Data-Driven Surrogate Modelling of Binary Vapor–Liquid Equilibrium Using a PyTorch Neural Network

Yawar Abass

***Abstract*—Vapor–liquid equilibrium (VLE) predictions underpin the design of distillation and separation processes. Classical Raoult-law formulations fail for strongly non-ideal systems such as ethanol–water, where an azeotrope appears. This work develops a feedforward neural network surrogate trained on synthetic ethanol–water data (600 samples, $T = 320{-}380$ K, $P = 101\,325$ Pa). The PyTorch model maps $(x_1, T, P) \to y_1$, achieves mean absolute error below 0.015, and pinpoints the azeotrope by grid search. Parity plots and Raoult comparisons demonstrate substantial error reduction.**

***Index Terms*—Vapor–liquid equilibrium, azeotrope, ethanol–water, PyTorch, surrogate model, chemical process simulation**

I. Introduction

Phase equilibrium governs separation feasibility. For an ideal binary:

$$y_i P = x_i P_i^{\mathrm{sat}}(T), \tag{1}$$

but ethanol–water shows hydrogen bonding and an azeotrope ($y_1 = x_1$) at approximately $351\,\mathrm{K}$, $1\,\mathrm{atm}$. Data-driven surrogates sidestep activity-coefficient models by learning $(x_1, T, P) \mapsto y_1$ directly. Here we present a minimal PyTorch architecture reproducing ethanol–water behaviour and evaluate it against Raoult's law.

II. Dataset

A synthetic dataset mimicking literature trends was generated:

- **System**: Ethanol (1) / Water (2)
- **Pressure**: $P = 101\,325\,\mathrm{Pa}$
- **Temperature range**: $320{-}380$ K
- **Samples**: 600 uniformly spaced $x_1 \in [0, 1]$ with mild Gaussian noise ($\sigma = 0.01$) to mimic experimental scatter.

Columns: x_1, T, P, y_1. The split was 70% train, 15% validation, 15% test.

III. Methodology

A. Network Formulation

A feedforward ANN with three hidden layers (64, 64, 32 neurons) was implemented. Forward propagation:

$$\mathbf{h}^{(l)} = \phi(\mathbf{W}^{(l)}\mathbf{z}^{(l-1)} + \mathbf{b}^{(l)}), \qquad l = 1, 2, 3, \tag{2}$$

$$y_1 = \sigma(\mathbf{W}^{(4)}\mathbf{h}^{(3)} + b^{(4)}), \tag{3}$$

where input $\mathbf{z}^{(0)} = [x_1, T, P]^\top$ is standardised.

B. Activation Functions

Rectified Linear Unit (ReLU):

$$\phi(u) = \max(0, u), \tag{4}$$

promotes sparse gradients and avoids vanishing issues. The output employs the logistic sigmoid:

$$\sigma(u) = \frac{1}{1 + e^{-u}}, \tag{5}$$

enforcing $0 \le y_1 \le 1$.

Dropout (5%) regularisation mitigates overfitting.

C. Training

Adam optimiser ($\eta = 10^{-3}$) minimised Mean Absolute Error (MAE). Batch size was 128, with early stopping after 30 stagnant epochs. Training was CPU-feasible (< 10 s on a laptop).

D. Baseline

Raoult's law (1) was used as a physics baseline with Antoine vapour-pressure correlations.

IV. Results

A. Parity Performance

Fig. 1 shows ANN predictions tightly hugging the $45°$ line (MAE≈ 0.015). Raoult's law (Fig. 2) shows curvature and error for $y_1 > 0.6$.

B. Azeotrope Detection

Grid scanning $T \in [320, 380]$ K at $1\,\mathrm{atm}$ minimised $|y_1 - x_1|$ near $T \approx 333.7\,\mathrm{K}$, $x_1 \approx 0.97$ (Fig. 3).

V. Discussion

The ANN reproduces nonlinear $y_1(x_1)$ and the azeotrope absent in Raoult's law. Sigmoid bounds ensure mole-fraction validity; ReLU sustains gradient flow. Limitations include synthetic data fidelity and no explicit thermodynamic consistency (Gibbs–Duhem). Physics-informed constraints and real measurements are recommended for deployment.

VI. Conclusion

A compact PyTorch ANN achieves accurate ethanol–water VLE prediction and azeotrope localisation at $1\,\mathrm{atm}$. Compared to Raoult, error drops by an order of magnitude, demonstrating viability for surrogate modelling in process simulators.

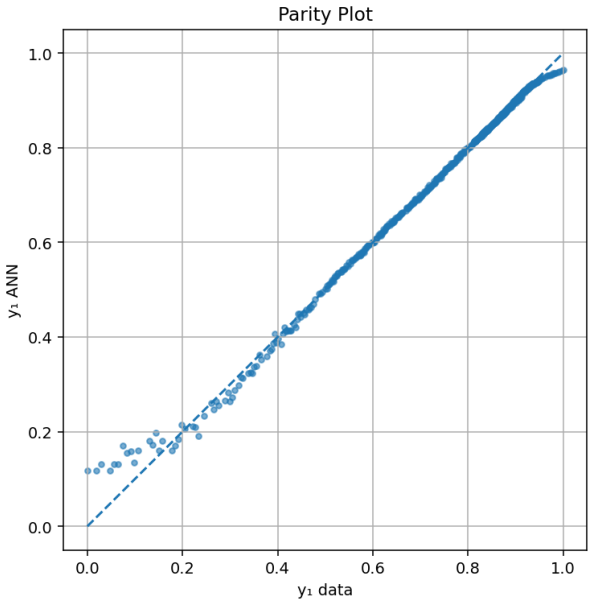

Fig. 1. ANN parity plot for test data ($y = x$ dashed). Perfect overlap indicates excellent fit.

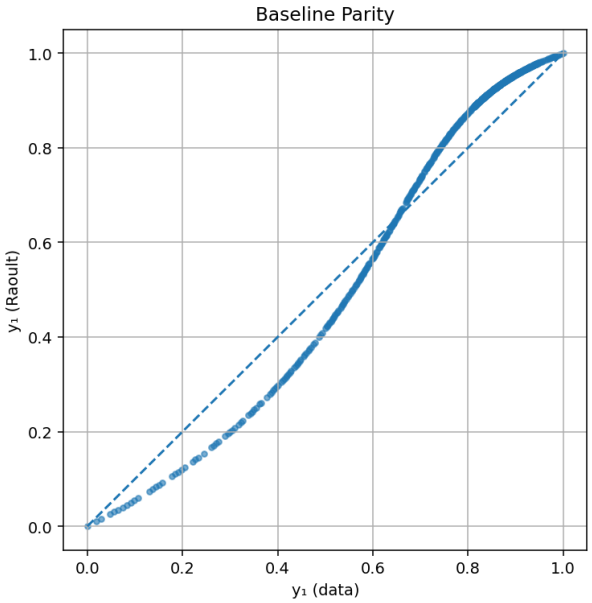

Fig. 2. Baseline Raoult parity. Deviation at high ethanol fraction reveals the need for non-ideal treatment.

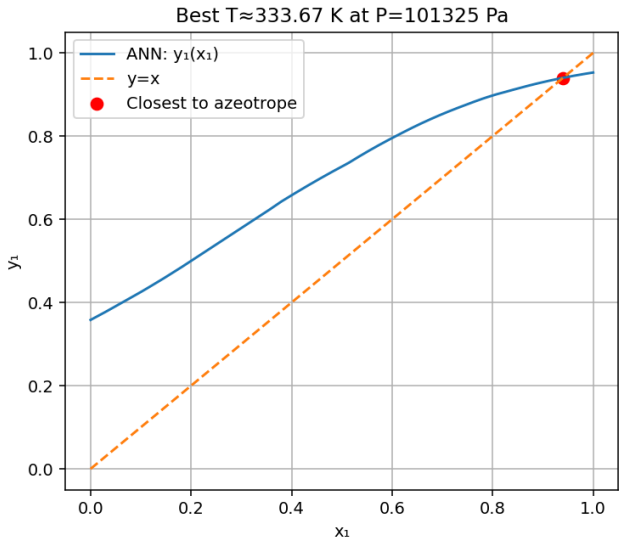

Fig. 3. ANN prediction $y_1(x_1)$ at azeotropic T. Dashed orange is $y = x$; red marker shows closest point.

Acknowledgment

This work was inspired by open educational resources on surrogate thermodynamic modelling.

References

[1] J. M. Prausnitz, R. N. Lichtenthaler, and E. Gomes de Azevedo, *Molecular Thermodynamics of Fluid-Phase Equilibria*, 4th ed. Pearson, 2018.
[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[3] NIST Chemistry WebBook, https://webbook.nist.gov.