Using fractional derivative in learning algorithm for artificial neural network: Application for salary prediction

Manisha Joshi
Department of Electronics and Telecommunication Engineering
Ramrao Adik Institute of Technology
Nerul, 400706, Navi Mumbai, India
manisha.joshi@ves.ac.in

Savita Bhosale
Department of Electronics Engineering
Ramrao Adik Institute of Technology
D Y Patil Deemed to be University
Nerul, 400706, Navi Mumbai, India
savita.bhosale@rait.ac.in

Vishwesh A. Vyawahare
Department of Electronics Engineering
Ramrao Adik Institute of Technology
D Y Patil Deemed to be University
Nerul, 400706, Navi Mumbai, India
vishwesh.vyawahare@rait.ac.in

***Abstract*—Fractional calculus has been adopted in the modelling of many scientific processes and systems. Due to the inherent feature of long term memory of fractional derivatives, it has been used in the learning process of neural networks. A fractional order derivative based back propagation learning algorithm in neural networks is proposed in this paper. Specifically, Riemann-Liouville (R-L), Caputo (C) and Caputo Fabrizio (CF) fractional Derivative based on the back propagation algorithms in a three layer feed-forward neural network employed. To get a faster learning rate without oscillation, momentum factor is incorporated. The effect of fractional order and momentum factor is investigated and compared. The performance of these fractional derivatives based algorithms with integer derivatives based algorithm in terms of mean square error (MSE), particularly the salary based on years of experience is predicted. Results demonstrate that fractional derivative based learning algorithms outperform the integer derivatives.**

***Index Terms*—back propagation, fractional derivative, momentum factor, perceptron**

I. Introduction

Artificial neural networks (ANNs) are computer systems that are modeled using the concept of biological neural networks. ANNs began as an attempt to use the architecture of the human brain to perform tasks that traditional algorithms failed to do well. These networks specifically worked on improving empirical results. Because of this ANNs can be used for various applications like image processing, system identification, classification of datasets, image recognition, mathematical modeling, data analysis, data mining, pattern recognition, etc. In order to improve performance further, perceptron, a type of supervised ANN and fractional calculus (FC) can be combined. Specifically, FC is used as a derivative in the back propagation algorithm of perceptrons. In this paper we employ R-L, C and CF fractional derivative in learning algorithm for salary prediction.

II. Literature Review

FC was first proposed by Gottfried Wilhelm Leibniz in 1695, and it was also mentioned in Niels Henrik Abel's works. Over the nineteenth and twentieth centuries, the theory and applications of FC grew significantly, and many contributions provided definitions for fractional derivatives and integrals. Caputo, R-L and G-L are the most commonly used definitions of FC [1]–[4]. Caputo fractional derivatives have advantages compared to R-L fractional derivatives. The initial conditions for time-fractional Caputo derivatives are the same as for integer-order differential equations [5], [6]. A literature survey reveals applications of new fractional derivatives like CF in various domains like modeling of diseases in animals [7], cancer treatment models [8], epidemic models for childhood diseases [9] etc.

Various methods are employed for prediction of the performance of phenomena such as college attrition [10], employee attrition [11], employee turnover [12], student's dropout rates using data mining [13] and Neural Network [14], bank failures using artificial intelligence methods [15] and many more.

Multilayer perceptrons are designed for prediction of the performance of various phenomena like finance [16], student's results of final exams [17], epidemiological data [18], employee attrition [19], etc.

A literature survey reveals applications of FC and perceptrons in various fields like energy [20], digital signal processing [21], bio-medical [22], [23], image processing [24], finance [25], system identification [25], controls [26], etc.

Salary is the most important criteria for selection of a job. However, if employee can accurately predict the salary of the job posting based on their year of experience, they can decide to continue the job or search for a new job.

A multilayer perceptron is employed for the prediction of the salary of employees based on years of experience. To train these perceptrons various back propagation algorithms such as delta rule, generalized delta rule etc are employed. In these algorithms, the weights are updated in order to achieve the desired target [27], [28].

Recently FC based perceptrons are proposed to update weights, these algorithms provide optimization procedures for finding local minima for differential functions. In [29]–[31],

978-1-6654-9291-1/22/$31.00 ©2022 IEEE

Caputo's fractional derivative is applied in the gradient descent method and optimization. Based on the literature survey, this work might be the first attempt at the application of Caputo-Fabrizio's fractional derivative in the back propagation learning algorithm.

In this paper, a perceptron based on Caputo's fractional derivative, R-L fractional derivative and CF fractional derivative in the back propagation learning algorithm is proposed to predict the salary based on the years of experience. Results are compared with the integer order back propagation learning algorithm. The performance of perceptrons is measured in terms of Mean Square Error (MSE). To prove the applicability of the model, an illustrative example for salary prediction is demonstrated. From this analysis it is proved that fractional based derivative provide less MSE in comparison to integer order derivative.

The paper is organized as follows: in Section III, the definitions and simple properties of FC are introduced. In Section IV, the proposed fractional-order three-layer Back Propagation neural networks with fractional derivatives are presented. Subsequently, Section V, presents simulation results to illustrate the precision and effectiveness of our proposed model. Finally, the proposed work is concluded in Section VI.

III. Preliminaries

A. Fractional Derivatives

Many of the real-world problems can be identified and described by the fractional order model. The FC has been discovered to be a very effective mathematical tool for modeling a wide range of real-world and engineering systems. It is observed that information can be processed efficiently in fractional order systems. They improve the simulation performance of the integer order system and find more accurate results. In this work, the RL, the Caputo and CF derivative definitions are used. In this section, these definitions are presented, which are further used further in the salary prediction model based on Artificial Neural Networks (ANN).

1) Riemann-Liouville fractional derivative: The Riemann-Liouville fractional derivative of order $\alpha > 0$ of a function $f(t) : (0, \infty \rightarrow R)$ is given as [32],

$$D^{\alpha} f(t) = \frac{d^n}{dt^n} \frac{1}{\Gamma(n-\alpha)} \int_0^t (t-T)^{n-\alpha-1} f(T) dT, \quad (1)$$

where n is integer, α is a real number and Γ is a gamma operator.

2) Caputo Fractional Derivative: The Caputo fractional derivative of order $\alpha > 0$ of a function $f(t) : (0, \infty \rightarrow R)$ is given as [32],

$$D^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t (t-T)^{n-\alpha-1} \frac{d^n}{dt^n} f(T) dT. \quad (2)$$

The Caputo operator is also linear, but this definition is more restrictive than the RL due to the requirement of absolute integrability of the n^{th} order derivative of the function $f(t)$. When $n = 1$, it is given as,

$$D^{\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-T)^{-\alpha} f^{'}(T) dT. \quad (3)$$

3) Modified Caputo's fractional time derivative by Caputo and Fabrizio: In [33], kernel $(t-T)^{-\alpha}$ and $\frac{1}{\Gamma(1-\alpha)}$ are replaced with $exp(-\frac{\alpha}{(1-\alpha)}t)$ and $\frac{M(\alpha)}{1-\alpha}$ respectively.

$$D^{\alpha} f(t) = \frac{Z(\alpha)}{(1-\alpha)} \int_0^t exp[-\alpha(\frac{t-T}{1-\alpha})] f^{'}(T) dT, \quad (4)$$

where $Z(\alpha)$: Normalization function such that $Z(0) = Z(1) = 1$. Expression 4 is referred as Caputo-Fabrizio fractional time derivative. It is also known as Modified fractional time derivative.

B. Multilayer Feed-forward Artificial Neural Network

The basic architecture of multilayer feed-forward artificial neural nets consisting of an input layer, a hidden layer and an output layer is illustrated in Fig.1. The training of these networks consists of three phases: Feed-forward, Error Back propagation, Weight and Bias updates.

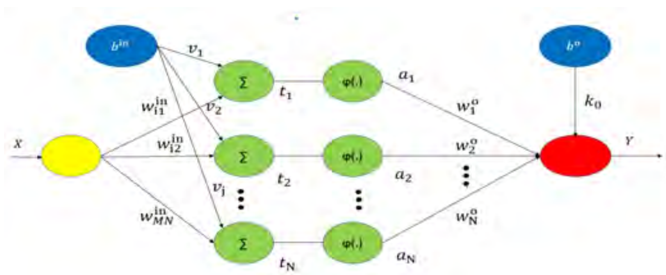

Fig. 1. Architecture of multi-layered perceptron [34]

1) Feed-forward phase: In this phase, input is passed through the neural network in forward direction. Let there be one input node, one hidden layer consisting of N number of neurons(nodes) and one output node in the neural network. In this paper nodes and neurons term will be used interchangeably. The input $X = [X_1, \cdots, X_M]$ is fed to neural network and output $Y(X, w_{ij}^{in}, w_j^o)$ is expressed as,

$$Y(X, w_{ij}^{in}, w_j^o) = \sum_{j=1}^{N} a_j w_j^0 + k_o b_o, \quad (5)$$

$$a_j = \varphi(t_j), j = 1, \cdots, N, \quad (6)$$

$$t_j = \sum_{i=1}^{M} X_i w_{ij}^{in} + v_j b^{in}, i = 1, \cdots, N, \quad (7)$$

$$\varphi(t_j) = \frac{2}{1+e^{-2t_j}} - 1, \quad (8)$$

where,

X_i: Input Vector,$X = [X_1, \cdots, X_M]$, where M is the size of input data

w_{ij}^{in}: The weight applied between input node in node j in hidden layer node, $i = 1, \cdots, M$, $j = 1, \cdots, N$
a_j: output of j^{th} Neuron in hidden layer after applying activation function. In literature many activation functions are defined. They are differentiable and increase monotonically.
w_j^o: The weight applied between node j in hidden layer and output layer, $j = 1, \cdots, N$
b^o, b^{in}: Bias applied at output layer and input layer respectively
k_o: The weight applied between output bias unit and node in output layer
v_j : The weight applied between input bias unit and j^{th} node in hidden layer
$\varphi(t_j)$: The activation function applied on t_j

2) Error Back-propagation phase: The output Y of a neural network is not the same as input X, so there will always be some error. In order to minimize this error and to get the correct output, error is feedback to input via hidden layer and this phase is called back propagation phase. Error back propagation learning algorithm has been used to update weights and thus minimizes the error function of the ANN. Here the gradient descent method has been used for updating the parameters [34].

3) Weight and bias updation: Let $w_{ij}^{in}(old)$ and $w_j^o(old)$ are the previous values of weights at input and output layers respectively. These old weights are updated by adding correction terms obtained from back propagation phase. Then $w_{ij}^{in}(new)$ and $w_j^o(new)$ are the new values of weights at input and output layers respectively, They are evaluated as follows,

$$w_j^o(new) = w_j^o(old) + \Delta w_j^o, \tag{9}$$

$$w_{ij}^{in}(new) = w_{ij}^{in}(old) + \Delta w_{ij}^{in}, \tag{10}$$

Similarly old bias weights are updated to and represented as follows,

$$b^o(new) = b^o(old) + \Delta b^o, \tag{11}$$

$$b^{in}(new) = b^{in}(old) + \Delta b^{in}, \tag{12}$$

Correction terms at output and input (Δw_j, Δw_{ij}^{in}, Δb^o and Δb^{in}) are evaluated by using backpropagation algorithm. These weights and bias are updated till a minimum error is achieved. MSE function is used to check performance of perceptrons [34].

IV. Proposed Model

In this paper, a novel salary prediction algorithm using ANN approach and fractional derivative is proposed. A three layer feed-forward architecture as described in the last section for salary prediction is implemented. Here, the neural network is trained to predict the salary based on the years of experience. Following modifications are proposed,

1) **Weight and bias update process based on fractional derivative**: In all neural network applications based on back propagation algorithms, weight and bias computation involves derivatives of activation function. Instead of choosing the first derivative, if fractional derivative of a_j is used, the characteristics of the neural network can be modified. Specifically, data can be processed efficiently by fractional-order systems and provide more accurate results. Therefore, R-L, Caputo derivative and CF derivatives are employed in the weight and bias update process by applying equations 1, 2 and 4.
2) **Momentum Factor** : The convergence of the back propagation algorithm depends upon learning rate. A larger value of it may converge the algorithm fast but may result in overshooting or oscillations. Whereas as a smaller value of it has the vice-versa effect. Addition of momentum factor to the normal gradient descent method is an efficient and commonly used method that allows a larger learning rate without oscillations. It is represented as β. Weights and bias are evaluated by adding β as follows,

$$w_j^o(new) = w_j^o(old) + \beta \Delta w_j^o, \tag{13}$$

$$w_{ij}^{in}(new) = w_{ij}^{in}(old) + \beta \Delta w_{ij}^{in}, \tag{14}$$

Similarly old bias weights are updated to and represented as follows,

$$b^o(new) = b^o(old) + \beta \Delta b^o, \tag{15}$$

$$b^{in}(new) = b^{in}(old) + \beta \Delta b^{in}, \tag{16}$$

Range of β is selected between 0 and 0.9 and the value of 0.9 is often used for the momentum factor. We perform the salary prediction analysis for β in the range $0, \cdots, 0.9$ instead of 0.9 only. Block diagram of proposed perceptron based on fractional derivative and momentum factor is illustrated in Fig 2.

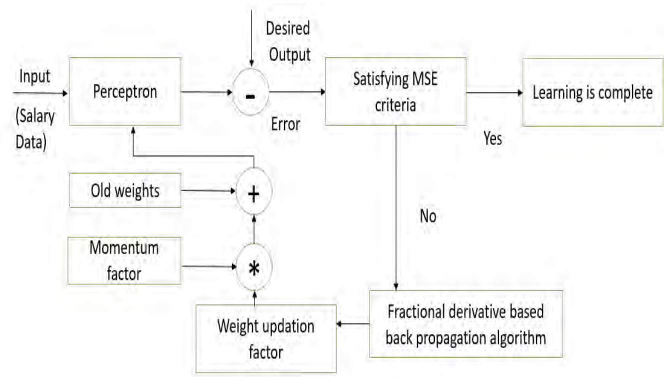

Fig. 2. Block diagram of proposed perceptron based on fractional derivative and momentum factor

V. Results and Discussions

In this section, the results are illustrated and the performance of our proposed model with fractional derivatives is evaluated. The salary of an employee based on years of experience is predicted and performance is evaluated in terms of MSE and compare the results with integer order derivatives.

The dataset for our work is downloaded [35]. In this data set, no. of samples are $M = 30$, consisting of two attributes: salary and years of experience. The data set is divided into training and testing sets, in a ratio of $80 : 20$.
Initially, a random value of the initial weights at input and output layers is selected. The rand function in MATLAB is used to generate random weights. The input values $X = [X_1, \cdots, X_{24}]$ from the training set are propagated in the further layer as a vector function of the weights and input data. At hidden layer, weighted input is passed through the activation function to introduce non-linearity in predicting the accurate output.
The output in a particular iteration is compared with the actual output and the weights are adjusted accordingly, till minimum error is obtained. The analysis considering Caputo, CF and RL fractional derivatives in the weight update process of the back propagation algorithm for various fractional order and momentum factors are performed and the results with integer derivatives are compared. Table I, II and III present MSE for Caputo, R-L and CF derivatives for various momentum factors (β), fractional order (α) and integer order (IO).
From Table I, it is observed that as β of Caputo derivative based learning algorithm is increased for a particular momentum factor, say α = 0.1, MSE decreases from the value of 19 to 10^{-23}. Also for the values of β in the range of 0.4-0.8, MSE is of the order of 10^{-23}. Further on increasing the value of α from 0.1-0.9, not much change is observed in MSE. For example, say for the value of β = 0.4, MSE is of the order of 10^{-23} , this means that there is not much impact of fractional order in learning process. This can be observed from the plot of MSE vs Fractional order for various momentum factors in the Caputo derivative based algorithm as illustrated in Fig.3.

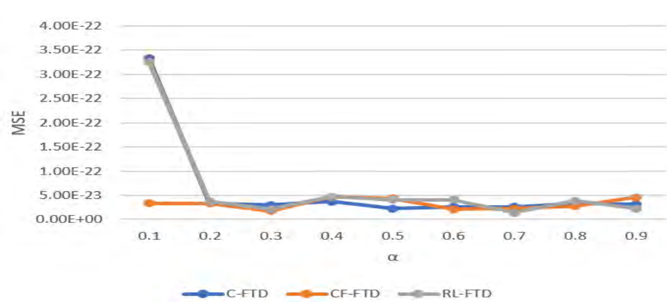

Fig. 3. MSE vs Fractional order plot for Caputo Derivative based learning algorithm for momentum factors in the range of $0.4 - 0.9$

In Table II of MSE for CF based learning algorithm, it is observed that except for the value of β= 0.1, MSE is of the order of 10^{-22} for β = 0.2 and 0.3 and it further decreases to 10^{-23} as β increases from 0.4 to 0.9. So overall MSE lies in the range of 10^{-22} to 10^{-23}. Fig.4 illustrates the plot of MSE vs Fractional order for various momentum factors in CF derivatives. It demonstrates that there is not much impact of fractional order on MSE.
In Table III of MSE for RL based learning algorithm, it is observed that except for the value of β = 0.1, MSE is of the order of 10^{-21} for β = 0.2 and 0.3 and it further decreases to 10^{-23} as β increases from 0.4 to 0.9. So overall MSE lies in the range of 10^{-21} to 10^{-24}. Fig.5 illustrates the plot of MSE vs Fractional order for various momentum factors in R-L derivatives, it also demonstrates that there is not much impact of fractional order on MSE.

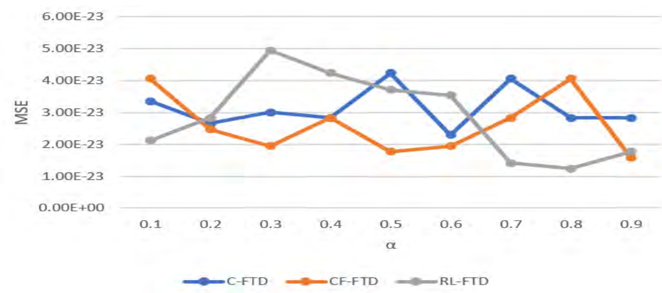

Fig. 4. MSE vs Fractional order plot for Caputo Fabrizio Derivative based learning algorithm for momentum factors in the range of $0.4 - 0.9$

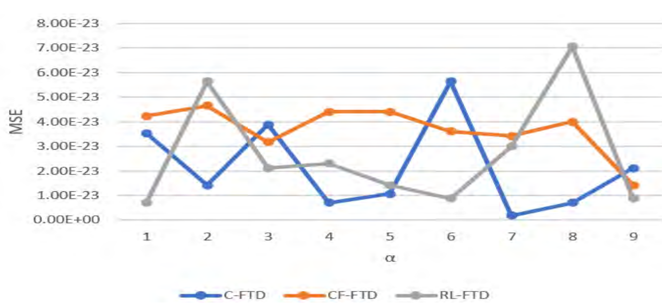

Fig. 5. MSE vs Fractional order plot for RL based learning algorithm for momentum factors in the range of $0.4 - 0.9$

From Table I, II and III and Fig. 3, 4 and 5, we can conclude that as β is increased there is decrease in MSE but α does not play any role in improving MSE. It is also observed that in all fractional derivative based learning algorithms, MSE is of the same order but it is very small in comparison to integer order derivative based learning as illustrated in Fig. 6. Thus, neural networks based on fractional derivative back propagation algorithm chosen for any fractional order can perform better than integer order (IO).
The plot exhibits the predicted salary of an employee based on years of experience as illustrated in Fig 7. It is observed that predicted salaries are overlapping with actual salaries which indicates that there is no error in the predicted output. Hence the fractional based derivative learning algorithms produce output with 100 % accuracy.

VI. Conclusion

In this paper, a modified learning algorithm based on fractional derivatives in three layer neural networks is presented. Specifically, this neural network is employed to predict the salary based on years of experience. The performance of three fractional derivative algorithms, Caputo, CF and R-L based back propagation algorithm in terms of MSE are investigated.

TABLE I
MSE AND FRACTIONAL ORDER FOR CAPUTO DERIVATIVE BASED LEARNING ALGORITHM FOR MOMENTUM FACTORS IN THE RANGE OF $0.1 - 0.9$

α/β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	19.0618	6.58E-09	1.61E-19	3.34E-22	2.47E-23	3.35E-23	2.82E-23	2.29E-23	3.53E-23	2.59E-12
0.2	9.7444	1.45E-09	1.19E-20	3.35E-23	2.65E-23	2.65E-23	3.00E-23	3.00E-23	1.41E-23	1.65E-13
0.3	9.574805	4.51E-10	1.56E-21	3.00E-23	2.65E-23	3.00E-23	3.18E-23	2.82E-23	3.88E-23	1.01E-14
0.4	3.9122	1.88E-10	3.72E-22	3.71E-23	3.88E-23	2.82E-23	3.35E-23	3.71E-23	7.06E-24	7.17E-16
0.5	3.2548	1.26E-10	3.41E-22	2.29E-23	3.35E-23	4.24E-23	4.24E-23	1.41E-23	1.06E-23	8.01E-17
0.6	3.3367	1.32E-10	3.83E-22	2.65E-23	1.41E-23	2.29E-23	2.29E-23	2.82E-23	5.65E-23	3.00E-17
0.7	4.3138	2.35E-10	4.18E-22	2.65E-23	3.00E-23	4.06E-23	3.00E-23	6.71E-23	1.76E-24	1.15E-16
0.8	7.1209	7.28E-10	3.51E-21	3.35E-23	4.24E-23	2.82E-23	4.06E-23	7.06E-24	7.06E-24	1.68E-14
0.9	15.1941	3.94E-09	6.55E-20	3.18E-23	2.82E-23	2.82E-23	2.12E-23	5.12E-23	2.12E-23	4.82E-11
IO	2.92E+6	963.88	0.22	3.59E-5	4.05E-9	2.95E-13	1.27E-17	1.31E-20	1.06E-20	1.29E-21

TABLE II
MSE AND FRACTIONAL ORDER FOR CAPUTO FABRIZIO DERIVATIVE BASED LEARNING ALGORITHM FOR MOMENTUM FACTORS IN THE RANGE OF $0.1 - 0.9$

α/β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	3.48E-08	4.29E-22	3.03E-22	3.35E-23	2.82E-23	4.06E-23	2.65E-23	7.76E-23	4.24E-23	9.94E+7
0.2	4.01E-09	3.39E-22	3.56E-22	3.39E-23	3.00E-23	2.47E-23	2.65E-23	2.12E-23	4.65E-23	1.39E+6
0.3	5.36E-10	3.39E-22	3.02E-22	1.76E-23	2.12E-23	1.94E-23	2.65E-23	3.53E-23	3.18E-23	1.77E+4
0.4	5.91E-11	4.46E-22	3.44E-22	4.76E-23	3.00E-23	2.82E-23	6.00E-23	5.29E-23	4.41E-23	205.02
0.5	8.63E-12	5.79E-22	3.34E-22	4.41E-23	1.06E-23	1.76E-23	4.24E-23	7.41E-23	4.41E-23	2.1412
0.6	2.50E-12	4.43E-22	3.44E-22	2.12E-23	2.65E-23	1.94E-23	5.12E-23	5.82E-23	3.61E-23	0.02
0.7	3.03E-12	4.48E-22	3.28E-22	2.29E-23	1.94E-23	2.82E-23	3.53E-23	1.06E-23	3.41E-23	1.69E-4
0.8	3.85E-11	4.52E-22	3.78E-22	2.82E-23	3.18E-23	4.06E-23	1.06E-23	5.29E-24	4.00E-23	1.26E-6
0.9	3.25E-09	5.17E-22	4.28E-22	4.59E-23	2.47E-23	1.59E-23	2.29E-23	8.82E-24	1.41E-23	8.32E-9
IO	2.92E+6	963.87	0.22	3.59E-5	4.05E-9	2.95E-13	1.27E-17	1.31E-20	1.06E-20	1.29E-21

TABLE III
MSE AND FRACTIONAL ORDER FOR RL DERIVATIVE BASED LEARNING ALGORITHM FOR MOMENTUM FACTORS IN THE RANGE OF $0.1 - 0.9$

α/β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	5.64E-08	1.36E-21	1.38E-21	3.26E-22	3.71E-23	2.12E-23	3.53E-24	2.12E-23	7.06E-24	19.08
0.2	1.46E-08	2.36E-21	2.66E-21	3.71E-23	1.06E-23	2.82E-23	3.00E-23	5.12E-23	5.65E-23	6.67E-9
0.3	4.96E-09	3.36E-21	3.36E-21	2.12E-23	3.35E-23	4.94E-23	3.53E-23	2.82E-23	2.12E-23	1.59E-19
0.4	2.35E-09	1.36E-21	1.48E-22	4.76E-23	1.59E-23	4.24E-23	4.24E-23	2.82E-23	2.29E-23	2.65E-21
0.5	1.63E-09	2.36E-21	2.36E-21	4.06E-23	2.29E-23	3.71E-23	2.29E-23	1.41E-23	1.41E-23	4.41E-21
0.6	1.71E-09	4.18E-22	5.18E-22	4.06E-23	4.06E-23	3.53E-23	1.06E-23	7.06E-24	8.82E-24	3.18E-21
0.7	2.86E-09	3.36E-21	3.66E-21	1.41E-23	3.00E-23	1.41E-23	1.94E-23	8.82E-24	3.00E-23	2.82E-21
0.8	7.82E-09	3.36E-21	3.62E-21	3.88E-23	4.41E-23	1.24E-23	1.94E-23	8.47E-23	7.06E-23	2.12e-21
0.9	3.54E-08	4.16E-22	4.26E-22	2.29E-23	3.18E-23	1.76E-23	2.82E-23	1.41E-23	8.82E-24	4.94E-21
IO	2.92E+6	963.87	0.22	3.59E-5	4.05E-9	2.95E-13	1.27E-17	1.31E-20	1.06E-20	1.29E-21

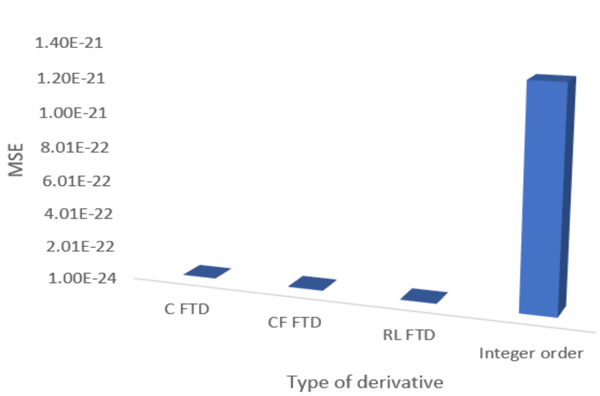

Fig. 6. MSE plot for Caputo, Caputo Fabrizio, RL and integer based learning algorithm for minimum MSE

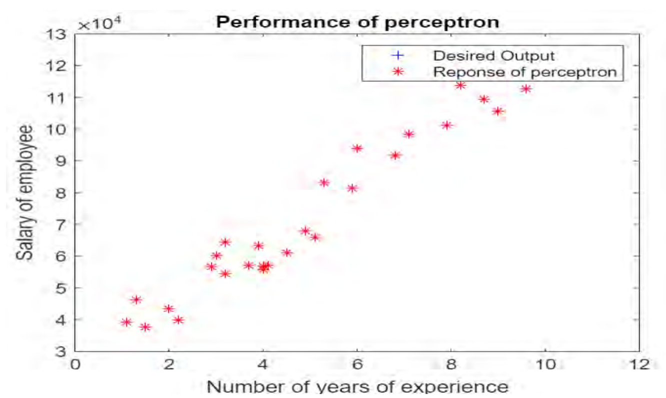

Fig. 7. Performance of perceptron

The impact of fractional order and momentum factor on MSE is compared. It is observed that in all three fractional derivative based algorithms shows less MSE in comparison to the integer order derivative. It is also demonstrated that as fractional derivatives process the information efficiently, so there is no impact of fractional order in the weight update process of the learning algorithm. Results illustrate that any of the fractional derivative can be used in a gradient based on back propagation learning algorithm. The MSE of order of 10^{-24} exhibits that fractional derivatives incorporated in the neural network process the information in a better way and provide the desired output.

References

[1] Valerio, D. Trujillo, J. Rivero, Margarita, T. Machado, J. Baleanu, and Dumitru, "Fractional calculus: A survey of useful formulas," *European Physics Journal ST (JCR)*, vol. 222, pp. 1827–1846, 09 2013.

[2] A. Loverro, "Fractional calculus: History, definitions and applications for the engineer," *University of Notre Dame, Notre Dame, in,USA*, vol. 46556, 08 2004.

[3] J. A. T. Machado and E. C. de Oliveira, "A review of definitions for fractional derivatives and integral," *Mathematical Problems in Engineering*, vol. 2014, pp. 1–6, 06 2014.

[4] R. E. Gutiérrez, J. M. Rosário, and J. T. Machado, "Fractional order calculus basic concepts and engineering applications," *Mathematical Problems in Engineering*, vol. 2010, 04 2010.

[5] I.Podlubny, *Fractional Differential Equations*. USA: Academic Press, 1998.

[6] Z. Odibat, "Computing eigenelements of boundary value problems with fractional derivatives," *Applied Mathematics and Computation*, vol. 215, pp. 3017–3028, 12 2009.

[7] S. Rezapour, S. Etemad, and H. Mohammadi, "A mathematical analysis of a system of caputo-fabrizio fractional differential equations for the anthrax disease model in animals," *Advances in Difference Equations*, vol. 2020, p. 2020:481, 09 2020.

[8] M. Dokuyucu, E. Çelik, H. Bulut, and H. M. Baskonus, "Cancer treatment model with the caputo-fabrizio fractional derivative," *The European Physical Journal Plus*, vol. 133, p. 92, 03 2018.

[9] D. Baleanu, S. M. Aydogn, H. Mohammadi, and S. Rezapour, "On modelling of epidemic childhood diseases with the caputo-fabrizio derivative by using the laplace adomian decomposition method," *Alexandria Engineering Journal*, vol. 59, no. 5, pp. 3029–3039, 2020.

[10] L. C. B. Martins, R. N. Carvalho, R. S. Carvalho, M. C. Victorino, and M. Holanda, "Early prediction of college attrition using data mining," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2017, pp. 1075–1078.

[11] A. D. and A. A. B., "Analyzing employee attrition using decision tree algorithms," *Computing, Information Systems and Development Informatics*, vol. 5, pp. 17–28, 3 2013.

[12] P. Rohit and P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms," *International Journal of Advanced Research in Artificial Intelligence*, vol. 5, 10 2016.

[13] B. Pérez, C. Castellanos, and D. Correal, *Predicting Student Drop-Out Rates Using Data Mining Techniques: A Case Study: First IEEE Colombian Conference, ColCACI 2018, Medellín, Colombia, May 16-18, 2018, Revised Selected Papers*, 05 2018, pp. 111–125.

[14] S. Dutta and S. K. Bandyopadhyaya, "Employee attrition prediction using neural network cross validation method," *International Journal of Commerce and Management Research*, vol. 6, pp. 80–85, 2020.

[15] Y. Zhao, M. Hryniewicki, F. Cheng, B. Fu, and X. Zhu, "Employee turnover prediction with machine learning: A reliable approach," vol. 869, 11 2018, pp. 737–758.

[16] V.-E. Neagoe, A.-D. Ciotec, and G.-S. Cucu, "Deep convolutional neural networks versus multilayer perceptron for financial prediction," in *2018 International Conference on Communications (COMM)*, 2018, pp. 201–206.

[17] F. Widyahastuti and V. U. Tjhin, "Predicting students performance in final examination using linear regression and multilayer perceptron," in *2017 10th International Conference on Human System Interactions (HSI)*, 2017, pp. 188–192.

[18] J. Gaudart, B. Giusiano, and L. Huiart, "Comparison of the performance of multi-layer perceptron and linear regression for epidemiological data," *Computational Statistics and Data Analysis*, vol. 44, pp. 547–570, 01 2004.

[19] E. Ramalakshmi and S. Kamidi, *Prediction of Employee Attrition and Analyzing Reasons: Using Multi-layer Perceptron in Spark*, 01 2020, pp. 183–192.

[20] V. Vyawahare, G. Espinosa-Paredes, G. Datkhile, and P. Kadam, "Artificial neural network approximations of linear fractional neutron models," *Annals of Nuclear Energy*, vol. 113, pp. 75–88, 03 2018.

[21] P. Kadam, G. Datkhile, and V. Vyawahare, *Artificial Neural Network Approximation of Fractional-Order Derivative Operators: Analysis and DSP Implementation*, 01 2019, pp. 93–126.

[22] V. Joshi, R. B. Pachori, and A. Vijesh, "Classification of ictal and seizure-free eeg signals using fractional linear prediction," *Biomedical Signal Processing and Control*, vol. 9, pp. 1–5, 2014.

[23] A. Sharafian, A. Sharifi, and W. Zhang, "Fractional sliding mode based on rbf neural network observer: Application to hiv infection mathematical model," *Computers and Mathematics with Applications*, vol. 79, no. 11, pp. 3179–3188, 2020.

[24] H. Krouma, Y. Ferdi, and A. Taleb-Ahmedx, "Neural adaptive fractional order differential based algorithm for medical image enhancement," pp. 1–6, 2018.

[25] M. Bohner and I. Stamova, "An impulsive delay discrete stochastic neural network fractional-order model and applications in finance," *Filomat*, vol. 32, pp. 6339–6352, 01 2018.

[26] F. Juntao and W. Huan, "Recurrent neural network fractional-order sliding mode control of dynamic systems," *Journal of the Franklin Institute*, vol. 357, no. 8, pp. 4574–4591, 2020.

[27] S. Khan, I. Naseem, M. Ammar, R. Togneri, and M. Bennamoun, "A fractional gradient descent-based rbf neural network," *Circuits, Systems, and Signal Processing*, vol. 37, pp. 5311–5332, 12 2018.

[28] K. S, A. J, Naseem, and M. Moinuddin, "A novel fractional gradient-based learning algorithm for recurrent neural networks." *Circuits Syst Signal Process 37*, pp. 593–612, 2018.

[29] C. Bao, Y.-F. PU, and Y. Zhang, "Fractional-order deep backpropagation neural network," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–0, 07 2018.

[30] J. Wang, Y. Wen, Y. Gou, Z. Ye, and H. Chen, "Fractional-order gradient descent learning of bp neural networks with caputo derivative," *Neural Networks*, vol. 89, pp. 19–30, 05 2017.

[31] Y. Chen and G. Zhao, "A caputo-type fractional-order gradient descent learning of deep bp neural networks," in *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 2019, pp. 546–550.

[32] S. Das, *Functional Fractional Calculus for System Identification and Controls*. Springer, 01 2008.

[33] M. Caputo and M. Fabrizio, "A new definition of fractional derivative without singular kernel," vol. 1, pp. 73–85, 04 2015.

[34] S. N. Sivanandam and S. N. Deepa, *Principles of Soft Computing*. India: WILEY publication, 2007.

[35] "Salary data with experience," https://www.kaggle.com /datasets/karthickveerakumar/salary-data-simple-linear-regression.