Dean&Francis

ISSN 2959-6157

# Research on Comparisons with three survival models Comparisons with Decision Tree, Random forests and CNNs

**Xinyang Liu***

Xi'an Jiao tong-Liverpool University

* Corresponding author.

Email:3445025593@qq.com

**Abstract:**

Decision tree, random forests and CNN are three advanced algorithms that can solve classification tasks. This article will compare the performance of the three different algorithms in classification tasks. Here, CNN extracts features through convolution operations to detect objects or recognize images, other two achieve this through building a tree structure. They could help to solve so many problems. Decision tree have been widely applied in fields such as radar signal classification, character recognition, remote sensing, medical diagnosis, expert systems, and speech recognition. CNN can be used to detect individual plants or pixels as well as for vegetation classification. Nevertheless, they still encounter challenges in the classification tasks, such as low classification accuracy.

**Keywords:** Decision tree, random forests, CNN, classification algorithm

## 1. Introduction

The main purpose of classification algorithms is to assign data samples to predefined categories based on the input features. The FCM algorithm was first proposed by Dunn in 1974 and later popularized by Bezdek. [7]Bezdek also conducted relevant research on the algorithm in 1973. Based on the theory of fuzzy sets, this algorithm allows a data point to belong to multiple clusters with different membership degrees, and is widely applied in various fields such as image segmentation, pattern recognition, data analysis, market segmentation, and bioinformatics. [8]However, the simple structure and lack of nonlinear transformation in the early models may result in the early algorithms having insufficient fitting ability for nonlinear, high-dimensional or complex and rich data. They are unable to handle simple nonlinear problems or have weak ability to capture the interaction relationships among multiple features.[9] The subsequent development of deep learning further incorporates designs such as deep structures and local connections, introducing multi-layer structures and nonlinear activations, to enhance the fitting of complex data.

Efficient algorithms of different structures and complexities can solve various classification tasks, but they also have different problems that require targeted improvements. This article will better understand the applicability and improvement potential of algorithms through comparisons among different algorithms. [3]

ISSN 2959-6157

## 2.1 . Definition

A decision tree is a tree-shaped structure that is easy to understand where different leaves represent different types, and the branches leads to different leaves. And the process of a leaf moving to another leaf can be regarded as a test.[10]

Random forest is a collection of decision trees which can obtain more complex results through multiple decision trees. [4]It's like asking a diverse group of experts for their opinion and going with the most popular answer.[5]

A Convolutional Neural Network (CNN) is a class of deep neural networks, most commonly applied to analysing visual imagery. [1]Their architecture is inspired by the organization of the animal visual cortex and is designed to automatically and adaptively learn spatial hierarchies of features from input data. [2]CNNs use special layers called convolutional layers that apply filters to the input data to create feature maps. [3]

## 2.2 . When to use

Decision tree can be applied when interpretability is key, simple data and initial data exploration: When you need to explain the decision-making process to nontechnical stakeholders, the transparency of a decision tree is valuable; For smaller, simpler datasets with clear decision boundaries, a decision tree can be a quick and effective solution; They can be useful for initial data analysis to understand feature importance.

3.Conclusion

As shown in table 1, decision tree is a tree-like structure that is easy to understand. It requires relatively little data but can handle various types of data, including numerical and categorical data. However, it also has some drawbacks, such as poor generalization of data, overfitting phenomenon, and instability caused by minor changes.

As shown in table 1, random forests can significantly reduces overfitting compared to a single decision tree, it has generally high accuracy and could deal with missing values and maintains accuracy even with missing data. However, it has disadvantages: It's more complicated than a decision tree; Training can be computationally expensive and time-consuming for lots of decision trees.

As shown in table 1, CNN has exceptional performance on unstructured data, especially images. It automatically learns features from the data, reducing more human cost. And it can be highly accurate and achieve state-of-the-art results on many benchmarks. But it also has disadvantages: It requires a large amount of data for training; It computationally intensive and often requires specialized hardware like GPUs; The "black box" nature makes them difficult to interpret.

**Table 1. At a Glance: A Comparative Overview**

| Feature | Decision Tree | Random Forest | Convolutional Neural Network (CNN) |
|---|---|---|---|
| Underlying Structure | A single tree-like model of decisions. | An ensemble of multiple decision trees. | A deep learning architecture with layers of neurons. |
| Interpretability | High :The decision-making process is transparent and easy to visualize. | Moderate: The collective decision of many trees can be complex to interpret directly. | A deep learning architecture with layers of neurons. |
| Performance | Moderate: Prone to overfitting on the training data. | High: Generally provides higher accuracy and better generalization than a single decision tree. | Very High: State-of-the-art performance, especially on unstructured data like images. |
| Data Suitability | Structured (tabular) data | Structured (tabular) data | Unstructured data (images, audio, text) and time-series data |
| Computational Cost | Low: Relatively fast to train. | Moderate: Training time increases with the number of trees. | High: Requires significant computational resources, often GPUs, for training. |
| Feature Engineering | Requires some feature engineering. | Less sensitive to feature scaling. | Automatically learns features from raw data. |

## REFERENCES

[1] 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. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, 8, 53. https://doi.org/10.1186/s40537-021-00444-6

[2] Chauhan, R., Ghanshala, K. K., & Joshi, R. C. (2018). Convolutional neural network (CNN) for image detection and recognition. In 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC) (pp. 278–282). IEEE. https://doi.org/10.1109/ICSCCC.2018.8703283

[3] Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on convolutional neural networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010

[4] Belgiu, M., & Drăgut, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

[5] Paul, A., Mukherjee, D. P., Das, P., Gangopadhyay, A., Chintha, A. R., & Kundu, S. (2018). Improved random forest for classification. IEEE Transactions on Image Processing, 27(8), 4012–4023. https://doi.org/10.1109/TIP.2018.2835141

[6] Biau, G., & Scornet, E. (2015). A random forest guided tour. arXiv preprint arXiv:1511.05741. https://arxiv.org/abs/1511.05741

[7] Karita, S., Chen, N., Hayashi, T., Hori, T., Inaguma, H., Jiang, Z., Someki, M., Soplin, N. E. Y., Yamamoto, R., Wang, X., Watanabe, S., Yoshimura, T., & Zhang, W. (2019). A comparative study on transformer vs RNN in speech applications. arXiv preprint arXiv:1909.06317. https://doi.org/10.48550/arXiv.1909.06317

[8] Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306. https://doi.org/10.1016/j.physd.2019.132306

[9] Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv:1702.01923. https://doi.org/10.48550/arXiv.1702.01923

[10] Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. Journal of Chemometrics, 18(5), 275–285. https://doi.org/10.1002/cem.873

[11] Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 21(3), 660–674. https://doi.org/10.1109/21.97458