ISSN: 2455 - 1597

ijergs
Open Access

International Journal of Engineering Research and Generic Science (IJERGS)
Available online at: https://www.ijergs.in

Volume - 11, Issue - 3, May – June - 2025, Page No. 61 - 65

Comparison of Neural Networks with Traditional Machine Learning Models - XGBoost, Random Forest

[1]Sumit Rajput, Research Scholar, Department of Computer Science & Engineering, Arya Institute of Engineering & Technology, Kukas, Jaipur, Rajasthan.

Abstract

The rapid advancement of machine learning has led to the development of various modeling techniques, each with its strengths and limitations. This study compares the performance of neural networks (NNs) with traditional machine learning models such as XGBoost and Random Forest across multiple datasets and tasks. While neural networks, particularly deep learning architectures, have gained significant attention for their ability to capture complex, non-linear relationships in large-scale data, traditional models like XGBoost and Random Forest remain highly effective for structured data and tabular datasets

The comparison focuses on key metrics such as accuracy, computational efficiency, interpretability, and scalability. Neural networks often excel in tasks involving unstructured data (e.g., images, text) and large datasets with high dimensionality, leveraging their ability to learn hierarchical features. However, they typically require substantial computational resources and extensive hyperparameter tuning. In contrast, XGBoost and Random Forest are more interpretable, computationally efficient, and often outperform neural networks on smaller, structured datasets.

This study also highlights the trade-offs between model complexity and performance, emphasizing the importance of selecting the right model based on the problem domain, data characteristics, and resource constraints. The findings suggest that while neural networks are powerful tools for specific applications, traditional machine learning models remain highly competitive and practical for many real-world scenarios. This comparison provides valuable insights for practitioners and researchers in choosing the most appropriate modeling approach for their specific use cases.

Keywords: XGBoost, Random Forest, Images, Text. Modelling.

Introduction

The field of machine learning has witnessed remarkable progress over the past few decades, with a proliferation of algorithms and techniques designed to tackle a wide range of predictive and analytical tasks. Among these, neural networks (NNs) particularly deep learning models, have emerged as a dominant paradigm, achieving state-of-the-art performance in domains such as computer vision, natural language processing, and speech recognition. However, traditional machine learning models like XGBoost and Random Forest continue to play a critical role, especially in structured data applications such as tabular data analysis, fraud detection, and recommendation systems.This study aims to provide a comprehensive comparison between neural networks and traditional machine learning models, focusing on their performance across various datasets and tasks. By evaluating key metrics such as accuracy, computational efficiency, interpretability, and scalability, we seek to identify the strengths and limitations of each approach. Additionally, we explore the trade-offs between model complexity and performance, offering practical insights for practitioners and researchers in selecting the most appropriate modeling technique for their specific use cases.The remainder of this paper is organized as follows: Section 2 provides an overview of neural networks, XGBoost, and

Random Forest, highlighting their underlying principles and architectures. Section 3 describes the experimental setup, including datasets, evaluation metrics, and implementation details. Section 4 presents the results and analysis, comparing the performance of the models across different tasks. Finally, Section 5 discusses the implications of the findings and offers recommendations for future research. Through this comparison, we aim to bridge the gap between modern and traditional machine learning approaches, fostering a deeper understanding of their respective roles in the evolving landscape of data science.

Neural Networks (NNs)

Neural networks are a class of machine learning models inspired by the structure and function of the human brain. They consist of interconnected layers of nodes (or neurons), which process input data through weighted connections and non-linear activation functions. Key characteristics of neural networks include:

Architecture: NNs are composed of an input layer, one or more hidden layers, and an output layer. Deep neural networks (DNNs) have multiple hidden layers, enabling them to learn hierarchical features from data.

Learning Process: NNs use gradient-based optimization techniques (e.g., backpropagation) to minimize a loss function, adjusting weights iteratively to improve performance.

Strengths

1. Excel at capturing complex, non-linear relationships in data.
2. Highly effective for unstructured data like images, text, and audio.
3. Scalable to large datasets and high-dimensional spaces.

Challenges

1. Require significant computational resources and training time.
2. Often need extensive hyper parameter tuning.
3. Lack interpretability, making them "black-box" models.

Traditional Machine Learning Models

Traditional machine learning models, such as XGBoost and Random Forest, are widely used for structured data and tabular datasets. These models are based on decision trees and ensemble learning techniques, which combine multiple weak learners to create a strong predictive model.

XGBoost (Extreme Gradient Boosting)

- A scalable and efficient implementation of gradient boosting.
- Builds trees sequentially, with each tree correcting errors made by the previous one.
- Known for its high accuracy, speed, and ability to handle missing data.
- Often outperforms other models on structured data competitions (e.g., Kaggle).

Random Forest

- An ensemble method that builds multiple decision trees and aggregates their predictions.
- Uses bagging (bootstrap aggregating) to reduce overfitting and improve generalization.
- Provides interpretability through feature importance scores.

- Robust to noise and outliers in the data.

Strengths

- Highly interpretable and easy to implement.
- Computationally efficient and require less tuning compared to NNs.
- Perform well on small to medium-sized structured datasets.

Limitations

- Less effective for unstructured data (e.g., images, text).
- May struggle with high-dimensional data compared to NNs.

Importance of Choosing the Right Model for Specific Tasks

Selecting the appropriate machine learning model for a given task is crucial for achieving optimal performance, efficiency, and interpretability. Different tasks, such as classification, regression, and prediction, have unique requirements and challenges, and the choice of model can significantly impact the outcomes. Below are key reasons why choosing the right model is essential:

Task-Specific Performance

Classification Tasks**:**

- ○ Models like Random Forest, XGBoost, and Support Vector Machines (SVMs) are often effective for binary or multi-class classification problems.
- ○ Neural networks, particularly Convolutional Neural Networks (CNNs), excel in image classification tasks

Regression Tasks:

- ○ Linear regression, decision trees, and ensemble methods like Gradient Boosting are commonly used for predicting continuous outcomes.
- ○ Neural networks can also be applied to regression tasks, especially when dealing with complex, non-linear relationships.

Data Characteristics

Structured vs. Unstructured Data:

- ○ Traditional models like XGBoost and Random Forest are well-suited for structured, tabular data.
- ○ Neural networks are more effective for unstructured data such as images, text, and audio.

Data Size and Dimensionality

- ○ For small to medium-sized datasets, traditional models often perform well and are computationally efficient.
- ○ Neural networks require large amounts of data to generalize effectively and are better suited for high-dimensional data.

Computational Efficiency

Resource Constraints:

○ Traditional models like Random Forest and XGBoost are generally faster to train and require less computational power compared to deep neural networks.

○ Neural networks, especially deep learning models, demand significant computational resources and time, making them less practical for resource- constrained environments.

Scalability

- Neural networks can scale to large datasets and complex problems but may require specialized hardware (e.g., GPUs).
- Traditional models are more scalable for smaller datasets and can be run on standard hardware.

Interpretability and Transparency

Model Interpretability

- Traditional models like decision trees and linear regression offer high interpretability, making it easier to understand and explain the model's decisions.
- Neural networks are often considered "black-box" models, with limited interpretability, which can be a drawback in applications requiring transparency (e.g., healthcare, finance).

Regulatory and Ethical Considerations

In regulated industries, interpretable models are often preferred to ensure compliance and ethical considerations.

Generalization and Overfitting

Overfitting Risks

- Neural networks, particularly deep learning models, are prone to overfitting, especially with small datasets. Techniques like dropout and regularization are required to mitigate this risk.
- Traditional models like Random Forest and XGBoost have built-in mechanisms (e.g., bagging, boosting) to reduce overfitting and improve generalization.
- Domain-Specific Requirements

Industry-Specific Needs

- In domains like healthcare, interpretability and accuracy are paramount, making traditional models or simpler neural networks more suitable.
- In domains like computer vision or natural language processing, the complexity of data often necessitates the use of deep learning models.

Challenges and Limitations

Both neural networks and traditional models have inherent challenges and limitations, which must be considered when designing hybrid approaches or selecting a model for a specific task.

Neural Networks

High Computational Cost

- Training deep neural networks requires significant computational resources (e.g., GPUs, TPUs) and time.
- Inference can also be slow, especially for large models.

Difficulty in Interpreting Results

- NNs are often considered "black-box" models, making it challenging to explain their predictions.
- Techniques like SHAP and LIME help but are not as intuitive as traditional model explanations.

Requires Expertise in Hyper parameter Tuning: NNs have many hyper parameters (e.g., learning rate, number of layers, activation functions), and tuning them requires expertise and experimentation.

Data Requirements: NNs typically require large amounts of labeled data to generalize effectively, which can be costly and time-consuming to obtain.

Conclusion

Neural Networks:

- Excel in handling unstructured data (e.g., images, text, audio) and modeling complex, non-linear relationships.
- Require significant computational resources, large labeled datasets, and expertise in hyperparameter tuning.
- Often considered black-box models, though techniques like SHAP and LIME are improving interpretability.

Traditional Models

- Highly effective for structured/tabular data, offering interpretability, efficiency, and robustness.
- Require manual feature engineering and may struggle with highly complex tasks or unstructured data.
- Examples like XGBoost and Random Forest dominate in structured data applications (e.g., Kaggle competitions).

References

1. Pillai, A. S. (2022). A natural language processing approach to grouping students by shared interests. Journal of Empirical Social Science Studies, 6(1), 1-16.
2. Pillai, A. (2023). Traffic Surveillance Systems through Advanced Detection, Tracking, and Classification Technique. International Journal of Sustainable Infrastructure for Cities and Societies, 8(9), 11-23.
3. Pillai, A. S. (2023). Advancements in natural language processing for automotive virtual assistants enhancing user experience and safety. Journal of Computational Intelligence and Robotics, 3(1), 27-36.
4. Pillai, A. S. (2022). Cardiac disease prediction with tabular neural network.
5. Nanda, A. S. (2024). AI in Treasury Management: EnhancingBank'sTreasury System for Budget Execution in the MediumTerm. International Journal of Science and Research (IJSR), 13(3), 1906–1912.