World Journal of Engineering and Technology, 2024, 12, 836-850
https://www.scirp.org/journal/wjet
ISSN Online: 2331-4249
ISSN Print: 2331-4222

Fusion of Activation Functions: An Alternative to Improving Prediction Accuracy in Artificial Neural Networks

Justice Awosonviri Akodia[1], Clement K. Dzidonu[2], David King Boison[3], Philip Kisembe[4]

[1]Accra Institute of Technology, Advanced School of Systems and Data Studies (ASSDAS), Accra, Ghana
[2]Accra Institute of Technology, Accra, Ghana
[3]Knowledge Web Centre, Accra, Ghana
[4]Department of Computer Engineering, Ghana Communication Technology University, Accra, Ghana
Email: Justiceaja1@gmail.com, dzidonu@ait.edu.gh, dvboison@gmail.com, pkisembe@gctu.edu.gh

How to cite this paper: Akodia, J.A., Dzidonu, C.K., Boison, D.K. and Kisembe, P. (2024) Fusion of Activation Functions: An Alternative to Improving Prediction Accuracy in Artificial Neural Networks. *World Journal of Engineering and Technology*, **12**, 836-850.
https://doi.org/10.4236/wjet.2024.124051

Received: July 29, 2024
Accepted: September 1, 2024
Published: September 4, 2024

Abstract

The purpose of this study was to address the challenges in predicting and classifying accuracy in modeling Container Dwell Time (CDT) using Artificial Neural Networks (ANN). This objective was driven by the suboptimal outcomes reported in previous studies and sought to apply an innovative approach to improve these results. To achieve this, the study applied the Fusion of Activation Functions (FAFs) to a substantial dataset. This dataset included 307,594 container records from the Port of Tema from 2014 to 2022, encompassing both import and transit containers. The RandomizedSearchCV algorithm from Python's Scikit-learn library was utilized in the methodological approach to yield the optimal activation function for prediction accuracy. The results indicated that "ajaLT", a fusion of the Logistic and Hyperbolic Tangent Activation Functions, provided the best prediction accuracy, reaching a high of 82%. Despite these encouraging findings, it's crucial to recognize the study's limitations. While Fusion of Activation Functions is a promising method, further evaluation is necessary across different container types and port operations to ascertain the broader applicability and generalizability of these findings. The original value of this study lies in its innovative application of FAFs to CDT. Unlike previous studies, this research evaluates the method based on prediction accuracy rather than training time. It opens new avenues for machine learning engineers and researchers in applying FAFs to enhance prediction accuracy in CDT modeling, contributing to a previously underexplored area.

Keywords

Artificial Neural Networks, Container Dwell Time, Fusion of Activation

Functions, Randomized Search CV Algorithm, Prediction Accuracy

1. Introduction

As the quest to extract the richest possible insights from data proliferates, a vast arsenal of statistical tools has been deployed, ranging from the conventional to the profoundly innovative. Traditional forms of regression and time series analysis, bounded by their inherent parametric or non-parametric assumptions, have been fundamental in our data modeling approaches [1]. However, the advent of Artificial Neural Networks (ANN), inspired by the intricate workings of human neurons, provides a radical departure from statistical dogma, fostering a broad spectrum of applications including natural language processing, medical diagnoses, agriculture, criminology, statistics, among others ([2]-[4]). Other areas of application include sediment load prediction [5] and emission prediction [6] At the heart of ANN lies the paramount process of learning, governed by two main types of parameters - parameters and hyperparameters. While the former is updated by the network, the latter is under the researcher's control, creating a crucial gateway for modulating ANN performance [7]. Consequently, the meticulous selection of hyperparameters has resulted an array of studies, each endeavoring to optimize prediction or classification outcomes [8] [9]. Among these hyperparameters, Activation Functions (AFs) merit particular attention. Defined as the transfer function enabling signal transmission from one neuron to another, AFs have a profound impact on network training [1] [10]-[13]. While numerous AFs have been developed, each carrying unique strengths and weaknesses, their efficacy is sometimes compromised by the complexity of the data being modeled or their inherent limitations, as in the case of the constant AF [1]. This has stimulated the concept of Fusion of Activation Functions (FAF), which amalgamates different AFs with an aim to create more robust functions for improved signal transfer [14]. However, these attempts have sometimes been met with suboptimal results, as discretionary choices lead to less-than-optimal training and prediction outcomes. Therefore, this research presents an exploration into the depths of FAFs, with a specific focus on the prediction of Container Dwell Time—The length of time freight containers occupie a space from reception to delivery or transfer to the next transport mode. By delving into a more comprehensive suite of FAFs, this study strives to pioneer a path towards improved performance in such predictive tasks, ultimately enriching the potential of ANNs in data modeling and prediction.

2. Literature Review

The scholarly research dissected underscores the indispensable role and multifaceted nature of Activation Functions (AFs) within Artificial Neural Networks (ANNs). These functions serve as the communication mechanism between neurons, aiding the propagation of signals through the network [1]. Over time, several distinctive types of AFs have been introduced, each with unique strengths and weaknesses. For instance, the step function, sigmoid function, hyperbolic tangent, Rectified Linear Unit (ReLU), Leaky ReLU (LReLU), Parametric ReLU (PReLU), Exponential Linear Unit (ELU), Gaussian Error Linear Unit (GELU), and Swish are among the well-documented Afs [15]-[21]. However, the application of singular AFs across diverse problems and datasets has presented challenges, primarily due to their limitations in different scenarios. The phenomenon of the vanishing gradient syndrome is one such challenge observed in complex datasets when using sigmoid AFs [13]. This predicament led to the exploration of a more versatile approach: The Fusion of Activation Functions (FAFs) [14]. FAFs, essentially the combination of multiple AFs, have been found to improve performance, particularly with larger and more complex datasets, and promote faster network convergence [21]-[25]. However, a significant portion of the extant literature focuses mainly on classification tasks, neglecting regression problems. Furthermore, many studies have prioritized network training speed, overlooking the paramount importance of prediction accuracy in various applications. Whilst training speed focuses on the amount of time required for convergence (completion of training), prediction or classification accuracy focuses on the size of the deviation between the desired outcome and the predicted outcome (usually measured in percentage) [26] [27]. Obviously. This significant gap in the literature serves as the impetus for the current investigation, aiming to enhance the application of FAFs for improving the prediction accuracy of ANNs in modeling Container Dwell Time (CDT). The focus on prediction accuracy is due to two things, first, due to the limited literature and two, its importance in terms of value. CDT is a crucial parameter in the administration and management of container terminals. In addition to the research on AFs and FAFs, other studies indirectly support the broader potential of ANNs in dealing with intricate tasks. These works, including those by [28], underscore the potential of leveraging strategies like FAFs to enhance ANN performance and accuracy. Although these studies do not directly delve into FAFs, their findings nevertheless offer corroborative evidence of the potential applicability and value of enhanced predictive models in various contexts. While the literature has seen extensive research into the discovery and fusion of new AFs, it has notably fallen short in their application to regression problems and assessing their impact on prediction accuracy. This critical review of the literature presents an opportunity for future research to bridge this gap. It provides a promising prospect for the exploration of FAFs in these areas, potentially paving a new path towards enhancing prediction accuracy in ANNs.

3. Methodology

The study considers 307,594 container records from 2014 to 2022 at the Port of Tema made up of import and transit containers delivered within the period under review. The main dataset is main of 307,594 container records extracted from the TOS system of Ghana Ports and Harbours Authority. The data is made up of the dwell days of containers as the dependent variable as well as their characteristics as the main independent variables. From [29] [30], it is believed that these features influence the dwell time of the container.

Below are the characteristics of the container under consideration (**Table 1**).

Table 1. Attributes of variables.

No	Variable Name	Classification	Scale of Measurement	Variable Type
1	Container Number	Subject	Nominal	String
2	Trade	Shipment-Level	Categorical (nominal)	String
3	Commodity	Shipment-Level	Categorical (nominal)	String
4	Size	Shipment-Level	Categorical (nominal)	String
5	Weight	Shipment-Level	Ratio	Decimal
6	Shipment Type	Shipment-Level	Categorical (nominal)	String
7	Day of Delivery	Shipment-Level	Categorical (nominal)	String
8	Date of Discharge	Shipment-Level	Ordinal	Date
9	Date of Delivery	Shipment-Level	Ordinal	Date
10	Last Port of call	Shipment-Level	Categorical (nominal)	String
11	Region of Origin	Shipment-Level	Categorical (nominal)	String
12	Fiscal Regime	Shipment-Level	Categorical (nominal)	String
13	Density of Value	Shipment-Level	Numeric	Decimal
14	Shipping Agency	Non-Shipment-Level	Categorical (nominal)	String
15	Freight Forwarder	Non-Shipment-Level	Categorical (nominal)	String
16	Carrier	Non-Shipment-Level	Categorical (nominal)	String
17	Trucking Company	Non-Shipment-Level	Categorical (nominal)	String
18	Risk Level	Non-Shipment-Level	Categorical (nominal)	String
19	Post-Entry	Non-Shipment-Level	Categorical (nominal)	Boolean
20	Scan	Non-Shipment-Level	Categorical (nominal)	Boolean
21	Customs Inspection	Non-Shipment-Level	Categorical (nominal)	Boolean

The method used for this study is adopted from [31] who, after reviewing about 97 Machine Learning (ML) research articles in 10 different application areas recommended for the adoption of the following life cycle in conducting ML projects. Data Collection, Data Pre-processing, Model Training, Model Testing and Model Evaluation. This involved the extraction of Container Dwell Time (CDT) data from the TOS system of Ghana Ports and Harbours Authority. This extraction was done to include all the required features of the container after which the dwell days were computed for each record. To be able to make the data machine learning-ready, the following pre-processing steps were taken:

1) Outliers were detected and replaced with the mean CDT.

2) The data was grouped under categorical and numeric features.

3) To achieve standardization, one-hot encoding was applied to the categorical features.

4) Standard Scaling was applied to the numeric features.

5) Principal Component Analysis was used to drop non-informative numeric features.

6) The dataset was then reorganized and divided into input and output feature sets.

Model Training and Testing

The pre-processed data was split into Training, Test and Validation sets.

The main library used was the Random Search CV from the scikit-learn library of Python.

Random Search CV takes the default hyperparameter arguments, (*estimator*, *param_distributions*, ***, *n_iter* = *10*, *scoring* = *None*, *n_jobs* = *None*, *refit* = *True*, *cv* = *None*, *verbose* = *0*, *pre_dispatch* = '*2*n_jobs*', *random_state* = *None*, *error_score* = *nan*, *return_train_score* = *False*) which are variable.

The data was then trained with sets of the set hyperparameters in Table 2 till convergence.

Table 2. Hyperparameter space using randomized search CV.

Hyperparameter	Choices
Algorithm	RandomizedSearchCV
Training Cross Validations	10
Estimator	MLPRegressor
Number of Iterations	200
Activation Functions	[Logistics, Relu, identity, Tanh, ajaILR, ajaIL, ajaIT, ajaIR, ajaLT, ajaLR, ajaTR]
Number of Hidden Layers	[3, 5, 10, 15, 20, 25, 30]
Learning Rate	[Constant, Adaptive, Invscaling]
Momentum	[Adam, SGD, ibfgs]
Scoring	MSE

The Mean Squared Error (MSE) is the loss function used to evaluate the performance of the model during training and testing. It measures the average of the squared deviations between the predicted and the expected. Represented by:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y - \hat{y}\right)^2$$

N = Sample Size

y = Actual output

$\hat{y}$ = Predicted output

4. Results

The Principal Component Analysis (PCA), a statistical procedure used to identify the principal components that best capture the variability in a dataset, conducted on three numerical features (BL Type, BL Version, and BOE Version), showed variance levels of 30%, 33%, and 37% respectively (see **Figure 1**). This significant variation suggests that these features hold valuable information and should not be eliminated during data preprocessing. Instead, they should be included in the feature set for training the ANN. This underscores the importance of feature selection and the value of PCA in identifying influential variables in a dataset.

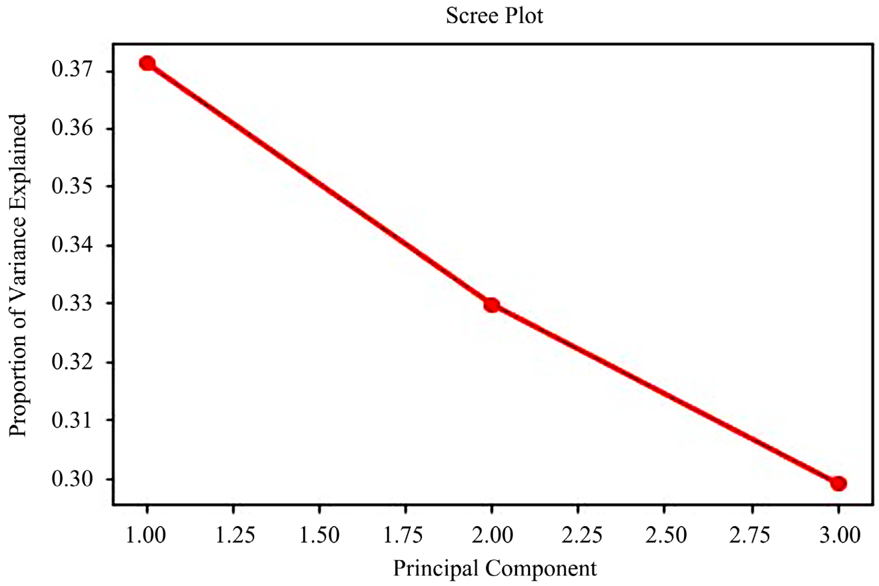

Figure 1. Scree plot of principal components.

The application of the RandomizedSearchCV function from the scikit-learn library in Python was employed to identify the optimal hyperparameters for the ANN model. This function uses a randomized search on hyperparameters (see **Table 3**), which is both time-efficient and effective, particularly when the dimensionality of the parameter space is high. The search converged on three hidden layers, a constant learning rate, the ajaLT activation function, and stochastic gradient descent (SGD) momentum. This optimal configuration suggests a potential direction for the construction of ANNs for similar problems.

Table 3. Optimum hyperparameters after convergence.

Hyperparameter Name	Best Value
Activation Functions	ajaLT
Number of Hidden Layers	3
Learning Rate	Constant
Momentum	SGD

In Table 4, the comparison between the actual CDT and predicted CDT shows a high degree of accuracy, reinforcing the effectiveness of the chosen hyperparameters and the ANN model itself. The activation function, ajaLT, which is a fusion of different activation functions, appears to contribute significantly to the model's performance, suggesting that this approach could be more widely applied in future ANNs (Figure 2). The high degree of similarity between actual and predicted CDT values confirms the validity of the approach used, both in terms of the ANN structure and the choice of hyperparameters.

Table 4. Comparison of actual and predicted CDT for ajaLT.

SN	CDT (Actual)	CDT (Predicted)
0	13.97	13.2
1	12.53	12.89
2	12.00	12.2
3	12.53	11.23
4	12.53	14
5	10.53	10.52
6	12.53	12.4
7	12.53	12.51
8	12.53	13
9	12.53	12.58
10	16.94	17
11	12.53	12.8
.	.	.
.	.	.
307,594	16.22	16.66

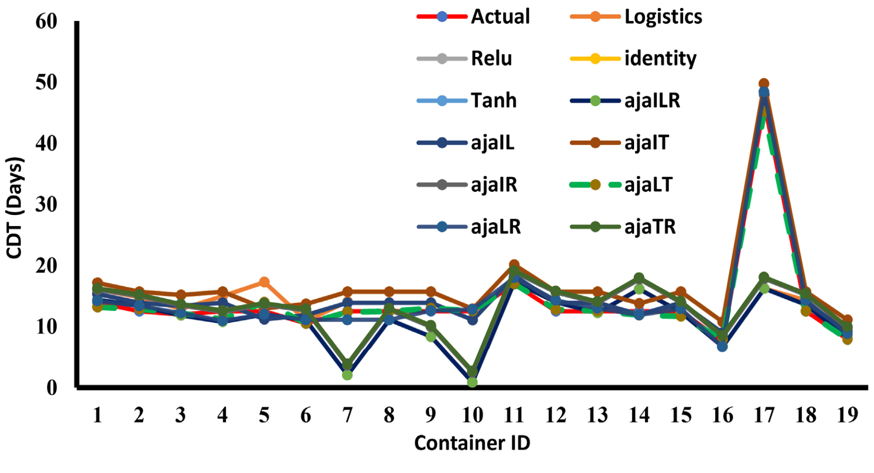

Figure 2. A comparison of the different activation functions.

A comparison of the fusion of activation functions (FAFs) with standard mono activation functions was also carried out (see **Figure 2**). This comparison showed that the ajaLT FAF (a fusion of the logistic function and the hyperbolic tangent function) outperformed other FAFs and mono activation functions, with an accuracy level of 82%. This provides strong evidence for the superiority of FAFs in this context, reaffirming previous findings [32] and advocating for their broader adoption in ANNs.

The results suggest that the use of FAFs in ANNs can lead to improved modeling outcomes when predicting CDT. In particular, the ajaLT FAF, which is a fusion of the logistic function and the hyperbolic tangent function, demonstrated superior performance in terms of prediction accuracy. Furthermore, the use of PCA helped to identify important features to include in the model, while the RandomizedSearchCV function provided an efficient means to determine the optimal hyperparameters for the ANN.

5. Discussions

Building upon a wealth of empirical literature that emphasizes the crucial role of Activation Functions (AFs) within Artificial Neural Networks (ANNs), the findings of our study demonstrate the application and performance of AFs within a real-world context [1] [15]-[21]. Our research reaffirms the mounting agreement within the scientific community that the fusion of various AFs, referred to as Fusion of Activation Functions (FAFs), can significantly improve network performance and promote quicker convergence, especially when working with larger or more complex datasets [14] [21]-[25]. The existing literature reveals a significant challenge when applying sigmoid AFs to complex datasets: the issue of vanishing gradients [13]. To address this prevalent problem, our study proposes a novel solution—applying FAFs, specifically the ajaLT function, which represents a fusion of the logistic function and the hyperbolic tangent function. This combination demonstrated a superior performance when compared to other singular AFs, contributing substantially to the high prediction accuracy observed in our results. Such an approach not only offers a fresh perspective on dealing with the vanishing gradient problem, but it also introduces a potentially game-changing methodology for future ANN applications. Another important contribution of our study lies in addressing the knowledge gap in the existing literature. The majority of previous research has primarily focused on the application of AFs and FAFs to classification tasks, largely overlooking regression problems. We tackled this underrepresented area by applying FAFs to model the Container Dwell Time (CDT), a key regression problem in the realm of container terminal administration and management [33]-[36]. Through this application, our study adds valuable knowledge to the yet under-explored area of FAFs application in regression tasks, thereby broadening the scope and usability of FAFs in various domains. Furthermore, our study takes a distinctive approach in comparison to the bulk of existing research. Many studies have prioritized network training speed, often neglecting the equally important aspect of prediction accuracy [37]. In contrast, we emphasized the fundamental aspect of prediction accuracy in our research, showcasing the impressive accuracy level achieved (82%) through the use of FAFs. This observation underscores the potential of FAFs in significantly enhancing prediction accuracy, thereby extending the applications of ANNs across a multitude of fields. Our findings not only echo but also significantly contribute to the ongoing discourse in the empirical literature regarding the role and advantages of FAFs in enhancing the performance of ANNs. We advance the discussion by demonstrating that FAFs, specifically the ajaLT function, can be effectively used in regression problems, such as predicting CDT, with high prediction accuracy. These insights add substantial value to the existing body of knowledge by emphasizing the potential of FAFs in ANNs, which in turn illuminates new research avenues.

The extensive exploration into the use of FAFs in ANNs reaffirms the potential of these functions in addressing complex problems within diverse domains. As a pivotal addition to the scientific literature, this study highlights the versatility of FAFs, especially in improving prediction accuracy, setting a promising trajectory for further research and applications in this dynamic and rapidly evolving field.

6. Conclusions

The conclusions of this study are multi-faceted and significantly contribute to the evolving body of knowledge on the application and performance of Activation Functions (AFs) in Artificial Neural Networks (ANNs). Our study aligns with and extends upon the extensive empirical literature that underscores the indispensable role of AFs in ANNs. First, our research confirmed the effectiveness of Fusion of Activation Functions (FAFs), a burgeoning concept in the field. We demonstrated that the ajaLT function, a fusion of the logistic function and the hyperbolic tangent function, showed superior performance compared to singular AFs. This fusion strategy effectively countered the common issue of vanishing gradients, typically associated with the use of sigmoid AFs in complex datasets. Therefore, our study provides a promising approach to tackle the vanishing gradient problem, offering a significant contribution to the ANNs field. Second, we addressed a gap in the current literature by demonstrating the application of FAFs in regression problems. Specifically, we applied FAFs to model the Container Dwell Time (CDT), an essential parameter in the administration and management of container terminals. By extending the application of FAFs to regression problems, our research diversifies the utility of ANNs and FAFs, broadening their scope and potential impact in various contexts. Third, our study highlighted the importance of prediction accuracy over network training speed, a priority often overlooked in the current literature. We achieved a high prediction accuracy (82%) through the use of FAFs, emphasizing their potential in enhancing the prediction accuracy of ANNs. Our research has reinforced the significance of FAFs in improving the performance of ANNs, particularly in complex datasets and regression problems. We demonstrated the versatility and efficacy of the ajaLT function and emphasized the importance of prediction accuracy. Our findings offer valuable insights and provide a robust foundation for future research, paving the way for more effective and accurate prediction models within the domain of ANNs. The potential applications of these findings extend across a myriad of fields, opening up new avenues for research and real-world applications.

7. Implications for Theory, Policy Relevance, Practice and Social Impact

Theory Implications

The results of this study provide valuable insights into the theoretical understanding of Artificial Neural Networks (ANNs) and the use of Activation Functions (AFs). The effectiveness of the Fusion of Activation Functions (FAFs), particularly the ajaLT function, contributes to the understanding of complex data modeling in ANNs. The research further extends the applicability of FAFs beyond classification tasks to regression problems. This broadens the scope of existing ANN theory, providing a richer understanding of how to tackle the common issue of vanishing gradients and improve prediction accuracy.

Policy Relevance

The successful application of FAFs to model Container Dwell Time (CDT) has policy implications, particularly for container terminal management and administration. Given that CDT is a crucial factor in the efficient functioning of these facilities, the ability to accurately predict it can inform strategic decision-making and policy development. Policies that improve logistical efficiency, reduce bottlenecks, and enhance overall terminal operations can be better informed by the predictive capabilities of ANNs leveraging FAFs.

Practice Implications

In practical terms, this study demonstrates the potential of FAFs to enhance the accuracy of ANNs in various applications. For professionals working with ANNs, particularly in data-heavy environments, the use of FAFs may lead to improved performance, more accurate predictions, and faster network convergence. Additionally, this research paves the way for further exploration and utilization of FAFs in different contexts, expanding the toolbox for data scientists, machine learning engineers, and other practitioners in the field.

Social Impact

From a social perspective, the study's findings could have wide-ranging impacts. Improved prediction accuracy in ANNs could enhance various sectors, including healthcare, logistics, and finance, among others. For instance, more accurate predictions of CDT could lead to more efficient container terminal operations, potentially reducing costs, minimizing delays, and enhancing overall productivity. These improvements could contribute to economic growth, job creation, and overall societal well-being. Further, the study could inspire future researchers and practitioners to innovate and expand upon the potential applications of FAFs, driving advancements in technology that benefit society at large.

8. Recommendations

Building on the successful application of Fusion of Activation Functions (FAFs) in the prediction of Container Dwell Time (CDT) observed in this study, it is crucial for future research to consider exploring the applicability of FAFs in a wider spectrum of regression problems. This could serve to provide a more comprehensive understanding of the potential use cases for FAFs, thus, illuminating novel avenues for predictive modeling. In addition, although the ajaLT function, a fusion of the logistic function and the hyperbolic tangent function, has demonstrated superior performance in this study, numerous other combinations of activation functions remain unexplored. Consequently, future studies are encouraged to delve into these combinations to assess their performance and potentially discover new effective FAFs that could contribute significantly to the field of predictive modeling. Given the priority of prediction accuracy over network training speed in this study, further research would benefit from a detailed investigation into the trade-offs between these two critical factors. Such understanding could provide valuable insights that would inform the practical decision-making process when it comes to choosing an appropriate AF or FAF for a specific problem or dataset. On a policy level, it is recommended that policymakers, especially those in the domain of container terminal management and administration, take into account the potential benefits of utilizing predictive models like the one presented in this study for strategic decision-making. The ability to accurately predict Container Dwell Time (CDT) could considerably enhance operational efficiency and, as such, policies should be established to encourage the utilization of such advanced predictive tools in practice. To maximize the potential of FAFs in improving prediction accuracy and network performance, it is necessary to offer training to practitioners in the field of data science and machine learning. Educational institutions, organizations, and learning platforms should strive to incorporate modules on FAFs in their training programs. This would foster capacity building and ensure the broad adoption of FAFs in the field. Beyond the scope of FAFs, there are numerous hyperparameters in neural networks that could influence the performance of the model. It is recommended for future research to explore these hyperparameters in the context of FAFs to further enhance the predictive power of these networks. Lastly, there is an urgent need to raise societal awareness about the potential benefits that can be derived from accurate predictive models, such as those empowered by FAFs. By informing the wider public of these benefits, support for further research and application of this technology could be enhanced, which in turn, would contribute to the advancement of predictive modeling.

9. Limitations and Future Research Direction

While this study offers significant insights into the application of Fusion of Activation Functions (FAFs) for predicting Container Dwell Time (CDT), it is not without its limitations, which serve to highlight future research directions. Firstly, the study was largely centered on the ajaLT activation function, a fusion of the logistic function and the hyperbolic tangent function, which outperformed the standard mono activation functions in the prediction of CDT. However, it is crucial to note that the performance of activation functions, whether singular or fused, can vary depending on the nature of the dataset and the problem at hand. Therefore, the superior performance of ajaLT in this study does not necessarily guarantee similar results across all use cases. Secondly, although the study achieved a relatively high prediction accuracy of 82%, there is still room for improvement. The use of other machine learning techniques, beyond the scope of this study, could potentially result in higher prediction accuracy. Additionally, this study's focus was primarily on prediction accuracy, rather than network training speed. While accuracy is indeed paramount in many applications, the trade-off between accuracy and training speed is an important factor to consider, particularly in use cases where real-time predictions are required. This study, however, did not delve into these trade-offs, which leaves room for future research. Lastly, the study was conducted using a specific dataset related to container dwell time. Although it demonstrated the potential of FAFs in the field of container terminal management, the application of these findings to other datasets and fields may not yield the same results. In light of these limitations, future research should explore the applicability of various FAFs across a broad range of datasets and problem types. There is also a need for comprehensive studies to understand the trade-offs between prediction accuracy and network training speed when applying FAFs. Additionally, further investigations into the influence of other hyperparameters on the performance of ANNs when using FAFs would be beneficial. Overall, this study paves the way for more extensive research on the application and optimization of FAFs in predictive modeling.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] McCulloch, W.S. and Pitts, W. (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, **5**, 115-133. https://doi.org/10.1007/bf02478259

[2] Mazzei, D. and Ramjattan, R. (2022) Machine Learning for Industry 4.0: A Systematic Review Using Deep Learning-Based Topic Modelling. *Sensors*, **22**, Article No. 8641. https://doi.org/10.3390/s22228641

[3] Obermeyer, Z. and Emanuel, E.J. (2016) Predicting the Future—Big Data, Machine Learning, and Clinical Medicine. *New England Journal of Medicine*, **375**, 1216-1219. https://doi.org/10.1056/nejmp1606181

[4] Rajkomar, A., Dean, J. and Kohane, I. (2019) Machine Learning in Medicine. *New England Journal of Medicine*, **380**, 1347-1358. https://doi.org/10.1056/nejmra1814259

[5] Heng, S. and Suetsugi, T. (2013) Coupling Singular Spectrum Analysis with Artificial Neural Network to Improve Accuracy of Sediment Load Prediction. *Journal of Water Resource and Protection*, **5**, 395-404. https://doi.org/10.4236/jwarp.2013.54039

[6] Deng, J., Zhong, S. and Ordys, A. (2013) Robustness and Accuracy Test of Particular Matter Prediction Based on Neural Networks. *Communications and Network*, **5**, 53-59. https://doi.org/10.4236/cn.2013.52b010

[7] Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning. MIT Press. https://www.deeplearningbook.org/

[8] Bergstra, J., Bardenet, R., Bengio, Y. and Kégl, B. (2011) Algorithms for Hyper-Parameter Optimization. In: Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F. and Weinberger, K.Q., Eds., *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 1-9. https://proceedings.neurips.cc/paper_files/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf

[9] Claesen, M. and De Moor, B. (2015) Hyperparameter Search in Machine Learning. *MIC* 2015: *The XI Metaheuristics International Conference*, Agadir, 7-10 June 2015, 14-1-14-5. http://arxiv.org/abs/1502.02127

[10] Agostinelli, F., Hoffman, M., Sadowski, P. and Baldi, P. (2014) Learning Activation Functions to Improve Deep Neural Networks. 3*rd International Conference on Learning Representations*, *ICLR* 2015, San Diego, 7-9 May 2015, 1-9. http://arxiv.org/abs/1412.6830

[11] A, S. (2017) Studying the Effect of Activation Function on Classification Accuracy Using Deep Artificial Neural Networks. *Journal of Remote Sensing & GIS*, **6**, Article ID: 1000203. https://doi.org/10.4172/2469-4134.1000203

[12] Emanuel, R.H.K., Docherty, P.D., Lunt, H. and Möller, K. (2023) The Effect of Activation Functions on Accuracy, Convergence Speed, and Misclassification Confidence in CNN Text Classification: A Comprehensive Exploration. *The Journal of Supercomputing*, **80**, 292-312. https://doi.org/10.1007/s11227-023-05441-7

[13] Jagtap, A.D. and Karniadakis, G.E. (2023) How Important Are Activation Functions in Regression and Classification? A Survey, Performance Comparison, and Future Directions. *Journal of Machine Learning for Modeling and Computing*, **4**, 21-75. https://doi.org/10.1615/jmachlearnmodelcomput.2023047367

[14] Asaduzzaman, M., Shahjahan, M. and Murase, K. (2009) Faster Training Using Fusion of Activation Functions for Feed Forward Neural Networks. *International Journal of Neural Systems*, **19**, 437-448. https://doi.org/10.1142/s0129065709002130

[15] Rosenblatt, F. (1958) The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, **65**, 386-408. https://doi.org/10.1037/h0042519

[16] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning Representations by Back-Propagating Errors. *Nature*, **323**, 533-536. https://doi.org/10.1038/323533a0

[17] Nair, V. and Hinton, G.E. (2010) Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the* 27*th International Conference on Machine Learning*, Haifa, 21-24 June 2010, 1-8.

[18] Maas, A.L., Hannun, A.Y. and Ng, A.Y. (2013) Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proceedings of the* 30*th International Conference on Machine Learning*, Vol. 28, 3.

[19] Clevert, D.-A., Unterthiner, T. and Hochreiter, S. (2015) Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). 4*th International Conference on Learning Representations*, *ICLR* 2016, San Juan, 2-4 May 2016, 1-14. http://arxiv.org/abs/1511.07289

[20] Hendrycks, D. and Gimpel, K. (2016) Gaussian Error Linear Units (GELUs).

http://arxiv.org/abs/1606.08415

[21] Ramachandran, P., Zoph, B. and Le, Q.V. (2017) Searching for Activation Functions.

[22] Jain, A., Kumar, A., Choubey, A., Balasubramanian, V. and Jawahar, C. (2019) Deep Fusion of Multiple Branches for High Resolution Image Classification. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 9097-9106.

[23] Chandra, B. and Sharma, R.K. (2016) Deep Learning with Adaptive Learning Rate Using Laplacian Score. *Expert Systems with Applications*, **63**, 1-7. https://doi.org/10.1016/j.eswa.2016.05.022

[24] Hu, W., Xie, S., Li, Y., Gu, C., Zhang, J. and Liu, J. (2020) Differentiable Neural Architecture Search with Fused Activation Functions. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 14179-14188.

[25] Das, P., Kadian, A., Torr, P. and Namboodiri, A. (2021) BBAFNets: Bidirectional Bottleneck Attention Fusion Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 7781-7790.

[26] Montesinos López, O.A., Montesinos López, A. and Crossa, J. (2022) Fundamentals of Artificial Neural Networks and Deep Learning. In: López, O.A.M., López, A.M. and Crossa, J., Eds., *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, Springer International Publishing, 379-425. https://doi.org/10.1007/978-3-030-89010-0_10

[27] López, O.A.M., López, A.M. and Crossa, J. (2022) Multivariate Statistical Machine Learning Methods for Genomic Prediction. Springer.

[28] Liu, H., Fu, Q., Du, L., Zhang, T., Yu, G., Han, S., *et al.* (2022) Learning Rate Perturbation: A Generic Plugin of Learning Rate Schedule towards Flatter Local Minima. *Proceedings of the* 31*st ACM International Conference on Information & Knowledge Management*, Atlanta, 17-21 October 2022, 4234-4238. https://doi.org/10.1145/3511808.3557626

[29] Kourounioti, I. and Polydoropoulou, A. (2017) Identification of Container Dwell Time Determinants Using Aggregate Data. *International Journal of Transport Economics*, **44**, 567-588.

[30] Kourounioti, I., Polydoropoulou, A. and Tsiklidis, C. (2016) Development of Models Predicting Dwell Time of Import Containers in Port Container Terminals—An Artificial Neural Networks Application. *Transportation Research Procedia*, **14**, 243-252. https://doi.org/10.1016/j.trpro.2016.05.061

[31] Kamiri, J. and Mariga, G. (2021) Research Methods in Machine Learning: A Content Analysis. *International Journal of Computer and Information Technology*, **10**, 78-91. https://doi.org/10.24203/ijcit.v10i2.79

[32] Al-Bakri, A.Y. and Sazid, M. (2021) Application of Artificial Neural Network (ANN) for Prediction and Optimization of Blast-Induced Impacts. *Mining*, **1**, 315-334. https://doi.org/10.3390/mining1030020

[33] Rusgiyarto, F., Sjafruddin, A., Frazila, R.B., Suprayogi and Burhani, J.T. (2018) Inland Container Depots Effect for Import Container Terminal Performance at Koja Container Terminal, Jakarta Based on Optimization-Simulation Model. *AIP Conference Proceedings*, **1977**, Article ID: 020038. https://doi.org/10.1063/1.5042894

[34] De Wibowo Muhammad Sidik, A., Ramdani, D., Sopandita, D., Zam Fadilah, A.Z. and Efendi, E. (2020) Modelling and Optimization Containers Dwell-Time in Tanjung

Perak Port Indonesia. 2020 6*th International Conference on Computing Engineering and Design* (*ICCED*), Sukabumi, 15-16 October 2020, 1-4. https://doi.org/10.1109/icced51276.2020.9415805

[35] De Armas Jacomino, L., Medina-Pérez, M.A., Monroy, R., Valdes-Ramirez, D., Morell-Pérez, C. and Bello, R. (2021) Dwell Time Estimation of Import Containers as an Ordinal Regression Problem. *Applied Sciences*, **11**, Article No. 9380. https://doi.org/10.3390/app11209380

[36] Yu, M., Liang, Z., Teng, Y., Zhang, Z. and Cong, X. (2021) The Inbound Container Space Allocation in the Automated Container Terminals. *Expert Systems with Applications*, **179**, Article ID: 115014. https://doi.org/10.1016/j.eswa.2021.115014

[37] Çetiner, H. and Metlek, S. (2023) Classification of Weather Phenomenon with a New Deep Learning Method Based on Transfer Learning. *International Conference on Recent Academic Studies*, **1**, 92-99. https://doi.org/10.59287/icras.678