An Improved Taylor Hyperbolic Tangent and Sigmoid Activations for Avoiding Vanishing Gradients in Recurrent Neural Nets

Tirupati Gullipalli
Department of Information Technology
Gayatri Vidya Parishad College of Engineering for Women, India
gtr@gvpcew.ac.in

Krishna Murali
Department of Computer Science and Engineering, University College of Engineering, India
krishnaprasad.mhm@gmail.com

Srinivasa Peri
Department of Computer Science and Engineering, Maharaj Vijayaram Gajapathi Raj College of Engineering, India
psr.sri@gmail.com

Abstract: *In deep learning, Hyperbolic Tangent (Tanh) and Sigmoid nonlinear activation functions can retain the complex relationship, which is more appropriate in Recurrent Neural Networks (RNNs). The gradients of these activation functions are vital in updating the weights during training the network. However, both functions are vulnerable to the vanishing gradient problem and expensive in exponent operations. It causes gradients to vanish during back propagation that leads to training overheads and low performance. Although most of the studies put forward methods to reduce exponent operations, there is not a viable solution to tackle the gradient issues. Hence, we propose a Taylor expansion of second order to realize Tanh and Sigmoid functions. In particular, Long Short-Term Memory (LSTM) network makes extensive use of these functions as well as gating mechanism to control the flow of information and gradients. In consequence, Taylor expansion Tanh and Sigmoid activation functions based parallel heterogeneous LSTM network integrated with Bayesian hyperparameter optimization is being proposed for multi-step time series prediction. The current model efficacy is evaluated on bench mark datasets Mackey-Glass Series (MGS), Electricity Transformer Temperature hourly 2 (ETTh2), coronavirus daily cumulative cases, Cumulative Deaths (CD-5) and (CD-7), daily New Cases (NC4), and Total Recovery Cases (TRC-8) in India. The model performance is compared with conventional models like the Auto Regressive Integrated Moving Average (ARIMA), Tree-based Pipeline Optimization Tool (TPOT) regressor, LSTM, Gated Recurrent Unit (GRU), transformer, and the proposed model with Tanh and Sigmoid activations. The analysis reveals that the current model achieves remarkable performance in terms of Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2 Score) when compared to existing models.*

Keywords: *Vanishing gradient, taylor expansion, functional approximation, sigmoid, hyperbolic tangent, neural networks.*

Received August 13, 2024; accepted May 26, 2025
https://doi.org/10.34028/iajit/22/5/15

1. Introduction

In Artificial Neural Networks (ANNs), the nonlinear behavior of activation functions is essential for investigating the mapping between input and output. Recently, machine learning has been modified to use ANNs for multi-labeling [4] and regression [14]. These include machine translation [17], image analysis [26], disease forecasting [1, 2, 3], and many more. A neuron performs two major operations: Product accumulation and activation function. The activation function incorporates nonlinearity to ensure that Neural Networks (NNs) are successful in a wide range of applications. The prominent activation functions are Hyperbolic Tangent (Tanh), Sigmoid, softmax and Rectified Linear Unit (ReLU) family. These activation functions play a crucial role in Recurrent Neural Networks (RNNs), often serving as gating mechanisms. The ReLU family is well-known for its unbounded activation capabilities, while softmax is commonly employed in the output layer, especially for multi-label classification and attention mechanisms. In regression tasks, Tanh and Sigmoid activation functions are typically used for both standard and recurrent activations. However, it's important to note that these functions can lead to vanishing gradient issues, which may result in small gradients that cause training instability, reliance on costly exponent calculations, and sluggish convergence [18]. During training, the vanishing gradient problem can occur when the weights of the derivatives start getting very close to zero. This makes it quite challenging to update the weights with the Backpropagation algorithm. Additionally, the hardware costs for executing exponent operations [43] can be significant. To tackle these issues, improving the derivatives of these functions can greatly reduce the vanishing gradient problem and enhance overall efficiency. Various methods have been proposed to address this, including adaptations for Sigmoid [46] and Tanh [9, 10] functions, cost reductions on Field Programmable Gate Arrays (FPGAs) [35], normalization techniques [31], and gradient clipping methods [47]. Most of these strategies aim to lessen the

reliance on exponent operations and prevent weights from nearing zero, rather than directly improving gradients. Moreover, performance can vary across different applications within the same field. These challenges can be effectively approached using well-known strategies that involve suitable activation functions and gated architectures.

Thus, a second-order Taylor expansion-based Tanh and Sigmoid activation function integrated with a parallel heterogeneous Long Short-Term Memory (LSTM) model is being put forward for multi-step series prediction. The work has been structured as follows: Part 2 designates the current work associated with the domain of vanishing gradient. In part 3, we describe the dataset in this work followed by the methods and experimental design in part 4 and results and discussions of study presented in part 5. Finally, in part 6 conclusions and future research directions have been forwarded.

2. Literature Survey

The vanishing gradients issue in ANNs can be addressed by promoting better derivative flow and simplifying the training process using Residual Neural Networks (ResNets). Recent studies have unveiled a ground breaking method incorporating a norm preservation mechanism [29] within ResNets, expanding upon this concept. The authors offer valuable perspectives on how keeping norm consistency can tackle the issues of vanishing gradients and enhance the performance of networks. However, it's worth noting that implementing this norm preservation may lead to increased computational demands, and the method could struggle to scale when applied to more complex models or larger datasets. A novel method [20] modifies the gradient flow for enhanced stability and effectiveness in learning by substituting the original derivative with an artificial one. This approach efficiently addresses the vanishing gradient issue for ReLU and Sigmoid functions at low processing costs, yet relies on a heuristic for the artificial derivative, necessitating a distinct design for each activation function. Recent study by Wei *et al*. [43] suggests using a Probability-based Sigmoid Function (P-SFA) approximation to minimize exponential operations, showing improvements in power consumption, processing speed, and recognition performance in certain datasets, though it doesn't directly address vanishing gradient problems.

The functional approximation has become popular for reducing costs and addressing issues of vanishing derivatives. According to De Ryck *et al*. [10], Tanh activation functions in NNs can be efficiently approximated using high-order Sobolev norms. The findings imply that function approximation rates can be attained by Tanh networks with only two hidden layers that are on par with or even higher than those of much deeper ReLU networks. Nonetheless, the main purpose of this discovery is to serve as a substitute for the ReLU function in prediction tasks.

Zhang *et al*. [47] tackles the vanishing gradients issue by introducing a framework to investigate the effects of gradient clipping on network training, although implementing adaptive clipping thresholds may complicate the process due to the need for additional computational resources. Moreover, Hochreiter [19] provided valuable theoretical insights explaining the struggles of conventional RNNs with vanishing gradients, illustrating how their proposed LSTM architecture effectively addresses these challenges through the utilization of gating mechanisms. This enhances gradient retention and overall learning performance. Additionally, a survey [24] covered the various activation functions employed across different NN architectures and tasks, shedding light on their impact on performance and training dynamics. Another comprehensive review [13] looked into a variety of activation functions, assessing their effectiveness through metrics like convergence speed, training stability, and overall performance against standard benchmarks. Zaki *et al*. [46] presented a simpler Sigmoid approximation for NNs on FPGAs, demonstrating improved performance over the original, although it fails to solve gradient issues.

Timmons and Rice [36] highlighted improvements in NNs training times through approximations of ReLU, Tanh, and Sigmoid across three architectures: the Modified National Institute of Standards and Technology (MNIST) classifier, MNIST auto encoder, and CharRNN. While these approximations led to reduced training times, they remain unsuitable for other network types.

Additionally, Chandra [9] investigated the use of Catmull-rom spline interpolation to approximate the Tanh function. The results indicated a smaller logic area requirement compared to the original Tanh, which effectively lowers costs but does not tackle the gradient issue.

Conversely, Cetin *et al*. [8] explored Sigmoid approximation via Taylor expansion in multilayer networks for diagnosing hepatitis. This approach employed three intervals to mimic the Sigmoid function and successfully achieved comparable accuracy to the original Sigmoid, though it came with the trade-offs of added complexity and hardware overhead. The investigation also addressed functional approximations related to softmax alternatives. In this context, research [42] examined periodic alternatives to softmax, particularly with attention mechanisms to address the gradient problem, suggesting these periodic activation functions offer advantages over traditional softmax.

Banerjee *et al*. [5] introduced soft-margin Taylor as a viable alternative to softmax, revealing that utilizing a Taylor expansion up to two terms yielded superior accuracy across three datasets (MNIST, Canadian Institute For Advanced Research (CIFAR10), and

CIFAR100) when compared to an expansion of up to ten terms. Temurtas *et al*. [35] estimated toluene gas concentrations through NN architectures that utilize a nine-term Taylor series expansion of the Sigmoid activation function based on transient sensor responses. The summary of other studies is given in Table 1 below.

Table 1. Summary of additional work related activation functions.

Ref.	Year	Type of activation function	Description
[15]	2024	Tanh Exponential Linear Unit (TeLU)	TeLU aims to mitigating issues such as vanishing and exploding gradients.
[39]	2024	Novel activation function	A novel activation function within the CNN enhances feature extraction capabilities compared to ReLU.
[24]	2024	Comprehensive survey of 400 activation functions for NN.	Describe the extensive compilation of fixed and adaptive activation functions.
[21]	2023	User Interface Activation (UIA) functions, which uses 3 hyperparameters allowing it to emulate existing activation.	The model achieves improvement of up to 5% compared to those using traditional activation functions.
[13]	2022	overview of activation functions	Understanding of various activation functions and their implications in DL.
[23]	2022	Scaled-Gamma-Tanh (SGT) activation	SGT activation designed to enhance the MRI classification using 3D CNNs.
[5]	2021	Soft margin-Taylor softmax	Results indicate that the SM-Taylor softmax can outperform the standard softmax
[10]	2021	Tanh activation function	Derive explicit error bounds for the approximation of Sobolev-regular and analytic functions using Tanh-activated NNs.
[20]	2021	Artificial derivatives are used in the back propagation algorithm.	Replacing the standard derivatives used in backpropagation with artificial derivatives.
[40]	2020	Optimization of activation functions within LSTM networks using a Differential Evolution Algorithm (DEA)	The DEA-based search systematically explores the space of possible activation functions, evaluating their performance within LSTM networks on specific tasks

3. Dataset

To assess the performance of the proposed model for multi-step time series forecasting, the author considers two bench mark datasets, Mackey-Glass (MG) series [33], Electricity Transformer Temperature hourly 2 (ETTh2) [48], and multiple coronavirus datasets [32, 37] have been taken from the Oxford Martin Programme on Global Development at Oxford Martin School [16].

3.1. Mackey-Glass Series (MGS)

In order to evaluate the efficiency of current model, a benchmark chaos dataset is considered as a first study. The time series is attained by resolving the Mackey-Glass (MG) equation [28], which is one of the frequently used benchmark datasets for evaluating the forecasting models. MG time series is shown in Figure 1, which is generated by a differential system as shown in Equation (1).

$$\frac{dx}{dt} = \frac{ax(t-\tau)}{1+x^c(t-\tau)} - bx(t) \quad (1)$$

a=0.2, c=10, and b=0.1 are the constant variables. $\tau \geq 17$ can set to generate the chaotic behavior of system. The time series of 1000 samples are generated, of which 80% is considered training; including 20% for validation and the remaining is used for testing.

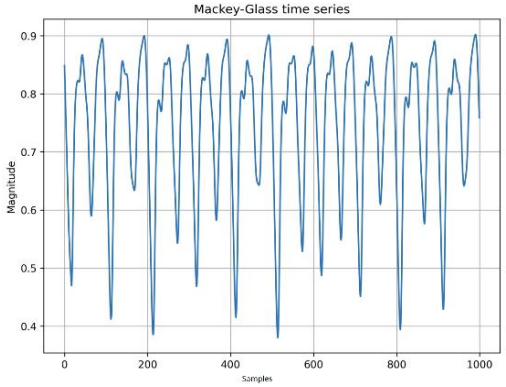


Figure 1. Graph of MG time series.

3.2. Electricity Transformer Temperature Hourly 2 (ETTh2)

The dataset is part of the Electricity Transformer Temperature (ETT) series, which is commonly used in time series forecasting and anomaly detection as displayed in Figure 2. It specifically contains temperature data from electricity transformers over time. The ETTh2 dataset monitors an electricity transformer from a region of a province in China [48], including oil temperature and variants of load from 1st July 2016 to 31st July 2016 at an hourly frequency.

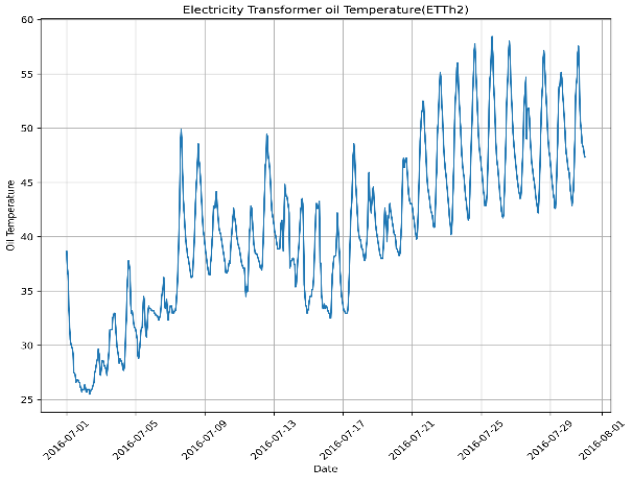


Figure 2. Graph of electricity transformer hourly oil temperature.

3.3. Coronavirus Dataset

In this study, different variants of coronavirus datasets have been taken from the Oxford Martin Programme to assess the proposed model. One set of datasets [32] consists of three variants such as daily cumulative confirmation cases, Cumulative Deaths (CD-5), and New Cases (NC4) in India between February 24, 2020, to May 20th, 2020, and another dataset [37] that consists of daily cumulative confirmation cases, CD-5 and recovery cases in India between January 30th, 2020 to August 11th, 2021.

4. Methods

In this section, examine the different phases, such as data preprocessing, the proposed methodology, and the experimental design.

4.1. Preprocessing

In the preprocessing stage, initially replace the missing values, and then the min-max transformation technique is applied for effective convergence of the model parameters. The *min-max* normalization is given in Equation (2), which converts the data from 0 to 1. The resultant scaled data is split into 80% for training and 20% for testing. Further, both are split into input and target with window size of 5 and 7.

$$y' = \frac{y - min_X}{max_X - min_X}(new\_max - new\_min) + new\_min \quad (2)$$

y'=Normalized value of y.
y=Observed value of x.
min_x=Minimal of x.
max_x=Maximal of x.
new_max=Maximal of normalized data.
new_min=Minimal of normalized data.

4.2. Proposed Methodology

The approach involves creating Tanh and Sigmoid activation functions based on Taylor expansion, followed by tuning the current model using Bayesian Optimization (BO).

4.2.1. Taylor Expansion Activation Functions

In LSTM networks, the Hyperbolic Tangent (*Tanh*) and Sigmoid activation functions are widely utilized to manage gradient flow, with the Tanh being a well-known activation function that outputs values between -1 and 1 for real number inputs, as shown in Equation (3).

$$f(x) = Tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

$x \in (-\infty, \infty)$

The Sigmoid function transforms any real number into a value between 0 and 1, as illustrated in Equation (4).

$$g(x) = sigmoid(x) = \frac{1}{(1 + e^{-x})} \quad (4)$$

Both *Tanh* and *Sigmoid* functions are susceptible to vanishing gradient issues [20], which can slow down convergence and decrease efficiency.

The Taylor series expresses a function as an infinite sum derived from its derivatives at a specific point. Vincent *et al.* [41] described a Taylor realization of softmax as sum of two terms for e^x i.e., $1+x+0.5x2$. The Taylor series for e^{-x} and e^x can be found in Equations (5) and (6).

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots + \frac{(-1)^n x^n}{n!} \quad (5)$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} \quad (6)$$

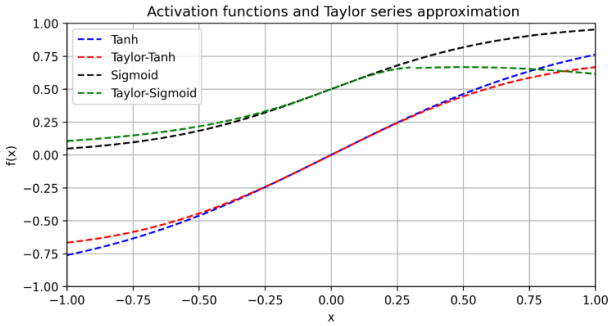


Figure 3. Graph of conventional and Taylor expansion approximations.

For single layer, the derivative loss of neuron is given by Equation (7).

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} * f'(x) * \frac{\partial x}{\partial w} \quad (7)$$

where L is the loss function, w is weights of neuron; x is the input to network and y is the output of neuron.

In n^{th} layer, gradient is accumulated and it is shown in Equation (8).

$$\frac{\partial L}{\partial w} = (\prod_{i=1}^{n} f'(x)) \frac{\partial L}{\partial yn} * \frac{\partial x}{\partial w} \quad (8)$$

Both $f'(x)$ and $g'(x)$ are less than one for Tanh and Sigmoid functions. Further, multiplying many small vales causes the gradient to decay exponentially leading to vanishing gradient.

Taylor based activation uses polynomial functions that is given by Equation (9)

$$f(x) = a_0 + a_1 x + a_2 x^2 + n a_n x^n \quad (9)$$

the derivative of $f(x)$ is as follows

$$f'(x) = (1 + O(x)) \quad (10)$$

Hence, accumulated gradient of n^{th} layer is given Equation (11)

$$\prod_{i=1}^{n} f'(x)) \approx (1 + 0(x))^n \quad (11)$$

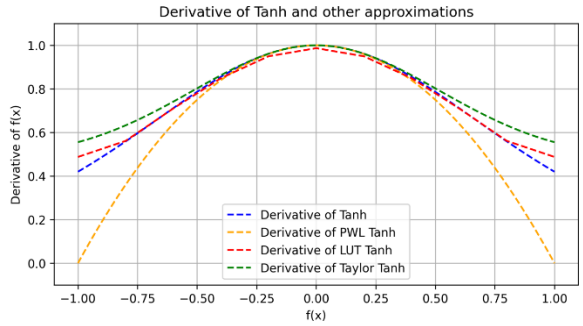
Equations (8) and (11) represent the gradients are preventing from vanishing issues.

$$g'(x) = sigmoid(x)(1 - sigmoid(x)) \quad (12)$$

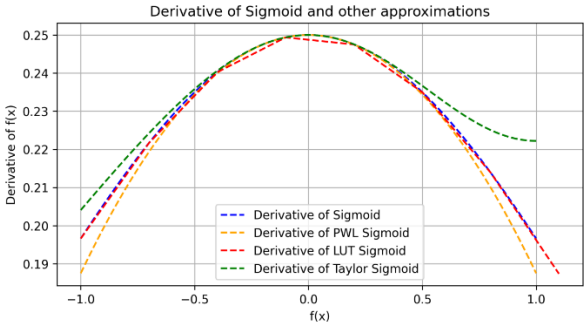
$$f'(x) = 1 - tanh^2(x) \quad (13)$$

In this study, n is set to 2 for implementing the Tanh and Sigmoid activation functions, with the Tanh activation function's Taylor expansion approximation illustrated in Figure 3. The derivatives of these determines the weights of NN using backpropagation algorithm. The derivatives of the Sigmoid and Tanh functions utilized in backpropagation are outlined in Equations (10) and (11). Taylor Tanh and Sigmoid functions offer improved gradients over their original counterparts, as illustrated

in Figure 4-a) and (b), potentially enhancing learning efficiency and mitigating the vanishing gradient issue in NN models.



a) Taylor expansion of Tanh and other approximations.



b) Taylor expansion of Sigmoid and other approximations.

Figure 4. Gradients of Taylor expansion and other approximations.

4.2.2. Proposed Model

The variants of RNNs are effective in many times series prediction [6, 12] tasks. One of the popular variants is the LSTM network, which can be used with Tanh and Sigmoid activation functions for the gating mechanism. However, LSTM and bidirectional network performance are not consistent across the different domains. Moreover, a bidirectional network consisting of two parallel networks processing input in forward and reverse order. As a result, these networks produce limited performance in many applications.

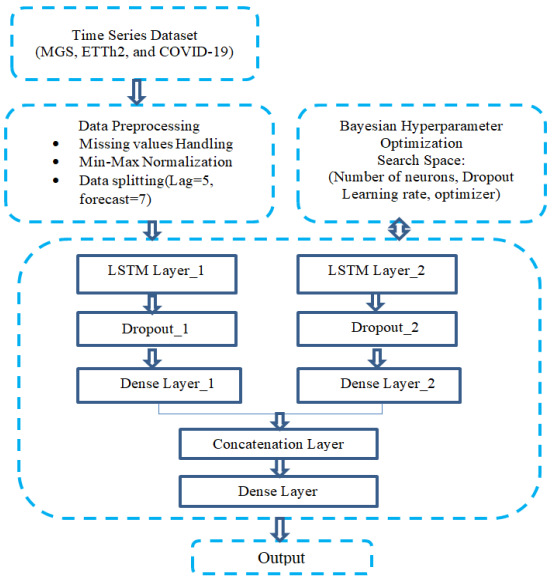


Figure 5. Architecture of proposed model.

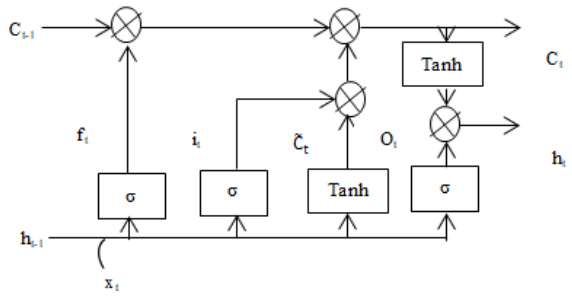


Figure 6. Structure of LSTM cell.

As a consequence, recent trends indicate that hybrid RNN models are popular in many regression tasks. In this work, a parallel arrangement of two LSTM networks as shown in Figure 5, utilizing Taylor expansion Tanh and Sigmoid activations is being proposed for multi-step time series prediction. Each LSTM uniquely captures the relationship and Taylor-based approximation activations limit vanishing gradient issues and enhance the prediction performance. The outputs from the parallel LSTM cells are concatenated and then processed by a dense network for prediction. The structure of LSTM cell is illustrated in Figure 6 and its internal operations governed by Equations (14) to (19).

$$f_t = \sigma(W_f x_t + W_f h_{t-1} + b_f) \quad (14)$$

$$i_t = \sigma(W_i x_t + W_i h_{t-1} + b_i) \quad (15)$$

$$\tilde{C}_t = tanh(W_c x_t + W_c h_{t-1} + b_c) \quad (16)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (17)$$

$$O_t = \sigma(W_O x_t + W_O h_{t-1} + b_O) \quad (18)$$

$$h_t = O_t * tanh(C_t) \quad (19)$$

Further, the model parameters were optimized using BO and determine optimal parameters in just a few iterations [11]. It is more effective than random search [7], grid, and manual tuning by experts [34]. Randomly initialized values in genetic algorithms often fail to identify optimal parameters [25], while the effectiveness of particle swarm optimization hinges on proper population initialization; in contrast, Bayesian hyperparameter optimization using Gaussian processes offers consistent performance and shorter run times compared to grid and manual search methods [45]. The current model parameters, as shown in Table 2, are optimized using BO [30] against other optimization [9, 31].

Table 2. Search space for proposed model.

Name of the parameter	Specifications
LSTM layer (number of neurons)	4-256
Dropout	0.1-0.5
Learning rate	0.001-0.1
Optimizer	Adam, RMSProp

4.2.3. Experimental Design

The proposed model with Taylor Tanh and Sigmoid activations is executed in the Google Colab Intel (R), Xenon (R) CPU@2.20GHz in the Python 3.10.12

environment, and other libraries including Tensorflow-2.15.0, Keras-2.15, BO-1.4.3, automatic Auto Regressive Integrated Moving Average (ARIMA) library pmdarima-2.0.4 and AutoML framework TPOT-0.12.2. Taylor expansion Tanh and Sigmoid functions are customized and same are utilized in the development of current model. The proposed model is fine-tuned with essential hyperparameters, including the number of neurons, learning rate, dropout, and optimizer type, while other parameters are set to 200 epochs, a batch size of 32, and a loss function of mean square error. Other models, LSTM, Gated Recurrent Unit (GRU), and transformer are tuned with the same search space by a BO with conventional Tanh and Sigmoid functions. The ARIMA model is developed using the pmdarima library, and p, d, and q values are selected using the auto_arima method for all datasets. An AutoML model is a Tree-based Pipeline Optimization Tool (TPOT) repressor, which uses genetic hyperparameter algorithm to select the best ML pipeline model for the dataset. The population and generation values are five considered for the TPOT regressor model fitting.

4.3. Evaluation Metrics

The current model and existing models' performance is assessed using metrics such as Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-Squared Score (R^2 Score).

- MAPE: it is used for measuring the accuracy of regression model by calculating the normalized difference between actual and predicted values. The range of MAPE is 0 to 1. MAPE is illustrated in Equation (20).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| * 100 \tag{20}$$

y_i, $\widehat{y_1}$ and n is the actual value, prediction value, and size of the data, respectively.

- MAE: it is another metric [22, 44] that evaluates regression models alongside MAPE. The range of MAE is from zero to infinity. *MAE* is computed as follows

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{21}$$

- R^2 score: the R^2 score, or coefficient of determination, indicates how well the data fits the regression curve [27], typically ranging from -∞ to 1, where a higher value reflects a more reliable prediction model, and can be calculated using Equation (22).

$$R^2 score = 1 - \left[\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right] \tag{22}$$

- RMSE, known as the residual, measures the prediction error by evaluating the difference between the best fit data and the actual data, as outlined in Equation (23).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{23}$$

5. Results and Discussions

The Taylor expansion Tanh and Sigmoid-based current model was assessed on bench mark chaotic Mackey-Glass Series (MGS) [33], ETT [48], Coronavirus Cumulative Confirmation (CCC-3) cases, daily NC4, CD-5 [32], CCC-6, CD-7, and Total Recovery Cases (TRC-8) [37, 38] in India.

5.1. Mackey-Glass Chaotic Series

The present model network parameters were determined using BO and the best set was used to predict the MGS a length of 7 samples.

Table 3. The performance metrics of proposed and existing on MGS for 7 sample ahead prediction.

Metrics / Model	MAPE (%)	MAE	RMSE	R^2 score
ARIMA (503)	14.1670	0.0958	0.1148	0.2062
TPOT	2.2318	0.0301	0.0235	0.9610
LSTM-BO	5.7279	0.0694	0.0508	0.8217
GRU-BO	4.4125	0.0561	0.0386	0.8961
Transformer	6.6530	0.0851	0.0571	0.7714
Pro.Model-Tanh and Sigmoid	3.3567	0.0452	0.0327	0.9231
Pro.Model-Taylor activation	3.1779	0.0423	0.0311	0.9302

The model obtains a MAPE of 3.1779, MAE of 0.0423, RMSE of 0.0311, and regression coefficient of 0.9302. The present model with Taylor expansion activation produces better performance metrics than conventional Tanh and Sigmoid functions as shown in Table 3. Further, the transition of training loss over epochs shown in Figure 8 is marginally good compared to traditional functions. The existing model, TPOT regressor produces low metrics on MGS but it is not consistent across other datasets when compared to a present model with Taylor expansion functions. For short length sequence, self-attention-based transformer unable to perform well [48] that results it yields high regression error than present model with Taylor expansion.

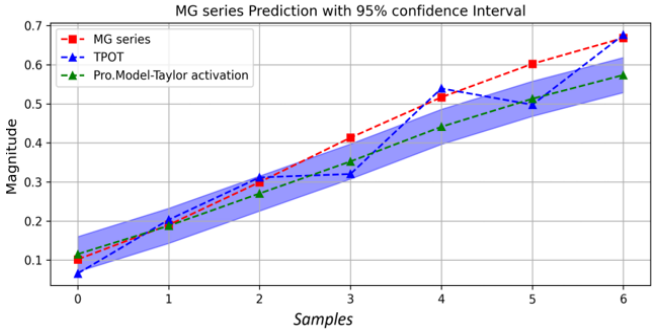


Figure 7. The proposed and TPOT model prediction on MGS.

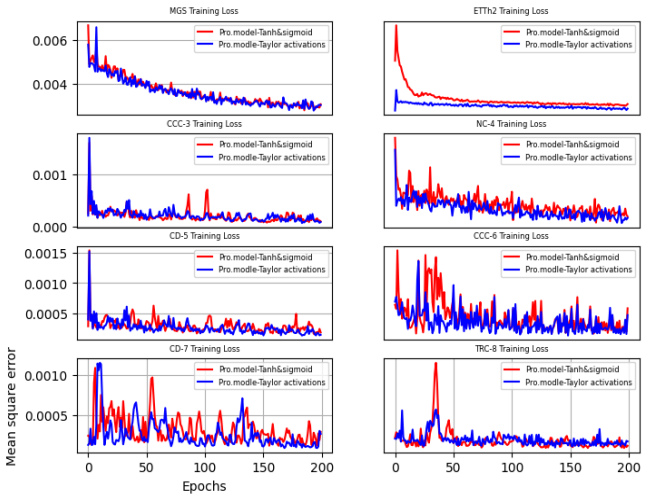


Figure 8. Training loss of Taylor expansion and conventional activation.

Additionally, the present model shows consistency in prediction on MGS, which is shown in Figure 7. The training loss of present model indicates low variation than conventional activation functions as shown in Figure 8. Hence, the present model helps to forecast the times series data with remarkable performance.

5.2. Electricity Transformer Temperature Hourly 2 (ETTh2)

Present model was assessed on another domain dataset, which is electrical transformer oil temperature on hourly basis. The Taylor expansion Tanh and Sigmoid model predicts 7-hour ahead oil temperature more effectively than other models including regular activation, TPOT and transformer. The current model achieves MAPE of 3.8234, MAE of 0.0579, RMSE of 0.0807, and regression coefficient of 0.6498.

Table 4. Current and existing models' regression metrics on ETTh2.

Metrics / Model	MAPE (%)	MAE	RMSE	R^2 score
ARIMA (212)	7.8006	3.7870	0.1333	0.0346
TPOT	5.9447	0.0915	0.1309	0.0767
LSTM-BO	5.6927	0.0882	0.1247	0.1638
GRU-BO	7.8169	0.1231	0.1639	-0.4410
Transformer	5.5644	0.0844	0.1357	0.063
Pro.Model-Tanh and Sigmoid	5.3719	0.0817	0.1130	0.3120
Pro.Model-Taylor activation	3.8234	0.0579	0.0807	0.6498

The Table 4 shows experimental metrics of present and existing models. The current model on 7-hour ahead prediction closely follows the actual data, which is appeared in Figure 9 than regular activation. The propagation loss of present model during training attains low and small fluctuation across the epochs when compared to regular activations as exhibited in Figure 8. Moreover, existing models like TPOT repressor and transformer could not attain good regression accuracy on 7-hour oil temperature prediction. Hence, proposed model can effectively use for short term prediction of electrical transformer oil temperature.

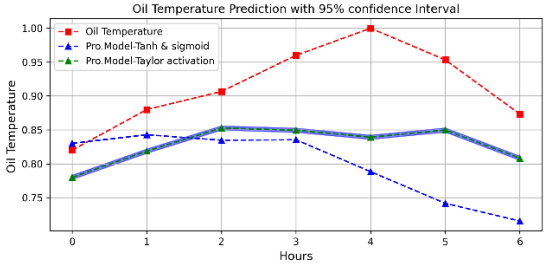


Figure 9. Current model with Taylor and conventional activation prediction on ETTh2.

5.3. Coronavirus Cumulative Confirmation (CCC-3) Cases

The present model was evaluated for cases of coronavirus in India i.e., between February 24, 2020, and May 20, 2020. For the Taylor expansion activations, the values of MAPE, MAE, RMSE, and regression coefficient obtained were 1.4570, 0.0109, 0.0130, and0.9703.

The performance metrics of the suggested and current approaches for CCC-3 prediction are displayed in Table 5. The results obtained from the current model demonstrate a notable advancement in prediction relative to the deep learning model utilizing the Grey Wolf hyperparameter Optimizer (GWO) [32]. Figure 10 illustrates the model with Taylor expansion activations and conventional activations on 7-day ahead CCC-3 forecasting. Existing models such as the ARIMA, TPOT regressor, LSTM, GRU, and transformer exhibit lower performance than the current model. Training loss of present model exhibits less fluctuations than regular activation as given in Figure 8.

Table 5. One week ahead CCC-3 prediction performance metrics of proposed and existing models.

Metrics / Model	MAPE (%)	MAE	RMSE	R^2 score
ARIMA (021)	17.0205	0.1273	0.1541	0.2986
TPOT	12.4757	0.0931	0.1769	-4.7532
LSTM-BO	2.3250	0.0163	0.0197	0.9220
GRU-BO	31.5538	0.2298	0.2388	-9.064
Transformer	5.0696	0.0379	0.0502	0.5470
Pro.Model-Tanh and Sigmoid	2.0671	0.0152	0.0167	0.9516
Pro.Model-Taylor activation	1.4570	0.0109	0.0130	0.9703

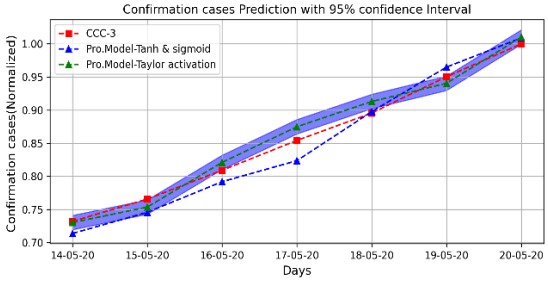


Figure 10. One week ahead CCC-3 prediction of proposed model with Taylor Tanh and Sigmoid.

5.4. Coronavirus Daily New Cases (NC-4)

The model was assessed on COVID-19 new cases in

India, the values of MAPE, MAE, RMSE, and regression coefficient were 9.2375, 0.0646, 0.0817, and 0.1756, respectively. The performance metrics of the existing and current approaches for 7-day ahead NC forecasting is displayed in Table 6. The results obtained from the current model demonstrate a notable advancement in prediction relative to the deep learning model utilizing the GWO [32]. Figure 11 illustrates the model with Taylor expansion activations and conventional activations on 7-day ahead NC-4 forecasting. Despite, the nonlinear behavior of new cases current model can closely approximate than conventional as well as existing models. Figure 8 shows the loss of the current and conventional activations.

Table 6. One week ahead NC-4 prediction regression metrics of proposed and existing models.

Metrics / Model	MAPE (%)	MAE	RMSE	R^2 score
ARIMA (013)	33.4896	0.2358	0.2738	-2.5984
TPOT	28.1645	0.2048	0.2729	-5.4537
LSTM-BO	10.3363	0.0703	0.0860	0.1156
GRU-BO	9.5522	0.0698	0.0966	- 0.1494
Transformer	18.3185	0.1144	0.1323	-1.5857
Pro.Model-Tanh and Sigmoid	9.3258	0.0648	0.0826	0.1737
Pro.Model-Taylor activation	9.2375	0.0646	0.0817	0.1756

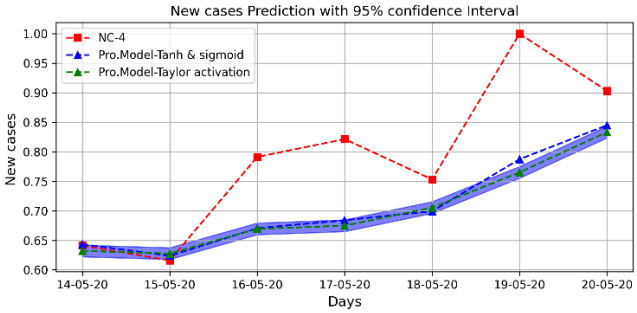


Figure 11. One week ahead NC-4 prediction of proposed model with Taylor Tanh and Sigmoid.

5.5. Cumulative Deaths (CD-5)

Another coronavirus univariate CD-5 dataset was used for interpreting the current network. The values of MAPE, MAE, RMSE, and regression coefficient obtained were 1.8657, 0.0144, 0.0180, and 0.9315. Table 7 represents the regression metrics of current and existing methods for forecasting 7-day ahead deaths. The model with Taylor obtains lower metrics when compared to existing models as shown in Figure 12.

Table 7. One week ahead CD-5 prediction metrics of proposed model and other models.

Metrics / Model	MAPE (%)	MAE	RMSE	R^2 score
ARIMA (220)	10.5459	0.0804	0.0931	0.7137
TPOT	21.0852	0.1687	0.2608	-13.002
LSTM-BO	3.1877	0.0237	0.0289	0.8072
GRU-BO	9.2664	0.0681	0.0838	-0.5595
Transformer	4.0115	0.0301	0.0408	0.6288
Pro.Model-Tanh and Sigmoid	2.1242	0.0169	0.0212	0.9156
Pro.Model-Taylor activation	1.8657	0.0144	0.0180	0.9315

The results obtained from the current model demonstrate a notable advancement in prediction relative to the deep learning model utilizing the GWO [32]. Figure 8 shows the loss of the current and conventional activations on CD-5.

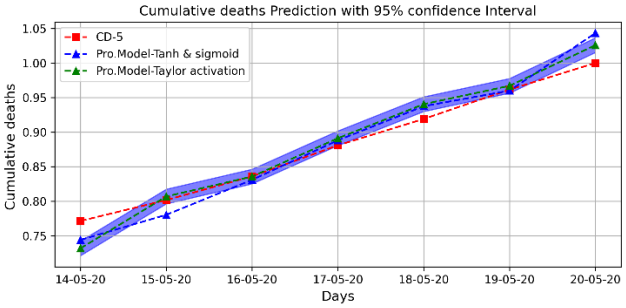


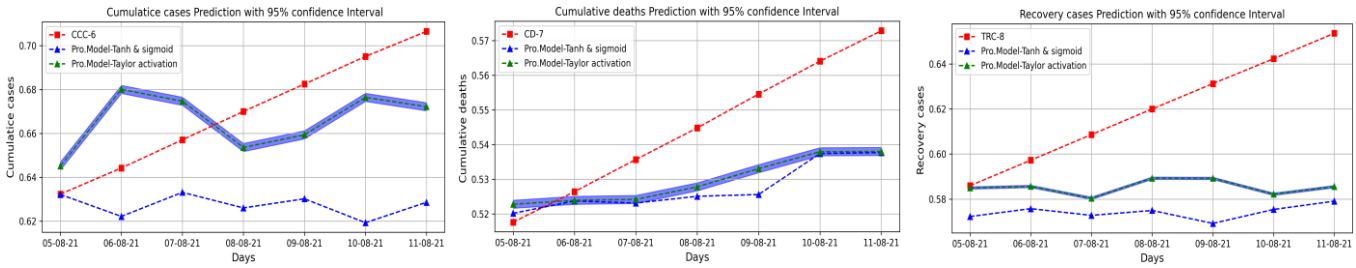
Figure 12. One week ahead prediction on CD-5 in India.

5.6. Coronavirus Cumulative Confirmation (CCC-6) Cases

One more univariate dataset CCC-6 was considered from 30th January, 2020 to 11th August 2021 for analyzing current network with Taylor expansion activations using Bayesian hyperparameter optimization. The model was obtained 2.0655, 0.0180, 0.0224, and 0.9555, MAPE, MAE, RMSE and R^2 score, respectively.

Table 8. One week ahead CCC-6 prediction metrics of proposed and existing models.

Metrics / Model	MAPE (%)	MAE	RMSE	R^2 score
ARIMA (020)	20.4712	0.1963	0.2740	-3.1329
TPOT	2.4388	0.0192	0.0261	0.9330
LSTM-BO	2.8126	0.0257	0.0348	0.8851
GRU-BO	15.3179	0.1362	0.1380	-0.8294
Transformer	9.3665	0.0823	0.0832	0.3293
Pro.Model-Tanh and Sigmoid	2.7505	0.0243	0.0297	0.9143
Pro.Model-Taylor activation	2.0655	0.0180	0.0224	0.9555



a) One week ahead CCC-6 prediction. b) One week ahead CD-7 prediction. c) One week ahead TRC-8 prediction.

Figure 13. Coronavirus cumulative cases one week ahead prediction.

Table 8 exhibits current and existing methodologies regression metrics for 7-day ahead forecasting. The model with Taylor approximation outperforms than existing methods, including conventional activations as displayed in Figure 13-a). Training loss of both current and conventional activations is displayed in Figure 8.

5.7. Cumulative Deaths (CD-7)

The univariate dataset CD-7 was used for the assessment of the current network with Taylor activations for 7-day ahead forecasting. On CD-7, model efficiency was determined in terms of MAPE, MAE, RMSE, and regression coefficient of 1.6314, 0.0123, 0.0159, and 0.9882, respectively. The present network and existing techniques metrics are shown in Table 9. It indicates notable regression accuracy when compared to existing methods. The model with Taylor approximation outperforms existing methods, including conventional activations as displayed in Figure 13-b). Figure 8 reveals the training loss of proposed and regular activation.

Table 9. One week ahead CD prediction metrics of proposed model and other models.

Metrics / Model	MAPE (%)	MAE	RMSE	R^2 score
ARIMA (021)	14.9573	0.1260	0.1389	0.3798
TPOT	41.9185	0.3640	0.4138	-6.5578
LSTM-BO	14.7701	0.1285	0.1371	0.1824
GRU-BO	20.3843	0.1731	0.1795	-0.4204
Transformer	3.7927	0.0327	0.0356	0.9441
Pro.Model-Taylor and Sigmoid	1.8060	0.0137	0.0175	0.9858
Pro.Model-Taylor activation	1.6314	0.0123	0.0159	0.9882

5.8. Total Recovery Cases (TRC-8)

On TRC-8, the current model produces MAPE, MAE, RMSE, and regression coefficient of 2.3395, 0.0189, 0.0247, and 0.9565, respectively. Table 10 resembles regression accuracy metrics of current as well as existing techniques for 7-day ahead recovery cases forecasting. It shows the scope of the current model in forecasting is remarkable when compared to others. As a consequence, the role of the present model in forecasting coronavirus diseases is remarkable. Figure 13-c) reveals the one-week ahead prediction by the present model and other models. Training loss of both current and conventional activations is displayed in Figure 8.

Table 10. Performance metrics of current and existing models on TRC-8.

Metrics / Model	MAPE (%)	MAE	RMSE	R^2 score
ARIMA (021)	30.6506	0.2954	0.4150	-6.2878
TPOT	3.1081	0.0231	0.0331	0.9237
LSTM-BO	5.4033	0.0477	0.0506	0.8377
GRU-BO(6gru)	11.7886	0.1031	0.1071	0.2102
Transformer	2.6536	0.0214	0.0311	0.9337
Pro.Model-Tanh and Sigmoid	2.6368	0.0204	0.0283	0.9426
Pro.Model-Taylor activation	2.3395	0.0189	0.0247	0.9565

5.9. Model Learning Overheads

Learning time of NN depends on number of model parameters and other specifications. The Training complexity of proposed Taylor expansion and conventional activations in shown in Figure 14. On MGS, learning time of Taylor expansion and regular activations are 471 and 349 seconds, respectively for 135804 model parameters. The model with regular activations shows reduction in learning time on majority of datasets than Taylor expansion approximation. It might be due to exponent operations of specific dataset during approximation.

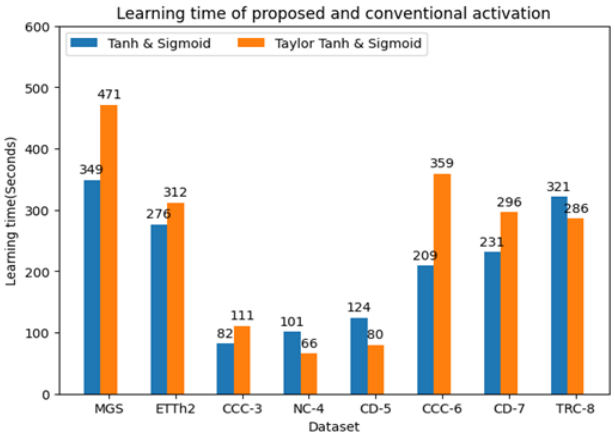


Figure 14. Computational overheads of proposed and conventional activation on datasets.

6. Conclusions and Future Scope

In this study, a new model based on Taylor expansion activation was developed and assessed on different univariate time series datasets. The findings indicate that the suggested framework utilizing the Taylor approximation activation function yields greater consistency in predictions compared to traditional activation functions like Tanh and Sigmoid. Additionally, the current methodology prevents the vanishing gradients issue in NNs. However, TPOT and transformer results reveal competitive performance on a few datasets. The transformer performance is more consistent than the TPOT regressor. The statistical ARIMA model produces high MAPE, MAE, RMSE, and low R^2 scores on all datasets. Hence, the present model can be utilized for short-term prediction where little past information is known and provides a diverse scope for developing the activation functions to avoid vanishing gradients. Conversely, the study provides a direction for improving the gradients of activation functions to prevent learning issues and reduce the exponent operations.

However, one of the challenges is hyperparameter tuning search space is limited due to the availability of processor and memory resources. On the other hand, computation overheads of current model with custom activation functions. In the future, the proposed methodology can be extended for multivariate time

series datasets and long-term prediction. Simultaneously, BO can be integrated with other well-known Hyper-Parameter Optimization (HPO) techniques to facilitate parallel processing.

References

[1] Abbasimehr H. and Paki R., "Prediction of COVID-19 Confirmed Cases Combining Deep Learning Methods and Bayesian Optimization," *Chaos, Solitons and Fractals*, vol. 142, pp. 110511, 2021. https://doi.org/10.1016/j.chaos.2020.110511

[2] Ajagbe S. and Adigun M., "Deep Learning Techniques for Detection and Prediction of Pandemic Diseases: A Systematic Literature Review," *Multimedia Tools and Applications*, vol. 83, no. 2, pp. 5893-5927, 2024. https://doi.org/10.1007/s11042-02315805-z

[3] Al-kateeb S., "Forecasting the Spread of Viral Diseases in Jordan Using the SARIMA Statistical Model," *The International Arab Journal of Information Technology*, vol. 21, no. 6, pp. 987-995, 2024. https://doi.org/10.34028/iajit/21/6/3

[4] Bamber S. and Vishvakarma T., "Medical Image Classification for Alzheimer's Using a Deep Learning Approach," *Journal of Engineering and Applied Science*, vol. 70, no. 1, pp. 1-18, 2023. https://doi.org/10.1186/s44147-023-00211-x

[5] Banerjee K., Vishak Prasad C., Gupta R., Vyas K., Anushree H., and Mishra B., "Exploring Alternatives to Softmax Function," *arXiv Preprint*, vol. arXiv:2011.11538v1, pp. 1-8, 2021. https://arxiv.org/abs/2011.11538

[6] Bekkar A., Hssina B., Douzi S., and Douzi K., "Air-Pollution Prediction in Smart City, Deep Learning Approach," *Journal of Big Data*, vol. 8, no. 1, pp. 1-21, 2021. https://doi.org/10.1186/s40537-021-00548-1

[7] Bergstra J., Bardenet R., Bengio Y., and Kegl B., "Algorithms for Hyper-Parameter Optimization," *in Proceedings of the Proceedings of the 25th International Conference on Neural Information Processing Systems*, Granada, pp. 2546-2554, 2011. https://dl.acm.org/doi/10.5555/2986459.2986743

[8] Cetin O., Temurts F., and Gulgonul E., "An Application of Multilayer Neural Network on Hepatitis Disease Diagnosis Using Approximation of Sigmoid Function," *Dicle Medical Journal*, vol. 42, no. 2, pp. 150-157, 2015. https://pdfs.semanticscholar.org/4617/fa90b37661e70b512fa7ff67e5d216b106bc.pdf

[9] Chandra M., "Hardware Implementation of Hyperbolic Tangent Function Using Catmull-Rom Spline Interpolation," *arXiv Preprint*, vol. arXiv:2007.13516v1, pp. 1-4, 2020. https://doi.org/10.48550/arxiv.2007.13516

[10] De Ryck T., Lanthaler S., and Mishra S., "On the Approximation of Functions by Tanh Neural Networks," *Neural Networks*, vol. 143, pp. 732-750, 2021. https://doi.org/10.1016/j.neunet.2021.08.015

[11] DeCastro-Garcia N., Castaneda A., Garcia D., and Carriegos M., "Effect of the Sampling of a Dataset in the Hyperparameter Optimization Phase over the Efficiency of a Machine Learning Algorithm," *Complexity*, vol. 2019, no. 1, pp. 1-16, 2019. https://doi.org/10.1155/2019/6278908

[12] Drewil G. and Al-Bahadili R., "Air Pollution Prediction Using LSTM Deep Learning and Meta Heuristics Algorithms," *Measurement: Sensors*, vol. 24, pp. 100546, 2022. https://doi.org/10.1016/j.measen.2022.100546

[13] Dubey S., Singh S., and Chaudhuri B., "Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark," *Neuro Computing*, vol. 503, pp. 92-108, 2022. https://doi.org/10.1016/j.neucom.2022.06.111

[14] Effrosynidis D., Spiliotis E., Sylaios G., and Arampatzis A., "Time Series and Regression Methods for Univariate Environmental Forecasting: An Empirical Evaluation," *Science of the Total Environment*, vol. 875, pp. 162580, 2023. https://doi.org/10.1016/j.scitotenv.2023.162580

[15] Fernandez A. and Mali A., "TELU Activation Function for Fast and Stable Deep Learning," *arXiv Preprint*, vol. arXiv:2412.20269v1, pp. 1-80, 2024. https://arxiv.org/html/2412.20269v1

[16] Global Development, Oxford Martin School, https://www.oxfordmartin.ox.ac.uk/global-development, Last Visited, 2024.

[17] Han Y. and Meng S., "Machine English Translation Evaluation System Based on BP Neural Network Algorithm," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, pp. 1-10, 2022. https://doi.org/10.1155/2022/4974579

[18] Hochreiter S. and Schmidhuber J., "Long Short Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[19] Hochreiter S., "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions," *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, vol. 6, no. 2, pp. 107-116, 1998. https://doi.org/10.1142/S0218488598000094

[20] Hu Z., Zhang J., and Ge Y., "Handling Vanishing Gradient Problem Using Artificial Derivative," *IEEE Access*, vol. 9, pp. 22371-22377, 2021. DOI: 10.1109/ACCESS.2021.3054915

[21] Hwang S. and Kim J., "A Universal Activation Function for Deep Learning," *Computers, Materials and Continua, Materials and Continua*,

vol. 75, no. 2, pp. 3553-3569, 2023. https://doi.org/10.32604/cmc.2023.037028

[22] Kaundal R., Kapoor A., and Raghava G., "Machine Learning Techniques in Disease Forecasting: A Case Study on Rice Blast Prediction," *BMC Bioinformatics*, vol. 7, no. 1, pp. 1-16, 2006. https://doi.org/10.1186/1471-2105-7-485

[23] Khagi B. and Kwon G., "A Novel Scaled-Gamma-Tanh (SGT) Activation Function in 3D CNN Applied for MRI Classification," *Scientific Reports*, vol. 12, no. 1, pp. 1-14, 2022. https://www.nature.com/articles/s41598-022-19020-y

[24] Kunc V. and Klema J., "Three Decades of Activations: A Comprehensive Survey of 400 Activation Functions for Neural Networks," *arXiv Preprint*, vol. arXiv:2402.09092v1, pp. 1-107, 2024. https://arxiv.org/abs/2402.09092

[25] Lessmann S., Stahlbock R., and Crone S., "Optimizing Hyperparameters of Support Vector Machines by Genetic Algorithms," *in Proceedings of the International Conference on Artificial Intelligence*, Las Vegas, pp. 74-80, 2005. file:///C:/Users/user/Downloads/Optimizing_Hyperparameters_of_Support_Vector_Machi%20(1).pdf

[26] Li M., Jiang Y., Zhang Y., and Zhu H., "Medical Image Analysis Using Deep Learning Algorithms," *Frontiers in Public Health*, vol. 11, pp. 1273253, 2023. DOI: 10.3389/fpubh.2023.1273253

[27] Lupon J., Gaggin H., De Antonio M., Domingo M., and et al., "Biomarker-Assist Score for Reverse Remodeling Prediction in Heart Failure: The ST2-R2 Score," *International Journal of Cardiology*, vol. 184, pp. 337-343, 2015. https://doi.org/10.1016/j.ijcard.2015.02.019

[28] Mackey M. and Glass L., "Oscillation and Chaos in Physiological Control Systems," *Science*, vol. 197, pp. 287-289, 1997. DOI: 10.1126/science.267326

[29] Mahaur B., Mishra K., and Singh N., "Improved Residual Network Based on Norm Preservation for Visual Recognition," *Neural Networks*, vol. 157, pp. 305-322, 2023. https://doi.org/10.1016/j.neunet.2022.10.023

[30] Nogueria F., A Python Implementation of Global Optimization with Gaussian Processes, GitHub, https://github.com/fmfn/BayesianOptimization, Last Visited, 2024.

[31] Ozturk M., "Hyperparameter Optimization of a Parallelized LSTM for Time Series Prediction," *Vietnam Journal of Computer Science*, vol. 10, no. 3, pp. 303-328, 2023. https://doi.org/10.1142/S2196888823500033

[32] Prasanth S., Singh U., Kumar A., Tikkiwal V., and Chong P., "Forecasting Spread of COVID-19 Using Google Trends: A Hybrid GWO-Deep Learning Approach," *Chaos, Solitons and Fractals*, vol. 142, pp. 110336, 2021. https://doi.org/10.1016/j.chaos.2020.110336

[33] Shahi S., Fenton F., and Cherry E., "Prediction of Chaotic Time Series Using Recurrent Neural Networks and Reservoir Computing Techniques: A Comparative Study," *Machine Learning with Applications*, vol. 8, pp. 100300, 2022. https://doi.org/10.1016/j.mlwa.2022.100300

[34] Snoek J., Larochelle H., and Adams R., "Practical Bayesian Optimization of Machine Learning Algorithms," *in Proceedings of the 26th International Conference on Neural Information Processing Systems*, Nevada, pp. 2951-2959, 2012. https://dl.acm.org/doi/10.5555/2999325.2999464

[35] Temurtas F., Gulbag A., and Yumusak N., "A Study on Neural Networks Using Taylor Series Expansion of Sigmoid Activation Function," *in Proceedings of the Computational Science and its Applications*, Assisi, pp. 389-397, 2004. DOI:10.1007/978-3-540-24768-5_41

[36] Timmons N. and Rice A., "Approximating Activation Functions," *arXiv Preprint*, vol. arXiv:2001.06370v1, pp. 1-10, 2020. https://arxiv.org/abs/2001.06370

[37] Tirupati G., Murali K., and Peri S., "An Improved Parallel Heterogeneous Long Short-Term Model with Bayesian Optimization for Time Series Prediction," *International Journal of Experimental Research and Review*, vol. 45, pp. 106-118, 2024. https://doi.org/10.52756/ijerr.2024.v45spl.009

[38] Tirupati G., Murali K., and Peri S., "COVID-19 Prediction Modeling Using Bidirectional Gated Recurrent Unit Network Model," *Journal of Webology*, vol. 18, no. 5, pp. 15-41, 2021. file:///C:/Users/user/Downloads/2022022603092 4pmwebology185-52.pdf

[39] Tun N. and Myat A., "Proposed Activation Function Based Deep Learning Approach for Real-Time Face Cover Detection System," *Authorea*, pp. 1-7, 2024. DOI:10.22541/au.172449788.82658502/v1

[40] Vijayaprabakaran K. and Sathiyamurthy K., "Towards Activation Function Search for Long Short-Term Model Network: A Differential Evolution-based Approach," *Journal of King Saud University Computer and Information Sciences*, vol. 34, no. 6, pp. 2637-2650, 2020. https://doi.org/10.1016/j.jksuci.2020.04.015

[41] Vincent P., De Brebisson A., and Bouthillier X., "Efficient Exact Gradient Update for Training Deep Networks with very Large Sparse Targets," *arXiv Preprint*, vol. arXiv:1412.7091v3, pp. 1-15, 2015. https://arxiv.org/abs/1412.7091

[42] Wang S., Liu B., and Liu F., "Escaping the

Gradient Vanishing: Periodic Alternatives of Softmax in Attention Mechanism," *IEEE Access*, vol. 9, pp. 168749-168759, 2021. https://ieeexplore.ieee.org/document/9662308

[43] Wei L., Cai J., Nguyen V., Chu J., and Wen K., "P-SFA: Probability based Sigmoid Function Approximation for Low Complexity Hardware Implementation," *Microprocessor and Micro Systems*, vol. 76, pp. 417, 2020. https://doi.org/10.1016/j.micpro.2020.103105

[44] Willmott C. and Matsuura K., "Advantages of The Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance," *Climate Research*, vol. 30, no. 1, pp. 79-82, 2005. https://www.jstor.org/stable/24869236

[45] Wu J., Chen X., Zhang H., Xiong L., Lei H., and Deng S., "Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization," *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26-40, 2019. https://doi.org/10.11989/JEST.1674-862X.80904120

[46] Zaki P., Hashem A., Fahim E., Mansor M., and et al., "A Novel Sigmoid Function Approximation Suitable for Neural Networks on FPGA," *in Proceedings of the 15th International Computing Engineering Conference*, Cairo, pp. 95-99, 2019. DOI: 10.1109/ICENCO48310.2019.9027479

[47] Zhang J., He T., Sra S., and Jadbabaie A., Why Gradient Clipping Accelerates Training: A Theoretical Justification for Adaptivity, https://github.com/JingzhaoZhang/why-clipping-accelerates, Last Visited, 2024.

[48] Zhou H., Zhang S., Peng J., Zhang S., and et al., "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting," *in Proceedings of the AAAI Conference on Artificial Intelligence*, Virtual, pp. 11106-11115, 2021. https://doi.org/10.1609/aaai.v35i12.17325

Tirupati Gullipalli obtained his B.Tech degree from BVRIT affiliated to JNTUH, Hyderabad, India. He pursued a Master's degree in Computer Science and Technology at Andhra University. Currently, pursuing PhD in Computer Science and Engineering from JNTU, Kakinada, India. His research interests are Data Mining, Image Processing, and Machine Learning.

Krishna Murali is currently a Professor of the Department of Computer Science and Engineering, University College of Engineering Kakinada (Autonomous), JNTUK, Andhra Pradesh, India. He has about 50+ research papers in various International Journals and Conferences, and attended many national and international conferences in India and abroad. He is a member of Association for Computing Machinery (ACM), ISTE and IAENG (Germany) is an active member of the board of reviewers in various International Journals and Conferences. His research interests include Data Mining, Big Data Analytics and High-Performance Computing.

Srinivasa Peri is currently Associate Professor in the Department of Computer Science and Engineering, MVGR College of Engineering (A), India. He obtained a PhD degree from JNTU, Kakinada, India. He has made substantial contributions in the area of Big Data and Machine Learning. His research includes Data Warehousing and Mining, Distributed Computing and Image Processing.