

Deep Fuzzy Systems in regression applications: a view on interpretability

Jorge Sampaio Silveira Júnior

Advisors: Jérôme Mendes, Cristiano Premebida
DEEC-UC, ISR-UC

Introduction

- Regression models have become predominant in increasingly complex real-world systems due to the large availability of data, inclusion of nonlinear parameters, and other aspects depending on application area.

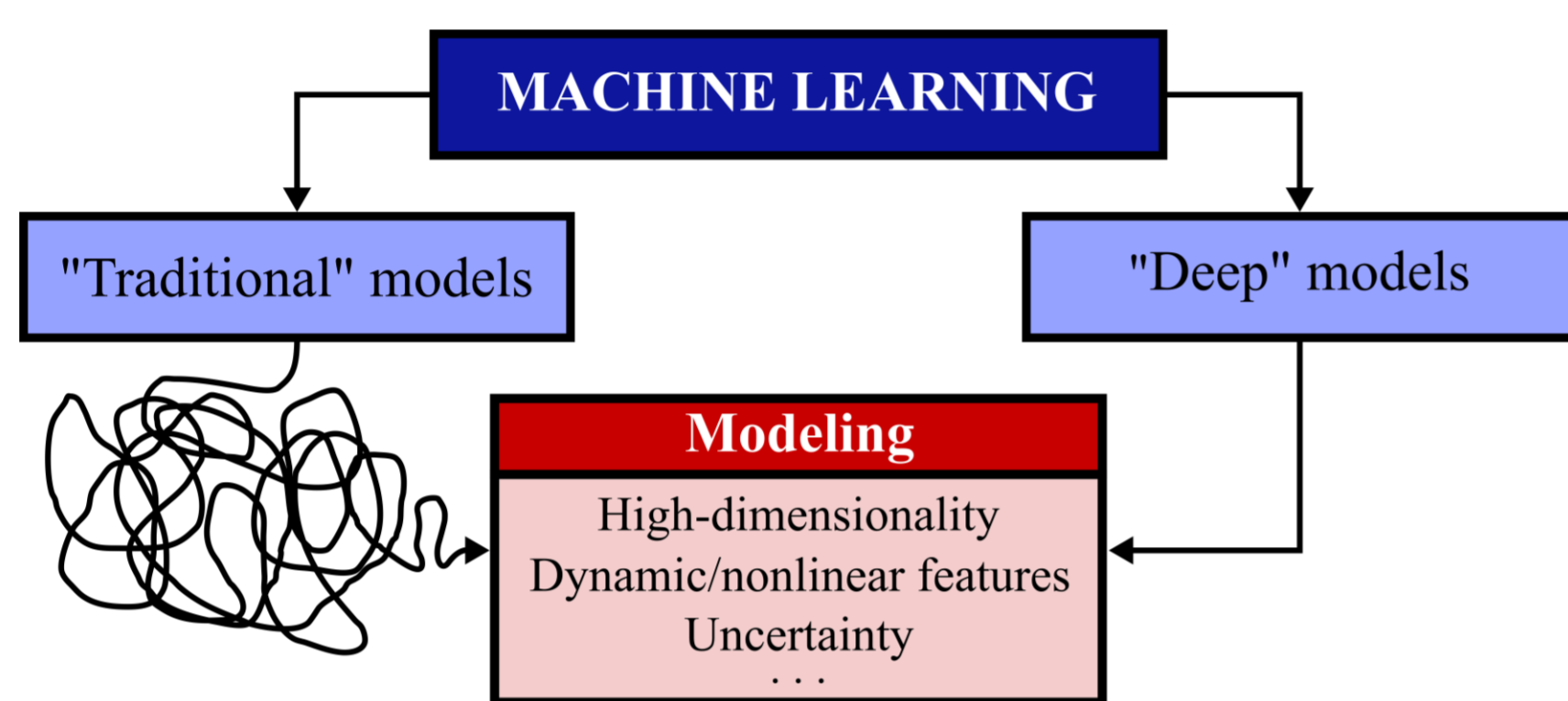


Fig.1 – Traditional vs. Deep learning techniques.

- Deep Learning (DL) disadvantages: need for a sufficiently large dataset for model training, sensitivity to hyperparameter selection, and lack of interpretability.

eXplainable Artificial Intelligence

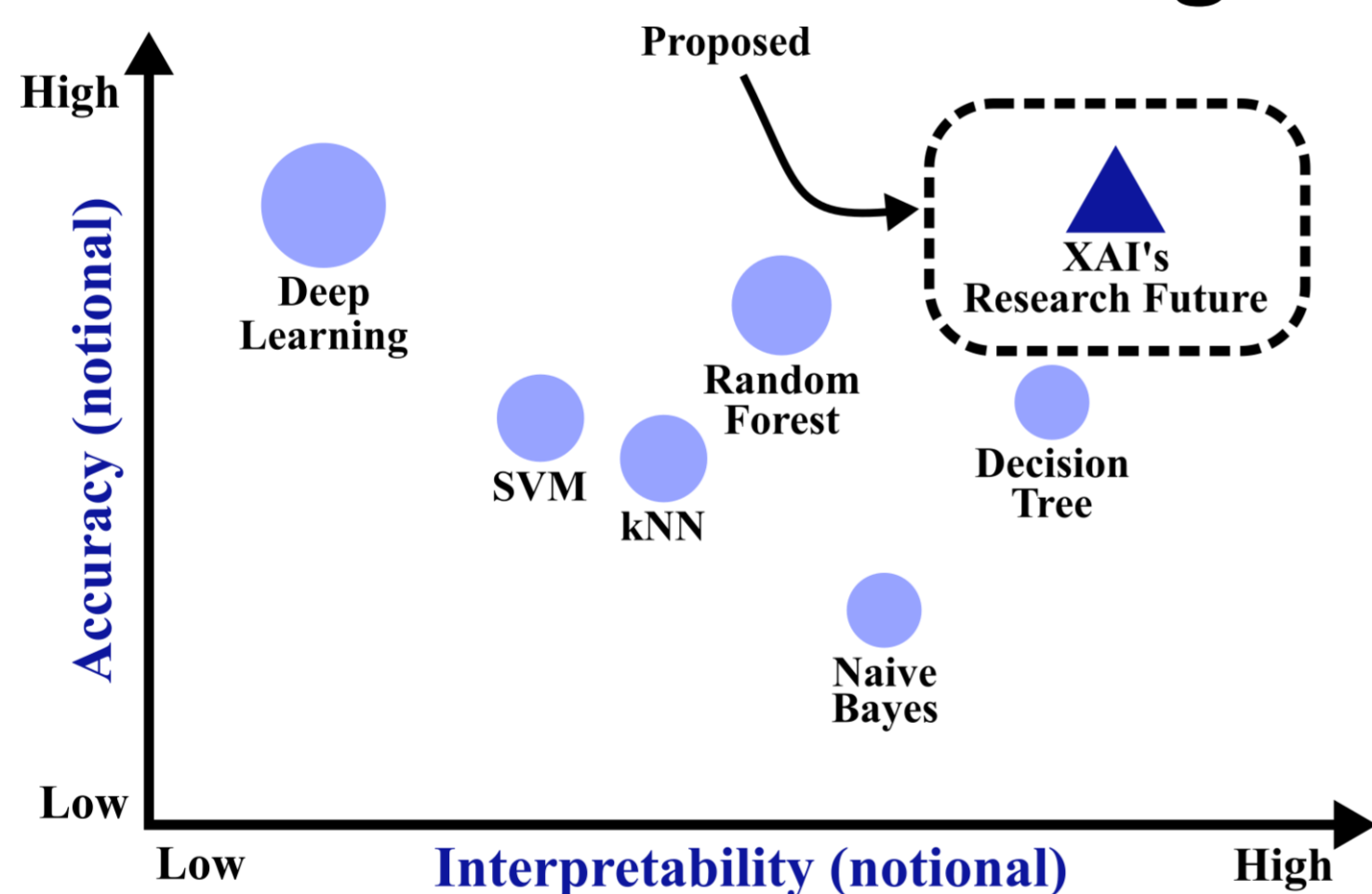


Fig.2 – Accuracy vs. interpretability for different machine learning models [1].

- Importance of developing inherently interpretable methodologies [2].
- Addressing the interpretability issue of DL techniques with fuzzy logic systems (FLS).

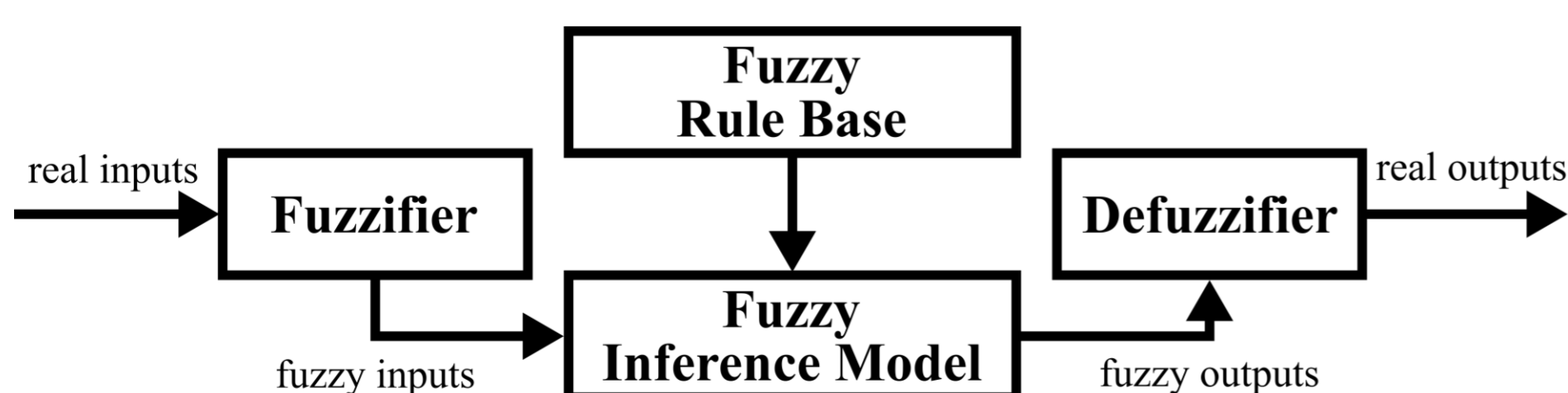


Fig.3 – Basic configuration of fuzzy logic systems.

Deep Fuzzy Systems (DFSs)

- Models built on top of DL structures with FLS.
- DFS can be summarized into two categories: **Standard DFS** and **Hybrid DFS**.

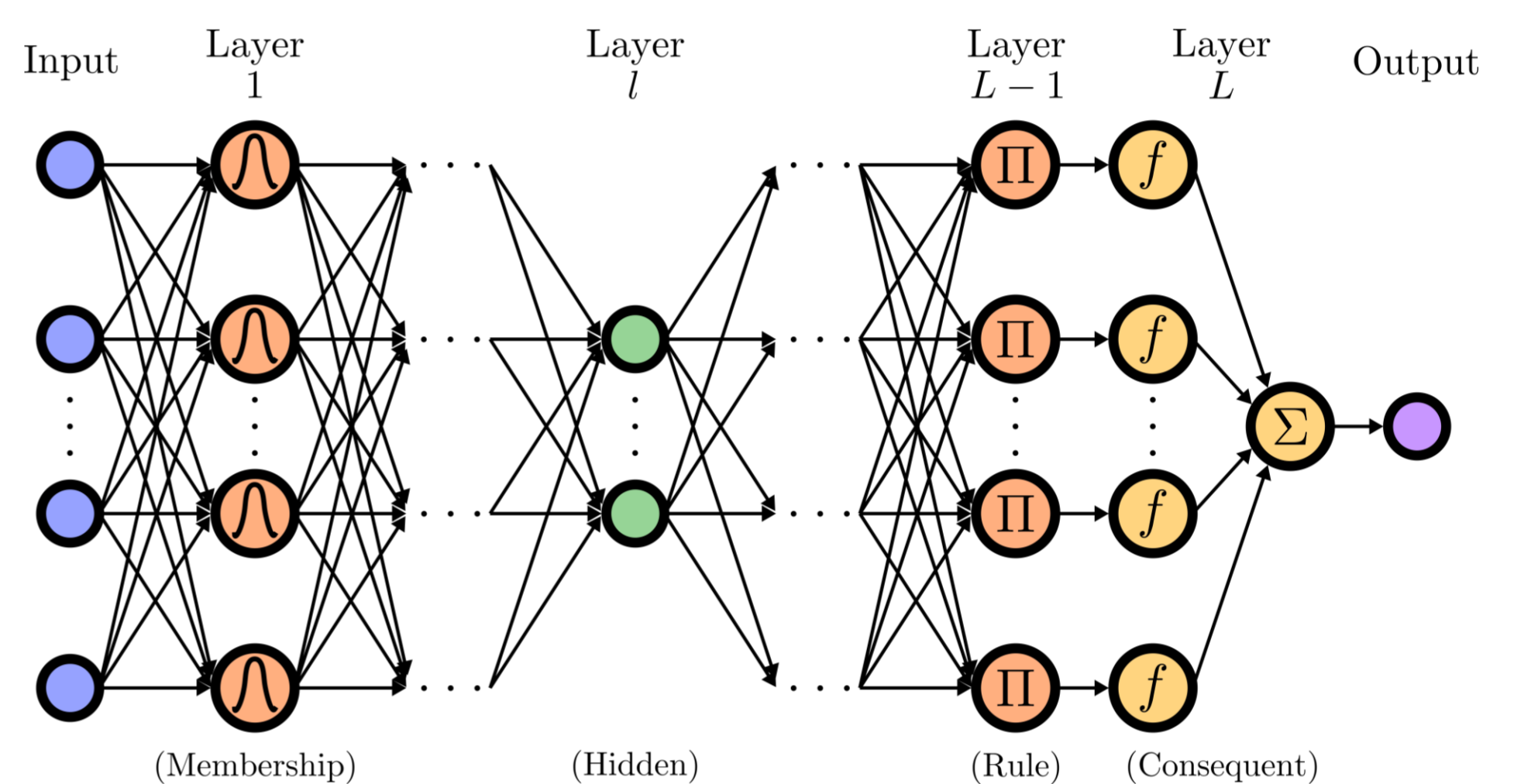


Fig.4 – Example of Standard DFS framework.

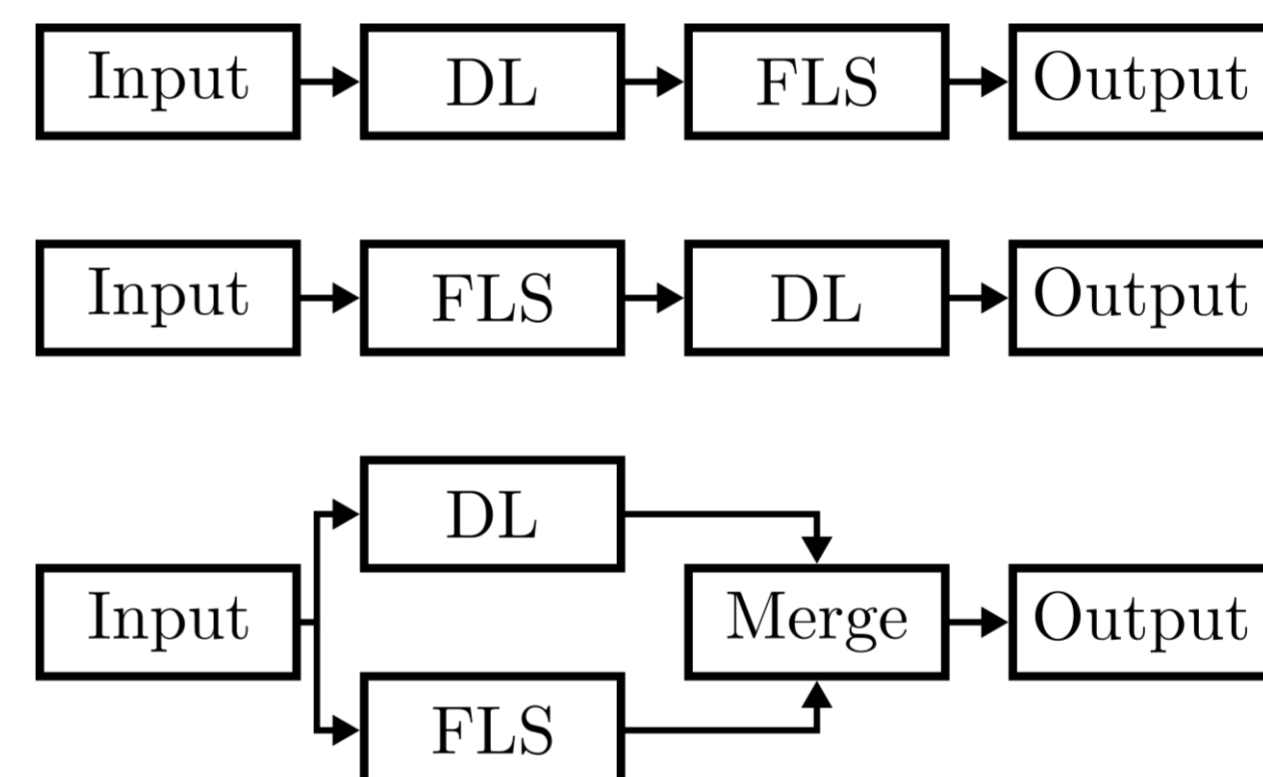


Fig.5 – Example of Hybrid DFS framework.

- Standard DFSs can promote more interpretability of their models when compared to Hybrid DFSs.
- Potential for developing trustworthy models for sensitive areas (e.g., cyber-security, medical, financial surveillance, and industrial processes).

References

[1] P. Angelov, E. Soares, R. Jiang, N. Arnold and P. Atkinson, "Explainable artificial intelligence: an analytical review", WIREs Data Mining and Knowledge Discovery, vol. 11, no. 5, 2021. doi: 10.1002/widm.1424.

[2] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead", Nature Machine Intelligence, vol. 1, no. 5, pp. 206-215, 2019. doi: 10.1038/s42256-019-0048-x.

Acknowledgement: This research was supported by Fundação para a Ciência e a Tecnologia (FCT) under the grant ref. 2021.04917.BD, and under the project UIDB/00048/2020.