A Fractional Computation Based Deep Learning Framework for Silicosis Detection

Rinki Sharma[1][0009−0003−0885−5548] and Priyanka Harjule[2][0000−0002−3481−5764]

Malaviya National Institute of Technology, Jaipur, India
rinkysharma0357@gmail.com
priyanka.maths@mnit.ac.in

Abstract. This study presents a new fractional computational approach applied to a new dataset, silicosis. This is a scalable and flexible approach for training neural networks using fractional computation, which conveniently use the conformable fractional derivative. During the training process, the method includes an independent variable, α, which provides additional degree to the framework. Fractional variants of the sigmoid and relu activation functions are explored and compared to conventional activation functions. This method builds on earlier approaches by employing the conformable fractional derivative. The fractional activation functions notably converge to the actions of their standard version when $\alpha = 1$, guaranteeing a smooth integration with conventional neural network models. The study also tackles the problem of managing both positive and negative inputs, which is a crucial prerequisite for the derivative but has been mainly disregarded in earlier studies, underscoring the originality of the current work. The experimental framework incorporates both feedforward neural network and convolutional neural network using fractional activation functions. The findings indicate that the suggested framework performs better and is more accurate for particular values of α. The efficiency of the suggested computational approach is demonstrated by showing that fractional activation on Convolutional Neural Network when paired with transfer learning, performs better for silicosis chest X-ray classification than conventional transfer learning models.

Keywords: Convolutional Neural Network · Activation · Fractional computation · Silicosis Data

1 Introduction

Silicosis is an irreversible and potentially fatal lung disease, which is, entirely preventable. It results from exposure to respirable crystalline silica. Many workers worldwide are at risk of contracting this illness across a variety of industries. Pneumoconiosis refers to a diverse group of occupational interstitial lung diseases resulting from the prolonged inhalation and accumulation of mineral dust in the lungs. Early detection of this is crucial for detect the disease in its prior stages, enabling timely interventions to improve outcomes for affected workers while also facilitating the evaluation of shortcomings in workplace safety

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

protocols[18]. This ailment is very common among Indian workers, particularly in Rajasthan, and early finding is the only way to treat it but there are very few studies for computer aided detection of this [16]. A similar occupational disease, Pneumoconiosis, was also spread in China. A detection model for this condition has been developed using a deep convolutional diagnostic approach [8]. In state of the art the transfer learning models are very helpful for classification of lungs X- rays [20]. Although state-of-the-art research has explored models for silicosis detection, the available datasets often suffer from imbalances and may lack proper authorization[19]. In contrast, this study utilizes a properly labeled, balanced, and authorized dataset to develop a custom model specifically for silicosis medical X- ray images classification. The proposed approach achieves an accuracy of 0.8276, demonstrating superior performance compared to existing methods in the literature.

Activation functions are pivotal in determining the stability of neural networks and significantly impact their performance in modeling and interpreting physical phenomena. They assess the relevance of neuron inputs and decide whether a neuron should be activated through mathematical computations[10]. The study of fractional activation functions for neural networks is a rapidly growing area of research. By extending common functions like sine, cosine, and the logistic function based on the Mittag-Leffler function concept, Ivanov presented a novel method for creating fractional activation functions[12]. The impact of these fractional activation functions on neural network learning and prediction accuracy was assessed through experiments. Optimal parameter settings which occcurs by random search $\alpha \approx 0$ and $\beta = 1$ improved accuracy, occasionally achieving 100% in over epochs.

Recently, the fractional approach in optimization and training of neural networks gained more highlight [9,11,21]. Fractional Adaptive Linear Units, a new generalization of adaptive activation functions, are introduced and the method expands on earlier effective activation function research [26]. Common activation functions can be categorized into families using fractional calculus, which enables the creation of sets using a fractional derivative with a new parameter α denoting the derivative order[27]. The fractional adaption in step and multiquadratic functions make it possible to select one and generate the other by calculating its fractional derivative. Expanding on this idea, a thorough explanation of the three primary activation function families in this way is given[6]. A new definition of the fractional derivative is proposed and considered a natural extension of the conventional derivative and satisfies the arithmetic properties of the classical derivative and also aligns with the well-known fractional derivatives of polynomials [14].

An improved conformable fractional derivative is presented and a sort of historical memory parameter is added to the formulation of this improved conformable fractional derivative, which is also local by definition [7]. The contribution to physics and its physical interpretations validate the conformable fractional derivative [24,1,22].

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

A family of fractional activation functions using improved fractional derivative and its impact with different $\alpha \in (0, 1]$ is represented on some datasets [15]. By using the ideal conformable fractional derivative to create fractional activation functions, this work expands on the earlier approach. This method's ability to guarantee that the fractional activation function acts similarly to the conventional activation function when $\alpha = 1$. This approach helps to compare $\alpha \in (0, 1]$ in a single framework. In the experiments it will also make sure that it can handle positive and negative inputs, which is required by the derivative's definition and hasn't been covered in earlier research, underscoring the novelty of this work.

The Highlights of the proposed work are given below:

1. A custom transfer learning model for classification of silicosis dataset has been given.
2. Analysis of effect of fractional values on the computation of learning paradigm of Neural Network.
3. The fractional activation definition is enhanced from other studies as it generalizes with standard one.

The structure of the article is as follows: section 2 provides mathematical preliminaries about conformable fractional derivative and their fundamental properties. Fractional activation functions viz. fractional sigmoid and fractional relu function with the effect of variable $\alpha \in (0, 1]$ is presented in section 3. The experimental work with these proposed functions on wine and silicosis datasets is shown in section 4. Finally, the conclusion is given in section 5.

2 Conformable Fractional Derivative:

In 2014, Khalil et al.[18] introduced the conformable fractional derivative, a type of local fractional derivative that retains many properties of classical derivatives. This formulation not only exhibits greater similarity to classical derivatives but also preserves the same arithmetic operations and other properties like Rolle's theorem and Mean value theorem providing a more consistent extension to fractional calculus.

Definition 1 : The "conformable fractional derivative" of order α of function Given a function $f : [0, \infty) \to \mathbf{R}$. is defined by

$$D^{\alpha}(f)(t) = \lim_{\epsilon \to 0} \frac{f(t + \epsilon t^{1-\alpha}) - f(t)}{\epsilon} \tag{1}$$

for all $t > 0$, $\alpha \in (0, 1]$. If f is α-differentiable in some $(0, a)$, $a > 0$, and

$$\lim_{t \to 0^+} f^{\alpha}(t) \text{ exists, then define } f^{\alpha}(0) = \lim_{t \to 0^+} f^{\alpha}(t).$$

here $f^{\alpha}(t) = D^{\alpha} f(t)$
Here are theorems related to conformable fractional derivatives, which provide

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

a foundation for the generalization and formal establishment of this derivative concept [28].

Theorem 1 Let $\alpha \in (0, 1]$ and f, g be α-differentiable at a point $t > 0$. Then

1. $D^{\alpha}(af + bg) = aD^{\alpha}(f) + bD^{\alpha}(g)$, for all $a, b \in \mathbb{R}$.
2. $D^{\alpha}(t^{p}) = pt^{p-\alpha}$ for all $p \in \mathbb{R}$.
3. $D^{\alpha}(\lambda) = 0$, for all constant functions $f(t) = \lambda$.
4. $D^{\alpha}(fg) = f(t)D^{\alpha}(g) + g(t)D^{\alpha}(f)$.
5. $D^{\alpha}\left(\frac{f}{g}\right) = \frac{g(t)D^{\alpha}(f) - f(t)D^{\alpha}(g)}{g^{2}}$.
6. If, in addition, f is differentiable, then $D^{\alpha}(f)(t) = t^{1-\alpha}\frac{df(t)}{dt}$.

Theorem 2 If a function $f : [0, \infty) \to \mathbb{R}$ is α-differentiable at $t_0 > 0$, where $\alpha \in (0, 1]$, then f is continuous at t_0.

3 Methodological Description for Fractional Activation

Fractional activation functions generalize conventional activation functions by introducing fractional exponents in the exponential term. This section presents the mathematical formulation of these functions and analyze their behavior across various fractional $\alpha \in (0, 1]$. To integrate conformable fractional derivatives into the activation functions of neural networks, a fractional generalization of the exponential term is employed which offers an additional degree of freedom that enhances the adaptability of the activation function during network training.

3.1 Fractional Sigmoid Function

The standard sigmoid function is given by $\sigma(x) = \frac{1}{1+e^{-x}}$. The fractional effect is introduced by incorporating the fractional exponent term with $D^{\alpha}e^{-x}$ in place of e^{-x}. The fractional exponent term increases generalization, adaptability, and memory efficiency in the activation function.

Therefore, the fractional sigmoid function is defined as

$$F^{\alpha}_{sig}(x) = \frac{1}{1 + D^{\alpha}e^{-x}} = \frac{1}{1 - e^{-x}x^{1-\alpha}} \tag{2}$$

To guarantee the generalization with $\alpha = 1$ and, to avoid the undefined input for layers the fractional sigmoid activation is modified as in Eq. (3):

$$F^{\alpha}_{sig}(x) = \frac{1}{1 + e^{-x}|x|^{1-\alpha}} \tag{3}$$

Now, it is well-defined for all x and for $\alpha \in (0, 1]$. And for $\alpha = 1$ it converges to the standard sigmoid function.

Derivative of $F^{\alpha}_{sig}(x)$ to be used in backpropogation is defined in Eq. (4) :

$$D^{\alpha}F^{\alpha}_{sig}(x) = F^{\alpha}_{sig}(x)(1 - F^{\alpha}_{sig}(x)) \tag{4}$$

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

The above derivative is well-defined and retains the properties of the original sigmoid derivative, this approach eliminates the complexity introduced by $|x|^{1-\alpha}$in the derivative computation. Additionally, this derivative enhances stability during backpropagation and provides a gradient that facilitates learning while leveraging the benefits of fractional calculus.

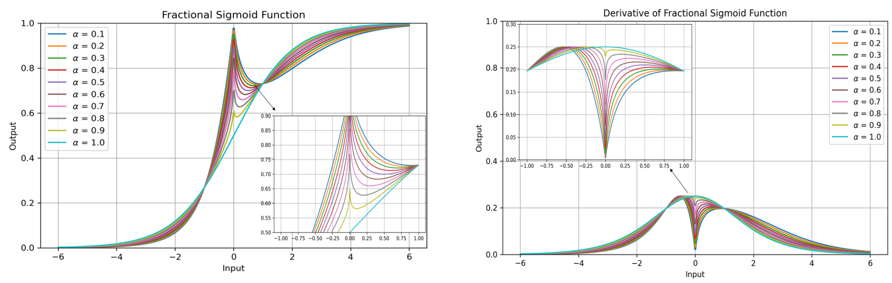


Fig. 1. Fractional sigmoid function and its derivative geometrical shape for $\alpha \in (0, 1]$

The geometrical representation of $F^{\alpha}_{sig}(x)$ and its derivative is shown in Figure 1. For each value of α the curve is different and the curvature of the curve varies for different α values. In the conventional case, as the sigmoid function approaches its extreme values (0 or 1), its gradient vanishes. This problem is mitigated by introducing the fractional term in sigmoid function. By doing so, because of different fractional values of α function reaches to saturation slowly as compared to the standard sigmoid function and as a result the gradient does not vanish easily[5]. As seen in Figure 1, the higher values of α near 1 provide a smoother transition as compared to its lower values, which have sharper slopes around 0, this accelerates the training even when neurons receive smaller inputs. Steeper gradients accelerate learning by enabling larger updates to the parameters during backpropagation, reducing the number of iterations required for convergence. However, excessively steep gradients can lead to instability or divergence, necessitating careful selection of activation functions and learning rates [25]. The fractional sigmoid allows fine-tuning of gradient behavior, making it beneficial for deep networks where standard sigmoid suffers from saturation using parameter α. The $F^{\alpha}_{sig}(x)$ allows the variable order alpha to have more adaptability as α near 0 ($0.1 - 0.3$) can accelerate the training due to its steep gradients and whenever the training is unstable getting large updates ($\alpha = 0.8{-}1.0$) can smoothen the gradients while other α values (0.4-0.7) provides moderate gradients generally.

3.2 Fractional relu Activation

The relu activation is very popular in the training of neural networks because of its simplicity and faster computation. It provides nonlinearity in the model despite having the simplest form among the activations. Since relu does not have exponential terms, the derivative of relu in backpropagation will be calculated

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

as a fractional conformable derivative to introduce the fractional effect in the training of the network. The standard relu activation function has a constant gradient for positive inputs, which limits its adaptability in learning [4]. As seen in Figure 2, the fractional derivative of relu, provides the dynamic gradients rather than flattening, allowing it to adapt smoothly to both large and small input values. This enhances gradient flow, improving the network's ability to learn more effectively across different scales of input. The different values of α will provide an additional degree of adaptability and a better generalization to the neural network. Fractional relu function is symbolically defined in Eq. (5), which is similar to the standard one for all alpha values. The effect of fractional parameter α is observed in its derivative given in Eq. (6). The geometrical shape of this fractional relu and fractional relu derivative is provided in Figure 2.

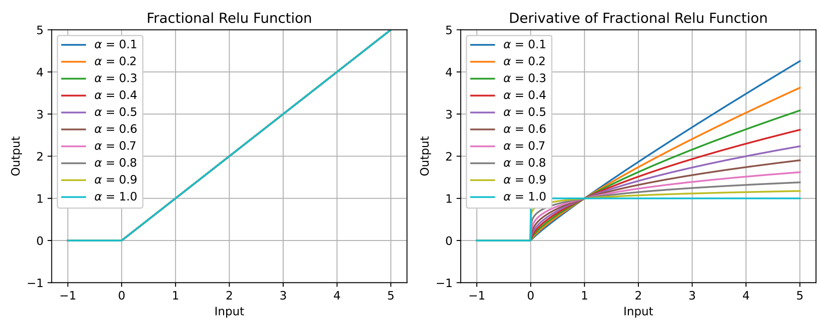


Fig. 2. Fractional relu and its conformable fractional derivative and its geometrical behaviour for $\alpha \in (0, 1]$

$$F^{\alpha}_{relu}(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} \tag{5}$$

And the conformable fractional derivative of fractional relu is given as:

$$D^{\alpha} F^{\alpha}_{relu} = \begin{cases} x^{1-\alpha}, x \geq 0 \\ 0,\ x < 0 \end{cases} \tag{6}$$

Figure 2 illustrate that derivative of fractional relu has different properties for different α values. The small α values provide higher slopes for larger inputs. The large α values near to 1 has flattening curve and slower growing derivative. This illustrates that, in learning the fine tuning of α can counter the problem of exploding and vanishing gradient in deep networks, as the term $x^{1-\alpha}$ in Eq.(6) provides a significant output for both smaller and larger inputs. Thereby making it a self adaptive activation function. The derivative of the activation has impact on the gradient flow during training yielding better training results for different α values in experiments.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

4 Results and Discussion

The proposed fractional activation functions are integrated in feedforfard neural network and convolutional neural network for experimental analysis. Two datasets have been used, one is a simple wine quality classification dataset and the second is the curated real dataset of X-ray images of silicosis disease. The results were compared through training and testing accuracies achieved for different α values. The details of the data set are given in Table 1.

Dataset	Type	Classes	Samples
Wine	Numeric	3	178
Silicosis	Image	2	421

Table 1. Details of Datasets

4.1 Experiment with wine Dataset

The wine dataset is a widely used dataset in machine learning, particularly for classification. It contains 178 samples with 13 features, representing the chemical properties of wines prevailing from three distinct cultivars. The data set is categorized into three classes, each representing a specific type of wine. Here, the classification of the wine dataset has been investigated using a neural network incorporating fractional sigmoid and fractional relu activation functions. The neural network architecture consists of two hidden layers, each containing 32 neurons, and remains consistent across training with both fractional activation functions.The learning rate, determined through random search, is set to 0.002, with 16 batch size, 200 epochs were taken for training, Adam optimizer was used for weight updation and loss minimization.

4.1.1 Wine classification using fractional sigmoid In this experiment, fractional sigmoid $F_{sig}^{\alpha}(x)$ was applied as activation to train the network. The results of achieved train and test accuracy are given in Figure 3. Figure 3 illustrates that for $\alpha = 0.3$ and $\alpha = 0.6$, the training accuracy attains highest value. However, $\alpha = 0.3$ also yields the highest testing accuracy among all alpha. Both training and testing accuracies reach 1.0, which can be genuine because dataset is very small and simple and the proposed model can learn the exact patterns of the data instead of general patterns. The dataset exhibits a simple structure and is almost linearly separable, making the results realistic and not an unusual occurrence. The reason for $\alpha = 0.3$ can be justified by the behaviour of fractional sigmoid which shows that, for small $\alpha < 1$ values, the training is faster because of sharper gradients. $\alpha = 0.1, 0.2$ exhibit steep gradients also, but they may lead to unstable gradients in an early phase. In contrast, $\alpha = 0.3$ provides a balance, ensuring both stable training and sufficiently steep gradients.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

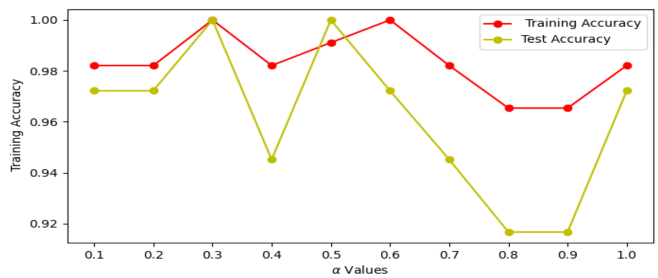


Fig. 3. Training and testing accuracies achieved for different α values with fractional sigmoid on wine dataset

For, $\alpha = 0.5$ results indicates some unusual conduct, with testing accuracy higher than training accuracy. However, this can be regarded as insignificant by considering the test dataset's small size and simple patterns[2]. Notably, $\alpha = 0.3$ and $\alpha = 0.6$ outperform $\alpha = 1$, which shows the standard training approach with the same parameters. This observation shows that the fractional effect on activation enhances generalization on unseen data as compared to the standard activation.

4.1.2 Wine classification using fractional relu In this experiment, the Fractional relu F^{α}_{relu} activation function and its conformable fractional derivative is used for classification of the wine dataset. All network parameters remain the same as the previous experiment with fractional sigmoid. The results for α values with achieved training and testing accuracies provided in Figure 4. For $\alpha = 0.2$, the training and testing accuracies both reach 1.0, which is the best result among all α.

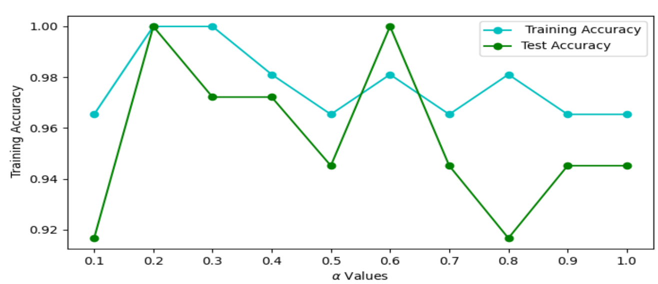


Fig. 4. Training and testing accuracies achieved for different α values with fractional relu on wine dataset

For $\alpha = 0.3$, the training accuracy remains 1.0, while the test accuracy is 0.9722, which is closer to the training accuracy. For $\alpha = 0.6$, the testing accuracy surpasses the training accuracy, which can be attributed to the small test size [2].

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

The training process has been highly effective, ensuring that the test data aligns well with the network without any errors. Here also, the best result is attained by $\alpha = 0.2$ other than $\alpha = 1$ therefore, it can be concluded that fractional activation behaves better than the standard activation function. The comparison of results with recent state of the arts is given in Table 2.

State of The Art	Accuracy	Method
[3](2024)	0.97	Generic Algorithm
[23](2023)	1.00	Deep Neural Network
[13](2024)	0.99	Fractional order Differential evolution
[17](2024)	0.98	BP Neural Network
Proposed	1.00	Fractional Activation

Table 2. Comparison of wine data accuracy with recent studies

4.2 Experiment on Silicosis Dataset

Silicosis identification in X-ray pictures is mostly dependent on radiologists' skill, which frequently causes delays in diagnosis. Computer-aided identification methods based on machine learning are being developed as a solution to this problem. However, the development of highly accurate deep-learning models for silicosis detection remains challenging due to the limited availability of large databases. This study proposes a novel approach with a new dataset for silicosis detection using transfer learning techniques applied to available X-ray radiographs. The Dataset was curated at Sawai Man Singh hospital Jaipur by a team of radiologists under the project no.1000114110 funded by the Government of Rajasthan. The Dataset division has been given below in Table 3. The dataset maintains a good balance among classes, effectively preventing class dominance. The test data remains unseen by the model throughout training. Due to limited data availability, no separate validation set was used. This study leverages transfer learning model VGG 19 from our previous studies[19] with additional pooling, flattening, two fractional dense layers. The sample images of silicosis dataset are

Silicosis Dataset	**Label**	**Count**
Train Data	Silicosis	195
	Normal	197
Test Data	Silicosis	14
	Normal	15
Total -		421

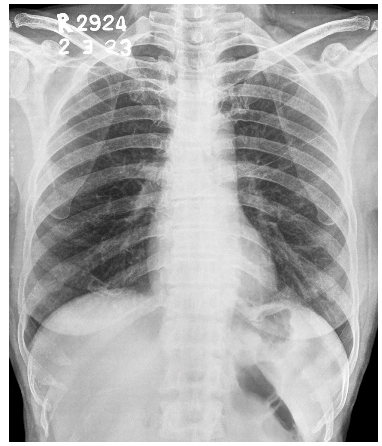
Table 3. Silicosis Dataset Distribution

shown in Figure 5.

The Architecture of the proposed model is given in Figure 6. This study proposes a custom model specifically for the regional chest X-ray dataset for

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

(a) Silicosis lungs



(b) Normal lungs

Fig. 5. Silicosis and normal lungs image of Dataset

silicosis classification using fractional sigmoid and fractional relu. The input function for the output layer can be represented by Eq. 7

$$Z(X) = F^{\alpha}(F^{\alpha}(GAP(V(X)) + b_i) + b_i') \tag{7}$$

Here $V(X)$ is the output of VGG 19 and GAP is global average pooling effect,

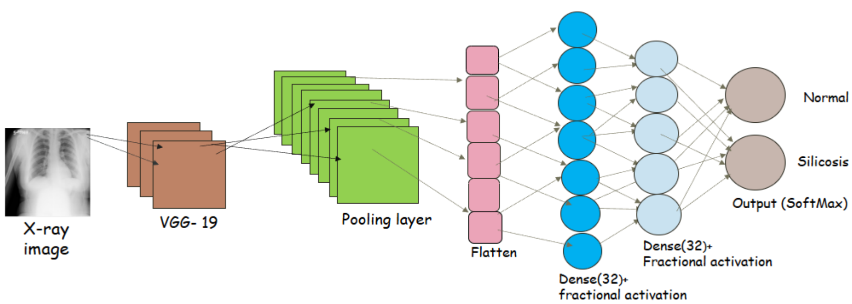


Fig. 6. Architecture of model for silicosis classification

b_i, b_i' are bias of dense layers and F^{α} is fractional activation for nonlinearity.

4.2.1 Silicosis classification using fractional sigmoid In this part the fractional sigmoid has been used in dense layers of the model to enhance the classification. The results have been shown in Figure 7. Graph shows achieved training and

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

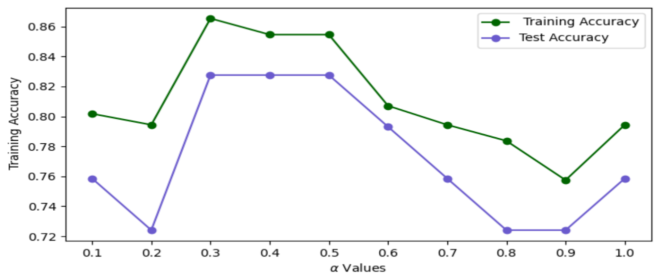


Fig. 7. Training and testing accuracies achieved for different α values with fractional sigmoid on silicosis dataset

testing accuracy for every $\alpha \in (0, 1]$. Training has been done for 50 epochs. The learning rate has been selected by random search to be 0.01, batch size is 16, dropout = 0.2 and Adam optimizer is used for weight update. Hyperparameter values were determined through a random search strategy applied to the model. As seen in Figure 7, $\alpha = 0.3, 0.4, 0.5$ shows best performance with $\alpha = 0.3$ giving the highest training accuracy so this can be considered the best α among others for this experiment, and the best test accuracy is achieved at $\alpha = 0.3$ which is 0.8276. It is observed that as α is approaching the value 1, performance of the model is decreasing. To show the model generalization performance on unseen test data the receiver operating characteristic (ROC) curve is given in Figure 9a in which the area under the curve (AUC) is 0.83. The AUC of 0.83 indicates that the model performs well, assigning a higher probability to a positive sample than a negative sample in 83 out of 100 cases. Also it is better than the random guessing which has 50% AUC. The model is showing good performance as it achieves high true positive rate by maintaining a low false positive rate.

4.2.2 Silicosis classification using fractional relu In this section, the fractional relu activation is involved in the custom model for silicosis identification. The model parameters remain consistent as the previous experiments i.e. learning rate of 0.01, batch size of 16, dropout of 0.2, and epochs set to 50. Figure 8 illustrates the training and testing accuracy achieved with fractional relu among different $\alpha \in (0, 1]$. The findings indicate that the best test accuracy of 0.7931 is attained with $\alpha = 0.3$ and 0.4, that is less than that achieved using the fractional sigmoid activation function. This outcome can be predicted because fractional sigmoid introduces fractional effects in both the function and its derivative, which provides fractional adaptability in forward pass as well as backward pass whereas the fractional relu has fractional value only in the derivative which has effect only in the backward pass. Additionally, fractional relu activation also has the issue of dying neurons for negative inputs like standard relu, which could explain why the performance is slightly lower compared to the fractional sigmoid for this silicosis dataset.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

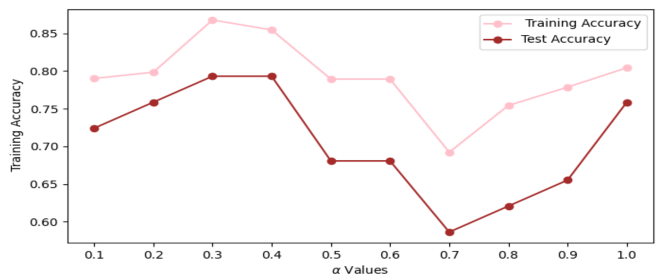
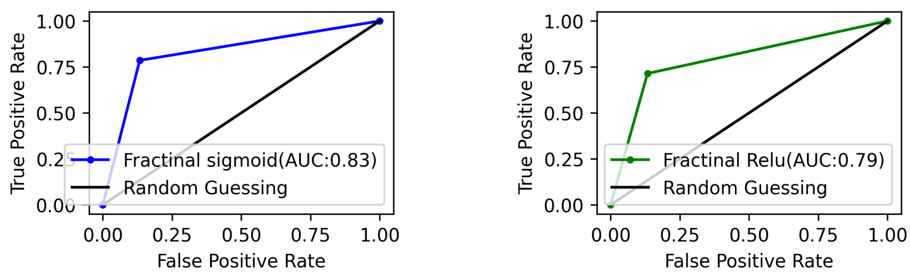


Fig. 8. Training and testing accuracy achieved for different α values with fractional relu on silicosis Dataset



(a) Roc curve using fractional sigmoid

(b) Roc curve using fractional Relu

Fig. 9. Roc curves for silicosis classification using fractional activations

The receiver operating characteristic (ROC) curve has been shown in Figure 9b with area under the curve value equals 0.79. This indicated that the the proposed model is performing better than the random guessing. The proposed model with fractional relu has classified the data by providing a high probability to a positive sample than a negative sample 79 cases out of 100. The performance of both activations with different performance metric is given in Table 4.

Activation	Accuracy	precision	sensitivity	F1-score
F^{α}_{relu}	0.7931	0.80	0.79	0.79
F^{α}_{Sig}	0.8276	0.84	0.83	0.83

Table 4. Evaluation of performance metrics for F^{α}_{relu} and F^{α}_{Sig} activation functions

5 Conclusion

This study introduces a generalized fractional variant of two widely used activation functions, sigmoid and relu. The fractional activation function adds another degree of freedom α to improve generalization capabilities. A promising approach is proposed to objectively improve training performance by providing the capacity to fine-tune α. Using the suitable α, the issue of vanishing and

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

exploding gradients can be mitigated, leading to a more flexible training process. The fractional sigmoid is less prone to the saturation problem, preventing vanishing gradients, while the fractional relu avoids flattening gradients and effectively adapts to inputs due to the varying α values. Additionally, this study experiments on a newly approved silicosis dataset to evaluate the effectiveness of the proposed approach. The empirical results for the silicosis and wine dataset using this approach demonstrate superior performance compared to the state of the art. For the silicosis dataset specifically, the fractional sigmoid outperforms the fractional relu. This may be due to the absence of the fractional α in the forward pass, as well as the inherent issue of dying neurons in relu, which can affect the training process for this data. Future research will focus on developing optimization algorithms for α and further improving silicosis classification, which could surely benefit patients in local communities. Also, a dynamic model that adapts the α as needed for slow or fast training can be made using different α values during training.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Data Availability The code and data will be provided as per the individual's request.

References

1. Alharbi, F.M., Baleanu, D., Ebaid, A.: Physical properties of the projectile motion using the conformable derivative. Chinese Journal of Physics **58**, 18–28 (2019)
2. Aliferis, C., Simon, G.: Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and ai. Artificial intelligence and machine learning in health care and medical sciences: Best practices and pitfalls pp. 477–524 (2024)
3. Chai, J., Bi, M., Teng, X., Yang, G., Hu, M.: A mixed mutation strategy genetic algorithm for the effective training and design of optical neural networks. Optical Fiber Technology **82**, 103600 (2024)
4. Ding, B., Qian, H., Zhou, J.: Activation functions and their characteristics in deep neural networks. In: 2018 Chinese control and decision conference (CCDC). pp. 1836–1841. IEEE (2018)
5. Dubey, S.R., Singh, S.K., Chaudhuri, B.B.: Activation functions in deep learning: A comprehensive survey and benchmark. Neurocomputing **503**, 92–108 (2022)
6. Esquivel, J.Z., Cruz Vargas, J.A., Lopez-Meyer, P.: Fractional adaptation of activation functions in neural networks. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 7544–7550 (2021). `https://doi.org/10.1109/ICPR48806.2021.9413338`
7. Gao, F., Chi, C.: Improvement on conformable fractional derivative and its applications in fractional differential equations. Journal of Function Spaces **2020**(1), 5852414 (2020)
8. Hao, C., Jin, N., Qiu, C., Ba, K., Wang, X., Zhang, H., Zhao, Q., Huang, B.: Balanced convolutional neural networks for pneumoconiosis detection. International Journal of Environmental Research and Public Health **18**(17), 9091 (2021)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

9. Harjule, P., Sharma, R., Kumar, R.: Fractional-order gradient approach for optimizing neural networks: A theoretical and empirical analysis. Chaos, Solitons & Fractals **192**, 116009 (2025)
10. Hayou, S., Doucet, A., Rousseau, J.: On the impact of the activation function on deep neural networks training. In: International conference on machine learning. pp. 2672–2680. PMLR (2019)
11. Herrera-Alcántara, O., Arellano-Balderas, S.: Adaptive morphing activation function for neural networks. Fractal and Fractional **8**(8), 444 (2024)
12. Ivanov, A.: Fractional activation functions in feedforward artificial neural networks. In: 2018 20th International Symposium on Electrical Apparatus and Technologies (SIELA). pp. 1–4. IEEE (2018)
13. Jin, T., Su, K., Gao, J., Xia, H., Dai, G., Gao, S.: Fractional-order differential evolution for training dendritic neuron model. Available at SSRN 4760944 (2024)
14. Khalil, R., Al Horani, M., Yousef, A., Sababheh, M.: A new definition of fractional derivative. Journal of computational and applied mathematics **264**, 65–70 (2014)
15. Kumar, M., Mehta, U., Cirrincione, G.: Enhancing neural network classification using fractional-order activation functions. AI Open **5**, 10–22 (2024)
16. Li, T., Yang, X., Xu, H., Liu, H.: Early identification, accurate diagnosis, and treatment of silicosis. Canadian respiratory journal **2022**(1), 3769134 (2022)
17. Liu, W., Liu, M., Yan, C., Qi, M., Zhang, L.: Corporate fraud detection based on improved bp neural network. Computing and Informatics **43**(3), 611–632 (2024)
18. Mushtaq, F., Bhattacharjee, S., Mandia, S., Singh, K., Chouhan, S.S., Kumar, R., Harjule, P.: Artificial intelligence for computer aided detection of pneumoconiosis: A succinct review since 1974. Engineering Applications of Artificial Intelligence **133**, 108516 (2024)
19. Sharma, G.K., Harjule, P., Agarwal, B., Kumar, R.: Silicosis detection using extended transfer learning model. In: International Conference on Recent Trends in Image Processing and Pattern Recognition. pp. 111–126. Springer (2023)
20. Sharma, G.K., Harjule, P., Sadhwani, T., Agarwal, B., Kumar, R.: Sequential transfer learning models with additional layers for pneumonia diagnosis. In: 2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3). pp. 1–6. IEEE (2023)
21. Sharma, R., Harjule, P.: Modified gradient descent approach involving caputo fractional derivative with metaheuristic optimizer. Journal of Vibration Testing and System Dynamics **8**(04), 443–453 (2024)
22. Stojiljkovic, V., et al.: A new conformable fractional derivative and applications. Selecciones Matemáticas **9**(02), 370–380 (2022)
23. Wang, D., Wang, T.: Federated ensemble algorithm based on deep neural network. In: International Conference on Soft Computing in Data Science. pp. 76–91. Springer (2023)
24. Ye, Y., Fan, H., Li, Y., Liu, X., Zhang, H.: Conformable bilinear neural network method: a novel method for time-fractional nonlinear partial differential equations in the sense of conformable derivative. Nonlinear Dynamics pp. 1–16 (2024)
25. Yu, X.H., Chen, G.A., Cheng, S.X.: Dynamic learning rate optimization of the backpropagation algorithm. IEEE Transactions on Neural Networks **6**(3), 669–677 (1995)
26. Zamora, J., Rhodes, A.D., Nachman, L.: Fractional adaptive linear units. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 8988–8996 (2022)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4

27. Zamora Esquivel, J., Cruz Vargas, A., Camacho Perez, R., Lopez Meyer, P., Cordourier, H., Tickoo, O.: Adaptive activation functions using fractional calculus. In: Proceedings of the IEEE/CVF international conference on computer vision workshops. pp. 0–0 (2019)
28. Zhao, D., Luo, M.: General conformable fractional derivative and its physical interpretation. Calcolo **54**, 903–917 (2017)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_4