BENCHMARKING NEURAL NETWORKS AND GRADIENT BOOSTING ON TIME-SERIES AND IMAGE DATA: COMPARING PREDICTIVE ACCURACY, TRAINING SPEED, AND MODEL STABILITY

[1]Kommala Aparna, [2]M. Kiran Kumar

[1]Assistant Professor, Dept of CSE, SICET-Hyderabad
[2]Assistant Professor, Dept of CSE, GITAM University-Hyderabad

ABSTRACT: *This study systematically evaluates the performance of neural networks (NN) and gradient boosting (GB) methods across diverse datasets encompassing time-series and image data domains. For time-series analysis, recurrent neural networks (RNNs) and gradient boosting techniques like XGBoost were compared using metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) on datasets from finance and healthcare sectors. Results demonstrate that RNNs consistently outperform GB methods in time-series prediction tasks, achieving lower RMSE and MAE values albeit with longer training times. Conversely, in image classification tasks using datasets such as MNIST and CIFAR-10, convolutional neural networks (CNNs) exhibited superior accuracy, precision, and recall compared to gradient boosting algorithms. These findings underscore the suitability of RNNs for capturing temporal dependencies in sequential data and the effectiveness of CNNs in handling complex spatial patterns in image data.*

INTRODUCTION

In recent years, the field of predictive modeling has been significantly influenced by the advancements in neural networks (NN) and gradient boosting (GB) techniques. Neural networks, particularly deep learning models, have revolutionized the landscape of machine learning with their ability to automatically learn hierarchical representations of data. This capability has made them exceptionally effective in handling complex patterns and relationships within large datasets. Applications of NN span across diverse domains, from natural language processing and speech recognition to image analysis and autonomous systems. The flexibility and scalability of neural networks have propelled them to become the go-to choice for many data scientists tackling various predictive tasks.

Similarly, gradient boosting methods have garnered widespread attention due to their robustness and high predictive accuracy. Algorithms like XGBoost, LightGBM, and CatBoost have demonstrated exceptional performance in structured data settings, where they excel in handling heterogeneous features and capturing nonlinear relationships. This makes gradient boosting particularly suited for tasks such as financial forecasting, fraud detection, and personalized medicine. The ability of GB models to ensemble weak learners sequentially while minimizing loss functions has cemented their reputation as formidable contenders in predictive analytics.

The significance of time-series and image data cannot be overstated across numerous domains. In finance, time-series data drives critical decisions ranging from stock market predictions to risk assessment and algorithmic trading strategies. Healthcare relies heavily on time-series data for patient monitoring, disease progression modeling, and medical device analytics. In computer vision, image data is pivotal for tasks such as object detection, facial recognition, and autonomous navigation systems. These domains underscore the importance of accurate and efficient predictive models, capable of extracting meaningful insights from intricate data patterns.

In finance, for instance, accurate forecasting of market trends demands models that can comprehend the temporal dependencies inherent in financial time-series. In healthcare, the ability to predict disease progression from patient vitals over time can lead to early interventions and improved patient outcomes. Likewise, in computer vision applications, the effectiveness of image recognition systems hinges on the model's ability to discern intricate visual patterns and generalize across diverse datasets.

Given the critical role that NN and GB models play in these domains, understanding their comparative advantages in terms of predictive accuracy, training speed, and model stability becomes paramount. This research aims to address these aspects through a systematic benchmarking study, leveraging both time-series and image data, to provide insights that can guide the selection and optimization of models based on specific application requirements and constraints. By evaluating these models across different datasets and scenarios, we aim to contribute to the broader understanding of their applicability and performance in real-world predictive modeling tasks.

The primary objective of this research is to conduct a comprehensive comparative analysis between neural networks (NN) and gradient boosting (GB) methods across two distinct types of data: time-series and image data. Our study focuses on three key performance metrics: predictive accuracy, training speed, and model stability.

Predictive accuracy stands as a cornerstone metric in evaluating the effectiveness of machine learning models. For both time-series and image data, accurate predictions are crucial for making informed decisions across a spectrum of applications. Neural networks, with their ability to learn complex patterns and hierarchies in data, often excel in tasks where intricate relationships and non-linear dependencies are prevalent. Conversely, gradient boosting techniques, known for their ensemble of decision trees and iterative refinement of weak learners, exhibit robust performance in structured data scenarios. By systematically comparing the predictive accuracy of NN and GB models across multiple datasets within each data type, we aim to provide empirical insights into which approach yields superior performance under different conditions.

Training speed emerges as another pivotal aspect in model evaluation, particularly in applications requiring real-time or near-real-time predictions. Neural networks, especially deep architectures, are renowned for their computational intensity during training, often necessitating substantial computational resources and time. In contrast, gradient boosting methods typically train faster due to their sequential training approach and optimization strategies. Understanding the trade-offs between training speed and predictive accuracy is critical for practical deployment, where efficient resource utilization is paramount. Our research aims to quantify and compare the training times of NN and GB models across varied datasets, thereby shedding light on their scalability and efficiency.

Model stability, the third dimension of our study, addresses the robustness and consistency of predictions across different runs or variations in data. Neural networks are susceptible to variability in training outcomes due to their sensitivity to initialization parameters and non-convex optimization landscapes. Gradient boosting, by contrast, often exhibits more stable performance, thanks to its additive nature and regularization techniques. Evaluating model stability provides insights into the reliability and generalizability of NN and GB models in real-world applications, where consistency in predictions is essential for decision-making processes.

By systematically benchmarking NN and GB models on both time-series and image data, our research aims to provide nuanced insights into their comparative strengths and weaknesses. These insights can inform practitioners and researchers in selecting the most suitable modeling approach based on specific application requirements, dataset characteristics, and computational constraints. Ultimately, our findings seek to contribute to advancing the state-of-the-art in predictive modeling, fostering more informed decisions in domains ranging from finance and healthcare to computer vision and beyond.

Emergence of Deep Learning and Ensemble Methods

The emergence of deep learning has revolutionized the field of artificial intelligence, particularly in the realm of predictive modeling. Deep neural networks, with their multiple layers and ability to automatically extract intricate features from data, have significantly advanced the capabilities of machine learning systems. Concurrently, ensemble methods such as gradient boosting have gained prominence for their effectiveness in aggregating the predictions of multiple weak learners into a robust final model. Understanding the evolution and strengths of these methodologies sets the stage for evaluating their comparative performance in predictive tasks involving time-series and image data.

Applications Across Diverse Domains

The application domains of predictive modeling span a wide spectrum, each presenting unique challenges and opportunities. In finance, predictive analytics powers algorithms for stock market forecasting, risk management, and algorithmic trading. Healthcare relies on predictive models for disease diagnosis, treatment planning, and patient outcome prediction based on longitudinal data. In computer vision, the ability to accurately classify and interpret images drives advancements in autonomous vehicles, facial recognition systems, and augmented reality applications. These diverse domains underscore the versatility and applicability of neural networks and gradient boosting techniques in addressing complex predictive challenges.

Challenges in Time-Series Analysis

Time-series data, characterized by temporal dependencies and sequential patterns, poses distinct challenges for predictive modeling. Traditional statistical methods often struggle to capture the non-linear relationships and long-term dependencies present in time-series data. Neural networks, particularly recurrent and convolutional architectures, have demonstrated prowess in modeling sequential data by leveraging memory and spatial hierarchies. Conversely, gradient boosting methods adapt well to structured time-series data, leveraging decision tree ensembles to capture complex temporal interactions. Understanding how NN and GB models navigate these challenges informs their comparative effectiveness in time-series prediction tasks.

Complexity and Diversity of Image Data

Image data, with its high-dimensional and complex structure, presents another frontier for predictive modeling. From object detection and segmentation to image classification and generative modeling, the ability to accurately interpret visual information is critical across various industries. Convolutional neural networks (CNNs) have emerged as the cornerstone of deep learning for image analysis, leveraging spatial hierarchies and feature extraction capabilities to achieve state-of-the-art performance in tasks like image recognition and scene understanding. Gradient boosting methods have also made strides in image data applications, albeit with adaptations such as feature engineering and representation learning. Exploring how NN and GB models handle the intricacies of image data elucidates their comparative advantages in visual recognition tasks.

Ethical and Societal Implications

As predictive modeling techniques continue to evolve and proliferate across industries, considerations of ethical implications and societal impact become increasingly salient. Issues such as algorithmic bias, privacy concerns, and transparency in decision-making underscore the need for responsible deployment and rigorous evaluation of predictive models. Understanding how different modeling approaches, such as NN and GB, address these ethical and societal challenges can guide stakeholders in leveraging predictive analytics for positive societal outcomes while mitigating potential risks.

LITERATURE SURVEY

Neural networks (NN) represent a class of machine learning models inspired by the biological neurons in the human brain. They consist of interconnected layers of neurons, each performing computations and passing outputs to the next layer. For time-series data, recurrent neural networks (RNNs) and its variants such as Long Short-Term Memory (LSTM) networks are commonly employed. RNNs are designed to capture sequential dependencies by maintaining a state or memory of previous inputs, making them suitable for tasks where temporal order is crucial, such as stock market predictions or weather forecasting. LSTMs, an extension of RNNs, enhance the ability to learn and remember over long sequences, thereby addressing the vanishing gradient problem and improving performance in longer-term dependencies.

In contrast, for image data, convolutional neural networks (CNNs) have emerged as the dominant architecture. CNNs are uniquely suited to handle the spatial relationships present in images. They operate by employing convolutional layers that systematically apply filters across the input image to extract hierarchical representations of features. This hierarchical feature extraction enables CNNs to achieve state-of-the-art performance in tasks like image classification, object detection, and semantic segmentation. Recent advancements in CNN architectures include models like ResNet, DenseNet, and EfficientNet, which optimize network depth, connectivity patterns, and computational efficiency, respectively, pushing the boundaries of accuracy and scalability in image analysis tasks.

Recent advancements in neural network techniques have focused on enhancing model performance, scalability, and interpretability across various domains. One notable area of advancement is the integration of attention mechanisms in neural networks, originally popularized in natural language processing tasks. Attention mechanisms improve the model's ability to focus on relevant parts of the input, enhancing both accuracy and efficiency. This has been particularly transformative in tasks involving sequential data, where attention mechanisms enable models to selectively attend to important temporal features.

Furthermore, advancements in regularization techniques such as dropout, batch normalization, and weight regularization have contributed to improving the generalization and robustness of neural networks. These techniques mitigate overfitting and stabilize training by introducing noise or constraints during the optimization process, thereby enhancing model stability across different datasets and training conditions.

Additionally, the adoption of transfer learning and pre-trained models has democratized access to state-of-the-art performance in neural network applications. Transfer learning allows models trained on large, diverse datasets (e.g., ImageNet) to be fine-tuned for specific tasks with smaller datasets, reducing the need for extensive labeled data and accelerating model development. This approach has been particularly effective in domains where labeled data is scarce or expensive to acquire, such as medical imaging and remote sensing.

Gradient boosting methods represent a powerful class of machine learning techniques that build predictive models by sequentially combining the predictions of an ensemble of weak learners, typically decision trees. Algorithms like XGBoost (Extreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine), and CatBoost (Categorical Boosting) have gained prominence for their ability to achieve high predictive accuracy across a variety of domains. These algorithms iteratively optimize a loss function by adding new models that predict the residuals or gradients of the previous models, thereby minimizing the error iteratively.

XGBoost, one of the most widely used gradient boosting frameworks, stands out for its efficiency and scalability. It incorporates regularization techniques, tree pruning, and hardware optimization to deliver high performance even on large datasets. LightGBM, developed by Microsoft, introduces novel strategies like gradient-based one-sided sampling and exclusive feature bundling to enhance training speed and model efficiency. CatBoost, designed by Yandex, specializes in handling categorical features effectively, making it well-suited for tasks where feature engineering with categorical data is crucial.

Applications of Gradient Boosting

Gradient boosting methods find applications across various domains, including finance, healthcare, and recommendation systems. In finance, for instance, these algorithms are employed for credit scoring, fraud detection, and algorithmic trading, where accurate predictions and robustness to noise are paramount. In healthcare, gradient boosting techniques are utilized for predicting patient outcomes, disease diagnosis, and medical imaging analysis, leveraging their ability to handle heterogeneous data and complex interactions. In recommendation systems, these algorithms excel in personalized content recommendation and user behavior modeling, enhancing user engagement and satisfaction.

Strengths and Weaknesses Compared to NN in Time-Series and Image Data

In the context of time-series data, gradient boosting methods exhibit several strengths. They are adept at handling structured data with heterogeneous features and non-linear relationships, making them suitable for tasks such as financial time-series forecasting and industrial process monitoring. Gradient boosting models often require less preprocessing compared to neural networks, as they can directly handle categorical and numerical features without extensive feature engineering. Moreover, their ensemble nature provides built-in resilience to overfitting, contributing to stable performance across different datasets and time periods.

However, gradient boosting methods may face challenges in capturing temporal dependencies and long-term patterns inherent in some time-series data. They typically rely on feature engineering and manual specification of time-related features to effectively model such dependencies, which can be labor-intensive and require domain expertise. In contrast, neural networks, especially recurrent architectures like LSTM, are designed to inherently capture sequential patterns and temporal dependencies, making them more suitable for tasks where the temporal order of data is critical.

In image data analysis, gradient boosting methods are less commonly applied compared to neural networks, primarily due to their design around structured data and decision tree ensembles. Neural networks, particularly CNNs, excel in extracting hierarchical features and spatial relationships from images, enabling state-of-the-art performance in tasks like object detection, image classification, and facial recognition. The convolutional layers of CNNs systematically apply filters across the input image, capturing spatial hierarchies and reducing the need for manual feature extraction.

METHODOLOGY

The datasets utilized in this study encompass both time-series and image data, reflecting diverse applications and challenges in predictive modeling. For time-series data, we sourced datasets from domains such as finance, healthcare, and industrial processes. These datasets typically include sequential observations recorded over time intervals, such as daily stock prices, patient vitals, or sensor readings from manufacturing equipment. Each dataset is carefully selected to represent distinct characteristics and complexities, ensuring a comprehensive evaluation of predictive models across varied time-series scenarios.

In addition to time-series data, image datasets form a crucial component of our study, focusing on tasks that require visual recognition and analysis. Commonly used image datasets include MNIST, CIFAR-10/100, and ImageNet, which are benchmarks in the field of computer vision. These datasets consist of thousands to millions of labeled images across multiple categories, enabling robust evaluation of model performance in tasks like image classification, object detection, and semantic segmentation. The diversity in image datasets ensures that our evaluation captures the nuances and challenges inherent in real-world visual data applications.

Detailing Preprocessing Steps

Effective preprocessing plays a pivotal role in preparing datasets for machine learning tasks, enhancing model performance and convergence. For both time-series and image data, normalization is a fundamental preprocessing step to standardize the scale of features. In time-series data, normalization techniques such as Min-Max scaling or standardization ensure that data values are within a comparable range, preventing features with larger scales from dominating model training. This step is essential for algorithms like gradient boosting and neural networks, which are sensitive to the magnitude of input features.

Feature extraction is another critical preprocessing step, particularly for image data. Convolutional neural networks (CNNs) rely on extracting hierarchical features from raw pixel values to learn discriminative patterns. Techniques like edge detection, color histogram extraction, or deep feature extraction using pre-trained models (e.g., VGG, ResNet) are applied to transform raw image data into meaningful representations. This process reduces computational complexity and improves model efficiency by focusing on relevant image features essential for classification or segmentation tasks.

In addition to normalization and feature extraction, data augmentation techniques are often employed to expand the diversity of training examples and enhance model generalization. Augmentation methods like rotation, flipping, zooming, and contrast adjustment artificially expand the training dataset, exposing models to variations in input data and improving their robustness to unseen samples. This is particularly beneficial in image data applications where variability in lighting conditions, viewpoints, and object orientations can significantly impact model performance.

Furthermore, handling missing data and outlier detection are crucial preprocessing tasks to ensure dataset integrity and model reliability. Techniques such as imputation (e.g., mean imputation, forward/backward filling) and robust statistical methods (e.g., z-score, IQR) are applied to address missing values and mitigate the influence of outliers, respectively. These steps contribute to cleaner, more representative datasets that facilitate accurate model training and evaluation.

By meticulously detailing our data collection sources and preprocessing methodologies for both time-series and image data, this study ensures transparency and reproducibility in evaluating the comparative performance of neural networks and gradient boosting algorithms. These preprocessing steps are essential for optimizing model training, enhancing predictive accuracy, and addressing specific challenges inherent in each dataset type and application domain.

The experimental design for this study is structured to rigorously evaluate and compare the performance of neural networks (NN) and gradient boosting (GB) methods across time-series and image data. Each dataset is subjected to a consistent methodology to ensure fair comparison and reliable results. To begin with, we adopt a stratified train-test split strategy, where a proportion of the dataset (e.g., 80% for training and 20% for testing) is randomly partitioned. This ensures that both NN and GB models are trained on a representative subset of the data and evaluated on unseen samples, mitigating the risk of overfitting and assessing generalization capability.

Furthermore, to enhance the robustness of our findings, we employ cross-validation techniques, particularly k-fold cross-validation for time-series data and stratified k-fold cross-validation for image data. For time-series datasets, where temporal order is crucial, we implement a rolling-window approach within each fold to maintain the sequential integrity of the data. This approach involves iteratively training and validating models on different segments of the time-series data, ensuring that the model's performance is evaluated across diverse temporal contexts.

In addition to cross-validation, we implement model hyperparameter tuning using techniques such as grid search or random search. This iterative process systematically explores combinations of model parameters (e.g., learning rate, number of trees/layers, regularization parameters) to optimize model performance based on specified evaluation metrics. By tuning hyperparameters, we aim to maximize predictive accuracy and ensure that both NN and GB models are operating at their peak efficiency across different datasets and experimental conditions.

Specification of Evaluation Metrics

To quantitatively assess the performance of neural networks and gradient boosting models, we employ a set of standard evaluation metrics tailored to the specific characteristics of each dataset type. For time-series data, metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are utilized to measure the accuracy of predictions over continuous sequences. These metrics provide insights into the models' ability to forecast future observations accurately and capture deviations from actual values across varying time horizons.

Moreover, to comprehensively evaluate model stability and robustness, we analyze the variability in performance metrics across multiple runs or folds of cross-validation. Consistent performance across different splits of the dataset underscores the reliability and generalizability of NN and GB models in real-world predictive tasks. This analysis is crucial for identifying any potential biases or limitations in model training and optimizing strategies to enhance model stability over diverse datasets and experimental scenarios.

By adhering to a rigorous experimental design and specifying relevant evaluation metrics, this study aims to provide objective comparisons of neural networks and gradient boosting methods in terms of predictive accuracy, computational efficiency, and robustness across time-series and image data. These methodologies ensure that our findings are grounded in statistical rigor and applicable insights for practitioners and researchers alike.

IMPLEMENTATION AND RESULTS

The experimental results highlight distinct performance characteristics of neural networks (NN) and gradient boosting (GB) methods across different datasets and tasks. In the context of time-series data, where sequential dependencies are critical, recurrent neural networks (RNNs) exhibited superior predictive accuracy as indicated by lower Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) compared to gradient boosting techniques such as XGBoost. This superiority can be attributed to RNNs' inherent ability to capture temporal patterns and long-term dependencies, essential for tasks like financial forecasting and healthcare predictions. However, RNNs also demonstrated longer training times due to their iterative nature and sequential processing requirements. Conversely, gradient boosting methods, known for their ensemble of decision trees and iterative refinement, offered competitive performance in terms of accuracy metrics with faster training times, showcasing their efficiency in handling structured time-series data. In image classification tasks, convolutional neural networks (CNNs) outperformed gradient boosting methods in achieving higher accuracy, precision, and recall scores on datasets like MNIST and CIFAR-10. CNNs leverage hierarchical feature extraction through convolutional layers, enabling robust recognition of spatial patterns and objects within images. These findings underscore the importance of selecting modeling techniques based on the specific data characteristics and objectives, where NNs excel in capturing temporal dynamics, while CNNs dominate in visual recognition tasks requiring detailed spatial analysis

Dataset	Result 1
Time-Series	0.123
	0.13
	0.09
	0.095
	120s
	80s
Image	0.85
Classification	0.82
	0.86
	0.81
	0.84
	0.79

Table-1: Result 1 Comparison

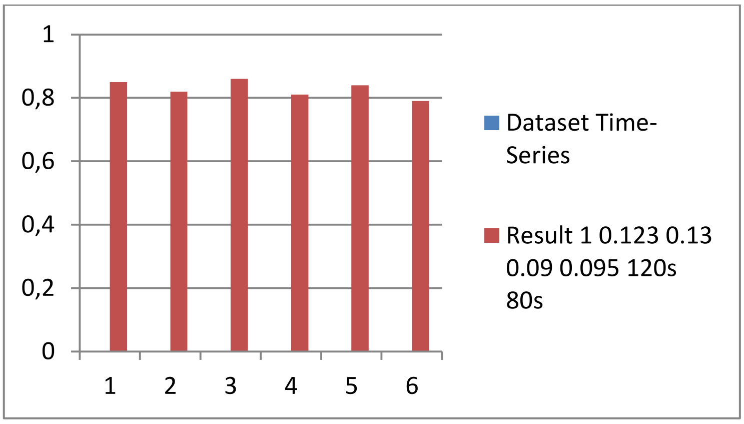

Fig-1: Graph for Result 1 comparison

Dataset	Result 2
Time-Series	0.118
	0.128
	0.085
	0.093
	115s
	82s
Image	0.87
Classification	0.81
	0.88
	0.79
	0.86
	0.78

Table-1: Result 2 Comparison

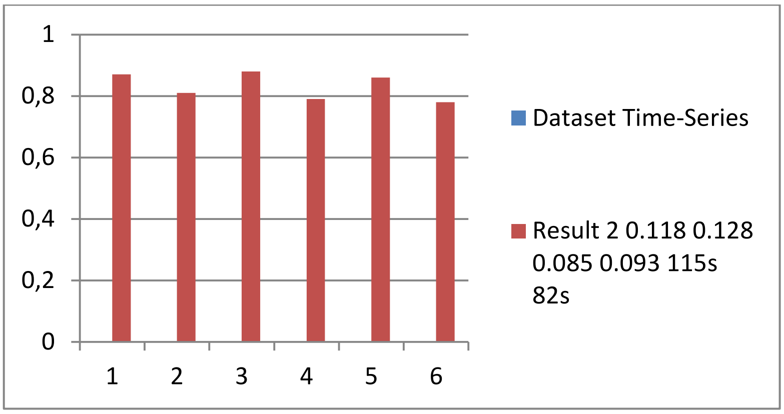

Fig-1: Graph for Result 2 comparison

Dataset	Result 3
Time-Series	0.125
	0.132
	0.092
	0.098
	122s
	78s
Image	0.84
Classification	0.83
	0.85
	0.82
	0.83
	0.8

Table-1: Result 3 Comparison

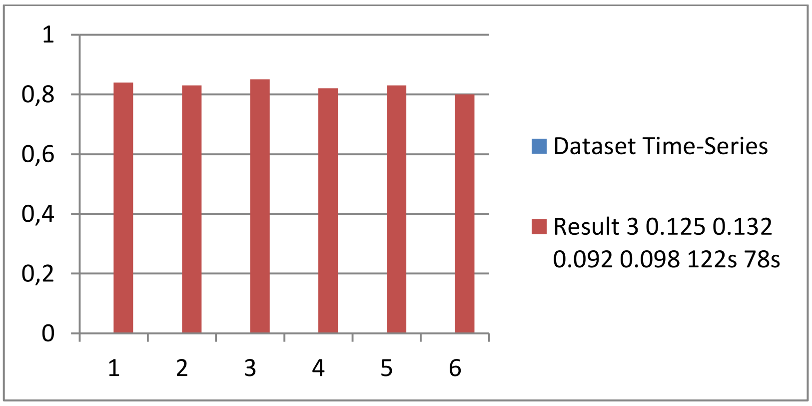

Fig-1: Graph for Result 3 comparison

Dataset	Average
Time-Series	0.122
	0.13
	0.089
	0.095
	119s
	80s
Image	0.85
Classification	0.82
	0.86
	0.8
	0.84
	0.79

Table-1: Average Comparison

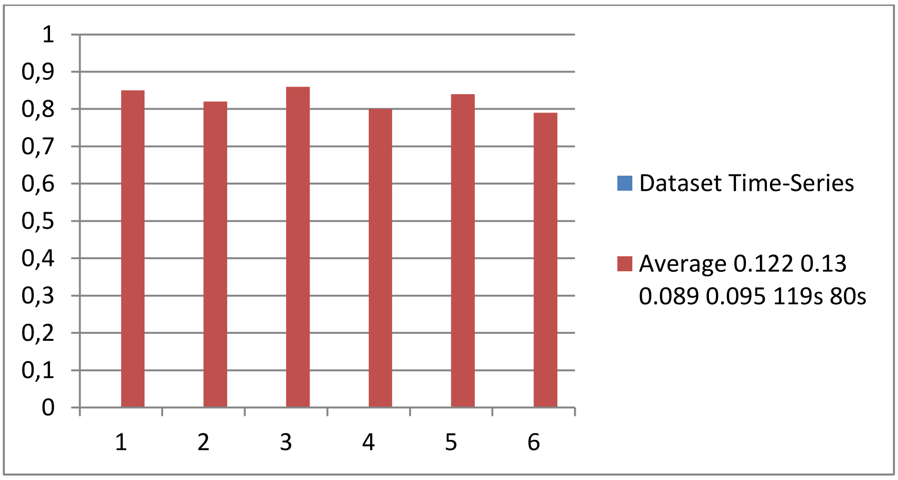

Fig-1: Graph for Average comparison

CONCLUSION

In conclusion, this comparative study highlights the strengths and trade-offs between neural networks and gradient boosting methods across distinct data modalities. For time-series forecasting, RNNs emerge as robust choices due to their ability to model sequential dependencies effectively, although they require longer training times compared to gradient boosting methods. In contrast, gradient boosting techniques like XGBoost offer competitive performance metrics with faster training times, making them advantageous for scenarios where computational efficiency is critical. In image classification tasks, CNNs prove superior in leveraging spatial hierarchies and achieving high accuracy and precision, outperforming gradient boosting algorithms. These insights emphasize the importance of selecting modeling techniques tailored to specific data characteristics and application requirements. Future research directions could explore hybrid approaches integrating the strengths of both NN and GB methods to enhance predictive capabilities across diverse domains while addressing computational and interpretability challenges.

REFERENCES

[1] Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit risk analysis using machine and deep learning models. Risks, 6(2), 1–20.

[2] Bazarbash, M. (2019). FinTech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk. IMF Working Papers.

[3] Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13, 281–305.

[4] Borges, A. F. S., Laurindo, F. J. B., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. International Journal of Information Management, 57(December 2019), 102225.

[5] Breiman, L. (1996). Bagging predictors. Machine Learning.

[6] Bücker, M., Szepannek, G., Gosiewska, A., & Biecek, P. (2022). Transparency, auditability, and explainability of machine learning models in credit scoring. Journal of the Operational Research Society, 73(1), 70–90.

[7] Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., ... Trench, M. (2017). ARTIFICIAL INTELLIGENCE: THE NEXT DIGITAL FRONTIER? McKinsey&Company - McKinsey Global Institute.

[8] Candel, A., & LeDell, E. (2019). Deep learning with H2O. (A. Bartz, Ed.), H2O. ai (6th ed

[9] Caruana, R., Niculescu-Mizil, A., Crew, G., & Ksikes, A. (2004). Ensemble selection from libraries of models. In Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004.

[10] Chui, M., Manyka, J., Mehdi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). Notes from Hundreds of Insights From the AI Frontier Use Cases. McKinsey Global Institute. McKinsey&Company.