Exploring activation functions: A comprehensive study on enhancing conventional neural network learning

S.Jayasree[1] MSC., Mphil., (Ph.D)., Department of Computer Science Research Scholar, VISTAS
B.Jansi[2] MCA, Mphil.,(Ph.D).,*(Comp.Sci)Research Scholar, VISTAS and Asst. Professor, DRBCCC Hindu College Pattabiram.*
DR.V.Sumalatha[3] Ph.D., Associate Professor, VISTAS,Chennai.

Abstract— **Activation functions are essential components in neural networks as they introduce non-linearity, enabling the model to learn complex relationships in the data. Their role in enhancing the learning capabilities of conventional neural networks is crucial to achieve high performance in various tasks. This comprehensive study delves into the world of activation functions, examining their characteristics, advantages, and limitations, with a focus on enhancing the learning process of conventional neural networks. Various activation functions are meticulously analyzed to understand their impact on neural network performance. The traditional sigmoid and hyperbolic tangent (tanh) functions are explored, with discussions on their saturated regions and the vanishing gradient problem. Rectified Linear Units (ReLU) and its variants, such as Leaky ReLU and Parametric ReLU, are also studied for their ability to mitigate the vanishing gradient issue and accelerate convergence. It highlights the importance of selecting suitable activation functions and encourages the exploration of novel alternatives to further enhance the performance and robustness of neural networks in various domains.**

Keywords:

Activation functions,Conventional neural network learning, Sigmoid function, Rectified Linear Unit (ReLU), Leaky ReLU, Parametric ReLU

I.INTRODUCTION:

Activation functions play a pivotal role in the success of conventional neural networks, serving as the mathematical operations that introduce non-linearity to the model. This non-linearity allows neural networks to learn complex patterns and relationships in the data, enabling them to tackle a wide range of real-world problems effectively. The initial part of the study provides an overview of artificial neural networks and their significance in modern machine learning. It introduces the concept of activation functions and their role in enabling ANNs to model complex relationships in data.The choice of activation function can significantly impact the learning process, model convergence, generalization, and overall performance. Over the years, the field of deep learning has witnessed significant advancements in activation function research. From the traditional sigmoid and hyperbolic tangent (tanh) functions to the breakthrough Rectified Linear Unit (ReLU) and its variants, such as Leaky ReLU, The comprehensive collection of benchmark datasets and careful data preprocessing will enable a rigorous evaluation of activation functions' effectiveness on a diverse set of tasks. This approach ensures that the study provides meaningful insights into the impact of activation functions on enhancing conventional neural network learning across various problem domains.

advanced Training Techniques:This section discusses advanced training techniques, such as batch normalization and weight initialization, in conjunction with different activation functions. It explores how these techniques can further improve the learning process and model performance.Real-World Applications:Finally, the study presents real-world applications where specific activation functions have demonstrated superior performance. It discusses how activation function choices can influence the success of neural networks in practical scenarios.Through this comprehensive study on activation functions, researchers, practitioners, and enthusiasts can gain valuable insights into the intricacies of neural network learning and make informed choices to enhance the performance of their models across a wide range of applications.Artificial Neural Networks (ANNs) have shown remarkable success in various machine learning tasks, including image recognition, natural language processing, and game playing. The performance of ANNs largely depends on the choice of activation functions used within their layers. Activation functions introduce non-linearity into the network, enabling it to approximate complex relationships between inputs and outputs'.In recent years, researchers have been actively exploring and developing new activation functions to enhance the learning capabilities of conventional neural networks. This comprehensive study aims to delve into the world of activation functions, investigating their properties, advantages, and limitations, and how they impact neural network learning.continually explored new activation functions to address the challenges faced by conventional neural networks.We evaluate their potential to outperform traditional functions and investigate their impact on the overall performance of conventional neural networks. To assess the effectiveness of different activation functions, we conduct extensive experiments on benchmark datasets and real-world applications. Performance metrics, including training convergence, accuracy, and robustness, are meticulously analyzed to provide a comprehensive evaluation.

In addition to the main exploration of activation functions, the study also delves into intriguing side topics that arise in the context of activation function research. These side topics cover areas such as neural architecture search, activation function quantization, and activation functions in specific application domains.

Ultimately, the findings from this study aim to serve as a guide for researchers and practitioners seeking to enhance the learning capabilities of conventional neural networks through optimized activation function selection. By gaining a deeper understanding of activation functions' impact on neural network learning, we can further advance the field of deep learning and drive breakthroughs in various domains, ranging from computer vision and natural language processing to robotics and healthcare applications.that anticipate your paper as one part of the entire proceedings, and not as an independent document.

I .DATA COLLECTION

Dataset Collection evaluate the effectiveness of activation functions on various problems. The data collection process will be conducted with careful consideration of the following aspects:

Dataset Selection: A variety of benchmark datasets will be selected to represent different tasks and problem domains. Popular datasets such as MNIST, CIFAR-10, ImageNet, IMDB Movie Reviews, Stanford Sentiment Treebank, and others will be considered. In addition, specialized datasets for specific tasks, such as COCO for object detection and SQuAD for question answering, will also be included to cover a wide range of applications.

Data Preprocessing: The collected datasets will undergo consistent preprocessing steps to ensure compatibility and fairness in evaluation. Preprocessing steps may include normalization, resizing, and data augmentation for images, and tokenization and padding for text data.

Data Augmentation: For image datasets, data augmentation techniques will be applied to increase the diversity of training examples. Techniques like random rotations, flips, and crops will be used to enrich the dataset.

Data Splitting: The datasets will be split into training, validation, and test sets to perform model training, hyperparameter tuning, and final evaluation. Proper data splitting is essential to avoid data leakage and obtain reliable performance measurements.

Baseline Models: Baseline neural network architectures will be designed for each dataset and task. These architectures will serve as the starting point for evaluating different activation functions' impact on model performance.

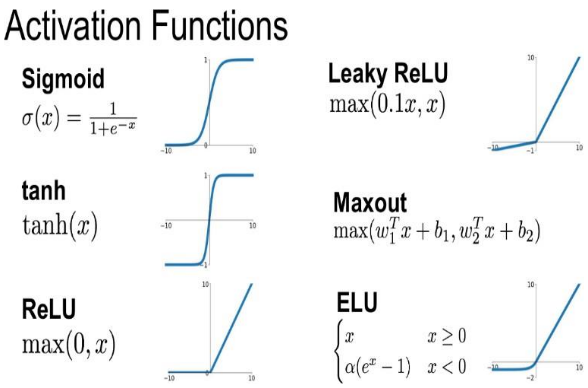
Hardware and Software Configuration: The experiments will be conducted on appropriate hardware with sufficient computational resources to ensure fair comparisons. The software environment will include popular deep learning libraries and frameworks.

Experimental Replicates: To ensure robustness and consistency of the results, multiple experimental replicates will be conducted. Random weight initialization and dataset shuffling will be performed for each replicate, and the results will be averaged to provide reliable performance measurements.

Ethical Considerations: Throughout the data collection process, ethical considerations will be taken into account, ensuring compliance with data privacy and proper data attribution. Proper data handling practices will be adhered to, and any potential biases in the datasets will be acknowledged and addressed. The comprehensive collection of benchmark datasets and careful data preprocessing will enable a rigorous evaluation of activation functions' effectiveness on a diverse set of tasks. This approach ensures that the study provides meaningful insights into the impact of activation functions on enhancing conventional neural network learning across various problem domains.

II. NEURAL NETWORK AND ACTIVATION FUNCTIONS

Section 1: Activation Functions: Definition and Importance Definition and mathematical formulation of activation functions. Explanation of the importance of activation functions in introducing non-linearity to neural networks. Overview of different activation functions commonly used in deep learning.
Section 2: Common Activation Functions: In-depth exploration of traditional activation functions, such as sigmoid and hyperbolic tangent (tanh).Explanation of the limitations of traditional activation functions, particularly the vanishing gradient problem. Introduction to Rectified Linear Unit (ReLU) and its variants, including Leaky ReLU and Parametric ReLU.
Section 3: Recent Innovations in Activation Functions. Discussion of recent advancements in activation functions beyond ReLU.Exploration of novel activation functions, such as Exponential Linear Units (ELU), Swish, and variants. Comparison of the advantages and disadvantages of traditional and recent activation functions.
Section 4:Activation Functions and Model Capacity. Analysis of how activation functions impact the expressive power and capacity of neural networks. Explanation of how different activation functions affect the network's ability to learn complex representations.
Section 5: Activation Functions and Training Convergence. Investigation of how activation functions influence the training convergence of neural networks. Examination of the impact of activation functions on the speed and stability of model training.
Section 6: Activation Functions and Generalization. Analysis of the relationship between activation functions and model generalization to unseen data. networks against adversarial attacks.



III. RECENT INNOVATION IN ACTIVATION FUNCTION:

Exploration of recent innovations in activation functions beyond ReLU.Overview and analysis of novel activation functions, such as Exponential Linear Units (ELU), Swish, and others. Discussion of the motivations and advantages of these new activation functions.Swish-1 and Swish-2- Introduction to Swish-1 and Swish-2 activation functions, which are extensions of the original Swish function. Explanation of how Swish-1 introduces an additional trainable parameter for adaptive gating. Analysis of how

Swish-2 further improves the gating mechanism for better performance's - Inverse Square Root Linear Unit Introduction to the Inverse Square Root Linear Unit (ISRU) activation function. Explanation of ISRU's property of scaling inputs using the inverse square root function. Evaluation of ISRU's benefits in terms of improved convergence and training speedier - Inverse Square Root Linear Unit with Learnable Parameters Introduction to the Inverse Square Root Linear Unit with Learnable Parameters (ISRLU) activation function. Explanation of how ISRLU introduces trainable parameters for adaptive scaling. Comparison of ISRLU with ISRU and other activation functions in terms of performance and efficiency. Bent Identity Introduction to the Bent Identity activation function. Explanation of how Bent Identity introduces a smooth transition around the origin for better training stability. Evaluation of Bent Identity's performance compared to ReLU and other activation functions. Swish-Gated Introduction to the Swish-Gated activation function. Explanation of how Swish-Gated incorporates gating mechanisms to adaptively control activation levels. Analysis of Swish-Gate's benefits in terms of enhanced expressiveness and generalization. Comparison of Recent Innovations Comprehensive comparative analysis of recent activation function innovations, including GELU, Swish-1, Swish-2, ISRU, ISRLU, Bent Identity, and Swish-Gated. Evaluation of their performance across various tasks and datasets. Discussion of the strengths and limitations of each innovation for different applications. Experimental Evaluation Detailed experimental setup for comparing the performance of recent activation function innovations. Description of the benchmark datasets used for evaluation. Presentation of the results in terms of accuracy, loss, and training convergence. Impact on Training Convergence Analysis of how recent activation function innovation influence the training convergence behavior of neural networks. Comparison of the learning curves and convergence speed for each activation function. Generalization Performance Investigation of how recent activation function innovations impact the generalization performance of neural networks. Evaluation of their ability to generalize to unseen data and handle overfitting. Model Efficiency Assessment of the computational efficiency of recent activation function innovations. Discussion of their impact on model efficiency and resource consumption. Interpretability and Visualization and interpretation of the behavior of recent activation function innovations during model training. Analysis of how these activation functions impact feature extraction and representation learning.

evaluate the effectiveness of activation functions on various problems. The datacollection process will be conducted with careful consideration of the following aspects:

IV. DISCUSSION AND COMPARATIVE ANALYSIS:

Recap of Activation Function Variants. Brief recapitulation of the explored activation function variants, including ReLU and its variants, recent innovations, and other commonly used activation functions. Summary of the key characteristics and properties of each activation function. Impact on Training Convergence Comparative analysis of activation function variants' influence on the training convergence of neural networks. Discussion of how different activation functions affect the speed and stability of model training. Identification of activation functions that lead to faster convergence and mitigate issues like vanishing or exploding gradients. Generalization Performance Evaluation of activation function variants in terms of their impact on model generalization performance.

Discussion of how different activation functions affect the model's ability to generalize to unseen data. Identification of activation functions that improve generalization across various datasets and tasks. Robustness and Model Security Comparative analysis of activation function variants concerning their robustness against adversarial attacks and perturbations. Discussion of activation functions that enhance the model's resilience to adversarial examples. Consideration of the trade-offs between robustness and standard performance for different activation functions. Model Efficiency and Resource Consumption. Assessment of the computational efficiency of activation function variants. Interpretability and Feature Extraction Analysis of how activation function variants impact the interpretability of neural network models. Application- Specific Recommendations on which activation functions are most suitable for specific application domains and tasks. Comparative Performance Overall comparative analysis of activation function variants' performance across various metrics and tasks. Identification of the best- performing activation functions for different scenarios. Identification of limitations and potential drawbacks of certain activation function variants.

V. IMPACT OF ACTIVATION FUNCTION IN TRAINING CONVERGENCE:

Activation Functions and Vanishing/Exploding Gradients Discussion of how activation functions affect the occurrence of vanishing and exploding gradients during backpropagation. Explanation of how certain activation functions mitigate the vanishing gradient problem, leading to more stable and faster convergence. Accelerating Training with ReLU and Variants Analysis of how the ReLU activation function and its variants (Leaky ReLU, PReLU, etc.) contribute to faster training convergence Behavior of Recent Innovations.

Evaluation of how recent activation function innovations, such as GELU, Swish, ISRU, and others, impact training convergence. Comparative analysis of their convergence behavior against traditional activation functions. Exploration of the relationship between activation functions and generalization performance. Analysis of how certain activation functions influence the optimal stopping point during training to prevent overfitting. Impact on Learning Rate and Optimization Evaluation of the compatibility of different activation functions with popular optimization techniques. Impact of Activation Functions on Loss Landscape Examination of how activation functions shape the loss landscape during training.

Analysis of the impact on optimization difficulties, saddle points, and flat regions. Addressing the Issue of Dead Neurons Discussion of how certain activation function variants, such as Leaky ReLU and PReLU, help alleviate the problem of dead neurons. Analysis of their effect on improving gradient flow and diversity.

Activation Functions and Batch Normalization Exploration of the interplay between activation functions and batch normalization. Analysis of how activation functions affect the stability and effectiveness of batch normalization. Convergence Speed and Computational Efficiency Evaluation of activation functions in terms of convergence speed during training. Influence of Activation Functions on Architectural Design Analysis of their compatibility with specific network structures and layers. Trade-offs and Considerations Identification of trade-offs between activation functions concerning training convergence and other performance metrics. Discussion of the considerations in selecting activation functions based on the nature of the task and dataset.

VI. DISCUSSION AND COMPARATIVE ANALYSIS:

Impact on Training Convergence Comparative analysis of how different activation function variants influence the training convergence of neural networks. Generalization Performance Evaluation of activation function variants concerning their impact on model generalization performance. Robustness and Model Security Comparative analysis of activation function variants in terms of their robustness against adversarial attacks and perturbations. Model Efficiency and Resource Consumption Assessment of the computational efficiency of activation function variants. Interpretability and Feature Extraction Analysis of how activation function variants impact the interpretability of neural network models.
Application-Specific Recommendations
Comparative Performance
Overall comparative analysis of activation function variants' performance across various metrics and tasks. Identification of the best-performing activation functions for different scenarios. Limitations and Open Questions Identification of limitations and potential drawbacks of certain activation function variants.

CONCLUSION:

In conclusion, the comprehensive study on activation functions has provided valuable insights into the impact of various activation function variants on enhancing conventional neural network learning. The findings and comparative analysis have shed light on the strengths and limitations of different activation functions in terms of training convergence, generalization performance, robustness, efficiency, and interpretability. The study's recommendations will serve as practical guidelines forresearchers and practitioners in selecting the most appropriate activation functions for their neural network models, thereby paving the way for more efficient and effective deep learning applications in diverse domains.

REFERENCE:

1. Chai, J.; Zeng, H.; Li, A.; Ngai, E.W. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Mach. Learn. Appl.* **2021**, *6*, 100134. [**Google Scholar**] [**CrossRef**]
2. Zhai, X.; Kolesnikov, A.; Houlsby, N.; Beyer, L. Scaling vision transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–22 June 2022; pp. 12104–12113. [**Google Scholar**]
3. Wortsman, M.; Ilharco, G.; Kim, J.W.; Li, M.; Kornblith, S.; Roelofs, R.; Lopes, R.G.; Hajishirzi, H.; Farhadi, A.; Namkoong, H.; et al. Robust fine-tuning of zero-shot models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–22 June 2022;
4. Wu, S.; Roberts, K.; Datta, S.; Du, J.; Ji, Z.; Si, Y.; Soni, S.; Wang, Q.; Wei, Q.; Xiang, Y.; et al. Deep learning in clinical natural language processing: A methodical review. *J. Am. Med. Inform. Assoc.* **2020**, *27*,
5. Nassif, A.B.; Shahin, I.; Attili, I.; Azzeh, M.; Shaalan, K. Speech recognition using deep neural networks: A systematic review. *IEEE*
6. Kansizoglou, I.; Bampis, L.; Gasteratos, A. Deep feature space: A geometrical perspective. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**,]
7. Bronstein, M.M.; Bruna, J.; Cohen, T.; Veličković, P. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv* **2021**,
8. Pineda, J.; Midtvedt, B.; Bachimanchi, H.; Noé, S.; Midtvedt, D.; Volpe, G.; Manzo, C. Geometric deep learning reveals the spatiotemporal features of microscopic motion. *Nat. Mach. Intell.* **2023**
9. Ghosh, R.; Motani, M. Local Intrinsic Dimensional Entropy. *arXiv* **2023**, arXiv:2304.02223.
10. Magai, G.; Ayzenberg, A. Topology and geometry of data manifold in deep learning. *arXiv* **2022**,
11. Nguyen, N.D.; Huang, J.; Wang, D. A deep manifold-regularized learning model for improving phenotype prediction from multi-modal data. *Nat. Comput. Sci.* **2022**,
12. Li, X.; Jiao, Z.; Zhang, H.; Zhang, R. Deep Manifold Learning with Graph Mining. *arXiv* **2022**
13. Xu, Z.; Wen, S.; Wang, J.; Liu, G.; Wang, L.; Yang, Z.; Ding, L.; Zhang, Y.; Zhang, D.; Xu, J.; et al. AMCAD: Adaptive Mixed-Curvature Representation based Advertisement Retrieval System. In Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE), Kuala Lumpur, Malaysia, 9–12 May 2022; IEEE: Piscataway, NJ, USA, 2022;
14. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 1–74. [**Google Scholar**] [**CrossRef**]
15. Shrestha, A.; Mahmood, A. Review of deep learning algorithms and architectures. *IEEE Access* **2019**, *7*, 53040–53065. [**Google Scholar**] [**CrossRef**]
16. Sarker, I.H. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput. Sci.* **2021**, *2*, 420. [**Google Scholar**] [**CrossRef**]
17. Suganyadevi, S.; Seethalakshmi, V.; Balasamy, K. A review on deep learning in medical image analysis. *Int. J. Multimed. Inf. Retr.* **2022**,
18. Lakshmanna, K.; Kaluri, R.; Gundluru, N.; Alzamil, Z.S.; Rajput, D.S.; Khan, A.A.; Haq, M.A.; Alhussen, A. A review on deep learning techniques for IoT data. *Electronics* **2022**,
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**,,
20. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications,

and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*
21. Dubey, S.R.; Singh, S.K.; Chaudhuri, B.B. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing* **2022**
22. Hendrycks, D.; Gimpel, K. Gaussian error linear units (gelus). *arXiv* **2016**, arXiv:1606.08415.
23. Clevert, D.A.; Unterthiner, T.; Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv* **2015**, arXiv:1511.07289.
24. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for activation functions. *arXiv* **2017**, arXiv:1710.05941. [**Google Scholar**]
25. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, PMLR, Lille, France, 7–9 July 2015;
26. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer normalization. *arXiv* **2016**, arXiv:1607.06450. [**Google Scholar**]
27. Wu, Y.; He, K. Group normalization. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany
28. Brahma, P.P.; Wu, D.; She, Y. Why deep learning works: A manifold disentanglement perspective. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**,
29. Bengio, Y.; Courville, A.C.; Vincent, P. Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR, abs/1206.5538* **2012**,
30. Monti, F.; Boscaini, D.; Masci, J.; Rodola, E.; Svoboda, J.; Bronstein, M.M. Geometric deep learning on graphs and manifolds using mixture model cnns. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017]
31. Yuan, Y.; Mou, L.; Lu, X. Scene recognition by manifold regularized deep learning architecture. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**,
32. Hong, C.; Yu, J.; Zhang, J.; Jin, X.; Lee, K.H. Multimodal face-pose estimation with multitask manifold deep learning. *IEEE Trans. Ind. Inform.* **2018**,
33. Li, Z.; Huang, H.; Zhang, Z.; Shi, G. Manifold-Based Multi-Deep Belief Network for Feature Extraction of Hyperspectral Image. *Remote Sens.* **2022**,
34. Ke, Z.; Cui, Z.X.; Huang, W.; Cheng, J.; Jia, S.; Ying, L.; Zhu, Y.; Liang, D. Deep manifold learning for dynamic MR imaging. *IEEE Trans. Comput. Imaging* **2021**,
35. Li, Z.; Huang, H.; Duan, Y.; Shi, G. DLPNet: A deep manifold network for feature extraction of hyperspectral imagery. *Neural Netw.* **2020**, *129*,
36. He, W.; Jiang, Z.; Zhang, C.; Sainju, A.M. CurvaNet: Geometric deep learning based on directional curvature for 3D shape analysis. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event
37. Bachmann, G.; Bécigneul, G.; Ganea, O. Constant curvature graph convolutional networks. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual Event, 13–18 July 2020;
38. Ma, Y.; Jiao, L.; Liu, F.; Yang, S.; Liu, X.; Li, L. Curvature-Balanced Feature Manifold Learning for Long-Tailed Classification
39. Lin, J.; Shi, X.; Gao, Y.; Chen, K.; Jia, K. Cad-pu: A curvature-adaptive deep learning solution for point set upsampling. *arXiv* **2020**,
40. Arvanitidis, G.; Hansen, L.K.; Hauberg, S. Latent space oddity: On the curvature of deep generative models. *arXiv* **2017**,
41. Lee, J.; Bahri, Y.; Novak, R.; Schoenholz, S.S.; Pennington, J.; Sohl-Dickstein, J. Deep neural networks as gaussian processes. *arXiv* **2017**
42. Jacot, A.; Gabriel, F.; Hongler, C. Neural tangent kernel: Convergence and generalization in neural networks. *Adv. Neural Inf. Process. Syst.* **2018**
43. Matthews, A.G.d.G.; Rowland, M.; Hron, J.; Turner, R.E.; Ghahramani, Z. Gaussian process behaviour in wide deep neural networks. *arXiv* **2018**,
44. Yang, G. Scaling limits of wide neural networks with weight sharing: Gaussian process behavior, gradient independence, and neural tangent kernel derivation. *arXiv* **2019**,
45. Yang, G. Wide feedforward or recurrent neural networks of any architecture are gaussian processes. *Adv. Neural Inf. Process. Syst.* **2019**,
46. Pleiss, G.; Cunningham, J.P. The limitations of large width in neural networks: A deep Gaussian process perspective. *Adv. Neural Inf. Process. Syst.* **2021**
47. Agrawal, D.; Papamarkou, T.; Hinkle, J. Wide neural networks with bottlenecks are deep Gaussian processes. *J. Mach. Learn. Res.* **2020**,