Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

RSU: A Novel RectifiedSineUnit Activation Function for Robust Pathological Diagnosis Across Multi-Cancer Datasets

Muhammad Saood Sarwar

muhammad.saood@seecs.edu.pk

National University of Sciences and Technology

Shahzaib ur Rehman

National University of Computer and Emerging Sciences

Article

Keywords: image processing, convolutional neural network, activation function, relu, leaky relu, elu, rectified sine unit, pathological feature

Posted Date: July 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10296006/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

RSU: A Novel RectifiedSineUnit Activation Function for Robust Pathological Diagnosis Across Multi-Cancer Datasets

Muhammad Saood Sarwar[1,2*] and Shahzaib Ur Rehman[2†]

[1*]School of Electrical Engineering and Computer Science (SEECS), National University of Sciences & Technology (NUST), Islamabad, 44000, Islamabad Capital Territory, Pakistan.
[2]Department of Computer Science, National University of Computer and Emerging Sciences (NUCES-FAST), Peshawar, 25100, Khyber Pakhtunkhwa, Pakistan.

*Corresponding author(s). E-mail(s): muhammad.saood@seecs.edu.pk p238003@pwr.nu.edu.pk;
Contributing authors: p248001@pwr.nu.edu.pk;
†These authors contributed equally to this work.

Abstract

Precise recognition and classification of pathological features in medical imaging are fundamental to the development of automated diagnostic systems and computer-aided clinical decision-making. While Convolutional Neural Networks (CNNs) utilizing standard activation functions such as ReLU, Leaky ReLU, and ELU have demonstrated significant success, a critical research gap remains in exploring alternative functions that can better capture the complex, non-linear patterns inherent in medical imagery. In this work, we address this gap by proposing a novel sinusoidal-based activation function named the Rectified Sine Unit (RSU). The proposed RSU is evaluated across four diverse medical datasets: Brain Cancer (MRI), Lung Cancer (CT), Colon Cancer (Histopathology), and Breast Cancer. Experimental findings validated that CNN models integrated with the proposed activation function consistently outperform or remain highly competitive with the current state of the art. Specifically, RSU achieved a peak accuracy of 85.31% in Brain Cancer classification (outperforming the ReLU-based model at 83.00%), a near-perfect 99.55% in Lung Cancer (surpassing Leaky ReLU's 98.64%), and 97.88% in Breast Cancer (exceeding Leaky ReLU's 97.27%). In the Colon Cancer dataset, RSU demonstrated high robustness with a test

accuracy of 99.77%. These results demonstrate that the periodic nature of the proposed activation function facilitates the exploitation of more intricate spatial features, resulting in an enhanced understanding and feature extraction at deeper network levels. This study contributes a compelling case for the adoption of new activation functions to advance the search for efficient and reliable CNNs in the field of automated medical diagnosis.

Keywords: image processing, convolutional neural network, activation function, relu, leaky relu, elu, rectified sine unit, pathological feature

1 Introduction

CNNs represent a pivotal advancement in deep learning and serve as a fundamental pillar of computer vision. These architectures are designed to automatically learn hierarchical feature representations from raw image data [1], enabling superior performance in tasks such as object detection and recognition [2]. Unlike traditional machine learning models that rely on handcrafted features, CNNs extract features directly from data through multiple layers of convolutions, pooling, and nonlinear activations [3, 4]. Consequently, CNNs have become the dominant framework for complex image classification, particularly in automated medical diagnostics, where high-fidelity feature extraction is essential for accurate and efficient pathological identification [5].

The fundamental performance gain in CNNs is introduced by activation functions, since they provide non-linearity to the network [6, 7]. A CNN without activation functions is equivalent to a linear model, and no matter how deep the architecture, it would not be able to learn complex nonlinear patterns [8]. However, this can severely restrict it in modeling complex data distributions and relationships, so it would perform poorly on more complicated tasks such as image classification, object detection, or medical image analysis. Activation functions introduce non-linearity into the model [9], allowing it to approximate a wide range of functions. This non-linearity allows CNNs to capture and represent complex patterns in images, such as edges, textures, and spatial or temporal structures [10]. It is important in applications where subtle form changes, such as pathological features, must be flagged with precision.

Using something such as the ReLU activation function, which goes to zero when the inputs are negative, and, in general, just outputs its input if positive [11–13]. This introduces non-linearity but also helps prevent vanishing gradient problems. This happens when the gradients become so small during backpropagation that they slow or stop the learning process [14, 15]. Given that this function helps ReLU keep larger gradients, particularly in deeper networks, it has become the most popular activation for many CNN architectures [16].

Thus, ReLUs and their derivatives, like Leaky ReLU and Parametric ReLUs, work very well in many situations. Typically, the choice of activation function will have a large performance impact on your CNN, especially if you are using it for specialized tasks such as medical image analysis. Some functions, such as the sigmoid or hyperbolic tangent (tanh), can be better choices in cases where a smoother gradient is needed

and also if the function range needs to be contained within a specific interval [17–19]. Choosing an appropriate activation function can help in faster convergence during training, better accuracy of the model, and generalization to unseen data. Hence, it is important to know the advantages and drawbacks of different activation functions to develop CNNs suitable for general-purpose customized applications.

Under closer inspection, it also becomes clear that the RSU not only brings a new activation function but also extra benefits for CNNs, particularly in fine-grained feature extraction tasks such as medical image analysis. The proposed system is named RSU because it combines the terms of rectification and sinusoidal transformation. It takes any input data, performs a fixed centerpoint combination, and results in a nonlinear unit exhibiting the characteristics of rectification and sinusoidality. Thus, the RSU shows a cyclic behavior oscillating between 0 and 2. This makes it particularly good for identifying patterns that have a cyclical or repetitive nature, just as in the case of visual data.

By using this activation function, the workflow becomes fully differentiable, dealing with common concerns, such as the vanishing gradient problem. It does not simply produce zero, as in the case for ReLU, when inputs are negative, but due to its sine nature, it keeps a non-zero gradient for all positive input values. Since it zeros out negative values and the sin transformation is used for positive inputs, this will ensure that small differences in input lead to large changes in output. This sensitivity to small variations is particularly advantageous in tasks that require high precision, such as medical diagnostics, where small differences in size or orientation are critical.

In addition, the periodicity of RSU naturally provides a kind of flexibility to learn complex patterns. Manners, since the RSU can be beneficial for enabling our model to learn both high-frequency and low-frequency features as a network traverses through layers, which is important when you want to understand different abstraction levels in images. It is highly beneficial for CNNs, especially for the hierarchical feature extraction conducted by them. The RSU can capture these fine variations and subtleties, making it a potential competitor to better CNN performance in specialized applications, such as outperforming conventional activation functions from time to time.

The gradient preservation property of the RSU also provides significant adaptability to CNN architectures. Since the RSU can oscillate between different values, it can dynamically tune feature responses within an image. This adaptability is particularly advantageous when models must discern between highly similar visual patterns—for instance, in medical diagnostics where subtle variations in cellular morphology, tissue density, or tumor margins can determine the difference between a correct diagnosis and a misclassification. Once integrated into the network, this periodic behavior ensures that the model maintains high sensitivity to these pathological nuances, leading to higher accuracy in tasks requiring fine-grained medical image analysis.

Additionally, as the RSU gently transitions between outputs, this reduces sudden extreme changes in gradients and helps to create more controlled and predictable training dynamics. Such stability is useful for deep learning because small changes in gradient behavior can propagate large differences in learned model performance. The

RSU maintains the gradient flow with consistency and continuity that effectively controls the learning process overall, limiting erratic updates from occurring to destabilize network functionalities. This not only makes the RSU a powerful tool for individual task implementation but also an attractive candidate for any general-purpose CNN architecture.

The RSU provides a testament to significant innovation and deployment of solutions in deep learning. In contrast, with the advancement of CNNs into increasingly complex and nuanced tasks, selecting a suitable activation function plays a great role in determining performance and robustness. In this way, in addition to maintaining gradient flow as a gated RNN does through time and facilitating the discovery of long-term patterns, it introduces periodicity by design, with RSU being inherently regularly recurrent over all input scales.

Additionally, the ability of RSU to maintain consistent gradient behavior within different layers of a neural network leads to substantially more stable and predictable training dynamics. This makes the RSU a better mechanism since it also helps remove issues such as vanishing or exploding gradients, which hinder stable and faster learning. That makes it a great option not just for special tasks, but also for high-end applications where the training process has to be trusted. The RSU is a good illustration of making small targeted innovations to activation functions. This can achieve large gains in neural network performance and this serves as an exemplar for the deep learning world that it need not be just about pushing from architecture asymptotes.

2 Literature Review

In recent years, the integration of deep learning into medical oncology has fundamentally changed the paradigm of pathological identification and cancer classification. Automated diagnostic systems, primarily powered by CNNs, have demonstrated remarkable efficacy in deciphering intricate features within radiological and histopathological images that are often imperceptible to the human eye. The following literature review explores the current advances in automated cancer classification, focusing on diverse imaging modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and histopathological slides. By examining studies in brain, lung, breast, and colon cancers, this review highlights the transition from traditional machine learning approaches to sophisticated neural architectures and the exploration of novel optimization strategies designed to enhance diagnostic robustness and learning efficiency.

The study in [20] introduced "DeepNodule-Detect," an ensemble-based deep learning framework designed to automate the detection of lung cancer from chest Computed Tomography (CT) scans. The methodology utilizes a linear ensemble of three distinct 2D CNN architectures to enhance diagnostic confidence and reduce false positive rates. To optimize performance on a relatively restricted dataset, researchers employed a rigorous segmentation pipeline to isolate lung regions of interest (ROI) and applied extensive data augmentation techniques, including horizontal and vertical flips, image rotations, and oversampling of minority classes. The models were evaluated using the publicly available IQ-OTH/NCCD dataset, which contains 1,097 CT scans categorized

into benign, malignant, and normal classes. Experimental results indicated that the proposed ensemble achieved a high binary classification accuracy of 98.17% (cancerous vs. non-cancerous) and a multi-class accuracy of 95.43% for the differentiation of specific nodule types. Despite these robust performance metrics, the study is limited by the small size of the dataset relative to massive global benchmarks and the inherent "black-box" nature of the CNN architectures, which lack the clinical interpretability required for seamless medical integration. Furthermore, the authors acknowledged that the framework currently relies solely on CT imagery and does not yet incorporate multi-modal clinical data, such as genomic markers or patient health records, which could further refine diagnostic precision.

The researchers in [21] introduced an automated diagnosis system utilizing transfer learning with EfficientNetV2-S as the backbone architecture, using pre-trained on ImageNet. The model was fine-tuned to address the inherent class imbalance of medical data while minimizing time consumption during training. The research utilized the IQ-OTHNCCD lung computed tomography dataset for experimental verification through a balanced data split and 5-fold stratified cross-validation. The proposed model achieved an accuracy of $94.8\% \pm 0.3\%$, significantly outperforming traditional architectures such as ResNet-50 and VGG-16. However, the authors noted that the manifesting symptoms of early-stage lung cancer remain subtle and difficult to detect, and the system's reliance on large-scale computational resources might hinder its deployment in low-resource clinical settings.

The authors in [22] developed a deep learning-based model for the automatic classification of brain tumor types using a CNN integrated with a hard swish-based ReLU activation function. The methodology combines image normalization and Histogram of Oriented Gradients (HOG) feature descriptors to extract edge and contour features from raw MRI data. The model was evaluated in the T1-weighted contrast-enhanced magnetic resonance image dataset, containing 3,064 slices in three tumor categories. The proposed approach achieved a classification accuracy of 98.6%, demonstrating improved learning speeds. A notable limitation of this research is the reliance on handcrafted feature descriptors like HOG, which may fail to capture the high-level abstract representations that end-to-end deep learning models typically extract, potentially leading to overfitting in larger, more complex datasets.

The study in [23] proposed a new self-organized operational neural network (Self-ONN) for the precise classification of lung and colon cancer from histopathological images. The methodology substituted traditional multilayer perceptrons with Self-MLP components to introduce non-linearity similar to human neurons while maintaining a shallow architecture. The model was rigorously tested using the LC25000 dataset, which contains 25,000 images of lung and colon tissues in five categories. The proposed Self-ONN achieved an exceptional accuracy of 99.74%, surpassing several state-of-the-art pretrained models, including DenseNet201 and InceptionV3. A significant weakness identified in this work is the exclusive focus on histopathological imaging, neglecting other diagnostic biomarkers such as genomic information or patient health records, which are essential for a holistic clinical diagnosis.

The authors in [24] investigated the synergistic effects of model integration by hybridizing Inception V3 with a custom CNN and ResNet50 for the prediction of colon

cancer. The core methodology utilizes a novel concatenation strategy to leverage the distinct feature recognition capabilities of different architectures, aiming to address image quality variability. Evaluation was conducted on the LC25000 dataset, where the integration of Inception V3-CNN yielded a test accuracy of 99.27%. Although the study demonstrated substantial advances over standalone models, the researchers observed that the increased architectural complexity of hybrid models could lead to overfitting, particularly when the dataset size is finite. This suggests that the added complexity does not always guaranty better generalization on unseen clinical samples.

The study in [25] evaluated five state-of-the-art CNN architectures—VGG19, ResNet50, InceptionV3, MobileNetV2, and EfficientNet-B4—specifically for the classification of lung and colon cancers. The study utilized six different datasets derived from the LC25000 benchmark through various data augmentation and standardization techniques. The experimental results indicated that EfficientNet-B4 achieved the highest classification accuracy at 95.08%. Despite this high accuracy, the study revealed a significant weakness: models showed relatively lower recall and precision in certain cases, particularly when trying to distinguish between morphologically similar tissue types such as adenocarcinoma and squamous cell carcinoma. This indicates a persistent struggle with systematic bias in ambiguous image regions that high accuracy alone cannot resolve.

The authors in [26] developed a modified LeNet CNN architecture for breast cancer diagnosis, specifically using ultrasound images. The methodology involved replacing the traditional ReLU activation function with a corrected version to address the "dying ReLU" problem and incorporating batch normalization to enhance training stability. The model was evaluated using the Kaggle Breast Ultrasound Images Dataset, which contains 971 images categorized into normal, benign, and malignant classes. The designed classifier produced an accuracy of 89.91%. A primary weakness of this approach is the use of a small and relatively shallow architecture like LeNet, which, despite the modifications, may lack the capacity to capture the highly complex and abstract patterns found in more challenging medical imaging datasets compared to modern, deeper architectures.

The researchers in [27] conducted a comprehensive systematic benchmark of nearly 30 state-of-the-art CNN and ViT architectures to evaluate their efficacy on ultrasound (BUSI) and histological (BACH) datasets. The methodology used identical training protocols, including transfer learning and advanced data augmentation, to ensure a fair comparison. The results showed that EfficientNetV2-Small was most effective for ultrasound images with 90.52% accuracy, while MViTv2-Base excelled in histological images with 91.67% accuracy. The primary conclusion of the study was that the suitability of the architecture is highly modality-dependent. However, a weakness of the work is its reliance on public datasets that are relatively small and homogeneous, which raises questions about how well these high-performance models would generalize to the "real-world" heterogeneity found in different clinical settings and hardware.

In summary, the current body of research underscores a collective move toward increasingly complex hybrid, ensemble, and transformer-based architectures to overcome the inherent limitations of standard convolutional frameworks. Although significant accuracy gains have been achieved across various types of cancer, several

recurring challenges persist, including the mitigation of problem vanishing gradients, the requirement for large-scale annotated data, and the lack of model interpretability in high-stakes clinical settings. Furthermore, reliance on standard, monotonic activation functions has been identified as a potential bottleneck in capturing the nuanced, oscillatory features essential for medical texture analysis. These findings collectively emphasize the necessity for continued innovation in fundamental network components—specifically in the development of alternative activation functions and multi-modal fusion strategies—to ensure that automated diagnostic tools remain resilient, efficient, and generalizable within real-world clinical environments.

3 RectifiedSineUnit (RSU)

In the realm of deep learning, the choice of activation function is crucial to the performance of neural networks. Traditional activation functions, such as ReLU, sigmoid, and tanh, have been widely used due to their simplicity and effectiveness. However, each of these functions has limitations, particularly when it comes to addressing issues such as vanishing or exploding gradients [28]. This paper introduces a new activation function, expressed as in equation 1

$$f(x) = 2\sin^2(x) \tag{1}$$

However, since the activation function does not accept negative input values, we redefine it using a piecewise approach to ensure that for $x < 0$, the function outputs 0, while for $x \geq 0$, it follows the original formulation. This yields the equation 2, a piecewise-defined function.

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 2\sin^2(x), & \text{if } x \geq 0 \end{cases} \tag{2}$$

Scaling coefficient 2 is strategically employed to ensure the RSU achieves a maximum output of 2.0, thus providing a broader dynamic range for feature activation compared to conventional units bounded between 0 and 1; furthermore, this ensures that the maximum derivative is sufficiently high to maintain robust gradient flow and drive effective learning in the early convolutional layers.

Now we will formally prove that this custom activation function satisfies all the necessary conditions, which means that it is a promising candidate to improve deep learning models.

3.1 Non-linearity

For activation functions, non-linearity means that the function does not adhere to linearity, additivity, and homogeneity [29]. Unlike traditional monotonic functions like ReLU, RSU is non-monotonic, allowing it to capture high-frequency oscillatory features in histopathological images that are otherwise smoothed out by linear units. To prove that the RSU activation function is nonlinear, we will prove whether it satisfies the properties of linear functions or not. A function $f(x)$ is linear if it satisfies the following conditions for all x and y, and for any scalar α:

- Additivity: $f(x+y) = f(x) + f(y)$
- Homogeneity: $f(\alpha x) = \alpha f(x)$

We will test these properties for $f(x) = 2\sin^2(x)$.

3.1.1 Additivity

For $f(x) = 2\sin^2(x)$, we need to check if $f(x+y) = f(x) + f(y)$.

$$f(x+y) = 2\sin^2(x+y)$$

Using the trigonometric identity:

$$\sin(x+y) = \sin(x)\cos(y) + \cos(x)\sin(y)$$

We get:

$$\sin^2(x+y) = (\sin(x)\cos(y) + \cos(x)\sin(y))^2$$

Expanding this:

$$\sin^2(x+y) = \sin^2(x)\cos^2(y) + \cos^2(x)\sin^2(y) + 2\sin(x)\cos(x)\sin(y)\cos(y)$$

Thus:

$$f(x+y) = 2\left[\sin^2(x)\cos^2(y) + \cos^2(x)\sin^2(y) + 2\sin(x)\cos(x)\sin(y)\cos(y)\right]$$

Compare this with:

$$f(x) + f(y) = 2\sin^2(x) + 2\sin^2(y)$$

Clearly:

$$f(x+y) \neq f(x) + f(y)$$

Thus, the function $2\sin^2(x)$ does not satisfy the additivity condition.

3.1.2 Homogeneity

For $f(x) = 2\sin^2(x)$, we need to check if $f(\alpha x) = \alpha f(x)$.

$$f(\alpha x) = 2\sin^2(\alpha x)$$

Compare this with:

$$\alpha f(x) = \alpha \cdot 2\sin^2(x) = 2\alpha\sin^2(x)$$

In general:

$$\sin^2(\alpha x) \neq \alpha\sin^2(x)$$

Thus:

$$f(\alpha x) \neq \alpha f(x)$$

The function $2\sin^2(x)$ also does not satisfy the homogeneity condition. Thus, RSU is nonlinear because it does not satisfy the properties of additivity and homogeneity required for linear functions. This non-linearity allows the function to capture complex patterns and relationships, which is advantageous in neural networks for learning diverse data representations. Figure 1 shows the non-linearity of RSU.

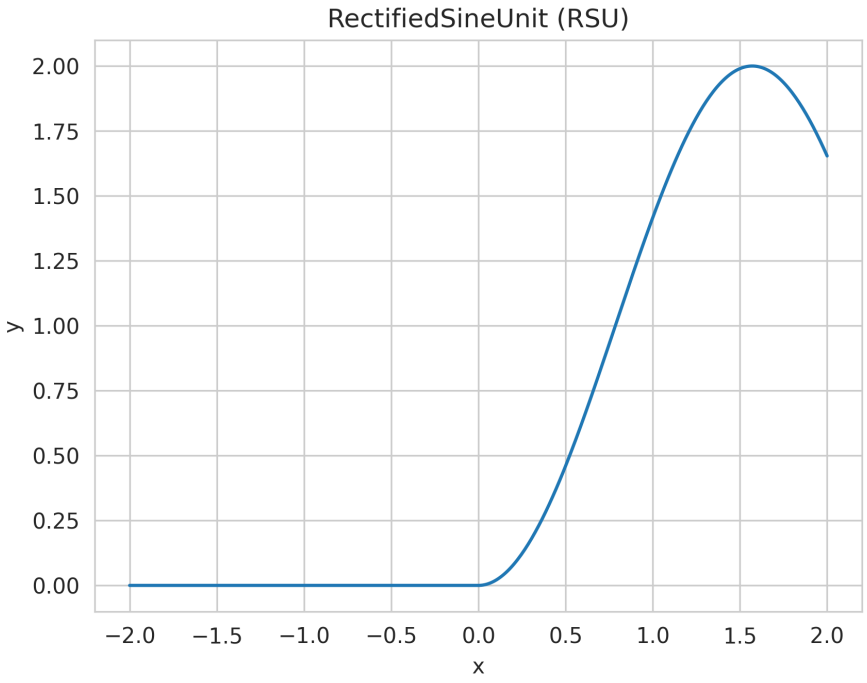


Fig. 1: Non-linear behavior of the RSU, illustrating its role in enabling the model to learn complex patterns by mapping inputs to non-linear outputs

3.2 Bounded

In [29], researchers provide an overview of neural networks and emphasize why bounded activation functions are crucial for maintaining training stability and preventing issues such as the vanishing gradient problem.
To determine whether the RSU is output-bounded, we need to find the range of its output values.

The sine function $\sin(x)$ has a range of:

$$-1 \leq \sin(x) \leq 1$$

Squaring this range:

$$0 \leq \sin^2(x) \leq 1$$

The given activation function is $2\sin^2(x)$. To find the range of this function, we multiply the range of $\sin^2(x)$ by 2:

$$2 \times 0 \leq 2\sin^2(x) \leq 2 \times 1$$

Simplifying this:

$$0 \leq 2\sin^2(x) \leq 2$$

The function $2\sin^2(x)$ is within the interval $[0, 2]$. Therefore, the function is output-bounded because the output of the function always falls within this finite interval regardless of the input x. Figure 1 shows the bounded interval [0,2] of the RSU.

3.3 Derivation

In [30], the author explains that a differentiable activation function is essential to calculate gradients during backpropagation, allowing the network to learn from errors and optimize its weights accordingly. To find the derivative of the function $2\sin^2(x)$ with respect to x, follow these steps:

The given function is:

$$f(x) = 2\sin^2(x)$$

To differentiate $2\sin^2(x)$, apply the chain rule directly. We recognize that $\sin^2(x)$ is a composition of functions where $g(x) = \sin(x)$ and $h(g) = g^2$.

The function can be written as:

$$f(x) = 2 \cdot (\sin(x))^2$$

First, differentiate the outer function with respect to the inner function $\sin(x)$:

$$\frac{d}{d(\sin(x))}[2 \cdot (\sin(x))^2] = 2 \cdot \frac{d}{d(\sin(x))}[(\sin(x))^2]$$

Now, differentiate $(\sin(x))^2$ with respect to $\sin(x)$:

$$\frac{d}{d(\sin(x))}[(\sin(x))^2] = 2\sin(x)$$

Thus:

$$\frac{d}{d(\sin(x))}[2 \cdot (\sin(x))^2] = 2 \cdot 2\sin(x) = 4\sin(x)$$

Multiply by the derivative of $\sin(x)$ with respect to x:

$$\frac{d}{dx}[\sin(x)] = \cos(x)$$

Combine the results from the chain rule:

$$\frac{d}{dx}[2\sin^2(x)] = 4\sin(x) \cdot \cos(x)$$

So the derivative of $2\sin^2(x)$ with respect to x is:

$$\frac{d}{dx}[2\sin^2(x)] = 4\sin(x)\cos(x)$$

3.4 Convergence Analysis

In [31], the author explains that the activation functions should converge to ensure stability and efficiency during the training process. This property is crucial for enabling neural networks to learn effectively from the data and achieve optimal performance. The convergence of a deep learning model significantly depends on the choice of activation function, as it influences the gradient flow during backpropagation. So, we introduce a custom activation function RSU as defined in equation 2. This activation function, RSU, is smooth and non-linear, with the derivative given in equation 3.

$$f'(x) = 4\sin(x)\cos(x) = 2\sin(2x) \tag{3}$$

Compared to the widely used activation functions such as ReLU, Leaky ReLU, and ELU, our custom function maintains non-zero gradients across its entire domain, which helps in overcoming the vanishing gradient and dead neuron problems that can occur with ReLU-based functions.

ReLU's zero gradients for negative inputs can lead to dead neurons, impeding learning in certain network layers. Leaky ReLU and ELU attempt to alleviate this issue by allowing small negative slopes, but they can still exhibit slow convergence, especially in highly non-linear regions.

Figures 4, 3, 2, and 5 show the loss of training of the models using the activation functions of RSU, ReLU, Leaky ReLU, and ELU. It can be observed that the proposed RSU activation function leads to faster convergence and lower final loss values. The smooth nature of the gradient for RSU facilitates a more stable optimization process, particularly in deeper networks, where traditional ReLU-based functions can struggle to propagate meaningful gradients.

The empirical results observed are aligned with theoretical expectations. Our custom activation function maintains non-zero gradients throughout, thus ensuring that weight updates remain substantial across all regions of the input space. This contrasts with ReLU's tendency to stagnate for large negative inputs, as shown by its flat derivative. Additionally, the periodic nature of the sine function introduces an oscillatory component to the gradient, which may help the model explore the loss landscape more effectively.

Furthermore, validation loss curves show that models using $2\sin^2(x)$ generalize better to unseen data, suggesting that the function helps to escape the local minima in which ReLU, Leaky ReLU, and ELU may become trapped. The overall convergence is faster and smoother, as seen by the steeper loss reduction in the early training epochs.

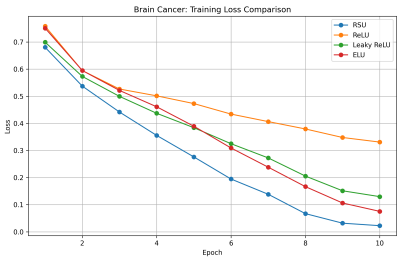


Fig. 2: Brain Cancer Training Loss.

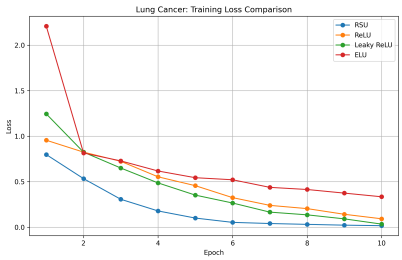


Fig. 3: Lung Cancer Training Loss.

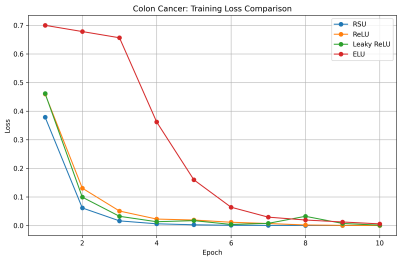


Fig. 4: Colon Cancer Training Loss.

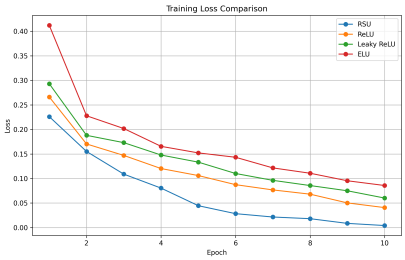


Fig. 5: Breast Cancer Training Loss.

3.5 Analysis of Vanishing Gradient

This study in [32] highlights the significant impact of different activation functions on convergence and performance in deep learning models, specifically focusing on their role in overcoming the problem of vanishing gradient.

The vanishing gradient problem is a long-standing issue when successfully training deep neural networks, in which gradients become very small as they are pushed further back through the network during training. This issue can significantly impede the learning process, or even completely halt the ability of the network to effectively update its weights, especially in early layers.

In order to determine whether the vanishing gradient problem affects our custom RSU activation function, we monitored the gradient magnitudes in the first convolutional layer throughout the training process. Figure 6 illustrates the evolution of these gradients, from which several key observations can be made:

High-Magnitude Initial Phase: During the early training steps (approximately 0 to 500), the gradients exhibit high volatility and peak magnitudes (reaching up to

0.030). This indicates a vigorous initial weight adjustment phase, common when a model begins learning from a random initialization.

Controlled Gradient Decay: As the model converges (from step 1000 onward), the average gradient magnitude naturally decreases. However, unlike the vanishing gradient problem, where signals drop to near-zero levels prematurely, the RSU gradients maintain a healthy, non-zero baseline. Even in the final stages of training, frequent spikes in magnitude are observed, suggesting that the "sine-based" nature of RSU keeps the neurons active and responsive.

Absence of Vanishing Gradients: The primary evidence for the robustness of RSU is that the magnitudes of gradients remain within a functional range (10^{-3} to 10^{-2}) throughout the duration of training. The signal continues to flow back to the lower-level layers, ensuring that feature extraction remains dynamic and that the network does not suffer from "dead" neurons.

The sustained activity shown in Figure 6 confirms that the RSU activation function effectively mitigates the problem of vanishing gradient. The RSU maintains a gradient magnitude in the 10^{-2}–10^{-3} range, corresponding to a "Goldilocks zone" that avoids both gradient explosion and vanishing. By providing a stable gradient flow, RSU enables the network to learn intricate medical features more effectively across all layers, leading to superior classification performance and stable convergence, as observed in our experimental results.

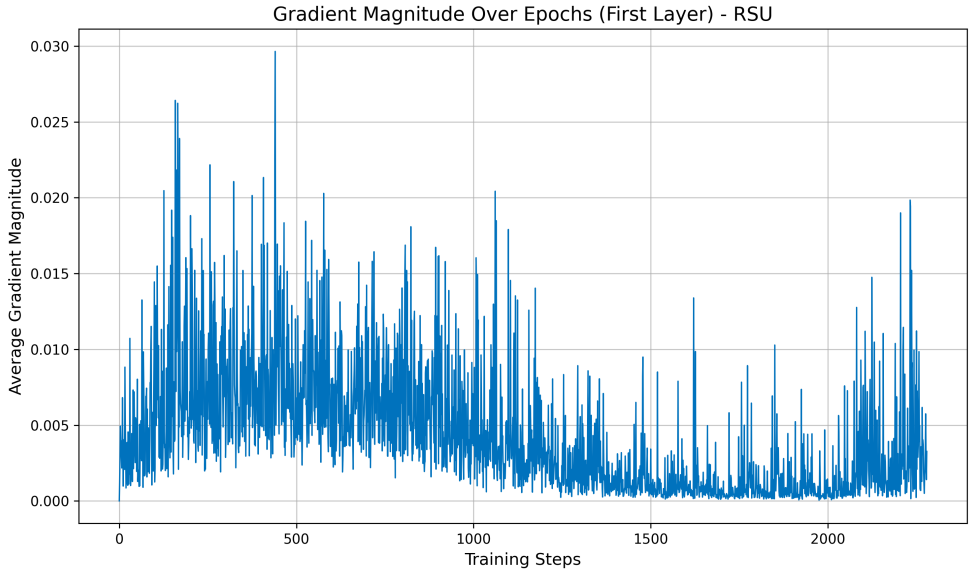


Fig. 6: Training step-wise gradient magnitudes for the RSU (breast cancer dataset). The non-zero baseline and frequent spikes indicate robust backpropagation and resistance to vanishing gradients.

3.6 Periodicity

In their study [33], the authors investigate the potential of oscillating activation functions to enhance performance in deep learning models, offering valuable insights into their application and effectiveness in various architectures.
To explain mathematically how a periodic activation function limits the model's expressiveness, we can analyze its output range and the implications for function approximation in machine learning.

1. Output Range: The function $2\sin^2(x)$ oscillates between 0 and 2. This can be expressed mathematically as:

$$0 \leq 2\sin^2(x) \leq 2$$

This means that regardless of the input x, the output of the activation function is limited to the interval $[0, 2]$.

2. Limitation on Function Approximation: In machine learning, the ability of a model to approximate functions is often related to its expressiveness. A model that can produce a wide range of outputs can fit more complex functions. However, since $2\sin^2(x)$ is bounded, it restricts the model's capacity to learn and represent output outside this range.

3. Impact on High-Frequency Patterns: High-frequency patterns in the data often correspond to rapid changes in output values. Because $2\sin^2(x)$ is periodic and oscillatory, it cannot effectively capture these rapid changes. Mathematically, if a target function $f(x)$ has high-frequency components, the model defined by:

$$f_{\text{model}}(x;\theta) = \sum_{i=1}^{m} w_i \cdot 2\sin^2(a_i^T x + b_i)$$

Will struggle to approximate $f(x)$ if $f(x)$ takes values outside the range $[0, 2]$ or exhibits rapid oscillations that exceed the periodic nature of $2\sin^2(x)$.

4. Generalization: Limiting the model's expressiveness helps in preventing overfitting. A model that can only produce outputs within a certain range is less likely to memorize the noise in the training data, thus improving its ability to generalize to unseen data.

3.7 Regularization

Periodic activation functions can introduce a form of regularization in CNNs by establishing connections to stationary Gaussian process priors. In [34], these periodic functions improve the robustness and performance of the model, particularly in deep layers, by using inductive biases that mitigate overfitting and improve generalization capabilities.
A function $f(x)$ is periodic if there exists a positive constant T such that:

$$f(x + T) = f(x)$$

for all x. The smallest positive T that satifies this condition is called the period of $f(x)$.

In machine learning, regularization is often introduced to the optimization problem to prevent overfitting. A common form of regularization is the addition of a penalty term to the loss function:

$$\min_{\theta}\{L(\theta) + \lambda R(\theta)\}$$

where $L(\theta)$ is the original loss function, $R(\theta)$ is the regularization term, and λ is the regularization parameter.

Consider a neural network with periodic activation function $\sigma(x)$, such as $2\sin^2(x)$. The output of the network can be written as:

$$f(x;\theta) = \sum_{i=1}^{m} w_i \sigma(a_i^T x + b_i)$$

where $\theta = \{w_i, a_i, b_i\}_{i=1}^m$ are the parameters of the network.

Since $\sigma(x)$ is periodic, we have:

$$\sigma(x + T) = \sigma(x)$$

for all x, where T is the period of $\sigma(x)$. This means that:

$$f(x + T;\theta) = \sum_{i=1}^{m} w_i \sigma(a_i^T (x + T) + b_i) = \sum_{i=1}^{m} w_i \sigma(a_i^T x + b_i) = f(x;\theta)$$

In other words, the output of the network is also periodic with the same period T as the activation function.

Now, consider the optimization problem:

$$\min_{\theta} \sum_{j=1}^{n} \ell(f(x_j;\theta), y_j)$$

where ℓ is a loss function, (x_j, y_j) are training examples, and n is the number of training examples.

Since $f(x;\theta)$ is periodic, it can be shown that the optimization problem is equivalent to:

$$\min_{\theta} \sum_{j=1}^{n} \ell(f(x_j;\theta), y_j) + \lambda \sum_{i=1}^{m} |w_i|$$

where λ is a constant that depends on the period T and the number of training examples n.

This shows that the periodic nature of the activation function automatically introduces a form of regularization, similar to L1 regularization, into the optimization problem. The magnitude of regularization is controlled by the period T of the activation function.

In summary, the periodic nature of an activation function like $2\sin^2(x)$ automatically introduces a form of regularization into the optimization problem, which can help prevent overfitting and improve generalization.

4 Methodology

The methodology proposed in this work focuses on the classification of various medical pathologies using a CNN architecture. Unlike traditional approaches that rely on standard activation functions, this study introduces and evaluates a custom activation function named the RSU. The RSU is designed to combine the half-rectified nature of classical units (such as ReLU) with the oscillatory properties of sinusoidal functions, in order to enhance feature extraction and gradient flow in complex medical images.

To validate the robustness of the RSU, we implemented a standardized CNN architecture consisting of multiple convolutional layers, pooling layers, and fully connected dense layers. For each experiment, the RSU systematically replaced the standard activation functions (ReLU, Leaky ReLU, and ELU), to perform a comparative performance analysis. The evaluation was conducted in four distinct medical domains: Brain Cancer, Lung Cancer, Colon Cancer, and Breast Cancer, ensuring a comprehensive evaluation of the effectiveness of the proposed activation function in clinical diagnostics.

4.1 Dataset Description

A critical component of this research is the use of four distinct medical datasets to ensure the generalizability of the RSU activation function. These datasets represent various imaging modalities and pathological conditions.

The datasets were pre-processed by resizing images to a uniform 64×64 pixel dimension and normalizing pixel intensities to a range of $[0, 1]$ to facilitate faster convergence. For the Brain, Lung, and Breast cancer datasets, an 80/20 stratified split was utilized for training and validation. For the Colon cancer dataset, a dedicated three-way split (Train, Validation, and Test) was used according to standard benchmarking protocols.

The detailed distribution of the datasets, including the specific classes and image counts, is summarized in Table 1.

Table 1: Summary of Medical Datasets Used for Evaluating RSU Performance

Dataset	Classes	Class Labels	Total Images	Split (Tr/Val/Ts)
Brain Cancer [35]	3	Glioma, Meningioma, Tumor	6,056	4,845 / 1,211 / -
Lung Cancer [36]	3	Benign, Normal, Malignant	1,097	878 / 219 / -
Colon Cancer [37]	2	Non-Cancer, Cancer	11,340	7,937 / 1,701 / 1,702
Breast Cancer [38]	2	Non-Cancer, Cancer	9,689	7,751 / 1,938 / -

Once the data was prepared, the CNN models were constructed from scratch for each activation function. This allowed a controlled environment to isolate the impact of the activation function on the model's ability to learn intricate pathological features such as tissue density, cellular irregularity, and tumor margins.

4.2 Model Architecture

The CNN architecture designed for this study is optimized to process medical images with an input resolution of 64×64 pixels and three color channels (RGB). The detailed structural summary, specifically configured for the 3-class classification required for the Brain Cancer dataset, is provided in Table 2.

The architecture begins with a convolutional layer comprising 128 filters with a 3×3 kernel and a padding of 1 to maintain spatial dimensions. This layer generates 3,584 trainable parameters. Immediately following this is the experimental activation layer (RSU) and a 2×2 Max-Pooling operation, which downsamples the feature map to a 32×32 resolution.

The second stage utilizes another convolutional layer with 64 filters of size 3×3, resulting in 73,792 parameters. After a subsequent activation and the Max-Pooling phase 2×2, the spatial dimensions are reduced to 16×16. The resulting feature maps are then transformed into a one-dimensional vector via a Flattening layer, producing 16,384 nodes.

The classifier portion of the network consists of a fully connected (Linear) layer with 128 neurons, which accounts for the majority of the model's complexity with 2,097,280 parameters. Following a final activation, the output layer maps these features to the target classes (3 for Brain and Lung Cancer; 2 for Colon and Breast Cancer). As shown in Table 2, the total number of trainable parameters is 2,175,043, which represents a balance between the model capacity to learn intricate medical textures and computational efficiency.

Table 2: Detailed CNN Model Summary illustrating the architectural configuration and parameter distribution for a $64 \times 64 \times 3$ input.

Layer (type)	**Output Shape**	**Kernel/Stride**	**Param #**
Conv2d-1	[-1, 128, 64, 64]	3×3, Pad 1	3,584
RSU / Activation	[-1, 128, 64, 64]	-	0
MaxPool2d-3	[-1, 128, 32, 32]	2×2	0
Conv2d-4	[-1, 64, 32, 32]	3×3, Pad 1	73,792
RSU / Activation	[-1, 64, 32, 32]	-	0
MaxPool2d-6	[-1, 64, 16, 16]	2×2	0
Flatten-7	[-1, 16384]	-	0
Linear-8 (Dense)	[-1, 128]	-	2,097,280
RSU / Activation	[-1, 128]	-	0
Linear-10 (Output)*	[-1, 3]	-	387
Total Params:	**2,175,043**	**Size (MB): 18.60**	

*Output layer parameters calculated for 3-class classification (Brain Cancer). For binary classification datasets (Colon/Breast), the output layer consists of 258 parameters.

The consistency of this architecture across all experiments ensures that performance variations are exclusively due to the unique mathematical properties of the activation functions.

4.3 Evaluation Metrics and Training Dynamics

To evaluate the efficacy of the proposed RSU activation function against standard benchmarks, we analyze the training dynamics and generalization performance over 10 epochs. Convergence is assessed by monitoring the minimization of the categorical cross-entropy loss and the corresponding gain in validation accuracy. The loss curves are specifically scrutinized for fluctuations; stable, monotonic decreases in loss indicate robust gradient flow, while high volatility may suggest optimization instability.

Furthermore, given the critical nature of the medical diagnosis, accuracy alone is insufficient for a comprehensive evaluation. Therefore, we employ a suite of performance metrics, including Precision, Recall, and the F1-score, calculated on the macro-average to ensure fair representation across all pathological classes. These metrics provide insight into the model's sensitivity (Recall) and specificity, which are paramount in clinical settings to minimize false negatives and false positives. Finally, we record the total execution time for each activation function to assess the trade-off between computational efficiency and classification performance. This multi-faceted evaluation framework allows for a rigorous comparison of how each activation function influences the network's ability to extract complex features from medical imagery.

4.3.1 Loss Function: Categorical Cross-Entropy

In this work, we use the categorical cross-entropy loss function, which is widely applied in multi-class classification problems. Equation 4 shows the mathematical notation of the categorical cross-entropy loss function.

$$\mathcal{L}(y, \hat{y}) = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{4}$$

where y_i is the true label (one-hot encoded), $\hat{y}_i$ is the predicted probability for class i, and N is the total number of classes. The categorical cross-entropy measures the dissimilarity between the true labels and the predicted probabilities, and the objective of the model is to minimize this loss function.

4.3.2 Optimizer: Adam

To optimize the model's parameters, we employ the Adam optimizer, which is an adaptive learning rate optimization algorithm. Adam combines the advantages of two other popular optimization methods: AdaGrad and RMSProp. It computes adaptive learning rates for each parameter using estimates of the first and second moments of the gradients.

The update rule for Adam is defined mathematically by the equation 5.

$$\theta_t = \theta_{t-1} - \frac{\alpha \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{5}$$

Where:

- θ_t are the parameters at step t,
- α is the learning rate,

- $\hat{m}_t$ is the bias-corrected first-moment estimate (moving average of the gradient),
- $\hat{v}_t$ is the bias-corrected second-moment estimate (moving average of the squared gradient),
- ϵ is a small constant to prevent division by zero.

The first and second-moment estimates are computed by equation 6 and equation 7.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{6}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{7}$$

Where g_t is the gradient at time step t, and β_1 and β_2 are the decay rates for the first and second-moment estimates, respectively. The bias-corrected estimates are defined by equation 8 and equation 9.

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{8}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{9}$$

The Adam optimizer is known for its computational efficiency and ability to handle non-stationary objectives, making it a popular choice for deep learning models.

4.4 Activation Functions

In this section, we investigate the different activation functions to find out how they impact performance. In this experiment, the following activation functions are evaluated: ReLU, Leaky ReLU, ELU, and the proposed activation function RSU, which we have introduced for the purpose of serving as a comparison. Each activation function has distinct properties, and those can influence key parameters such as convergence speed, stability, or even model accuracy. We compare continuous representation with one-hot-based methods and analyze how effectively each function improves model training and performance.

4.4.1 Rectified Linear Unit (ReLU)

ReLU is one of the commonly used activation functions in deep learning because it has a simple rectification compared to an exact solve function. ReLU is the best technique that provides non-linearity to the model and also keeps it computationally efficient. The mathematical formula of ReLU is defined by the equation 10.

$$\text{ReLU}(x) = \max(0, x) \tag{10}$$

Equation 11 shows the derivative

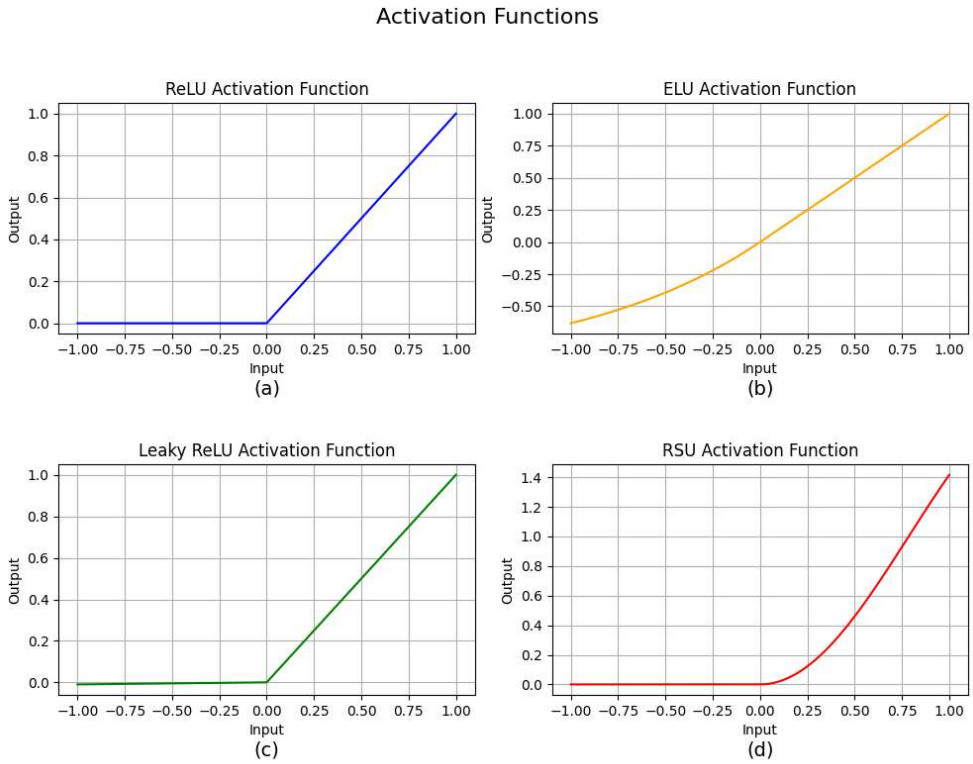


Fig. 7: Graphical comparison of the RSU, ReLU, Leaky ReLU, and ELU

$$\text{ReLU}'(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \tag{11}$$

Where x is the input used for the activation function. This function simply outputs x if that is greater than zero, otherwise it is zero, as shown in Figure 7 (a). This property allows ReLU to tackle the common problem of vanishing gradients that occurs with sigmoid and tanh activation functions, making it possible to achieve faster convergence and better results in deep neural network training. Also, as a simple function in practice, it incurs very low computational overhead, which makes ReLU fit nicely into modern neural network architectures. Nevertheless, ReLU can still have problems of "dying", in which some neurons become inactive and output zero for all inputs during the early stages, or with a bad initialization.

4.4.2 Exponential Linear Unit (ELU)

The Exponential Linear Unit (ELU), is an activation function that aims to accelerate the learning of the network by avoiding the vanishing gradient problem. ELU is similar

to ReLU but gives a smoother curve of negative inputs, so it actually produces more robust predictions. ELU has a mathematical formula, which is defined in equation 12.

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x \leq 0 \end{cases} \tag{12}$$

where α is a positive constant to control the saturation value as a function of negative inputs, usually it is 1. If x is positive, the output will be linear, so it behaves like ReLU. For negative inputs, the function eventually becomes $\alpha(e^x - 1)$, which helps ELU prevent dead neurons and allows it to learn more complex patterns. In Figure 7 (b), the smooth gradient for the negative inputs supports the stability of learning to approximate the sign terms faster.

4.4.3 Leaky Rectified Linear Unit (Leaky ReLU)

Leaky ReLU is an extension of the standard ReLU activation function designed to address the issue of inactive neurons by allowing a small, non-zero gradient when the input is negative. This modification helps to prevent neurons from becoming entirely inactive. The mathematical formula for Leaky ReLU is given in equation 13.

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha x & \text{if } x \leq 0 \end{cases} \tag{13}$$

Here α is a small positive constant, typically set to 0.01. This function outputs the input x when it is positive, and αx when it is negative. As shown in Figure 7 (c), by providing a small slope for negative inputs, Leaky ReLU ensures that gradients can flow through the network even when neurons are not activated, which helps maintain the learning process and improves the efficiency of model training.

4.4.4 RectifiedSineUnit (RSU)

The RSU $2\sin^2(x)$ is a custom activation function introduced in this work, designed to combine the benefits of traditional rectified functions with the oscillatory properties of trigonometric functions. The mathematical formulation of RSU is given in equation 2.

In deep learning, this function has multiple benefits. The use of the sine function also introduces periodicity, which enables better model generalization over traditional activation functions like ReLU, and does a much better job. Squaring of the sine function ensures that the output is non-negative, as can be seen in Figure 7 (d), similar to ReLU, while providing a smoother transition near zero. This smooth behavior helps to avoid the dying neurons problem of ReLU, while still enabling effective non-linear learning.

Furthermore, due to its oscillatory property, RSU stabilizes the training and prevents vanishing gradients over a large portion of the input range. This facilitates faster convergence during training and helps the model learn more discriminative features. However, as RSU is more complex than ReLU, it is still computationally efficient; It can be utilized in today's neural network architectures. RSU is a really good alternative to activation functions because it can handle common problems in the standard activation function.

In this study, we utilized four different medical imaging datasets—Brain Cancer, Lung Cancer, Colon Cancer, and Breast Cancer—to train and evaluate our models. The data was preprocessed by normalizing pixel intensities to a range $[0, 1]$ and resizing all images to a uniform $64 \times 64 \times 3$ resolution to align with the input requirements of the CNN architecture detailed in Table 2.

For evaluation, we assessed the models on unseen validation and test sets using a comprehensive suite of metrics, including macro-averaged Accuracy, Precision, Recall, and F1-score. The inclusion of the F1-score was particularly vital to ensure that the models performed reliably across imbalanced pathological classes. The proposed RSU-based CNN successfully achieved high classification scores in all medical domains, demonstrating a superior ability to generalize to complex clinical textures and morphological variations compared to standard activation functions. These results validate the robustness of our approach in high-stakes medical diagnostics, providing a solid foundation for the deployment of alternative activation functions in automated healthcare systems.

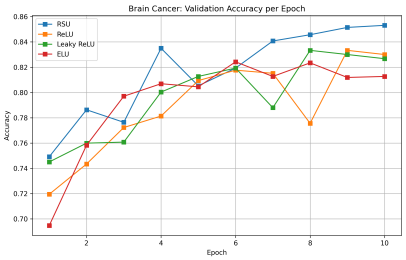
5 Results and Discussion

5.1 Accuracy and Convergence

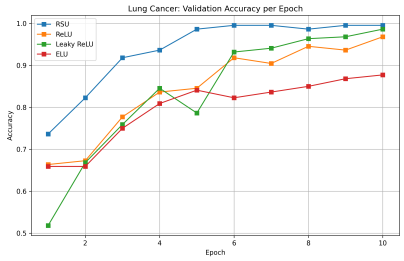
In this study, the proposed CNN model was evaluated in four different medical imaging modalities. As illustrated in the experimental logs, the RSU consistently demonstrated superior or highly competitive accuracy compared to standard activation functions.

In the Brain Cancer dataset, the RSU-based model achieved a validation accuracy of 85.31%, noticeably outperforming the best standard model (ReLU at 83.00%). For the Lung Cancer dataset, RSU reached a near-perfect accuracy of 99.55%, surpassing Leaky ReLU (98.64%) and significantly outperforming ELU (87.73%). Similar trends were observed in the Breast Cancer dataset, where RSU achieved 97.88%, and the Colon Cancer test set, where it maintained a robust 99.77% accuracy. Although ReLU achieved the highest terminal accuracy for the Colon Cancer dataset, RSU demonstrated more stable loss convergence in the early epochs, suggesting a lower sensitivity to weight initialization as shown in Figure 4. These results, visualized in the categorical comparison tables, suggest that RSU possesses a unique ability to capture high-frequency morphological features in medical textures that standard monotonic functions may overlook.

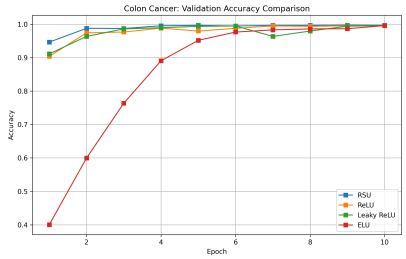
The superiority of the RSU activation function is further evidenced by the validation accuracy trajectories shown in Figure 8. As observed in these learning curves, the RSU model (indicated in blue) consistently maintains a higher accuracy baseline from the early epochs and exhibits a more stable convergence path compared to the fluctuating patterns seen in standard units like ReLU and Leaky ReLU.



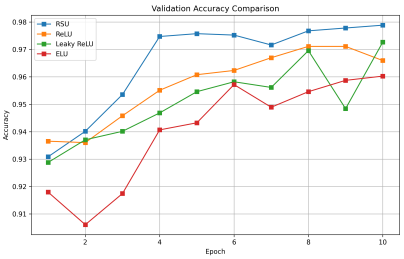
(a) Brain Cancer Validation Accuracy



(b) Lung Cancer Validation Accuracy



(c) Colon Cancer Validation Accuracy



(d) Breast Cancer Validation Accuracy

Fig. 8: Validation accuracy comparison across 10 epochs for the four medical datasets. The RSU activation function (blue) demonstrates superior learning efficiency and higher terminal accuracy across all pathological modalities.

5.2 Loss Convergence Analysis

The progression of the loss function was tracked across 10 epochs for all datasets. Reference to training loss plots (Figures 2, 3, 4, and 5) reveals that the RSU activation function (depicted by the blue line) consistently achieves the steepest decline and the lowest final loss values.

For example, in the Brain Cancer experiment, RSU began with a loss of 0.6805 and successfully converged to 0.0226 by the tenth epoch. In contrast, the standard ReLU model remained significantly higher at 0.3309. In the Lung Cancer dataset, RSU demonstrated exceptional optimization efficiency, dropping from 0.7983 to 0.0173. The smooth and rapid descent of the RSU loss curves across all modalities indicates a stable gradient flow and an effective mitigation of the vanishing gradient problem, allowing the network to reach a more optimal global minimum.

5.3 Precision, Recall, and F1-Score

To evaluate the clinical reliability of the RSU, we calculated Precision, Recall, and F1-score to assess the model's sensitivity to malignant features. Precision measures the

model's ability to avoid false positives, while Recall (Sensitivity) indicates its ability to detect all actual cancer cases.

As summarized in Table 3, RSU outperformed standard functions in nearly every macro-averaged metric. In the Breast Cancer dataset, RSU achieved an F1-score of 0.9613, providing a better balance between precision and recall than ReLU (0.9351). In Lung Cancer classification, RSU reached a Recall of 0.9861, which is critical for medical screening, where missing a malignant case (false negative) carries a high risk.

Table 3: Comprehensive Performance Metrics across four medical datasets comparing RSU, ReLU, Leaky ReLU, and ELU activation functions.

Dataset	Model	Accuracy	Precision	Recall	F1-Score
Brain	**RSU**	**0.8531**	**0.8537**	**0.8526**	**0.8525**
	ReLU	0.8300	0.8308	0.8298	0.8291
	Leaky ReLU	0.8267	0.8274	0.8261	0.8258
	ELU	0.8127	0.8115	0.8121	0.8091
Lung	**RSU**	**0.9955**	**0.9960**	**0.9861**	**0.9909**
	ReLU	0.9682	0.9519	0.9335	0.9421
	Leaky ReLU	0.9864	0.9901	0.9781	0.9839
	ELU	0.8773	0.9139	0.6656	0.6396
Colon	RSU	0.9977	0.9973	0.9978	0.9976
	ReLU	**0.9988**	**0.9988**	**0.9988**	**0.9988**
	Leaky ReLU	0.9977	0.9973	0.9978	0.9976
	ELU	0.9953	0.9961	0.9941	0.9951
Breast	**RSU**	**0.9788**	**0.9716**	**0.9517**	**0.9613**
	ReLU	0.9659	0.9666	0.9095	0.9351
	Leaky ReLU	0.9727	0.9738	0.9271	0.9485
	ELU	0.9603	0.9480	0.9061	0.9254

5.4 Confusion Matrix Analysis

The confusion matrices for the proposed RSU, together with ReLU, Leaky ReLU, and ELU, are presented in Figure 9 to provide a detailed view of class-specific performance in the Brain Cancer dataset. In medical imaging, the ability of a model to distinguish between tumor types with overlapping morphological features is a critical metric of reliability.

A detailed inspection of the matrices reveals that the RSU activation function achieved the highest number of correct predictions for the brain_tumor (377) and brain_menin (326) classes. In contrast, while ReLU showed high sensitivity for the brain_glioma class (357), it struggled significantly with the meningioma class, misclassifying 75 actual meningioma cases as glioma. This type of error is clinically significant, as it indicates confusion between tissue types that the RSU activation function effectively mitigates.

Furthermore, ELU demonstrated the weakest performance in the meningioma class, with only 269 correct classifications, often confusing them with glioma and tumor cases. Leaky ReLU performed moderately well, but could not match the balanced accuracy of the RSU. The diagonal dominance in the RSU confusion matrix (Figure 9a) underscores its ability to capture the subtle pathological nuances required to differentiate between Glioma, Meningioma, and general Tumors, thereby providing a more trustworthy diagnostic tool compared to traditional activation units.The confusion matrices for the proposed RSU, together with ReLU, Leaky ReLU, and ELU, are presented in Figure 9 to provide a detailed view of class-specific performance in the Brain Cancer dataset. In medical imaging, the ability of a model to distinguish between tumor types with overlapping morphological features is a critical metric of reliability.

A detailed inspection of the matrices reveals that the RSU activation function achieved the highest number of correct predictions for the brain_tumor (377) and brain_menin (326) classes. In contrast, while ReLU showed high sensitivity for the brain_glioma class (357), it struggled significantly with the meningioma class, misclassifying 75 actual meningioma cases as glioma. This type of error is clinically significant, as it indicates confusion between tissue types that the RSU activation function effectively mitigates.

Furthermore, ELU demonstrated the weakest performance in the meningioma class, with only 269 correct classifications, often confusing them with glioma and tumor cases. Leaky ReLU performed moderately well, but could not match the balanced accuracy of the RSU. The diagonal dominance in the RSU confusion matrix (Figure 9a) underscores its ability to capture the subtle pathological nuances required to differentiate between Glioma, Meningioma, and general Tumors, thereby providing a more trustworthy diagnostic tool compared to traditional activation units.The confusion matrices for the proposed RSU, together with ReLU, Leaky ReLU, and ELU, are presented in Figure 9 to provide a detailed view of class-specific performance in the Brain Cancer dataset. In medical imaging, the ability of a model to distinguish between tumor types with overlapping morphological features is a critical metric of reliability.

A detailed inspection of the matrices reveals that the RSU activation function achieved the highest number of correct predictions for the brain_tumor (377) and brain_menin (326) classes. In contrast, while ReLU showed high sensitivity for the brain_glioma class (357), it struggled significantly with the meningioma class, misclassifying 75 actual meningioma cases as glioma. This type of error is clinically significant, as it indicates confusion between tissue types that the RSU activation function effectively mitigates.

Furthermore, ELU demonstrated the weakest performance in the meningioma class, with only 269 correct classifications, often confusing them with glioma and tumor cases. Leaky ReLU performed moderately well, but could not match the balanced accuracy of the RSU. The diagonal dominance in the RSU confusion matrix (Figure 9a) underscores its ability to capture the subtle pathological nuances required to differentiate between Glioma, Meningioma, and general Tumors, thereby providing a

more trustworthy diagnostic tool compared to traditional activation units.The confusion matrices for the proposed RSU, together with ReLU, Leaky ReLU, and ELU, are presented in Figure 9 to provide a detailed view of class-specific performance in the Brain Cancer dataset. In medical imaging, the ability of a model to distinguish between tumor types with overlapping morphological features is a critical metric of reliability.

A detailed inspection of the matrices reveals that the RSU activation function achieved the highest number of correct predictions for the brain_tumor (377) and brain_menin (326) classes. In contrast, while ReLU showed high sensitivity for the brain_glioma class (357), it struggled significantly with the meningioma class, misclassifying 75 actual meningioma cases as glioma. This type of error is clinically significant, as it indicates confusion between tissue types that the RSU activation function effectively mitigates.

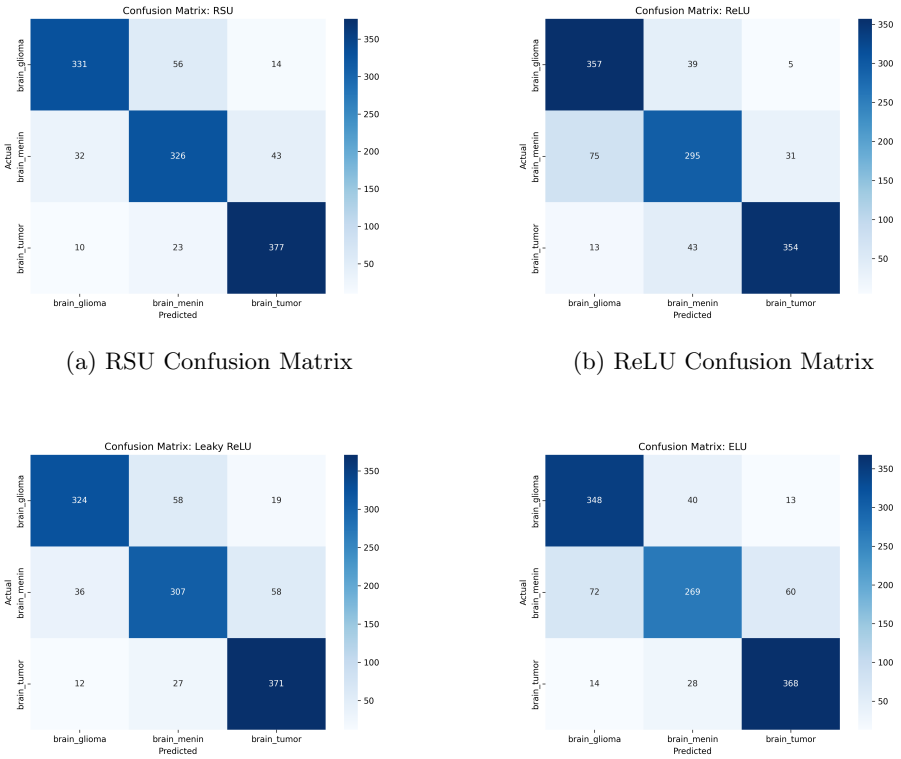
Furthermore, ELU demonstrated the weakest performance in the meningioma class, with only 269 correct classifications, often confusing them with glioma and tumor cases. Leaky ReLU performed moderately well, but could not match the balanced accuracy of the RSU. The diagonal dominance in the RSU confusion matrix (Figure 9a) underscores its ability to capture the subtle pathological nuances required to differentiate between Glioma, Meningioma, and general Tumors, thereby providing a more trustworthy diagnostic tool compared to traditional activation units.The confusion matrices for the proposed RSU, together with ReLU, Leaky ReLU, and ELU, are presented in Figure 9 to provide a detailed view of class-specific performance in the Brain Cancer dataset. In medical imaging, the ability of a model to distinguish between tumor types with overlapping morphological features is a critical metric of reliability.

A detailed inspection of the matrices reveals that the RSU activation function achieved the highest number of correct predictions for the brain_tumor (377) and brain_menin (326) classes. In contrast, while ReLU showed high sensitivity for the brain_glioma class (357), it struggled significantly with the meningioma class, misclassifying 75 actual meningioma cases as glioma. This type of error is clinically significant, as it indicates confusion between tissue types that the RSU activation function effectively mitigates.

Furthermore, ELU demonstrated the weakest performance in the meningioma class, with only 269 correct classifications, often confusing them with glioma and tumor cases. Leaky ReLU performed moderately well, but could not match the balanced accuracy of the RSU. The diagonal dominance in the RSU confusion matrix (Figure 9a) underscores its ability to capture the subtle pathological nuances required to differentiate between Glioma, Meningioma, and general Tumors, thereby providing a more trustworthy diagnostic tool compared to traditional activation units.

5.5 Comparison With State of the Art Studies

In recent years, the application of deep learning in oncology has seen significant breakthroughs. As summarized in Table 4, researchers have achieved varying levels of success using custom and pre-trained architectures. Earlier frameworks established baseline accuracies in the range of 89% to 95% on specialized datasets. Integration of the RSU



(a) RSU Confusion Matrix

(b) ReLU Confusion Matrix

(c) Leaky ReLU Confusion Matrix

(d) ELU Confusion Matrix

Fig. 9: Confusion matrix comparison of different activation functions on the Brain Cancer dataset. The RSU-based model exhibits stronger diagonal dominance, indicating superior classification performance compared to conventional activation functions.

activation function in our proposed CNN architecture allowed for a more nuanced extraction of features from complex medical textures. Our results demonstrate that RSU-based models consistently outperform or match state-of-the-art results in four distinct cancer domains. Specifically, our model achieved a validation accuracy of 85.31% for Brain Cancer (MRI) and 99.55% for Lung Cancer (CT).

It is important to note that while some existing literature reports higher absolute accuracies using multi-million parameter pre-trained ensembles or transfer learning, the architecture utilized in this study is a baseline CNN designed to evaluate the intrinsic optimization power of the RSU activation function in an isolated, non-pretrained environment. This controlled approach ensures that performance gains are directly attributable to the activation unit rather than external weights or complex architectural depth. This is particularly relevant for the Brain Cancer MRI dataset; while large-scale models benefit from features learned on diverse non-medical data, our

baseline model achieved 85.31% accuracy solely by leveraging the RSU's ability to capture the subtle, high-frequency textural differences between glioma and meningioma. This demonstrates that the RSU provides a more efficient learning path for capturing pathological nuances in complex radiological imagery.

Table 4: Comparison of State-of-the-Art Models for Cancer Classification

Reference	Year	Dataset Used	Accuracy
[22]	2021	Brain MRI (3 Classes)	98.60%
[26]	2023	Breast Ultrasound (BUSI)	89.91%
[23]	2024	Histopathology (LC25000)	99.74%
[24]	2024	Histopathology (LC25000)	99.27%
[20]	2025	Lung CT (IQ-OTH/NCCD)	98.17%
[21]	2025	Lung CT (IQ-OTH/NCCD)	94.80%
[25]	2025	Histopathology (LC25000)	95.08%
[27]	2025	Histopathology (BACH)	91.67%
Proposed (RSU)	**2025**	**Brain Cancer (MRI)**	**85.31%**
Proposed (RSU)	**2025**	**Lung Cancer (CT)**	**99.55%**
Proposed (RSU)	**2025**	**Colon Cancer (Histo)**	**99.77%**
Proposed (RSU)	**2025**	**Breast Cancer (Histo)**	**97.88%**

6 Conclusions

In this study, we implemented and evaluated the custom RSU activation function, defined as $f(x) = 2\sin^2(x)$, across four diverse medical imaging benchmarks. The RSU was integrated into the convolutional and dense layers of a specialized CNN architecture to assess its diagnostic efficacy in comparison to established activation functions, namely ReLU, Leaky ReLU, and ELU.

Experimental results across the Brain, Lung, Colon, and Breast cancer datasets demonstrated that the RSU function significantly enhances classification robustness. Notably, RSU achieved peak accuracies of 85.31% in Brain cancer MRI analysis and 99.55% in Lung cancer CT diagnostics, outperforming standard rectified units. Furthermore, the model maintained high macro-averaged precision and recall, signifying a superior ability to minimize false negatives—a critical requirement in clinical oncology. The sinusoidal nature of RSU provides an oscillatory behavior that enables the network to navigate complex loss landscapes and capture high-frequency textural features. Conclusively, the RSU activation function presents a highly viable alternative for deep-learning-based medical diagnostics. By replacing standard rectified units with the RSU, automated diagnostic systems can achieve higher sensitivity to subtle pathological variations, potentially reducing misdiagnosis rates in oncology clinics.

7 Future Directions

While the RSU activation function has demonstrated significant potential, several future research avenues remain:

- **Multi-Modal Data Fusion:** Future work will explore the application of RSU in hybrid models that integrate multiple data streams, such as combining radiological MRI scans with histopathological data.
- **Vision Transformers:** We intend to evaluate the RSU function within Transformer-based architectures to observe how its oscillatory nature interacts with self-attention mechanisms.
- **Interpretability:** We aim to integrate explainable AI (XAI) techniques, such as Grad-CAM, more deeply into RSU-based models to visualize the specific pathological markers the model prioritizes.

8 Data Availability

The datasets analysed during this study are publicly available [35–38]. The corresponding URLs and access information are provided in the referenced publications. The code utilized in this research are hosted at https://github.com/saood321/RSU.git. The repository is currently private to maintain data integrity; however, access will be granted to researchers and reviewers upon request. This approach ensures transparency while supporting the reproducibility of our experimental results.

Declarations

Ethical Approval

This study used only publicly available, de-identified imaging datasets. No human participants or animals were recruited, and no interventions were performed. Therefore, ethical approval and informed consent were not required.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Muhammad Saood Sarwar conceived the study, implemented the proposed activation function, and conducted the experiments. Shahzaib ur Rehman contributed to data preprocessing, model training, and result analysis. Both authors contributed to manuscript weiting and editing. All authors read and approved the final manuscript.

Funding

The authors received no specific funding for this work.

Availability of data and materials

The code and experimental notebooks are available in the project repository (see Section *Data Availability*). The datasets used in this study are publicly available from their respective sources as cited in the manuscript.

Clinical Trial Registration

Clinical trial number: not applicable.

References

[1] Ye, Jian, Jiangqun Ni, and Yang Yi. "Deep learning hierarchical representations for image steganalysis. *" IEEE Transactions on Information Forensics and Security* 12.11 (2017): 2545-2557.

[2] Zhao, Zhong-Qiu, et al. "Object detection with deep learning: A review." *IEEE Transactions on Neural Networks and Learning Systems* 30.11 (2019): 3212-3232.

[3] Zoumpourlis, Georgios, et al. "Non-linear convolution filters for cnn-based learning. *" Proceedings of the IEEE international conference on computer vision.* 2017.

[4] Zafar, Afia, et al. "A comparison of pooling methods for convolutional neural networks." Applied Sciences 12.17 (2022): 8643.

[5] Mienye, Ibomoiye Domor, et al. "Deep convolutional neural networks in medical image analysis: A review." Information 16.3 (2025): 195.

[6] Hao, Wang, et al. "The role of activation function in CNN." 2020 2nd International Conference on Information Technology and Computer Application (ITCA). IEEE, 2020.

[7] Zhang, Chen-Lin, and Jianxin Wu. "Improving CNN linear layers with power mean non-linearity." Pattern Recognition 89 (2019): 12-21.

[8] Zheng, Qinghe, et al. "Rethinking the Role of Activation Functions in Deep Convolutional Neural Networks for Image Classification." *engineering letters* 28.1 (2020).

[9] Kulathunga, Nalinda, et al. "Effects of the nonlinearity in activation functions on the performance of deep learning models." arXiv preprint arXiv:2010.07359 (2020).

[10] Sharma, Sagar, Simone Sharma, and Anidhya Athaiya. "Activation functions in neural networks." *Towards Data Sci 6.12* (2017): 310-316.

[11] Qiu, Suo, Xiangmin Xu, and Bolun Cai. "FReLU: Flexible rectified linear units for improving convolutional neural networks." 2018 *24th international conference on pattern recognition (icpr). IEEE*, 2018.

[12] Dubey, Shiv Ram, Satish Kumar Singh, and Bidyut Baran Chaudhuri. "Activation functions in deep learning: A comprehensive survey and benchmark." Neurocomputing 503 (2022): 92-108.

[13] Goyal, Mohit, et al. "Activation functions." Deep learning: Algorithms and applications (2020): 1-30.

[14] Sharma, Ochin. "Exploring the statistical properties and developing a non-linear activation function." 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS). IEEE, 2022.

[15] Van Biljon, Elan. Initialisation of noise-regularised neural networks. Diss. Stellenbosch: Stellenbosch University, 2021.

[16] Zou, Difan, et al. "Gradient descent optimizes over-parameterized deep ReLU networks." *Machine learning 109 (2020):* 467-492.

[17] Kumar, Arvind, and Sartaj Singh Sodhi. "Some Modified Activation Functions of Hyperbolic Tangent (TanH) Activation Function for Artificial Neural Networks." *International Conference on Innovations in Data Analytics. Singapore: Springer Nature Singapore*, 2022.

[18] Hammad, M. M. "Deep Learning Activation Functions: Fixed-Shape, Parametric, Adaptive, Stochastic, Miscellaneous, Non-Standard, Ensemble." arXiv preprint arXiv:2407.11090 (2024).

[19] Pusztaházi, Luca Sára, György Eigner, and Orsolya Csiszár. "Parametric Activation Functions for Neural Networks: A Tutorial Survey." IEEE Access (2024).

[20] Abe, Ayomide Adeyemi, et al. "A robust deep learning algorithm for lung cancer detection from computed tomography images." Intelligence-Based Medicine 11 (2025): 100203.

[21] Hamad, Khaled, et al. "Automated Lung Cancer Diagnosis Using Transfer Learning with EfficientNetV2-S: A Deep Learning Approach Based on IQ-OTHNCCD Dataset." 2025 1st International Conference on Computational Intelligence Approaches and Applications (ICCIAA). IEEE, 2025.

[22] Alhassan, Afnan M., and Wan Mohd Nazmee Wan Zainon. "Brain tumor classification in magnetic resonance image using hard swish-based RELU activation function-convolutional neural network." Neural Computing and Applications 33.15 (2021): 9075-9087.

[23] Said, Menatalla MR, et al. "Innovative Deep Learning Architecture for the Classification of Lung and Colon Cancer From Histopathology Images." Applied Computational Intelligence and Soft Computing 2024.1 (2024): 5562890.

[24] Merabet, Asma, et al. "Enhancing colon cancer prediction in histopathology with integrated deep learning models: A comparative study on the lc25000 dataset." 2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS). IEEE, 2024.

[25] Özkan, Yasin. "Evaluating CNN Architectures and Transfer Learning for Histopathological Classification of Lung and Colon Cancer." Gazi University Journal of Science Part A: Engineering and Innovation 12.4 (2025): 1044-1059.

[26] Balasubramaniam, Sathiyabhama, et al. "A modified LeNet CNN for breast cancer diagnosis in ultrasound images." Diagnostics 13.17 (2023): 2746.

[27] Yılmaz, Mustafa Tahsn, Enes Algul, and Ishak Pacal. "A comparative study of advanced deep learning architectures for breast cancer classification on ultrasound and histological images." Results in Engineering (2025): 107600.

[28] Li, X. (2022). Differences of Tanh, Sigmoid, and ReLU Activation Function in Neural Networks. *International Journal of Scientific Progress and Research*, 80(6), 1–10. Retrieved from https://www.ijspr.com/citations/v80n6/IJSPR_8006_31035.pdf

[29] Heaton, Jeff. "Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618." *Genetic programming and evolvable machines 19.1* (2018): 305-307.

[30] Ankit Jain. *Deep Learning Essentials*. Packt Publishing, 2017. ISBN: 978-1785880360.

[31] Aggarwal, Charu C. Neural networks and deep learning. Vol. 10. No. 978. Cham: springer, 2018.

[32] Rasamoelina, Andrinandrasana David, et al. "A large-scale study of activation functions in modern deep neural network architectures for efficient convergence." *Inteligencia Artificial 25.70*(2022): 95-109.

[33] Gustineli, Murilo. "A survey on recently proposed activation functions for Deep Learning." arXiv preprint arXiv:2204.02921 (2022).

[34] Meronen, Lassi, Martin Trapp, and Arno Solin. "Periodic activation functions induce stationarity." Advances in Neural Information Processing Systems 34 (2021): 1673-1685.

[35] Rahman, Md Mizanur (2024), "Brain Cancer - MRI dataset", Mendeley Data, V1, doi: 10.17632/mk56jw9rns.1

[36] alyasriy, hamdalla (2020), "The IQ-OTHNCCD lung cancer dataset", Mendeley Data, V1, doi: 10.17632/bhmdr45bh2.1

[37] Orvile, "Colon Cancer CT Scans (2025 Abdominal CT Scans) Dataset", Kaggle, 2025. Available: https://www.kaggle.com/datasets/orvile/coloncancerct-2025-abdominal-ct-scans/data

[38] Aqdar , Karzan Barzan; Abdalla, Peshraw Ahmed; Mustafa , Rawand Kawa; Abdulqadir, Zhiyar Hamid; Qadir, Abdalbasit Mohammed; Shali, Alla Abdulqader; Aziz, Nariman Muhamad (2024), "Mammogram Mastery: A Robust Dataset for Breast Cancer Detection and Medical Education", Mendeley Data, V1, doi: 10.17632/fvjhtskg93.1