Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

An Overview on Effectiveness of Activation Functions in Processing Medical Images

Suhad Al-Shoukry[1], Zalili Binti Musa[2]

[1] Department of Computer System Techniques, Al-Najaf Technical Institute, Al-Furat Al-Awsat Technical University, Kufa, Iraq. E-mail: inj.suhadaalzhra2010@atu.edu.iq
[2] Faculty of Computing, University Malaysia Pahang Al-Sultan Abdullah, Pekan 26600, Malaysia.
E-mail: zalili@ump.edu.my
[*] Corresponding author Suhad Al-Shoukry E-mail: inj.suhadaalzhra2010@atu.edu.iq

https://doi.org/10.46649/fjiece.v3.2.13a.22.5.2024

Abstract. *This study studies the impact of activation functions in the field of machine learning and deep learning in general and especially on medical images for different aims such as classification, clustering, feature engineering. Important components that add nonlinearity and allow networks to learn intricate patterns are activation extraction, training and etc. The study first starts by explaining various activation function types that are commonly used in NN applications and fields. Subsequently, a comprehensive comparative analysis is conducted, to evaluate how activation functions perform in terms of accuracy and their impact on speed convergence. Understanding how activation functions impact the categorization of medical imagery is crucial to the study's findings. In additions, the study illustrates overview of which activation functions yield optimal results. The Main results contribute how to select best activation functions that suits most accurate and efficient medical image classification based on this overview, any researcher can choose best activation function after reading this overview,*

Keywords: *Activation Functions, Medical Image Classification, Neural Networks, Comparative Analysis, Accuracy and Convergence*

1. INTRODUCTION

In today's healthcare system, medical imaging is essential for precise diagnosis, planning of treatments, and disease monitoring. The examination of images showing brain tumours is one of the most important aspects of medical imaging. A patient's health may be significantly impacted by brain tumours, thus prompt and accurate diagnosis is essential for successful treatment. The field of analysis medical images had witnessed significantly advancement after the introduction of Neural Networks (NNs) and Deep learning (DL), presenting new opportunities for the automated and precise detection, classification and segmentation of tumours. In Machine Learning (ML) branch, NNs, showed to be highly adept at handling complex, high-dimensional medical image processing. None-the-less, activation function selection significantly impacts the performance of those networks. NNs utilize activation functions as non-linear transformation components for imparting non-linearity and facilitating recognition of the complicated patterns that are

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

found in the medical images. In the area of medical image analyses, selecting the correct activation functions is highly important for the production of reliable and accurate results. [1][2].

In that regard, the objective of the present survey is the thorough investigation of efficiency of different activation functions in brain tumour image processing. The space of NN activation functions is explored and the way that they impact some tasks, such as classification, segmentation, and brain tumour detection. The main objective is shedding light on the distinct challenges and opportunities that are presented by the processing off medical images through the concentration on the brain tumour images. Additionally, we hope to provide information that is applicable to more general medical imaging applications.

This survey article's structure can be described as follows: We present some background insight on the processing of medical images and discuss the role that is played by NNs in this area in Section 2. A review of the activation functions as well as their function in the NNs will be provided in Section 3. In Section 4, we address the specific difficulties of brain tumour image processing and present the several approach that are utilized in this field. The proposed process for the selection and classification of papers for the survey will be described in Section 5. More information on factors that influence activation function selection will be provided in Section6.

Our objective is giving practitioners, academics, and medical professionals a detailed grasp of the impact of activation functions on brain tumour image processing throughout the present detailed study. This information may be helpful for the NN architecture designers make better judgments for better results in the area of medical image processing.

As a result of its capability of provide non-invasive viewing, analyses, and interpretation of the physiological and anatomical data, the processing of medical images became one of the most vital areas of study in health-care industry. Those images, ranging from the MRIs and PET scans to the X-rays and CT scans, offer very valuable insights into internal human body structures and workings. Many medical areas, like oncology, radiology, neurology, and cardiology, utilize the medical imaging. Finding valuable data in those photos is highly important for precise and timely diagnoses. [3][4].

The field of medical image processing witnessed a revolution because of the NNs, which provide effective and automated solutions for many challenging issues. Neural network had a significant impact in advancing image segmentation, disease detection, and treatment planning as a result of their ability in learning complicated patterns from massive datasets. NNs can develop representations that have the ability of capturing subtle anomalies and subtleties in images, which may be rather hard to detect for conventional approaches as a result of availability of numerous medical image libraries.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

In neural network designs, activation functions represent non-linear transformation elements. Through adding non-linearity to the model, providing NNs with the capability of making sophisticated relationship approximations in the data. For accurately analysing images by using the networks, it requires having the ability to capture features at a variety of levels of abstraction, which had been made possible through activation functions. Generalization capacity of the network, in addition to its expressiveness, and rate of convergence are all influenced by the selections of activation function [5].

It remains highly vital to carry out studies that are focused on the efficacy of the activation function in processing medical images, in particular, the cases of brain tumour images. The ability of a network in the differentiation between the healthy tissues, detection of subtle tumour characteristics, and operating with a high level of accuracy all can be highly impacted by the chosen activation functions. Considering the wide range of accessible activation functions, there is a necessity in the thorough assessment of everyone's effectiveness on the brain tumour images.

Additionally, the complicated spatial linkages and dynamic nature of the medical images require a deeper comprehension of behaviours of different activation functions. The increasing dependence of the medical image analyses on DL approaches necessitates the determination of the activation functions that produce the optimal results in terms of resilience, precision, and computational effectiveness.

The objective of the present overview is closing knowledge gap between the approaches of medical image processing and NN theory. The objective of our study is offering guidance to the practitioners as well as the researchers in the cases of NN architecture constructions for the medical image analyses by the methodical analysis of performances of various activation functions on tasks that are associated with the processing of brain tumour images.

Choosing the right activation function is crucial as it directly affects a neural network's capacity to learn and generalize. Different tasks and data characteristics may require specific activation functions. For instance, sigmoid and tanh activations can work well when outputs need to be bounded, while ReLU-like activations often show better performance in deeper networks due to their reduced vanishing gradient problem.

In the field of medical image analysis, the selection of appropriate activation functions can significantly impact the network's ability to accurately detect and classify pathological features within images. Understanding activation function strengths and limitations is vital in the construction of effective NN architectures that are tailored to tasks of medical image analyses.

2. MATERIALS AND METHODS

Activation Functions:

NNs' essential building blocks, activation functions add non-linearity to network's computations. They're vital in deciding, in response to weighted summation of its inputs, whether to activate a neuron or not. NNs may learn and reflect complex patterns that the linear functions on their own are incapable of grasping by the use of activation functions to model complicated relationships in the data [6].

Neural networks can approximate a wide range of functions primarily due to the non-linearity introduced by activation functions. This non-linearity makes neural networks highly effective tools for tackling complex tasks, such as medical image processing. The objective of the activation functions is ascertaining the output of a neuron's, which is helpful for the network's learning and prediction formation. [7].

Commonly Used Types:

Many activation function types are often utilized in the NN architectures [8][9][10]:

- Sigmoid: this activation function maps input values to a range within 0 to 1. It can be identified by a smooth S-shaped curve, allowing it to squash the input values into probability-like output, which makes it proper for the tasks of binary classification.
- ReLU (Rectified Linear Unit): this activation function returns input value in the case where that value is positive, whereas it returns 0 otherwise. ReLU has gained popularity due to its simplicity and efficiency in training deep networks. However, it might sometimes suffer from "dying ReLU" problem where neurons may become inactive throughout the process of training.
- Leaky ReLU: This variant of ReLU allows a small negative slope for input values less than 0, mitigating the dying ReLU problem.
- Softmax: Softmax is often utilized in the output layer for multi-class classification tasks. It converts the network's raw scores into probability distributions over classes.
- Tanh (Hyperbolic Tangent): Tanh maps input values to a range between -1 and 1. Like the sigmoid, it is useful for squashing input values, but it has the advantage of a symmetric output range.

Characteristics of Activation Functions:

Activation functions possess several characteristics that impact their suitability for specific tasks [10][11]:

- Linearity vs. Non-linearity: Nonlinear activation functions enable neural networks to capture complex relationships between inputs and outputs. Linear activation functions result in linear models regardless of network's depth.
- Differentiability: Differentiability is essential for backpropagation, a fundamental training algorithm for neural networks. Most activation functions used are differentiable, allowing gradients to be computed for weight updates.

Medical Image Processing: Challenges and Techniques:

The medical images present particular difficulties requiring sophisticated approaches of processing for the accurate analyses and interpretation. This is particularly right with brain tumour images. Those difficulties result from some things, such as image noise, anatomical diversity, and the requirements for accurately identifying minute abnormalities. [12][13].

Unique Challenges of Medical Images:

In comparison with the processing of the natural images, medical image processing poses distinctive issues that result from:

- Noise and Artefacts: medical image quality and precision may be affected by noise introduced throughout the process of acquisition.

- Anatomical Variability: as a result of large inter- and intra-subject variability that is found in the human anatomy, reliable approaches that are capable of adjusting to various structural configurations are necessary.

- Scale and Complexity: Algorithms which are capable of capturing coarse as well as fine features are required due to the fact that the medical images often include complicated structures at many scales.

Techniques for Medical Image Analysis:

Different approaches are utilized for the extraction of meaningful information from the medical images, addressing challenges that are specific to the domain. Those approaches include [14] [15]:

- Image Enhancement: improving anatomical structure visibility, pre-processing approaches are implemented for noise removal, contrast boosting, and image quality improvement.
- Classification: this approach divides the image to distinctive classes (such as diseased vs. healthy) based upon areas or the entire image. NNs and other ML forms are usually utilized for job classification.

- Segmentation: which includes the identification of areas of interest, such as tumours or organs, within an image. Advanced DL-based approaches, thresholding and region growth are utilized.
- Registration: which enables meaningful comparisons and disease development tracking through aligning several photos of one patient, taken at various times or with various modalities.

Role of Neural Networks:

Concerning the tackling difficulties that are related to the area of medical image processing, like brain tumour segmentation and identification, NNs have proven to be revolutionary tools. The innovations below were made possible as a result of their ability in extracting complex features from large data-sets [16][17][18]:

- Feature Learning: through the automatic extraction of pertinent features from the medical images, NNs can do away with human feature extraction requirement. Which is especially helpful in the case of working with varied and intricate image structures.
- Accuracy of Classification: concerning medical image classification, NNs have exceptionally good performance, showing improved specificity and sensitivity in the identification of healthy from tissues from the pathological ones.
- Precision in the Segmentation: Anatomical structures and malignancies may be segmented accurately by the convolutional NNs (CNNs). Precise delineation is facilitated by its hierarchical architecture that collects the features at several scales.
- Adaptability: NNs have been considered as flexible tools for the interpretation of medical images, accommodating inherent heterogeneity in the medical images as a result of their capability of adapting to different situations of imaging and anatomy.

NNs have shown high potential in the processing of the brain tumour images by the automatic identification and classification of tumours, relieving radiologists of some of their work and enhancing early diagnoses. We may additionally increase those methods' effectiveness and precision by skilfully using the activation function power within NN topologies, which will lead ultimately to better treatment outcomes and patient care.

3. RESULTS AND DISCUSSION

In the case when saturating activation functions, like Sigmoid or Tanh, are utilized in the network's training phase, vanishing gradient problems arise. Those functions restrict the response range of neurons to a specified range, say [0,1]. Yet, non-saturating functions like Exponential Linear Unit (ELU) and ReLU only handle the positive component of the gradient for avoiding the vanishing gradient problem. Leaky-ReLU, for instance, is a modified version of ReLU that addresses this problem in both negative and positive ways; yet, it still needs pre-specified parameters that need to be carefully chosen based upon the training set and application [19]. Parametric activation functions, such as PReLU, PELU, and so on, could be used as a workaround since they enable the network to modify input data by learning the parameter through network training. It is preferable to combine a sub-set of functions with various behaviours in order to take use of their distinct characteristics. When at least two functions are combined in a nonlinear or linear way to generate a more complex function, the combination coefficients are learned throughout the network's training phase and changed to account for the input data. The research in [20] shows how networks could adapt to get over the limitations of fixed activation functions. Recent research by the authors of [21] has demonstrated how adding adaptive functions to deep networks quickens convergence. Those results encourage the researchers to use adaptive functions in pre-built, well-known deep networks like U-Net and look into the ways where such functions can significantly enhance the performance of the networks. It is clear that the shapes of the learned and fundamental functions could differ. The adaptive blending unit (ABU) [22] is an example that utilizes a linear combination of 6 fundamental functions (Identity, ReLU, Tanh, SELU, and ELU) with trainable parameters for scaling. The work done in [23] is another effective experiment in this field. The authors described three different types of combinations of activation functions as follows:

- Mixed activation function

$$f_{mix}(x) = pf_1(x) + (1-p)f_2(x) \qquad (1)\ [23]$$

where $f_1$and f_2 are main activation functions and p is a set of coefficients that should be learned in the phase of training.

- The gated activation functions

$$f_{gate} = \sigma(wx)f_1(x) + \big(1 - \sigma(wx)\big)f_1(x) \qquad (2)\ [23]$$

where:

- σ represents the sigmoid function.
- w is the gating mask acquired during network training.
- $\sigma(wx)$ serves as the combination coefficient and represents the percentage of information change for a given input x. It is generated by multiplying the input by the gating mask w.

Utilizing a hierarchical tree-like structure, the hierarchical activation function may learn non-linear transformations.

Current studies in [24] show how such AFFs can improve deep network performance. It goes without saying that the kind of basic activation functions utilized could affect capabilities of resulting AAF; for instance, if the adaptive activation functions mentioned above are drawn from ReLU family, then the basic functions employed will also come from that family. Which is why, variability in the family of basic functions could have a significant impact on variance in the resulting combined functions.

The classification of brain tumour types is an essential step in creating a treatment approach and increasing patient survival rates. Those days, brain tumour types could be accurately detected using MRI, which is routinely employed. However, manual brain tumour classification requires a limited number of images and is quite time-consuming. This is addressed in [25] using a state-of-the-art DL-based model for automatic brain tumour classification. The authors created brain images using the T1-weighted, contrast-enhanced MRI scans in the dataset. In order to create feature vectors and enhance the clarity regarding raw brain images, a histogram of directed gradients as well as normalization methods are employed. Along with other feature descriptors, the histogram descriptor performs a decent job of capturing edge and contour characteristics. Following this, the acquired features are utilized to train CNNs, which are after that utilized for classifying brain tumours into gliomas, meningiomas, and pituitary tumours. Furthermore, ReLU activation function with a hard-swish basis improves CNN performance, leading to faster learning. With regard to the testing phase, the suggested model showed an amazing accuracy rate of 98.6%. This confirms the effectiveness of the proposed model in brain tumour classification, outperforming state-of-the-art algorithms such as deep CNNs with transfer learning and fine-tuning techniques. Unlike other activation functions such as the sigmoid and hyperbolic tangent function (tan h), the hard swish-based RELU activation function is utilized to increase classification performance and learning speed. For preventing overfitting, hard swish-based RELU activation function has advantages in precisely managing the number of hidden neurons. This activation function outperforms RELU by a little margin since their graph morphologies are comparable. However, issues like lack of saturation, the vanishing gradient problem, and computing simplicity are all successfully addressed by hard swish-based RELU activation function. It effectively prevents the significant shift in weight values observed in this investigation, which results in the disappearance of the minimal gradient. Through carefully adjusting the number of hidden neurons, saturation is reduced and the NN's ability to process information and learn is enhanced. The general hard-swish based ReLU activation function equation is represented by equation (3).

$$f(x) = \max(\beta * x, x) * x \qquad (3)\ [25]$$

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

The parameter β is in range [0-1], this is an adaptive parameter. x is the input.

The study [26] demonstrated how doctors' clinical experience plays a major role in brain tumour diagnosis. Nevertheless, since the development of computer-assisted diagnostics, the precision of determining the type of tumour has significantly improved. Consequently, a CNN depending on complex networks (CNNBCN) with a modified activation function was developed specifically for the use of MRI in the classification of brain tumours. Unlike conventional approaches, CNNBCN network structure is not built or optimized manually. Instead, it is constructed by using randomly generated graph approaches, which are subsequently changed through a network generator into a format compatible with computing NNs. The enhanced CNNBCN model achieves an astounding 95.49% accuracy rate for brain tumour classification, outperforming several models from previous attempts. In addition, the upgraded CNNBCN model outperformed popular models, such as DenseNet, ResNet, and MobileNet in terms of test loss for brain tumour classification during the course of the tests. This shows that the upgraded CNNBCN model achieves exceptional performance in brain tumour image categorization while concurrently advancing NN design methodologies. The authors used a new combination of activation functions for optimizing the resulting model. Many model modules are subjected to the combination of activation functions made up of Gaussian error linear units (GeLUs) and ReLUs following multiple rigorous tests. ReLU activation function is utilized by the classifier and the first two CNNBCN modules; the GeLU activation function is used by the remaining CNNBCN modules. The GeLU non-linearity is the expected transformation of a stochastic regularizer that randomly applies the identity or zero map to a neuron's input. GeLU nonlinearity weights are input by their magnitude, as opposed to gates input by their sign as in ReLU. An approximate mathematical expression for GeLU(x), which is thought to represent a regular normal distribution, can be found in equation (4). The actual formula is displayed in equation number three:

$$f(x) = 0.5 * x\left(1 + \tanh\left[\sqrt{\frac{2}{\pi}}\,(x + 0.044715x^3)\right]\right) \quad (4)\ [26]$$

$$f(x) = 0.5 * x\left(1 + \frac{2}{\sqrt{\pi}}\int_0^{\frac{x}{\sqrt{2}}} e^{-t^2}dt\right) \quad (5)\ [26]$$

Where x is the input and t is a dummy variable for integration.

The research [27] focused on the automatic segmentation regarding brain tumours from MRI using computer vision. The authors brought up the subject of utilizing DNNs for image segmentation since they have demonstrated that DNNs could accurately segment images of brain tumours and since the gradient

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

diffusion problem as well as intricacy of DNN training could make the process computationally and time-consuming. They introduced brain tumour segmentation utilizing the Deep Residual Learning Network (ResNet) to overcome the gradient problem with DNN. ResNets are faster and more accurate than traditional DNN models because they introduce shortcut skip connections parallel to the layers of the CNN. Simulations and BRATS 2015 dataset were used to assess the suggested method, proving its superiority. The enhanced accuracy of 83%, 90%, and 85% for the entire, core, and enhancing regions of brain tumours, respectively, is validated by the results. Furthermore, compared to previous DNN methods, our method shows an average computation time that is three times faster. In DNN's final layer, they employed the Softmax activation function after using the ReLU activation function in the convolution layers. They were dependent on other research projects that used ReLU as an activation function [28].

The study [29] employs a modified deep CNN to particularly address the segmentation task inside medical image processing. The proposed approach aims to accelerate learning and enhance network performance using fewer parameters. Adaptive activation functions, which depends on the well-known U-Net architecture and are renowned for their effectiveness in segmentation, are applied to achieve customization. In this tailored network, U-Net's complexity is drastically reduced—by around 200 times—by employing mitigating factors to accelerate learning. One significant modification is the addition of adaptive activation functions, where every convolution layer's data-driven activation function is determined as a linear combination of 16 widely-used fundamental functions. With this adjustment, the accuracy drops those results from parameter reduction is effectively countered, enabling the model to be tuned even when the training data is scarce. Following evaluation on widely-used retinal imaging datasets, including STARE, DRIVE, HRF, CHASE, and ARIA, the recommended customized U-Net achieved accuracy levels of 97%, 96%, 97%, 96%, and 95% in segmenting blood vessels. Furthermore, on ISIC skin lesion dataset, the proposed network achieves 98% accuracy in lesion area segmentation. These collected data clearly demonstrate that the customized network outperforms previous successful study attempts in medical segmentation. They also show how same customization tactics may be applied to other popular deep networks, transforming them into efficient small models which could overcome challenges caused by lack of available data. The authors' use of adaptive activation functions had a major impact on the DL networks' performance. Training a deep network, even with cutting-edge hardware, usually takes time, therefore it's necessary to reduce the network's complexity and parameters while keeping the deep network's learning performance and capabilities [30]. A linear combination of 16 well-known activation functions—combination coefficients that are learned throughout network training—was used as an adaptive function. Consequently, every convolution layer is able to use the learnt activation function.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

Details about **adaptive activation function (AAF)** is shown below in **Figure 2** and the equation (6)

$$AAF = F \times P \qquad (6)\ [30]$$

Where: $F: [f_i(x)]_{1\times16}$ is the vector of 16 different basic activation functions, and $P: [P_i]_{16\times1}$ represents the coefficient vector

It is significant to remember that the AAF is more flexible because each element of P might have a value between -1 and +1. According to Figure 4's block diagram, the AAF learning block can be seen of as a little perceptron network, with a mixture of frequently used activation functions as the output. It is significant to notice that each layer of the network adds 16 extra parameters when this recommended AAF is added, both to that layer and to the network overall. In this case, there are two possible methods: (a) using one AAF for all of the network's convolution layers, or (b) using distinct AAFs for every convolution layer. While the network performs better than traditional ones when a single universal AAF is used across all convolution layers, learning distinct AAFs for each layer results in more flexible performance. The 14 convolution layers in the proposed design means that 14 distinct AAFs—or, more precisely, 14 distinct P vectors—will need to be learned during network training. Compared to the 120825 parameters of the suggested architecture, this adds only 224 more parameters to the network, making it negligible. One can initialize P vector values to $\frac{1}{16}$

Using deep learning (DL) approaches for intelligent screening of cervical anomalies and early detection of precancerous phases has been studied by authors in [31]. The main goal is to stop cervical abnormalities from developing into cancer. In a number of biomedical applications, such as disease prediction, medical image analysis, and image segmentation, DL has demonstrated encouraging outcomes. The authors of this work intentionally created very deep residual learning based networks for cervical cancer screening. Their study was primarily concerned with how activation functions affected ResNet performance. Three residual networks with identical architectures and distinct activation functions were constructed. They utilized a dataset made up of colposcopy cervical images to assess these models' performance. The findings of the experiment showed that the residual networks which were created to screen cervical images with high accuracy were able to use leaky and parametric rectified linear units (LEaky-RELU and PRELU) as activation functions. PRELU network attained a flawless accuracy of 100%, whereas Leaky-RELU network only managed to acquire an accuracy of 90.2%. When these accuracy levels were compared to the findings of other relevant studies, they demonstrated better performance in terms of correctly identifying healthy and pre-cancerous colposcopy cervical images.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

The authors emphasize that such early and accurate diagnosis can significantly contribute to the prevention of cervical cancer transformation. (PReLU formula is shown in equation (7), Leaky ReLU formula is shown in equation (8)

$$f(x) = \begin{cases} x \; if \; (x > 0) \\ ax \; if (x \leq 0) \end{cases} \qquad (7) \; [30]$$

Or:

$$f(x) = \max(0, x) + a * \min(0, x)$$

$$a: is \; trainable \; parameter$$

$$f(x) = \max(0.1x, x) \qquad (8) \; [30]$$

The results of study [31] are showed in the **Table 1**. The results illustrated that ReLU costs much time with high MSE, PReLU achieved best MSE and Leaky ReLU achieved learning epochs with less training time. The testing for three activation functions was made in the same conditions for the three activation functions. In [32], the authors investigated the hierarchical framework of Capsule Networks (CapsNets) and its application in brain tumour classification tasks. The initial convolution layer of CapsNets typically employs an activation function, with the rectified linear unit (ReLU) being commonly used in such tasks. ReLU does have certain drawbacks, though, such as the possibility of neuron activation failure owing to zero derivatives and poor performance accuracy in brain tumour classification. The authors suggested parametric scaled hyperbolic tangent (PSTanh), a unique activation function, as a solution to such drawbacks. Through addressing the problem of vanishing gradients, offering a small gradient by adding the β and λ parameters, and facilitating faster optimization, PSTanh improves the conventional hyperbolic tangent. The authors examined eight common activation functions, including Memrister-Like Activation Function (ReLU), tanh, PReLU, Leaky-ReLU, SELU, ELU, Swish, and ReLU-Memrister-Like Activation Function (RMAF), in order to assess PSTanh's effectiveness and comparing it with other activation functions. Many data-sets, like MNIST, CIFAR100, fashion-MNIST, ImageNet and CIFAR10, have been utilized in extensive researches. deep CNN and CapsNets models such as AlexNet, SqueezeNet, DenseNet-121 and ResNet-50 have been amongst models that are used throughout the training. Results of experiments that are performed on the CapsNets and CNN models for the tasks of the classification of brain tumour have shown that the proposed PSTanh activation function had a better performance compared with other functions. This result emphasized PSTanh's advantages compared to the present activation functions in the tests of brain tumour classification. The formulas for additional activation functions which are discussed by the authors are Swich in eq8 and scaled exponential linear unit (SELU) in eq9, eq10. Eq9 represents the PSTanh.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

$$f(x) = \begin{cases} x \; if \; (x > 0) \\ a(e^x - 1) \; if (x \leq 0) \end{cases} \qquad (9)\ [32]$$

$$Swish(x) = x.\sigma(\beta x) \qquad (10)\ [32]$$

$$\beta \; is \; trainable \; parameter \; and \; \sigma(x) is \; sigmoid \; function$$

$$PSTanh(x) = f(\beta, \gamma, x) = x.\gamma(1 + \tanh(\beta x)) \qquad (11)\ [32]$$

$$\beta \; and \; \gamma \; are \; trainable \; parameters$$

Their results showed that $\beta = 1.50$ and $\gamma = 0.50$ are optimal values which have led to a high level of accuracy.

In [32], The authors explored the use of deep learning techniques to address various health issues, including the diagnosis of pneumonia, a significant global health problem and leading cause of death. Convolutional neural networks (CNNs), a widely used artificial intelligence technique, have applications in segmentation, classification, and signal processing. In deep networks, activation functions play a crucial role. The development of activation functions (AFs) in CNN architectures considers features such as avoiding local minima in the CNN model and enhancing training performance. Although ReLU is commonly used in most studies, it has the drawback of not incorporating negative weights into the network. To address this issue, the impact of newly proposed AFs on real-world problems has been investigated. In this study, a novel method for detecting pneumonia based on a newly proposed activation function and various CNN models is suggested. The experimental studies involved detecting pneumonia from chest X-ray images using several activation functions, including ReLU, LReLU, Mish, Sigmoid, Swish, Smish, Logish, Softplus, and the proposed activation function Superior Exponential (SupEx). Additionally, experimental studies were conducted using the MNIST and CIFAR-10 datasets to further validate the effectiveness of the proposed SupEx AF. The results demonstrated that the CNN models employing the proposed SupEx activation function achieved superior performance in both pneumonia detection from chest X-ray datasets and in traditional benchmark datasets like MNIST. The notable aspect of this study is the introduction of a new activation function for pneumonia detection and its successful application in traditional benchmark datasets, highlighting its potential for broader use in various deep learning tasks. Away from medical images and unstructured data, Relu activation function was adopted by authors in [34] for training neural network in the field of Unmanned Aerial Vehicles to get real measurements of Magnetometers sensors from noised measurements, authors achieved high accuracy results.

4. CONCLUSIONS

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

In conclusion, the investigation into the efficiency of activation functions in the processing of medical images, highlighted their crucial role in neural network topology. Through a thorough comparative analysis, we gained valuable insights into how different activation functions impact convergence speed, accuracy, and overall performance across various tasks in medical image analysis. This study establishes a foundation for selecting optimal activation functions to maximize results by understanding the correlations between activation functions and medical image analyses. One significant factor to consider as medical image analysis evolves is the careful selection of activation functions. The potential to transform the precision and effectiveness of medical image analysis through the synergy of activation functions and neural networks could significantly impact patient care, disease management, and therapeutic decision-making. This study underscores the importance of bridging theoretical understanding with real-world applications, motivating professionals and scholars to conduct further research that expands on this critical area. Ultimately, this study enhances the capability to use state-of-the-art technologies to improve patient care and outcomes.

CONFLICTS OF INTEREST

The author declares that there are no conflicts of interest regarding the publication of this paper. This research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

There are no known conflicts of interest (political, personal, religious, ideological, academic, intellectual, commercial or any other) to declare in relation to this manuscript. The research was conducted independently and was not commissioned by any organization or group with an interest in the outcome of the research.

5. REFERENCES

[1] X. Wang, Y. Zhao, and F. Pourpanah, “Recent advances in deep learning,” *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 4, pp. 747–750, Feb. 2020, doi: https://doi.org/10.1007/s13042-020-01096-5

[2] S. Shajun Nisha and M. Nagoor Meeral, “9 - Applications of deep learning in biomedical engineering,” ScienceDirect, Jan. 01, 2021. https://www.sciencedirect.com/science/article/abs/pii/B9780128230145000089 (accessed Aug. 13, 2023).

[3] Castiglioni et al., “AI applications to medical images: From machine learning to deep learning,” Physica Medica, vol. 83, pp. 9–24, Mar. 2021, doi: https://doi.org/10.1016/j.ejmp.2021.02.006.

[4] S. Mishra, Hrudaya Kumar Tripathy, and B. Acharya, “A Precise Analysis of Deep Learning for Medical Image Processing,” Studies in computational intelligence, pp. 25–41, Jul. 2020, doi: https://doi.org/10.1007/978-981-15-5495-7_2.

[5] S. Scardapane, S. Van Vaerenbergh, S. Totaro, and A. Uncini, “Kafnets: Kernel-based non-parametric activation functions for neural networks,” *Neural Networks*, vol. 110, pp. 19–32, Feb. 2019, doi: https://doi.org/10.1016/j.neunet.2018.11.002

[6] Y. Wang, Y. Li, Y. Song, and X. Rong, “The Influence of the Activation Function in a Convolution Neural Network Model of Facial Expression Recognition,” Applied Sciences, vol. 10, no. 5, p. 1897, Mar. 2020, doi: https://doi.org/10.3390/app10051897.

[7] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, “Activation functions in deep learning: A comprehensive survey and benchmark,” Neurocomputing, vol. 503, pp. 92–108, Sep. 2022, doi: https://doi.org/10.1016/j.neucom.2022.06.111.

[8] Sharma, S., Sharma, S. and Athaiya, A., 2017. Activation functions in neural networks. *Towards Data Sci*, *6*(12), pp.310-316.

[9] E. Chigozie, W. Nwankpa, A. Ijomah, S. Gachagan, and Marshall, “Activation Functions: Comparison of Trends in Practice and Research for Deep Learning,” 2018. Available: https://arxiv.org/pdf/1811.03378.pdf

[10] Feng, J. and Lu, S., 2019, June. Performance analysis of various activation functions in artificial neural networks. In *Journal of physics: conference series* (Vol. 1237, No. 2, p. 022030). IOP Publishing.

[11] B. Ding, H. Qian, and J. Zhou, “Activation functions and their characteristics in deep neural networks,” 2018 Chinese Control And Decision Conference (CCDC), Jun. 2018, doi: https://doi.org/10.1109/ccdc.2018.8407425.

[12] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, “A Survey of MRI-based Medical Image Analysis for Brain Tumour Studies A Survey of MRI-based Medical Image Analysis for Brain Tumour Studies 2.” Accessed: Aug. 13, 2023. [Online]. Available: https://boris.unibe.ch/15734/2/BauerPMB2013.pdf

[13] H. Shaari et al., “Deep Learning-Based Studies on Pediatric Brain Tumours Imaging: Narrative Review of Techniques and Challenges,” Brain Sciences, vol. 11, no. 6, p. 716, May 2021, doi: https://doi.org/10.3390/brainsci11060716.

[14] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, “Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges,” Journal of Digital Imaging, vol. 32, no. 4, pp. 582–596, May 2019, doi: https://doi.org/10.1007/s10278-019-00227-x.

[15] M. Razzak, S. Naz, and A. Zaib, “Deep Learning for Medical Image Processing: Overview, Challenges and Future.” Available: https://arxiv.org/ftp/arxiv/papers/1704/1704.06825.pdf

[16] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, “Medical image retrieval using deep convolutional neural network,” Neurocomputing, vol. 266, pp. 8–20, Nov. 2017, doi: https://doi.org/10.1016/j.neucom.2017.05.025.

[17] M. Razzak, S. Naz, and A. Zaib, “Deep Learning for Medical Image Processing: Overview, Challenges and Future.” Available: https://arxiv.org/ftp/arxiv/papers/1704/1704.06825.pdf

[18] A. Selvikvåg Lundervold and A. Lundervold, “An overview of deep learning in medical imaging focusing on MRI,” Zeitschrift für Medizinische Physik, vol. 29, no. 2, Dec. 2018, doi: https://doi.org/10.1016/j.zemedi.2018.11.002.

[19] B. Hanin, “Which Neural Net Architectures Give Rise To Exploding and Vanishing Gradients?,” *arXiv.org*, Oct. 26, 2018. https://arxiv.org/abs/1801.03744 (accessed Aug. 12, 2023).

[20] S. Flennerhag, H. Yin, J. Keane, and M. Elliot, “Breaking the Activation Function Bottleneck through Adaptive Parameterization,” arXiv.org, Nov. 22, 2018. https://arxiv.org/abs/1805.08574 (accessed Aug. 12, 2023).

[21] A. D. Jagtap, K. Kawaguchi, and G. E. Karniadakis, “Adaptive activation functions accelerate convergence in deep and physics-informed neural networks,” *Journal of Computational Physics*, vol. 404, p. 109136, Mar. 2020, doi: https://doi.org/10.1016/j.jcp.2019.109136.

[22] L. R. Sütfeld, F. Brieger, H. Finger, S. Füllhase, and G. Pipa, “Adaptive Blending Units: Trainable Activation Functions for Deep Neural Networks,” arXiv.org, Jun. 26, 2018. https://arxiv.org/abs/1806.10064 (accessed Aug. 12, 2023).

[23] S. Qian, H. Liu, C. Liu, S. Wu, and H. S. Wong, “Adaptive activation functions in convolutional neural networks,” Neurocomputing, vol. 272, pp. 204–212, Jan. 2018, doi: https://doi.org/10.1016/j.neucom.2017.06.070.

[24] R. ZahediNasab and H. Mohseni, “Neuroevolutionary based convolutional neural network with adaptive activation functions,” Neurocomputing, Dec. 2019, doi: https://doi.org/10.1016/j.neucom.2019.11.090.

[25] A. M. Alhassan and W. M. N. W. Zainon, “Brain tumour classification in magnetic resonance image using hard swish-based RELU activation function-convolutional neural network,” *Neural Computing and Applications*, Jan. 2021, doi: https://doi.org/10.1007/s00521-020-05671-3.

[26] Z. Huang et al., "Convolutional Neural Network Based on Complex Networks for Brain Tumour Image Classification With a Modified Activation Function," in IEEE Access, vol. 8, pp. 89281-89290, 2020, doi: https://doi.org/10.1109/ACCESS.2020.2993618

[27] L. H. Shehab, O. M. Fahmy, S. M. Gasser, and M. S. El-Mahallawy, “An efficient brain tumour image segmentation based on deep residual networks (ResNets),” *Journal of King Saud University - Engineering Sciences*, vol. 33, no. 6, pp. 404–412, Sep. 2021, doi: https://doi.org/10.1016/j.jksues.2020.06.001.

[28] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” arXiv.org, May 18, 2015. https://arxiv.org/abs/1505.04597.

[29] A. Farahani and H. Mohseni, “Medical image segmentation using customized U-Net with adaptive activation functions,” Neural Computing and Applications, Oct. 2020, doi: https://doi.org/10.1007/s00521-020-05396-3.

[30] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” Lecture Notes in Computer Science, vol. 9351, pp. 234–241, 2015, doi: https://doi.org/10.1007/978-3-319-24574-4_28.

[31] K. M. A. Adweb, N. Cavus, and B. Sekeroglu, “Cervical Cancer Diagnosis Using Very Deep Networks Over Different Activation Functions,” IEEE Access, vol. 9, pp. 46612–46625, 2021, doi: https://doi.org/10.1109/access.2021.3067195.

Al-Furat Journal of Innovations in Electronics and Computer
Engineering (FJIECE)
ISSN -2708-3985

[32] K. Adu, Y. Yu, J. Cai, I. Asare, and J. Quahin, “The influence of the activation function in a capsule network for brain tumour type classification,” International Journal of Imaging Systems and Technology, vol. 32, no. 1, pp. 123–143, Aug. 2021, doi: https://doi.org/10.1002/ima.22638.

[33] Kiliçarslan, S., Közkurt, C., Baş, S., & Elen, A. (2023). Detection and classification of pneumonia using novel Superior Exponential (SupEx) activation function in convolutional neural networks. Expert Systems with Applications, 217, 119503.

[34] Assad, A., Serikov, S., & Al Bitar, N. (2024). Using an artificial neural network to improve the orientation accuracy of unmanned aerial vehicles. *International Journal of Open Information Technologies*, *12*(3), 36-44.

Figure and tables

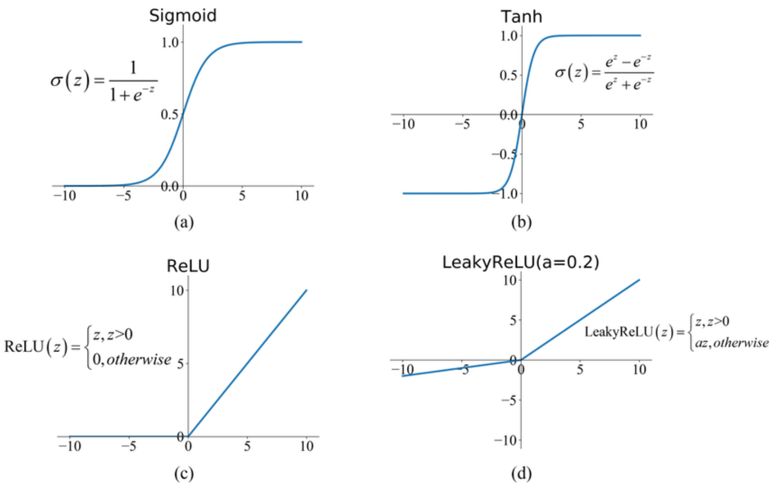

Figure 1: curves of most common activation functions

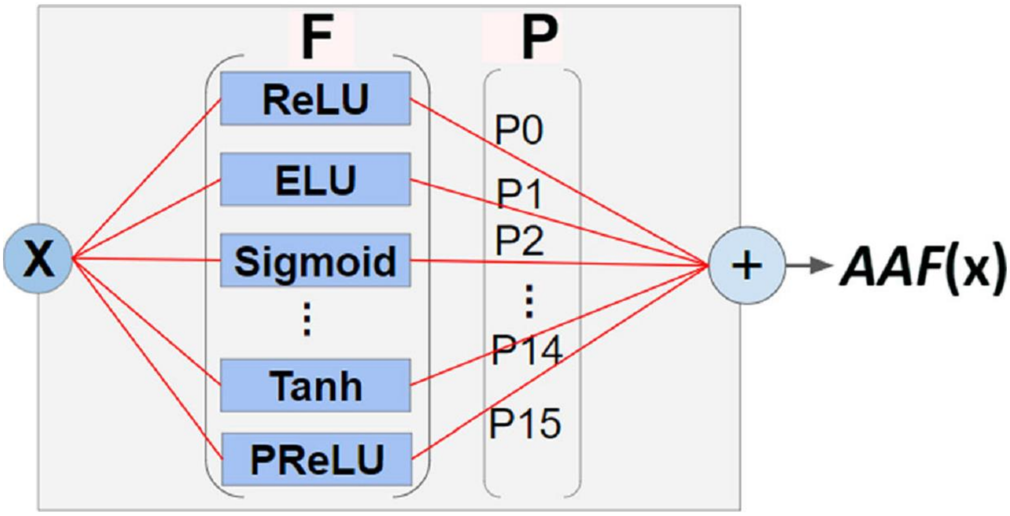

Figure 2: adaptive activation function [29]

Table 1. The results illustrated that ReLU costs much time with high MSE

Table 1. Results in [31]

Parameter	ReLU	Leaky ReLU	PReLU
Training time (Minutes)	125	119	122
Mean Square Error	0.0074	0.0014	0.0001

