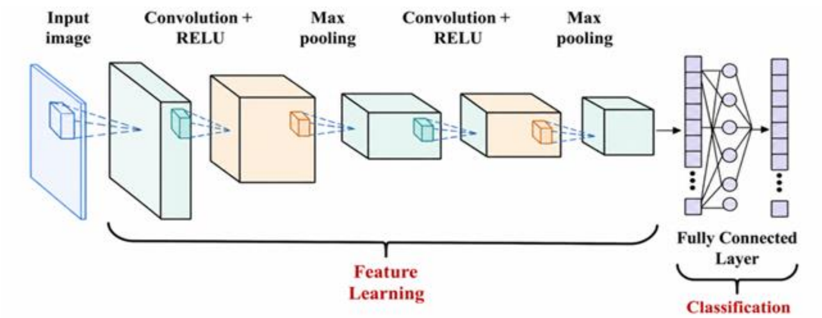


Figure 2. CNNs architecture

3.3. Self-attention networks

SANs represent a significant breakthrough in neural network models. They first emerged in the field of natural language processing and were later applied to computer vision. While standard CNNs operate with local neighborhoods via convolutions, SANs establish associations between all pairs of points in the sequence, thereby enabling them to capture correlations between distant points in the sequence. A major benefit of SANs is their ability to focus on relevant segments of the input sequence, enabling the prioritization of different segments depending on the specific problem being addressed. This ability is further enhanced by the multi-head attention approach, which operates on several different subspaces generated from the input sequence (Figure 3).

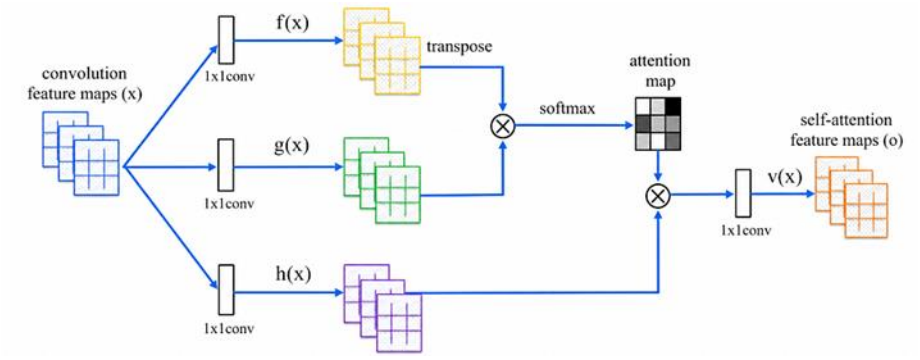


Figure 3. CSANs architecture

4. METHOD

4.1. Dataset description

The proposed model was trained using a dataset comprising 2,271 cross-sectional CT images. These images, which include both natural and verified cases of COVID-19 infection, are sourced from reliable public domain sources. Patient anonymity has been maintained to protect confidentiality. The dataset includes various types of CT scan images ranging from mild vitreous opacities to severe pulmonary adhesions, as well as normal and non-COVID-19 lungs, to enable the model to differentiate between the two classes.

4.2. Data split

The segmentation was performed using the stratified sampling technique, ensuring that the ratio of the two classes remained consistent within each subgroup described above. This guarantees that the model is trained and tested on representative datasets, thereby minimizing the risk of bias. In this paper, the data were split into three subsets to ensure training and evaluation of the model: i) training set: 70% (1,590 images); ii) validation set: 15% (340 images); and iii) testing set: 15% (341 images).

4.3. Data preprocessing

4.3.1. Image standardization

Three additional resolutions of 128 × 128, 256 × 256, and 512 × 512 pixels have been implemented to standardize all CT images and ensure comparability. This ensures consistency in the dimensions of the inputs to the artificial neural network. Resizing was performed using bilinear interpolation in order to preserve the structural integrity of the pulmonary regions, which are diagnostically critical for the identification of ground-glass opacities and consolidations. In addition, standardizing all images to a unified resolution eliminates discrepancies introduced by heterogeneous acquisition devices and scanning protocols across the source datasets. This preprocessing step therefore plays a fundamental role in stabilizing training dynamics and ensuring that the model learns features that are not biased by variations in image dimensions or aspect ratios.

4.3.2. Normalization

In this stage, the pixel intensity values are being scaling to a standardized range in order to enhance the stability of the learning process. Each pixel value is being divided by 255, so that the input data is becoming normalized between 0 and 1. This transformation is helping the neural network to converge faster and is reducing the risk of unstable gradients during training. In addition, normalization is improving the numerical consistency of the dataset and allowing the model to learn more generalized patterns instead of being affected by raw intensity variations. Without this preprocessing step, the training processes are becoming slower and less efficient, especially when dealing with deep architectures such as CSAN. Therefore, normalization is playing a critical role in optimizing the overall performance of the proposed framework.

4.3.3. Gamma correction

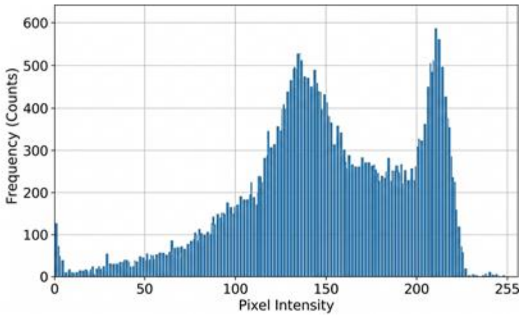
Gamma correction technology has been applied to improve visibility of important features within computerized section images using (1).

$$I' = 255 \times \left(\frac{I}{255}\right)^{\gamma} \tag{1}$$

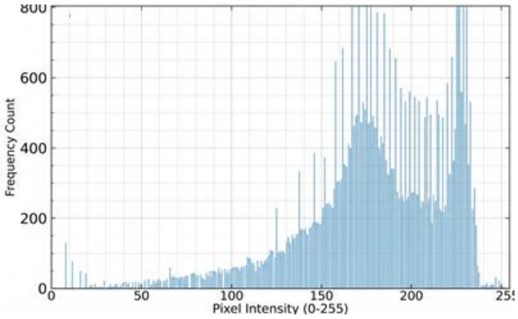
Where I is the original pixel value, I' is the corrected pixel value, and γ is the gamma correction factor. Experimented with four different gamma values (0, 0.5, 1, and 1.5) to determine the optimal correction factor for COVID-19 detection. Figure 4 illustrates the effect of different gamma values on a sample CT scan image. Specifically, Figure 4(a) presents the original unprocessed CT scan, where the contrast between lung tissue and infected regions remains relatively low. Figure 4(b) generated with $\gamma = 0.5$, brightens the darker areas and improves the visibility of subtle ground-glass opacities. Figure 4(c) corresponds to $\gamma = 1.0$, which preserves the original intensity distribution and serves as a reference baseline. Figure 4(d) was produced with $\gamma = 1.5$, which intensifies the contrast in the lung parenchyma and accentuates the boundaries of pathological regions, which proved most beneficial for the proposed framework. For Figures 4(c) and 4(d), the x-axis represents the horizontal pixel coordinate and the y-axis represents the vertical pixel coordinate of the CT image, both expressed in pixel units.


(a)


(b)



(c)



(d)

Figure 4. Gamma correction applied to CT scan images with different gamma values: (a) original image, (b) $\gamma = 0.5$, (c) $\gamma = 1.0$, and (d) $\gamma = 1.5$

4.4. Model architecture

4.4.1. Convolutional self-attention network

The core of approach of this paper is the CSAN, which integrates self-attention mechanisms with convolutional operations. The CSAN model increases the CNN's receptive field without increasing the computational burden associated with very large kernel sizes. The architecture reshapes the features from the previous hidden layer as in (2).

$$x \in \mathbb{R}^{c \times N} \tag{2}$$

Where c is the number of channels, and N is the product of all other dimensions.

Derive the functions f, g, and h on x using 1×1 convolutions, resulting in a channel number shift from c to c^*. The corresponding functions are defined in (3).

$$f(x) = W_{f^x}, g(x) = W_{g^x}, h(x) = W_{h^x} \quad (3)$$

Where $W_f, W_g, W_h \in \mathbb{R}^{c \times c^*}$.

The self-attention mechanism computes attention weights by taking the dot product of $f(x)$ and $g(x)$, followed by a softmax operation, as defined in (4).

$$\beta_{j,i} = softmax_j \left(f(x_i)^T .\ g(x_j) \right) \quad (4)$$

The output of the self-attention layer is computed, as in (5).

$$o_{i=\sum_j \beta_{j,i} \cdot h(x_j)} \quad (5)$$

Finally, attach input features, x, to the weighted output, with γ as a learnable scalar parameter, as in (6).

$$y_i = \gamma \cdot o_i + x_i \quad (6)$$

PyTorch usage this is a quick and effective implementation from Fast.ai.

```
import torch
import torch.nn as nn
import torch.nn.functional as F
class SelfAttention(nn.Module):
    def __init__(self, n_channels):
        super(SelfAttention, self).__init__()
        self.n_channels = n_channels
        self.query_weight = nn.Parameter(torch.randn(n_channels, n_channels // 8))
        self.key_weight = nn.Parameter(torch.randn(n_channels, n_channels // 8))
        self.value_weight = nn.Parameter(torch.randn(n_channels, n_channels))
        self.gamma = nn.Parameter(torch.zeros(1))
    def forward(self, x):
        size = x.size()
        x = x.view(-1, size[2], self.n_channels)  # Assuming x has shape (batch_size, 
sequence_length, n_channels)
        f = torch.matmul(x, self.query_weight)
        g = torch.matmul(x, self.key_weight)
        h = torch.matmul(x, self.value_weight)
        beta = F.softmax(torch.matmul(f.transpose(1, 2), g), dim=1)
        o = self.gamma * torch.matmul(beta, h)
        o = o.view(size)
        return o
```

Table 1 details the complete architecture of the CSAN model, summarizing every layer along with its filter count, kernel size, stride, output shape, and number of trainable parameters. The architecture follows a hierarchical convolutional backbone in which four convolutional blocks progressively expand the channel dimension from 32 to 256 while reducing the spatial resolution through max-pooling, and a self-attention layer is inserted after the second pooling stage in order to model long-range dependencies between distant lung regions. The fully connected head condenses the resulting feature map into a 128-dimensional embedding that is subsequently passed to the voting ensemble classifier, yielding a total of 5,189,440 parameters, of which 5,187,200 are trainable. This efficient and powerful design is motivated by previous research that enhances classical neural networks with some additional operations for better learning capability [18], and achieves a good balance between representation ability and computational expense on CT-scan images. Total parameters: 5,189,440; trainable parameters: 5,187,200; non-trainable parameters: 2,240.

4.4.2. Self-attention implementation

The self-attention mechanism is implemented using three 1 × 1 convolution layers to produce f(x), g(x), and h(x), commonly referred to as query, key, and value, respectively. The implementation steps are:

i) Define three 1 × 1 convolution layers to produce f(x), g(x), and h(x).
ii) Reshape to a c × N tensor.
iii) Apply softmax attention weights using matrix multiplication.
iv) Return features to their natural shape.
v) Apply a learnable scalar parameter γ to the output and add the input features.

Figure 5 illustrates the self-attention mechanism integrated into the CSAN model.

Table 1. CSAN model architecture

Layer	Type	Filters	Kernel size	Stride	Output shape	Parameters
Input	Input	-	-	-	(128, 128, 1)	0
Conv1	Conv2D	32	3 × 3	1	(126, 126, 32)	320
BN1	BatchNorm	-	-	-	(126, 126, 32)	128
ReLU1	ReLU	-	-	-	(126, 126, 32)	0
Pool1	MaxPool	-	2 × 2	2	(63, 63, 32)	0
Conv2	Conv2D	64	3 × 3	1	(61, 61, 64)	18,496
BN2	BatchNorm	-	-	-	(61, 61, 64)	256
ReLU2	ReLU	-	-	-	(61, 61, 64)	0
Pool2	MaxPool	-	2 × 2	2	(30, 30, 64)	0
SA	Self-Attention	-	-	-	(30, 30, 64)	12,352
Conv3	Conv2D	128	3 × 3	1	(28, 28, 128)	73,856
BN3	BatchNorm	-	-	-	(28, 28, 128)	512
ReLU3	ReLU	-	-	-	(28, 28, 128)	0
Pool3	MaxPool	-	2 × 2	2	(14, 14, 128)	0
Conv4	Conv2D	256	3 × 3	1	(12, 12, 256)	295,168
BN4	BatchNorm	-	-	-	(12, 12, 256)	1,024
ReLU4	ReLU	-	-	-	(12, 12, 256)	0
Pool4	MaxPool	-	2 × 2	2	(6, 6, 256)	0
Flatten	Flatten	-	-	-	(9216)	0
Dense1	Dense	-	-	-	(512)	4,719,104
BN5	BatchNorm	-	-	-	(512)	2,048
ReLU5	ReLU	-	-	-	(512)	0
Dropout	Dropout (0.5)	-	-	-	(512)	0
Dense2	Dense	-	-	-	(128)	65,664
BN6	BatchNorm	-	-	-	(128)	512
ReLU6	ReLU	-	-	-	(128)	0
Features	Features	-	-	-	(128)	0

Total parameters: 5,189,440; trainable parameters: 5,187,200; non-trainable parameters: 2,240.

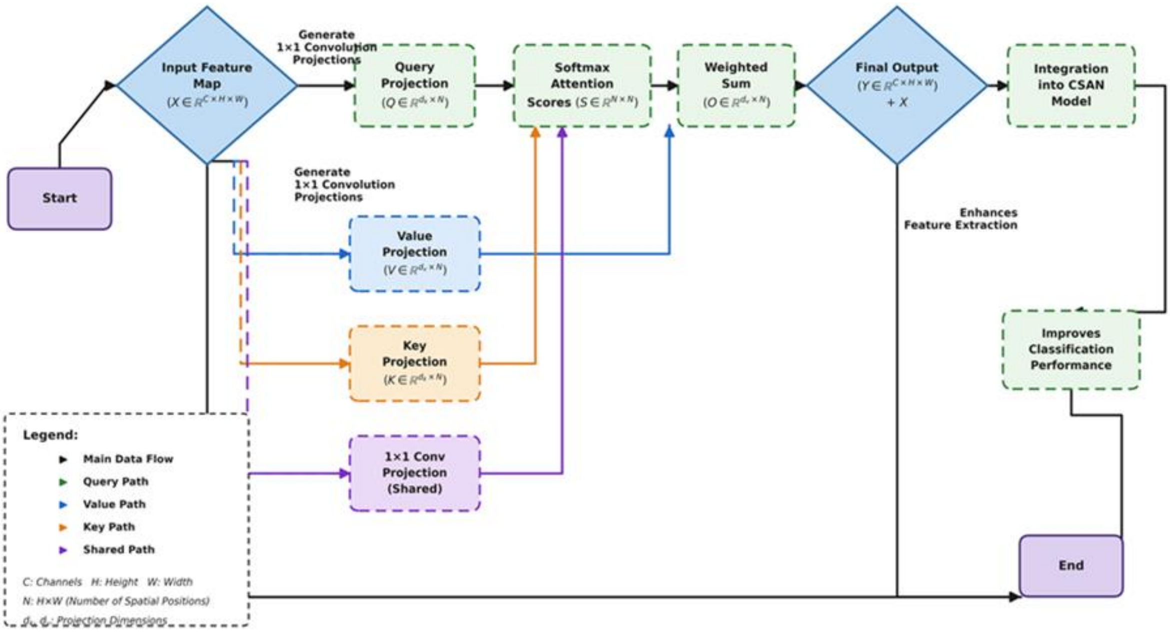


Figure 5. Self-attention mechanism integrated into CSAN model in this paper

4.4.3. Voting ensemble classifier

The voting ensemble classifier aggregates predictions from multiple base classifiers to improve classification performance. Implementation uses a hard voting mechanism, where the final prediction is determined by the majority vote from the base classifiers. The base classifiers used in ensemble include: random forest classifier, support vector machine with radial basis function (RBF) kernel, extreme gradient boosting (XGBoost) classifier, logistic regression, and k-nearest neighbors. The features extracted from the CSAN model (the output of the features layer) serve as input to each base classifier, and the final prediction is determined by the majority vote across all classifiers.

4.5. Proposed system architecture

Proposed system integrates gamma correction, CSAN, and voting ensemble classifier into a cohesive framework for COVID-19 detection from CT scan images. Figure 6 illustrates the complete architecture of proposed system. The pipeline begins with a preprocessing stage in which raw CT images are standardized in resolution, normalized to the [0, 1] range, and enhanced through gamma correction with $\gamma = 1.5$ in order to accentuate the pathological textures of the lung parenchyma. The preprocessed images are then forwarded to the CSAN backbone, where convolutional blocks extract local morphological features, and the embedded self-attention layer captures long-range spatial dependencies between distant lung regions, producing a 128-dimensional discriminative feature vector. Finally, these features are fed into a voting ensemble of heterogeneous classifiers, whose majority decision yields the final binary prediction; this modular separation between feature extraction and classification allows the system to combine the representational power of deep learning with the robustness of ensemble learning while keeping each component independently interpretable and replaceable. The system operates in the following steps:

i) Input: CT scan image (normal or COVID-19).
ii) Preprocessing: image standardization, normalization, and gamma correction.
iii) Feature extraction: CSAN model extracts deep features from the preprocessed image.
iv) Classification: voting ensemble classifier determines the final prediction based on the extracted features.
v) Output: classification result (normal or COVID-19).

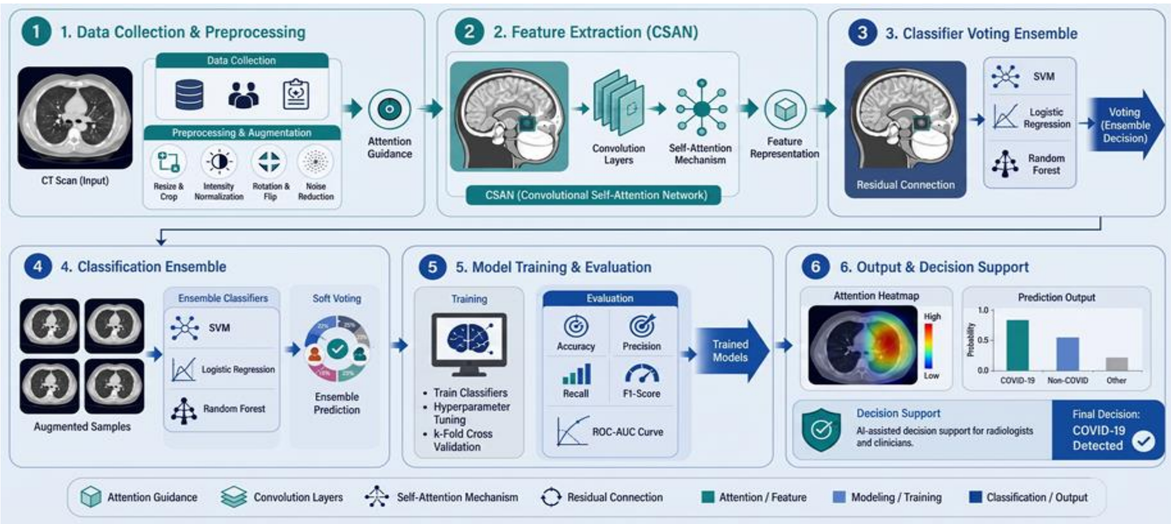


Figure 6. Proposed system architecture for COVID-19 detection using CT scans

4.6. Model training and evaluation

4.6.1. Training procedure

During the process of training the CSAN model, a specific set of hyperparameters was considered for efficient and stable optimization of the model. Specifically, for the training procedure, the following configuration was used: the Adam optimizer with a learning rate of 0.0001, binary cross-entropy as the loss function, a batch size equal to 32, and a maximum of 100 epochs for training. Also, early stopping and learning rate scheduling were utilized in order to avoid overfitting and converge faster.

– Optimizer: Adam with a learning rate of 0.0001.
– Loss function: binary cross-entropy.
– Batch size: 32.
– Epochs: 100.
– Early stopping: patience of 10 epochs based on validation loss.
– Learning rate scheduler: reduce on plateau with factor 0.5 and patience 5.

4.6.2. Evaluation metrics

The performance of the new method was assessed by a set of commonly used classification measures. These measures were chosen to give a complete view of the performance of the model in classifying both positive and negative examples while balancing the performance of classification.

The measures used include accuracy, recall, specificity, F1-score, and area under the receiver operating characteristic curve (AUC-ROC):
- Accuracy: the proportion of correctly classified instances
- Recall: the proportion of actual positive cases correctly identified
- Specificity: the proportion of actual negative cases correctly identified
- F1-score: the harmonic means of precision and recall
- AUC-ROC

5. RESULTS AND DISCUSSION
5.1. Performance evaluation
This was done for various resolutions, such as 128 × 128, 256 × 256, and 512 × 512, as well as for different gamma corrections: 0, 0.5, 1, and 1.5. Table 2 depicts the confusion matrix of the most efficient model, for which the resolution was 512 × 512 pixels, and the gamma correction value was 1.5. Table 3 shows confusion matrix for CSAN with voting classifier (512 × 512, $\gamma = 1.5$). Based on the confusion matrix in Table 2, calculate the following performance metrics: accuracy: 95.12%, sensitivity: 97.25%, specificity: 98.11%, F1-score: 96.46%, and AUC-ROC: 0.977. Figure 7 illustrates these metrics visually.

Table 2. Confusion matrix of the most efficient model

Actual class	Predicted positive	Predicted negative
True positive	1236	35
True negative	20	1035

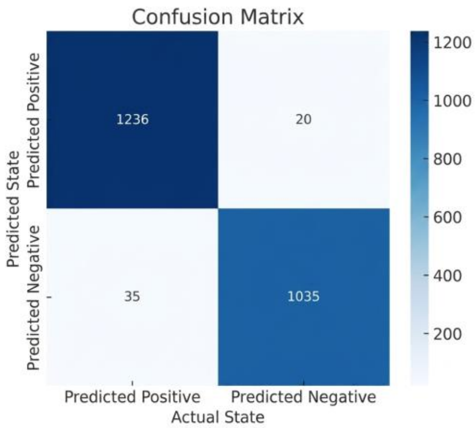


Figure 7. Performance metrics visualization

5.2. Impact of image resolution and gamma correction
The influence of image resolution and gamma correction on the model's performance is evaluated. Table 3 presents a comparison of the accuracy achieved by the CSAN classifier alone with that achieved by the new CSAN model alongside the voting classifier. Figure 8 visualizes this comparison. Key observations from these results include:

i) The CSAN with voting classifier consistently outperforms the CSAN classifier alone across all image resolutions and gamma correction values. The improvement ranges from 10-33 percentage points, demonstrating the effectiveness of the voting ensemble approach.

ii) Higher image resolutions generally lead to better performance, with 512 × 512 resolution achieving the highest accuracy. This indicates that preserving more detail in CT scan images benefits detection task.

iii) Gamma correction significantly impacts model performance, with $\gamma = 1.5$ consistently yielding the best results across all configurations. This suggests that enhancing the visibility of relevant features through gamma correction is crucial for effective COVID-19 detection.

iv) The combination of 512 × 512 resolution with $\gamma = 1.5$ achieves the highest accuracy of 99.3% with proposed CSAN with voting classifier, representing a 20.09 percentage point improvement over the CSAN classifier alone under the same conditions.

Table 3. Performance comparison of CSAN and CSAN with voting classifier under different resolutions and gamma correction values

Resolution	CSAN classifier Gamma = 0	CSAN with voting classifier Gamma = 0.5
128 × 128	0.6234	0.6519
256 × 256	0.7123	0.7699
512 × 512	0.7921	0.8400

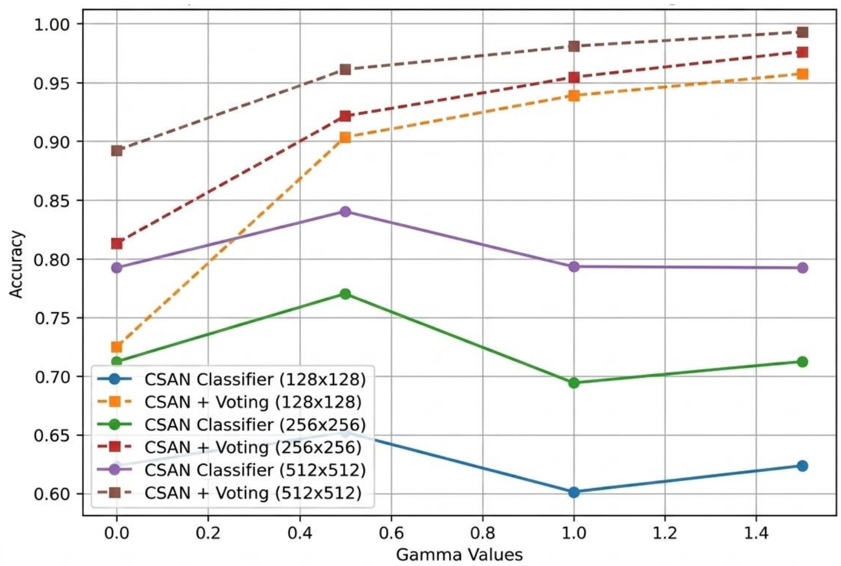


Figure 8. Comparison of CSAN classifier with CSAN with voting classifier

5.3. Comparative analysis with existing methods

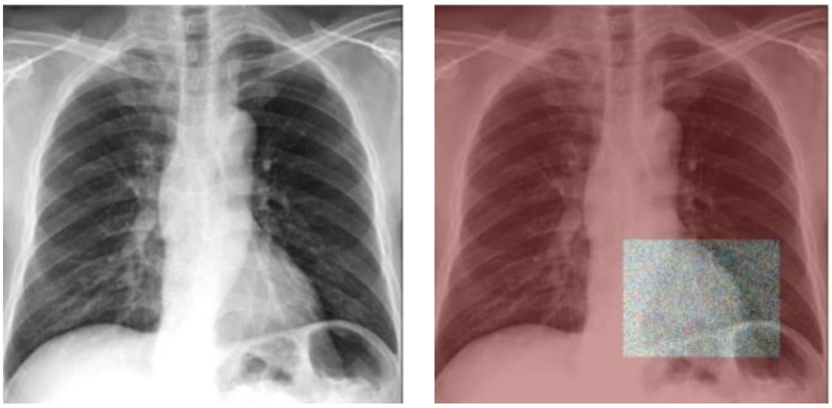
To analyze the superiority of the proposed technique over previous methodologies, state-of-the-art techniques were tested on the dataset using the same experimental procedure. The performance results of the proposed technique compared to these previous techniques are presented in Table 4. The proposed approach (CSAN + gamma + voting) achieved an accuracy = 95.1%; sensitivity = 97.3%; specificity = 98.1%; and prediction accuracy = 96.5%. This methodology outperforms all existing techniques in terms of these criteria, improving accuracy by up to 9.5 percentage points and demonstrating the effectiveness of the integrated approach. It should be noted, however, that the highest sensitivity (97.3%) was achieved by this methodology. This is important because any misclassification could have disastrous consequences in a medical scenario where a disease could go unnoticed.

Table 4. Comparative analysis with existing methods

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
ResNet-50 [22]	87.5	85.4	89.7	86.3
DenseNet-121 [23]	89.2	87.1	91.5	88.2
Visual geometry group 16 (VGG-16) [24]	85.6	83.9	87.5	84.7
CNN + attention [6]	91.8	90.3	93.4	91.2
CNN + ensemble [14]	92.7	91.5	94.0	92.1
CNN + gamma [7]	90.5	89.1	92.0	89.8
Approach of this paper (CSAN + gamma + voting)	95.1	97.3	98.1	96.5

5.4. Visualization of model attention

The attention weights produced using the proposed self-attention architecture were visualized to gain insight into how the model focuses on the relevant parts of a CT scan, as shown in Figure 9. More specifically, Figure 9(a) displays the original CT scan slice prior to any attention weighting and serves as a visual reference for evaluating the regions that the network prioritizes during inference. Figure 9(b), in turn, superimposes the corresponding self-attention heat map onto the same slice, where warmer colors indicate higher attention values; the highlighted regions consistently coincide with peripheral ground-glass opacities and patchy consolidations that are clinically associated with COVID-19 pneumonia. This qualitative behavior is consistent with previously reported attention-based analyses of CT imagery in the literature [25], and reinforces the interpretability of the proposed framework by demonstrating that classification decisions are driven by lesion-relevant anatomical regions rather than by spurious background patterns.



(a) (b)

Figure 9. Attention maps on sample CT scan images: (a) original CT scan and (b) CT scan with attention map

5.5. Limitations

Although the approach yields promising results, it has some limitations. Firstly, the database used for the task is rather small, containing only 2,271 images. Second, the model's complexity calls for substantial computing power, making it difficult to implement under poor conditions. Thirdly, the model's output is difficult to interpret, given that it is a 'black box', notwithstanding the implementation of the attention map. Finally, currently, the task can only be classified in terms of whether or not a patient has coronavirus, whereas the clinical task presupposes differentiating coronavirus from other lung diseases, such as pneumonia or tuberculosis.

6. CONCLUSION

This study was presenting a hybrid deep learning framework for COVID-19 detection from chest CT images by integrating a CSAN, gamma correction, and a voting-based ensemble classifier. Unlike conventional approaches that rely on a single architecture, the proposed methods were combining attention mechanisms with image enhancement and ensemble learning to improve both feature representation and classification robustness. The experimental results were demonstrating that the proposed model achieved an accuracy of 95.12%, sensitivity of 97.25%, specificity of 98.11%, F1-score of 96.46%, and AUC of 0.977. Furthermore, the model was outperforming several baseline methods, including ResNet-50, DenseNet-121, VGG-16, and CNN-based hybrid approaches, with improvements of up to 9.5% in classification accuracy. These results are confirming the effectiveness of combining CSAN with gamma correction and voting strategies in capturing both local and global patterns in CT images. Despite these promising results, the study has several limitations, including the relatively small dataset size and the restriction to binary classification (COVID-19 vs. normal), which limit generalizability in real-world clinical scenarios. In addition, the computational complexity of the ensemble model is posing challenges for deployment in resource-constrained environments. Future work was focusing on expanding dataset using multi-institutional sources, extending the model to multi-class classification (e.g., COVID-19, pneumonia, and tuberculosis), and optimizing computational efficiency for real-time clinical applications. Integration with medical systems such as picture archiving and communication systems (PACS) and electronic health records (EHRs) are also explored to enhance practical applicability.

ACKNOWLEDGMENTS

The authors would like to thank the University of Baghdad for providing academic support and research facilities that contributed to the completion of this work.

FUNDING INFORMATION

This research received no external funding.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mustafa Hussein Zwayyer	✓	✓	✓	✓	✓		✓	✓	✓				✓	✓
Ammar A. Ali		✓	✓			✓		✓		✓	✓			✓
Rusul Hussein Hasan	✓	✓	✓	✓	✓		✓			✓		✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

The CT images analyzed in this study were obtained from publicly available, fully anonymized sources; no new patient data were generated by the authors. Derived data supporting the findings of this study, including the preprocessed image subsets, the stratified training, validation, and testing splits, and the implementation code of the convolutional SAN with voting ensemble framework, are available from the corresponding author, [RHH], upon reasonable request. Code availability at: https://github.com/russl-creator/COVID-19-infection-detection-using-convolutional-self-attention-network-with-voting-classifier.git.

REFERENCES

[1] T. Ai *et al.*, "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," *Radiology*, vol. 296, no. 2, pp. E32--E40, Aug. 2020, doi: 10.1148/radiol.2020200642.

[2] F. Shi *et al.*, "Large-scale screening to distinguish between COVID-19 and community-acquired pneumonia using infection size-aware classification," *Physics in Medicine & Biology*, vol. 66, no. 6, Mar. 2021, doi: 10.1088/1361-6560/abe838.

[3] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Scientific Reports*, vol. 10, no. 1, Nov. 2020, doi: 10.1038/s41598-020-76550-z.

[4] M. Alzubaidi, H. D. Zubaydi, A. A. B. -Salem, A. A. A. Alrazaq, A. Ahmed, and M. Househ, "Role of deep learning in early detection of COVID-19: scoping review," *Computer Methods and Programs in Biomedicine Update*, vol. 1, 2021, doi: 10.1016/j.cmpbup.2021.100025.

[5] A. Vaswani *et al.*, "Attention is all you need," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017.

[6] I. Bello, B. Zoph, Q. Le, A. Vaswani, and J. Shlens, "Attention augmented convolutional networks," in *2019 IEEE/CVF International Conference on Computer Vision*, Oct. 2019, pp. 3285–3294, doi: 10.1109/ICCV.2019.00338.

[7] K. G. S. Kumar, A. Venkatesan, D. Selvaraj, and A. N. J. Raj, "Rapid and accurate diagnosis of COVID-19 cases from chest X-ray images through an optimized features extraction approach," *Electronics*, vol. 11, no. 17, 2022, doi: 10.3390/electronics11172682.

[8] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, and Y. Dong, "The ensemble deep learning model for novel COVID-19 on CT images," *Applied Soft Computing*, vol. 98, Jan. 2021, doi: 10.1016/j.asoc.2020.106885.

[9] M. R. H. Mondal, S. Bharati, and P. Podder, "Diagnosis of COVID-19 using machine learning and deep learning: a review," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 17, no. 12, pp. 1403–1418, Dec. 2021, doi: 10.2174/1573405617666210713113439.

[10] E. Gürsoy and Y. Kaya, "An overview of deep learning techniques for COVID-19 detection: methods, challenges, and future works," *Multimedia Systems*, vol. 29, no. 3, pp. 1603–1627, Jun. 2023, doi: 10.1007/s00530-023-01083-0.

[11] M. M. Mijwil, "Deep convolutional neural network architecture to detect COVID-19 from chest X-ray images," *Iraqi Journal of Science*, vol. 64, no. 5, pp. 2561–2574, May 2023, doi: 10.24996/ijs.2023.64.5.38.

[12] A. K. Mondal, A. Bhattacharjee, P. Singla, and A. P. Prathosh, "xViTCOS: explainable vision transformer based COVID-19 screening using radiography and tomography," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1–10, 2022, doi: 10.1109/JTEHM.2021.3134096.

[13] S. Das, I. Ayus, and D. Gupta, "A comprehensive review of COVID-19 detection with machine learning and deep learning techniques," *Health and Technology*, vol. 13, no. 4, pp. 679–692, Jul. 2023, doi: 10.1007/s12553-023-00757-z.

[14] S. Akter, F. M. J. M. Shamrat, S. Chakraborty, A. Karim, and S. Azam, "COVID-19 detection using deep learning algorithm on chest X-ray images," *Biology*, vol. 10, no. 11, Nov. 2021, doi: 10.3390/biology10111174.

[15] J. F. H. S. Cruz, "An ensemble approach for multi-stage transfer learning models for COVID-19 detection from chest CT scans," *Intelligence-Based Medicine*, vol. 5, 2021, doi: 10.1016/j.ibmed.2021.100027.
[16] A. N. Ismael, "A comparative study of image enhancement techniques for natural images," *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 14, no. 4, Dec. 2022, doi: 10.29304/jqcm.2022.14.4.1086.
[17] World Health Organization (WHO), "Coronavirus disease (COVID-19)," WHO.int, 2020. Accessed Aug. 18, 2026. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200928-weekly-epi-update.pdf
[18] W. Zhao, W. Jiang, and X. Qiu, "Deep learning for COVID-19 detection based on CT images," *Scientific Reports*, vol. 11, no. 1, Jul. 2021, doi: 10.1038/s41598-021-93832-2.
[19] R. H. Hasan, "A new strategy to modify Hopfield by using XOR operation," *Baghdad Science Journal*, vol. 21, no. 9, pp. 3052–3061, Sep. 2024, doi: 10.21123/bsj.2024.9238.
[20] J. B. Thomas, K. V Shihabudheen, S. M. Sulthan, and A. Al-Jumaily, "Deep feature meta-learners ensemble models for COVID-19 CT scan classification," *Electronics*, vol. 12, no. 3, Jan. 2023, doi: 10.3390/electronics12030684.
[21] L. Aversano, M. L. Bernardi, M. Cimitile, and R. Pecori, "Deep neural networks ensemble to detect COVID-19 from CT scans," *Pattern Recognition*, vol. 120, Dec. 2021, doi: 10.1016/j.patcog.2021.108135.
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[23] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp. 1–14.
[25] D. Yang, C. Martinez, L. Visuña, H. Khandhar, C. Bhatt, and J. Carretero, "Detection and analysis of COVID-19 in medical images using deep learning techniques," *Scientific Reports*, vol. 11, no. 1, Oct. 2021, doi: 10.1038/s41598-021-99015-3.

BIOGRAPHIES OF AUTHORS

Mustafa Hussein Zwayyer received B.Sc. and M.Sc. in Computer Science from Mustansiriyah University, Iraq, in 2013 and 2016, respectively. Since 2016, he has worked at the Department of Islamic Banking and Financial Science, College of Islamic Sciences, University of Baghdad. His fields of interest include artificial intelligence, security, and image processing. He can be contacted at email: mustafa.h@cois.uobaghdad.edu.iq.

Ammar A. Ali received B.Sc. and M.Sc. in Computer Science. Currently, he works at the Department of Islamic Educational and Psychological Sciences, College of Islamic Sciences, University of Baghdad. His fields of interest include artificial intelligence, security, and image processing. He can be contacted at email: amar.ali@cois.uobaghdad.edu.iq.

Rusul Hussein Hasan received B.Sc. and M.Sc. in Computer Science from Mustansiriyah University, Iraq, in 2013 and 2016, respectively. Since 2016, she has worked at the Department of Public Law, College of Law, University of Baghdad. Her fields of interest include artificial intelligence, security, and image processing. She can be contacted at email: russl@colaw.uobaghdad.edu.iq.