An Investigation of NN-FCM and NF-RVFL Networks for Mining Patterns in Data

Author: Iftesam Nabi
Student ID: 103-84-915
Supervisor: Dr. Pradeep Chowriappa
Course: CSC 550C
Quarter: Summer 2025
Date: August 07, 2025

Louisiana Tech University
Department of Computer Science

Executive Summary

Objective: Benchmarking Neural-Network–driven Fuzzy Cognitive Map (NN-FCM) against the Neuro-Fuzzy Random Vector Functional Link (NF-RVFL) on UCI datasets, comparing accuracy, interpretability, and efficiency, with a depth ablation (1–6 hidden layers) to quantify the effect of architecture depth on performance and computation.

Approach:

- Applied feature scaling and label factorization.
- Hyperparameters selected using Optuna with 5-fold cross-validation.
- Trained 5-member seed ensembles with ReduceLROnPlateau scheduling and early stopping.
- Results are summarized per dataset and depth variant.

Key NN-FCM Findings:

- **Credit Approval:** R^2 = 82.99% ± 4.03 (best 87.96%, nn-fcm-4).
- **Breast Cancer:** R^2 = 95.07% ± 2.40 (best 96.52%, nn-fcm-4).
- **Abalone:** R^2 = 66.86% ± 5.74 (best 74.91%, nn-fcm-5).
- **CRX:** R^2 = 45.28% ± 1.42 (best 47.14%, nn-fcm-2).
- Depth–accuracy relationship is non-monotonic: shallow depth works better on CRX, deeper capacity benefits Abalone.

Runtime and Practical Trade-Off:

- NN-FCM requires substantially longer training/inference time, especially on Abalone.
- NF-RVFL is comparatively fast due to randomized hidden mappings and a closed-form/ridge readout.
- NN-FCM's higher accuracy on nonlinear problems can be advantageous: once trained, its learned Gaussian memberships and linear readout define an explicit predictive equation that can be applied directly to similar datasets for faster deployment without retraining.

Direct comparison to NF-RVFL (R^2).

Dataset	NF-RVFL-R	NF-RVFL-K	NF-RVFL-C	NF-RVFL Best	NN-FCM Best (variant)	Δ Abs (pts)	Δ Rel
Abalone	63.8253	65.4586	64.0887	65.4586	74.91 (nn-fcm-5)	+9.4514	+14.44%
Breast Cancer	72.3351	72.6921	70.8953	72.6921	96.52 (nn-fcm-4)	+23.8279	+32.78%
Credit Approval	85.9420	85.9420	87.8581	87.8581	87.96 (nn-fcm-4)	+0.1019	+0.12%
CRX	55.2687	55.1957	55.2092	55.2687	47.14 (nn-fcm-2)	-8.1287	-14.71%

Table of Contents

1 Introduction

Operational decision systems require a balance among predictive performance, interpretability, and computational efficiency. Fuzzy Cognitive Maps (FCMs) encode causal relationships among interpretable concepts. Neural augmentations (NN-FCM) allow memberships and rule activations to be learned directly from data, providing modern capacity while retaining interpretability. Random Vector Functional Link (RVFL) networks transform inputs through randomized hidden features with a closed-form readout; NF-RVFL augments this with fuzzy logic. In this study, NN-FCM is benchmarked against NF-RVFL on three binary classification tasks (Credit Approval, Breast Cancer, CRX) and one regression task (Abalone), and the effect of depth is analyzed.

1.1 Background and Motivation

FCMs provide interpretability through graph structures and rule semantics. Fixed memberships may underfit complex data; NN-FCM addresses this by learning Gaussian centers and scales per rule and feature, followed by residual MLP stacks. NF-RVFL emphasizes efficiency via randomized expansions and ridge-regularized readouts. The goal is to identify conditions under which NN-FCM's additional capacity is justified relative to NF-RVFL.

1.2 Research Questions and Objectives

- How does NN-FCM compare to NF-RVFL in accuracy (R^2), interpretability, and efficiency?
- How does NN-FCM depth (1–6 hidden layers) affect accuracy and compute across datasets of varying complexity?
- What evaluation protocol yields reliable, reproducible comparisons?

1.3 Contributions

- A reproducible NN-FCM pipeline with Optuna search, early stopping, and seed ensembling.
- A depth ablation demonstrating non-monotonic accuracy gains vs. runtime.
- A direct R^2 comparison against NF-RVFL (R/K/C).
- Practical guidance mapping dataset traits to the appropriate NN-FCM depth.

1.4 Organization of the Report

Section 2 reviews related work. Section 3 describes datasets. Section 4 details methods. Section 5 presents experimental setup. Section 6 reports results, including the NF-RVFL comparison. Section 7 previews interpretability. Section 8 discusses implications and threats to validity. Sections 9–10 provide recommendations and conclusions. Section 11 documents the environment. Section 12 lists references. Appendices include logs and equations.

2 Related Work

FCMs model causal interactions among concepts, with learning strategies ranging from Hebbian updates to gradient-based optimization. NN-FCM variants introduce neural modules (e.g., MLPs, attention) atop memberships to capture higher-order interactions. RVFL employs randomized hidden nodes with linear readout for fast learning; NF-RVFL incorporates fuzzy basis functions or kernelized mappings. Interpretability methods include rule saliency, feature attribution (permutation/SHAP), and sparsity-inducing penalties.

3 Datasets

3.1 Abalone

Regression predicting Rings (age proxy) from shell measurements. Nonlinear relationships and noise justify deeper capacity and careful regularization. Runtime sensitivity makes this dataset informative for evaluating the accuracy–efficiency trade-off between NN-FCM and NF-RVFL.

3.2 Breast Cancer Wisconsin

Binary classification (benign vs. malignant) from morphological cell measurements. The dataset is widely used to assess model stability, calibration, and classification metrics. Strong signal and relatively low dimensionality make it suitable for examining depth vs. overfitting behavior.

3.3 Credit Approval

Binary approval decision with mixed numeric and categorical features. The dataset is used to evaluate classification performance under feature scaling and label factorization. Results are reported with leakage-free protocols in the final version (Accuracy, F1, AUROC in addition to R^2).

3.4 CRX

A commonly used variant of Credit Approval with different preprocessing/splits (e.g., missing-value handling, categorical encoding, train/test partition). It is treated as a separate dataset to enable direct comparison with NF-RVFL results reported under this CRX protocol.

4 Methods

4.1 NN-FCM Architecture

Gaussian memberships are learned per feature and rule, and membership activations are averaged across features to produce rule activations $\mu(x)$. These activations feed residual MLP blocks with BatchNorm, ReLU, and Dropout. Three parallel linear heads predict scalar outputs which are averaged at inference. Optimization is performed with AdamW, ReduceLROnPlateau, and early stopping.

4.2 Depth Ablation Design (nn-fcm-1…6)

- nn-fcm-1: 1 hidden layer (baseline).
- nn-fcm-2: 2 hidden layers (residual).
- nn-fcm-3: 3 hidden layers (residual stack).
- nn-fcm-4: 3 hidden layers (residual stack).
- nn-fcm-5: 5 hidden layers (residual stack).
- nn-fcm-6: 6 hidden layers (residual stack).

4.3 NF-RVFL Baseline

NF-RVFL augments RVFL with fuzzy logic while keeping randomized hidden mappings and ridge-regularized readouts. The R/K/C variants reported in the paper are used for comparison.

5 Experimental Setup

5.1 Environment and Tooling

Google Colab, Python 3.10+, NumPy/SciPy, scikit-learn, PyTorch, Optuna, and Matplotlib were used. GPU (T4/A100) resources were used when available.

5.2 Preprocessing and Data Management

Continuous features were standardized and categorical labels factorized when needed. The final protocol will wrap preprocessing and the estimator inside an sklearn Pipeline so that cross-validation and testing are leakage-free.

5.3 Hyperparameter Search and Early Stopping

Optuna with 5-fold cross-validation optimized mean R^2 over n_rules, learning rate, mlp_units, dropout, and L2. Training employed AdamW with ReduceLROnPlateau ($\gamma=0.5$, patience=20) and early stopping when validation loss plateaued.

5.4 Training Regime and Ensembling

Final runs used epochs=2000, batch_size=70, and patience=400. An ensemble of M=5 seed variants was trained, and predictions were averaged.

5.5 Evaluation Protocol and Validity Considerations

- Current logs compute R^2 on the full dataset after global scaling, creating an optimistic bias due to leakage.
- Leakage-free 5-fold CV (mean±SD, 95% CIs) or a clean hold-out test with tuning only on training folds will be reported.
- For classification tasks, a logit head will be used; Accuracy, F1, and AUROC will be reported alongside R^2.

6 Results

6.1 Summary by Dataset and Variant

Dataset	Runs	R^2 mean %	R^2 SD %	Best R^2 %	Worst R^2 %	Train time mean (s)	Train time SD (s)	ES epoch mean	ES epoch SD
abalone	6	66.86	5.74	74.91	60.80	446.58	81.52	473.33	68.88
breast_cancer_wisconsin	6	95.07	2.40	96.52	90.24	71.03	20.15	451.83	64.18
credit_approval	6	82.99	4.03	87.96	77.13	76.05	12.41	465.17	29.40
crx	6	45.28	1.42	47.14	43.37	64.62	14.74	448.83	14.99

6.1.1 R^2 (%) by Dataset × Variant

dataset	nn-fcm-1	nn-fcm-2	nn-fcm-3	nn-fcm-4	nn-fcm-5	nn-fcm-6
abalone	60.80	61.81	63.54	68.19	74.91	71.88
breast_cancer_wisconsin	90.24	95.72	96.24	96.52	95.53	96.19
credit_approval	77.13	80.40	81.52	87.96	86.18	84.73
crx	44.12	47.14	46.27	43.37	45.96	44.85

6.1.2 Train Time (s) by Dataset × Variant

dataset	nn-fcm-1	nn-fcm-2	nn-fcm-3	nn-fcm-4	nn-fcm-5	nn-fcm-6
abalone	600.2	419.0	363.5	410.2	425.7	460.9
breast_cancer_wisconsin	50.2	53.9	68.7	62.5	99.4	91.5
credit_approval	64.6	57.4	87.6	77.4	86.7	82.6
crx	50.4	49.2	62.2	63.8	74.0	88.1

6.1.3 Early-stop Epoch by Dataset × Variant

dataset	nn-fcm-1	nn-fcm-2	nn-fcm-3	nn-fcm-4	nn-fcm-5	nn-fcm-6
abalone	609	479	442	450	429	431
breast_cancer_wisconsin	582	421	416	430	425	437
credit_approval	486	495	444	494	433	439
crx	469	444	425	444	458	453

6.1.4 Best Variant per Dataset

dataset	best_variant	best_R²_%
abalone	nn-fcm-5	74.91
breast_cancer_wisconsin	nn-fcm-4	96.52
credit_approval	nn-fcm-4	87.96
crx	nn-fcm-2	47.14

6.1.5 Per-variant Mean R^2 Across Datasets

variant	R^2_mean_%
nn-fcm-5	75.65
nn-fcm-6	74.41
nn-fcm-4	74.01
nn-fcm-3	71.89
nn-fcm-2	71.27
nn-fcm-1	68.07

6.2 Depth vs. Accuracy and Runtime

The relationship between depth and accuracy is non-monotonic. CRX and Credit Approval tend to peak at 2–4 layers; Breast Cancer peaks at ~4 layers; Abalone benefits from deeper capacity (~5 layers). Runtime generally increases with depth, especially on Abalone. This cost must be weighed against the accuracy benefits when nonlinearity is significant.

6.3 Per-Dataset Analysis and Observations

6.3.1 Credit Approval (Credit_approval.csv)

The best variant is nn-fcm-4 (87.96% R^2). Variants 1–3 form a smooth ramp, indicating benefits from moderate depth; beyond 4 layers, returns diminish. Classification metrics will be added in the final study.

6.3.2 CRX Split (crx.csv)

This split yields lower R^2 overall, with the best at nn-fcm-2 (47.14%). Differences likely reflect preprocessing and the smaller/easier split used in the NF-RVFL paper. Nomenclature and protocol will be unified.

6.3.3 Breast Cancer Wisconsin

R^2 remains high across depths (best 96.52% at 4 layers). Classification metrics and calibration curves will complement R^2.

6.3.4 Abalone

Deeper models help (nn-fcm-5 = 74.91% R^2). To manage runtime, reductions in n_rules, stronger regularization, and dynamic plateau detection are proposed.

6.4 Head-to-Head with NF-RVFL (R^2)

Best NN-FCM variants are compared against NF-RVFL-R/K/C from the paper. NN-FCM surpasses NF-RVFL on Abalone and Breast Cancer, essentially ties on Credit Approval, and trails on the CRX split.

Dataset	NF-RVFL-R	NF-RVFL-K	NF-RVFL-C	NF-RVFL Best	NN-FCM Best (variant)	Δ Abs (pts)	Δ Rel
Abalone	63.8253	65.4586	64.0887	65.4586	74.91 (nn-fcm-5)	+9.4514	+14.44%
Breast Cancer	72.3351	72.6921	70.8953	72.6921	96.52 (nn-fcm-4)	+23.8279	+32.78%
Credit Approval	85.9420	85.9420	87.8581	87.8581	87.96 (nn-fcm-4)	+0.1019	+0.12%
CRX	55.2687	55.1957	55.2092	55.2687	47.14 (nn-fcm-2)	-8.1287	-14.71%

Efficiency note: NF-RVFL's training and inference are generally much faster due to closed-form readouts. NN-FCM can be computationally intensive; however, once trained, the learned memberships and linear heads define an explicit predictive equation that can be ported to similar datasets, enabling faster downstream scoring without retraining.

7 Interpretability Preview

NN-FCM exposes interpretable components through memberships and rule activations. Rule saliency can be computed by aggregating membership strengths per feature and rule; permutation or SHAP analyses at membership and MLP layers provide local/global attributions. For NF-RVFL, interpretability includes examination of readout coefficients and permutation analyses through randomized fuzzy features.

8 Discussion

8.1 When Does Depth Help?

Shallow to moderate depth suffices for modest nonlinearity and limited data (e.g., CRX). Moderate depth (~4 layers) balances accuracy and stability on Breast Cancer and Credit Approval. Deeper stacks (~5–6) benefit complex regression (Abalone) but incur higher runtime.

8.2 Error Sources and Threats to Validity

- Data leakage from global scaling before evaluation (inflates R^2).
- Metric mismatch for classification if only R^2 is reported; Accuracy/F1/AUROC and calibration are needed.

- Variance due to random seeds, mitigated by ensembling; bootstrap can provide confidence intervals.
- Protocol mismatch vs. NF-RVFL (dataset variants/splits); identical settings will be reproduced or differences clearly annotated.

8.3 Practical Implications

- Time-constrained deployments may prefer nn-fcm-2/3 for strong results at lower cost, or NF-RVFL for maximum speed when accuracy trade-offs are acceptable.
- Nonlinear regression tasks may justify 4–5 layers; runtime can be controlled by reducing n_rules and increasing regularization.
- For reuse across similar datasets, NN-FCM's learned memberships and readout can be frozen to yield a direct predictive equation, enabling rapid scoring without re-optimization.

9 Recommendations and Next Steps

- Adopt an sklearn Pipeline for leakage-free preprocessing and evaluation.
- Report 5-fold CV mean±SD (and 95% CIs) or a 70/30 train/test with tuning on train only.
- For classification, use a logit head with BCE/CE; report Accuracy, F1, AUROC, and calibration.
- Explore sparsity (L1/group-lasso), temperature-scaled memberships, monotonic constraints, calibration, and bootstrap ensembling.
- Reproduce NF-RVFL under identical splits and metrics; provide a side-by-side table including parameters and time.

10 Conclusion

A reproducible NN-FCM pipeline was developed and architectural depth was shown to require dataset-specific tuning. Against NF-RVFL, NN-FCM led on Abalone and Breast Cancer, tied on Credit Approval, and lagged on the CRX split using the paper's R^2 values. Although NN-FCM can be slower than NF-RVFL, the improved accuracy on nonlinear problems and the availability of an explicit predictive equation after training can make deployment practical for families of similar datasets.

11 System / Environment Configuration

Hardware: Google Colab VM (T4/A100 GPU when available), 12–25 GB RAM.
OS/Runtime: Colab Ubuntu base, Python 3.10+.
Software: NumPy, SciPy, scikit-learn, PyTorch, Optuna, Matplotlib.
Reproducibility: fixed seeds for ensembling; version-controlled code and data processing; exact versions to be pinned.

12 References

Sajid, Malik, Tanveer, Suganthan. Neuro-Fuzzy Random Vector Functional Link Neural Network for Classification and Regression Problems. IEEE Transactions on Fuzzy Systems, 32(5):2738–2749, 2024.

UCI datasets: Credit Approval, CRX, Breast Cancer Wisconsin, Abalone.

13 Appendix

13.1 Appendix A: NN-FCM Console Logs (abridged)

- credit_approval.csv — nn-fcm-4: Early stopping @ epoch 494; training 77.4 s; R^2 = 87.96%; first 5 preds: [0.0200 0.0089 0.0449 0.1086 0.3896].
- credit_approval.csv — nn-fcm-1: Early stopping @ epoch 486; training 64.6 s; R^2 = 77.13%.
- crx.csv — nn-fcm-2: Early stopping @ epoch 444; training 49.2 s; R^2 = 47.14%.
- breast-cancer-wisconsin.csv — nn-fcm-4: Early stopping @ epoch 430; training 62.5 s; R^2 = 96.52%.
- abalone.csv — nn-fcm-5: Early stopping @ epoch 429; training 425.7 s; R^2 = 74.91%.

13.2 Appendix B: Key Equations Used in the Code

B.1 Notation and Shapes:

$$X \in \mathbb{R}^{B\times d}, \quad C \in \mathbb{R}^{R\times d}, \quad \Sigma \in \mathbb{R}^{R\times d}, \quad \mu(x) \in \mathbb{R}^{R}, \quad u \in \mathbb{N}, \quad k = 1, \dots, 3$$

B.2 Gaussian Memberships:

$$rbf_{(r,f)(x)} = \exp\left(-\frac{1}{2} * \frac{\left(\left(x_f - c_{r,f}\right)^2\right)}{}\left(\sigma_{r,f}^2 + \varepsilon\right)\right), \;\; \mu_{r(x)} = \left(\frac{1}{d}\right)\sum_{f=1}^{d} rbf_{(r,f)(x)}$$

B.3 Residual MLP Blocks:

$$h_1 = Drop\left(ReLU\left(BN_{1(W_1\mu + b_1)}\right)\right), \;\; h_{\ell+1} = Drop\left(ReLU\left(BN_{(\ell+1)(W_{\ell+1}h_\ell + b_{\ell+1})}\right)\right) + h_\ell$$

B.4 Multi-Head Readout:

$$\hat{y}^k = w^{(k)^{\mathrm{T}}} h_L + b^k, \; k = 1, \dots, 3; \; \hat{y} = \left(\frac{1}{3}\right) \sum_{k=1}^{3} \hat{y}^k$$

B.5 Loss and R²:

$$MSE = \left(\frac{1}{B}\right) \sum_{i=1}^{B} (y_i - \hat{y}_i)^2, \; R^2 = 1 - \frac{\left(\sum_{i=1}^{B} (y_i - \hat{y}_i)^2\right)}{} \left(\sum_{i=1}^{B} (y_i - \bar{y})^2\right)$$

B.6 Standardization:

$$z_f = \frac{(x_f - \mu_f)}{} \sigma_f$$

B.7 AdamW (Decoupled Weight Decay):

$$\theta_{t+1} = \theta_t - \eta * \frac{\widehat{\ } m_t}{} \left(sqrt(\widehat{\ } v_t) + \varepsilon\right) - \eta \, \lambda \, \theta_t$$

B.8 ReduceLROnPlateau:

$$\eta \leftarrow \gamma \, \eta$$

B.9 Early Stopping:

$$\delta \approx 10^{-6}, \; patience \approx 400$$

B.10 K-Fold CV Objective (Optuna):

$$\min_{\Theta} J(\Theta) = -\left(\frac{1}{K}\right) \sum_{k=1}^{K} R^2_{(k)(\Theta)}, \; \Theta = \{ n_{rules}, lr, mlp_{units}, dropout, \ell_2 \}$$

B.11 Ensemble Averaging:

$$\hat{y}_{ens(x)} = \left(\frac{1}{M}\right) \sum_{m=1}^{M} \hat{y}^{(m)(x)}, \; M = 5$$

B.12 Classification Head (future):

$$BCE = -\left(\frac{1}{B}\right) \sum_{i=1}^{B} \left[y_i \log\left(\sigma(\hat{z}_i)\right) + (1 - y_i) \log\left(1 - \sigma(\hat{z}_i)\right) \right], \; \sigma(t) = \frac{1}{1 + e^{-t}}$$

13.3 Appendix C: GitHub

Repository: *NN-FCM vs NF-RVFL — Comparative Report*.

Available at:
https://github.com/iftesam/NN-FCM-vs-NF-RVFL-Comparative-Report