Faculty of Computer Science

OSDA Big Homework

Moscow 2024

OSDA Big Homework: Neural FCA

UCI Acute Inflammations Dataset
Exploring Interpretability and Efficiency in Classification Tasks

The presentation was prepared by Muhammad Zeeshan Asghar

Dataset Description

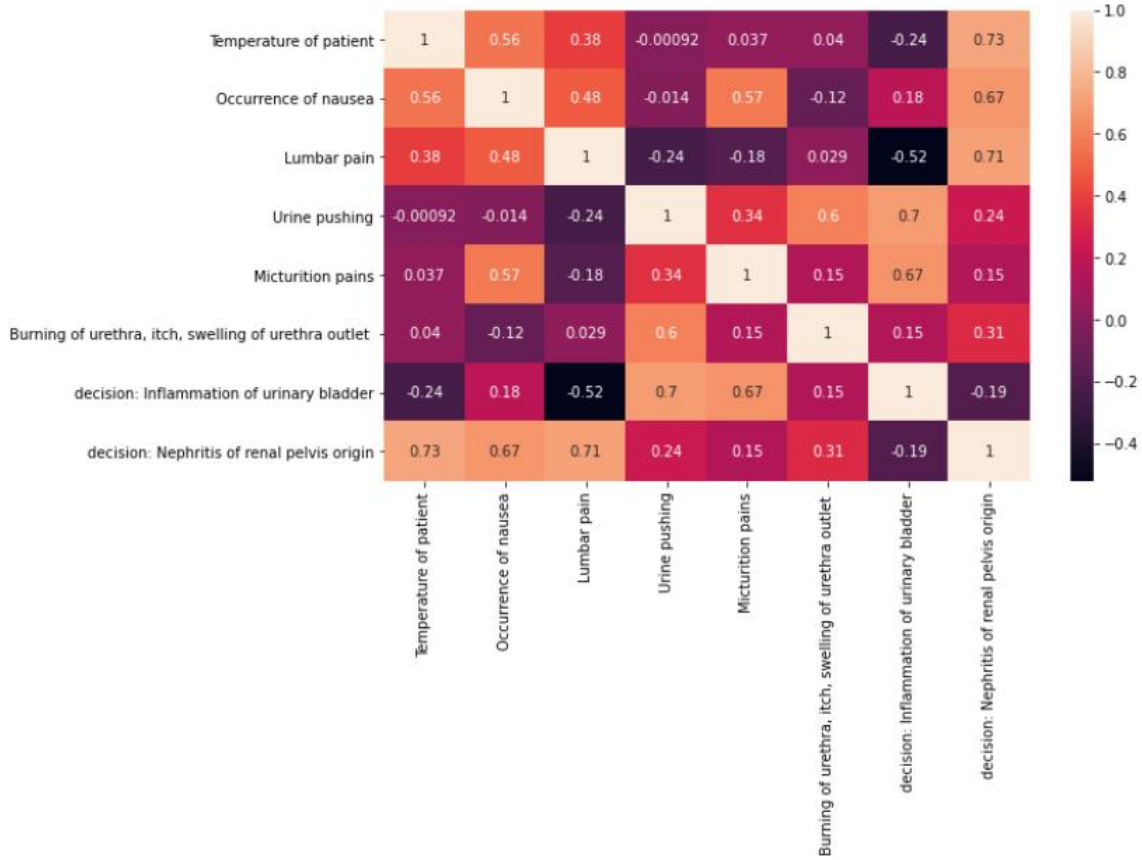
The UCI Acute Inflammations dataset is a collection of patient records, including six features (five binary and one numerical) and two target variables (bladder inflammation and nephritis).

Features: 6 attributes, 5 binary (e.g., nausea, lumbar pain) and 1 continuous (temperature)

Targets:
– Bladder Inflammation (Cystitis): Binary classification (0 = no, 1 = yes)
– Nephritis: Binary classification (0 = no, 1 = yes)

Dataset

Variable Name	Role	Type
temperature	Feature	Continuous
nausea	Feature	Categorical
lumbar-pain	Feature	Categorical
urine-pushing	Feature	Categorical
micturition-pains	Feature	Categorical
burning-urethra	Feature	Categorical
bladder-inflammation	Target	Categorical
nephritis	Target	Categorical



Prediction quality measure

Target 1:

Bladder Inflammation (Cystitis): 60 positive and 60 negative samples.

Balance Score = 1

We can use F1 score or accuracy for evaluation

Target 2:

Nephritis: 50 positive and 70 negative samples.

Balance Score = 0.714

We should use F1 score for evaluation

Approach Overview

1. Evaluate binarization methods for the temperature attribute.
2. Compare two concept selection metrics: F1 score and accuracy.
3. Investigate the impact of activation functions on Neural FCA performance.
4. Benchmark Neural FCA against traditional classifiers, like Logistic Regression and Random Forest.

First Binarization Approach

We see that the temperature readings are mostly between either 37-37.5 degrees or 40-41.5 degrees.

- Threshold: 38°C

- Categorization:

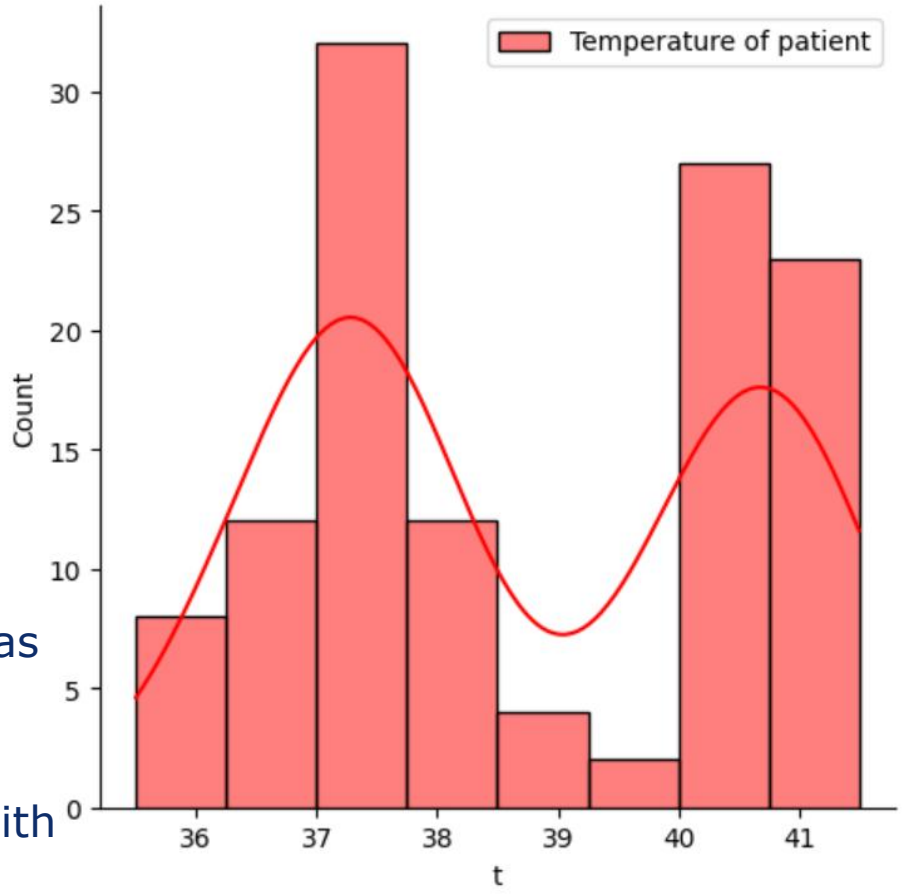
  - High Temperature: Temperatures > 38°C

  - Low Temperature: Temperatures ≤ 38°C

Rationale:

- Source: The threshold is based on the World Health Organization (WHO) definition of fever, which considers a temperature above 38°C as indicative of a fever. This provides a legitimate and widely accepted basis for the threshold.

- Data Distribution: The data shows a clear distinction around 38°C, with fewer samples around this point, making it a suitable cutoff.



First Binarization Result for diagnosing Bladder Inflammation

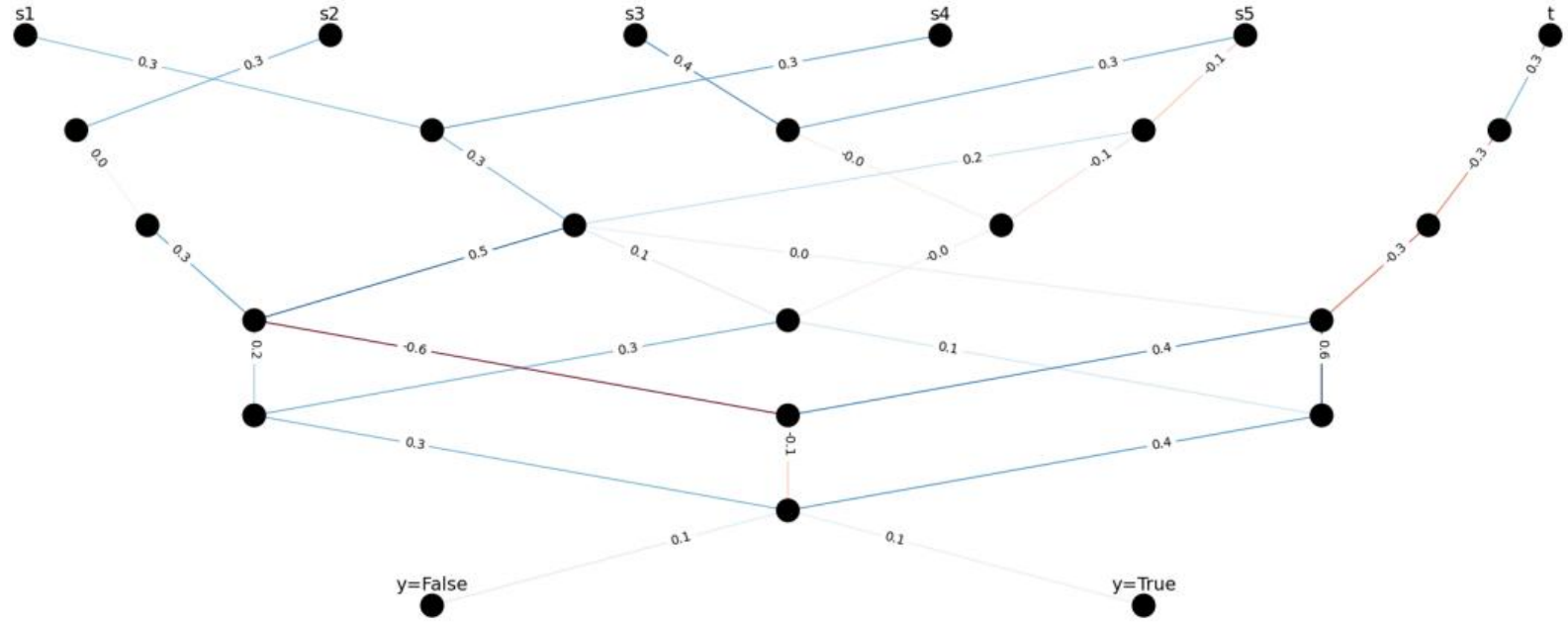


Figure 2: Neural network for Model 1 (Dichotomic scale, F1-score to select best concepts, ReLU activation and target attribute Bladder Inflammation).

First Binarization Result for diagnosing Nephritis

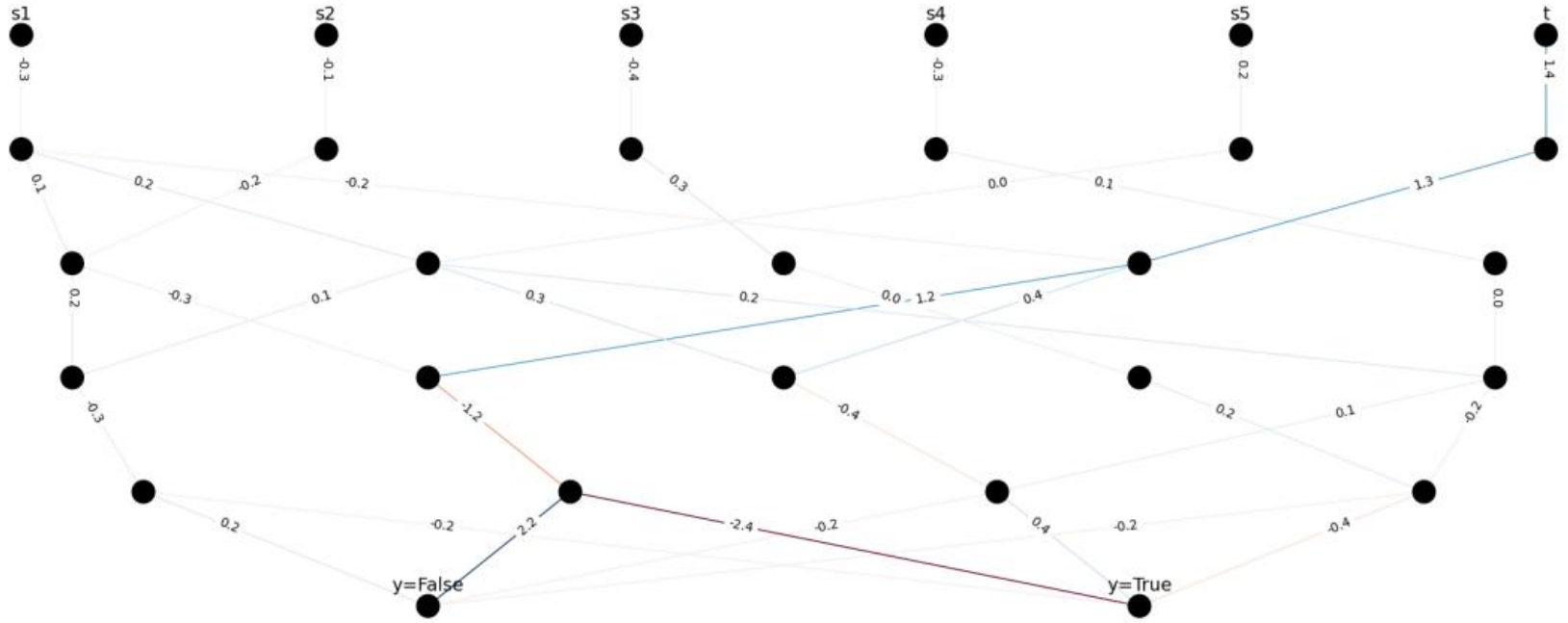


Figure 3: Neural network for Model 2 (Dichotomic scale, F1-score to select best concepts, ReLU activation and target attribute Nephritis).

Second Binarization Approach

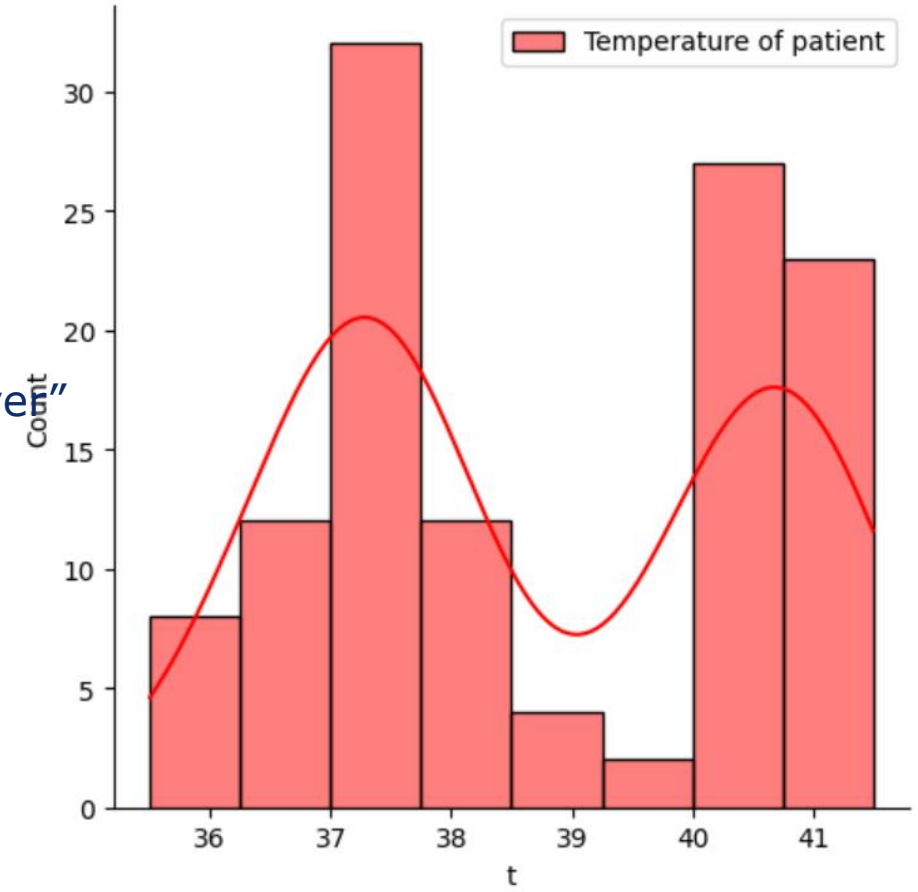
For the second approach we used ordinal scale to binarize the temperature.

- Temperature ≤ 36.5: Categorized as ”Normal”
- Temperature ≤ 37.5: Categorized as ”Normal” or ”Elevated”
- Temperature ≤ 41.5: Categorized as ”Normal” or ”Elevated” or ”Fever”

This approach captures intermediate states.

Rationale:

Human body temperature by wikipedia
https://en.wikipedia.org/wiki/Human_body_temperature



Second Binarization Result for diagnosing Bladder Inflammation

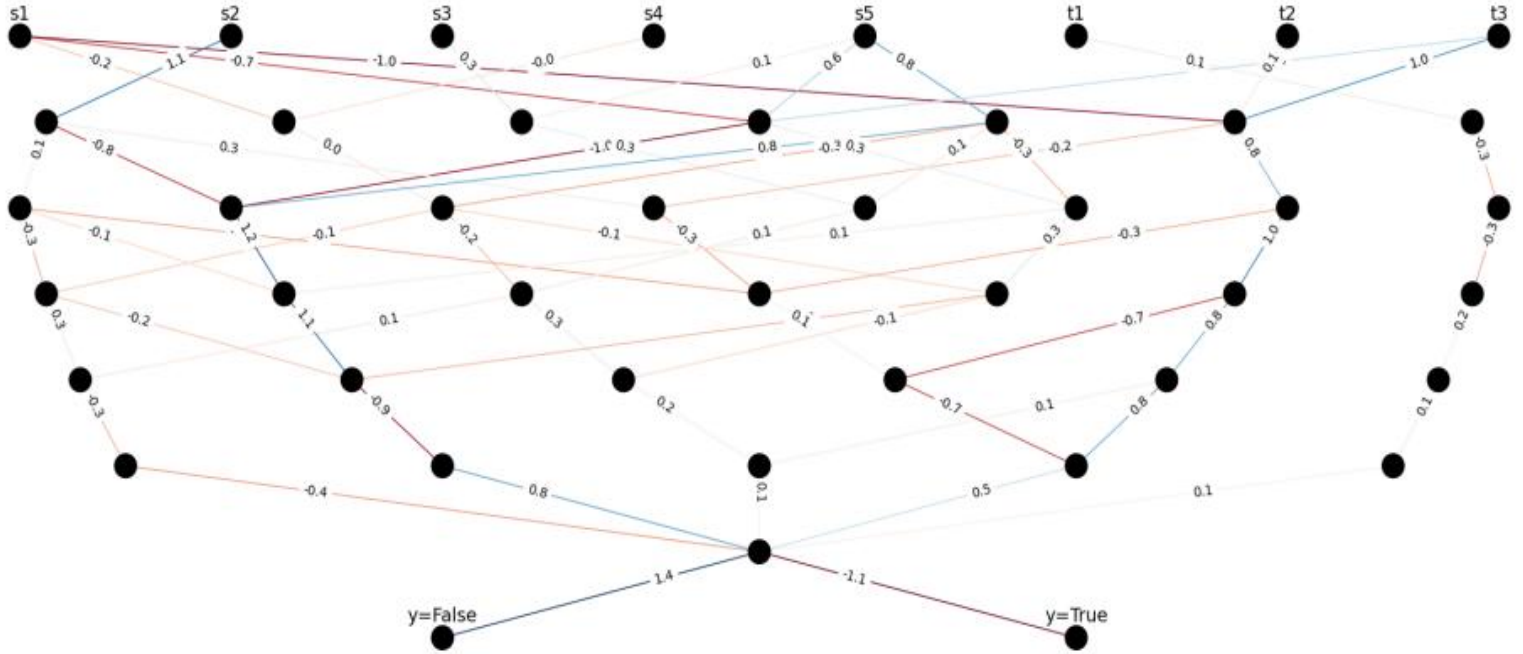


Figure 4: Neural network for Model 3 (Ordinal scale, F1-score to select best concepts, ReLU activation and target attribute Bladder Inflammation).

Second Binarization Result for diagnosing Nephritis

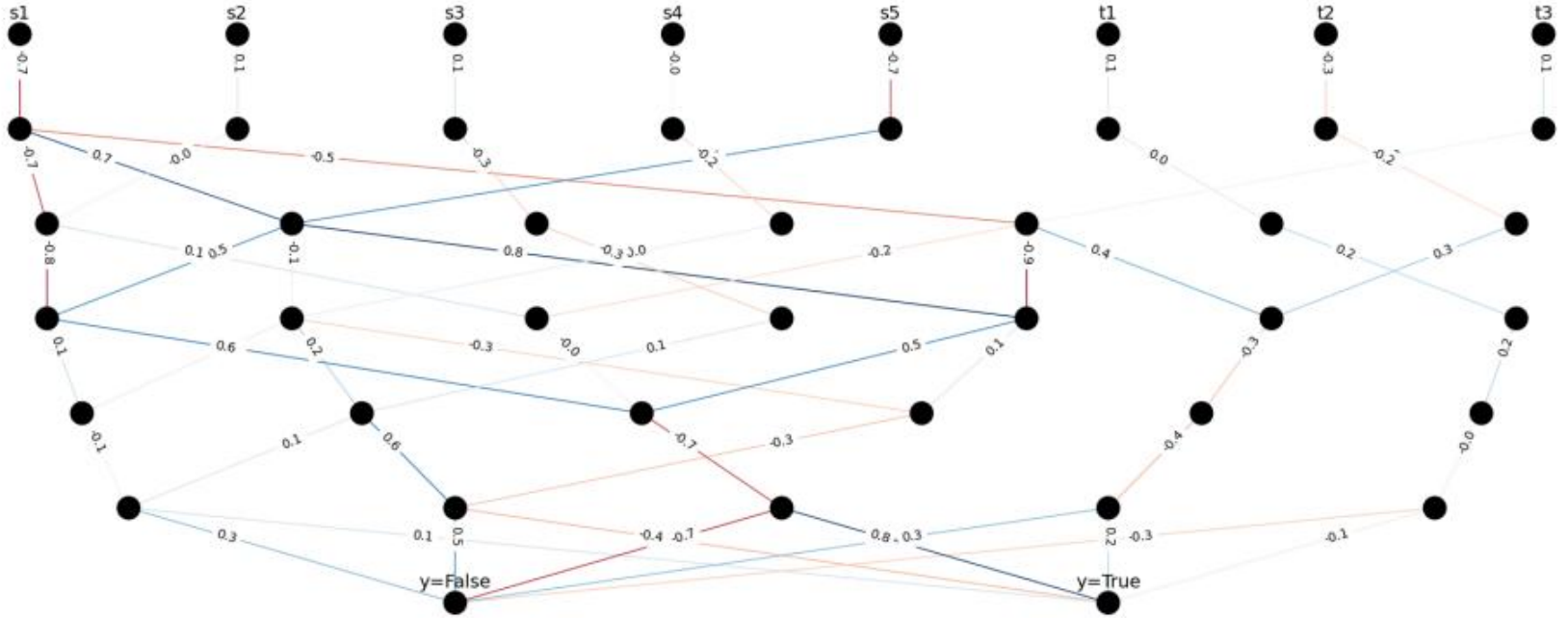


Figure 5: Neural network for Model 4 (Ordinal scale, F1-score to select best concepts, ReLU activation and target attribute Nephritis).

Another Approach for selecting best concepts

Let’s use accuracy instead of F1 score to select best concepts to apply neural network on. We will test this on both binarized techniques.

First Binarization Result for diagnosing Bladder Inflammation

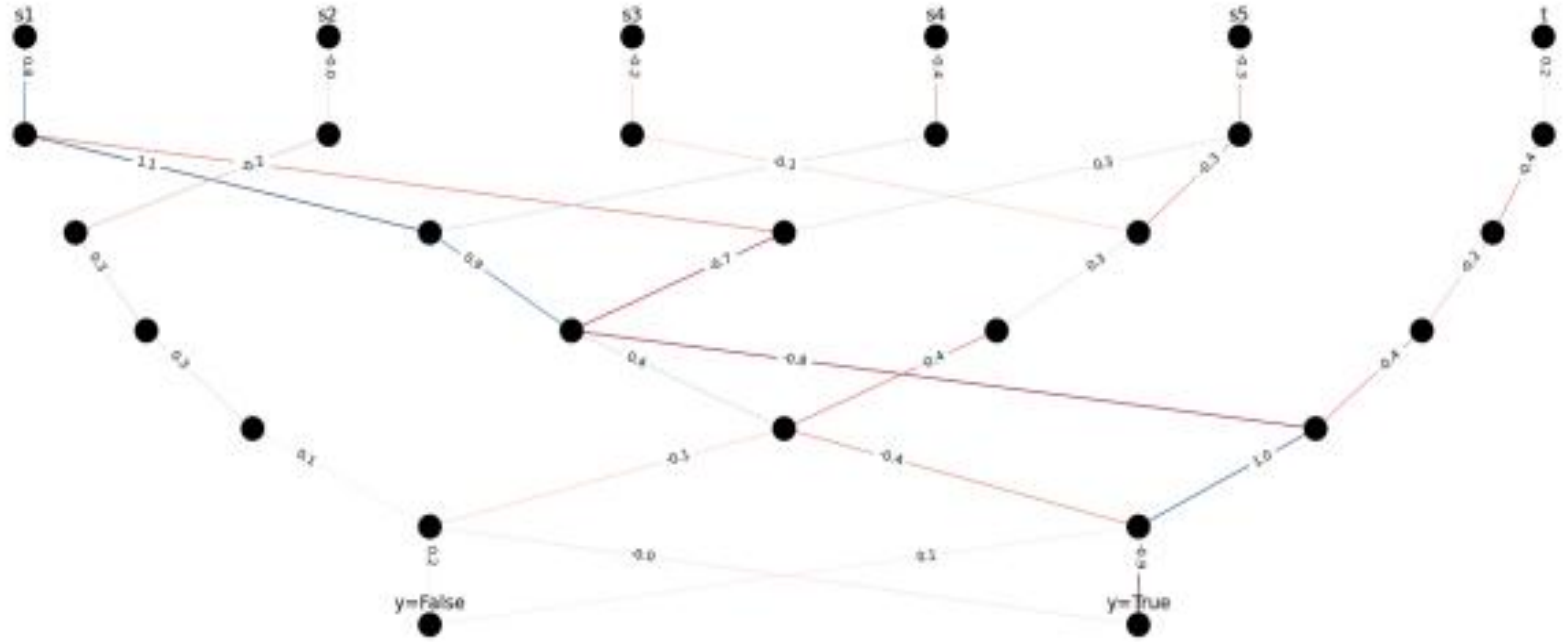


Figure 6: Neural network for Model 5 (Dichotomic scale, accuracy to select best concepts, ReLU activation and target attribute Bladder Inflammation).

First Binarization Result for diagnosing Nephritis

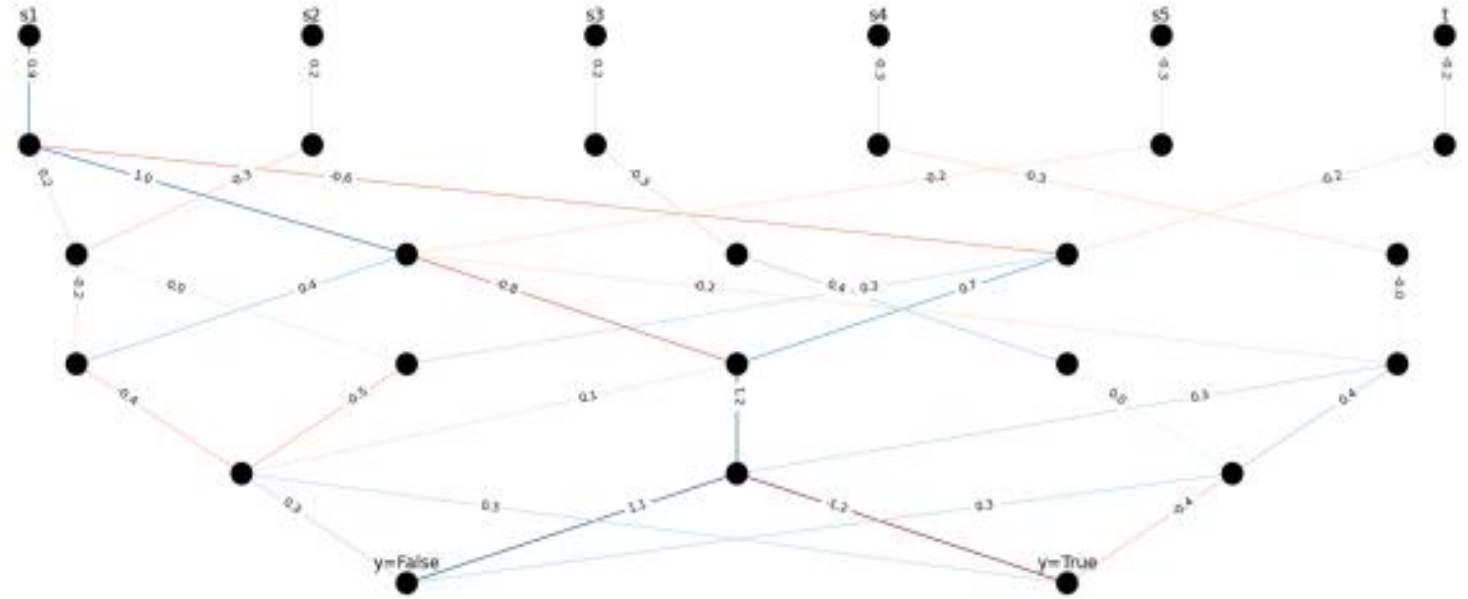


Figure 7: Neural network for Model 6 (Dichotomic scale, accuracy to select best concepts, ReLU activation and target attribute Nephritis).

Second Binarization Result for diagnosing Bladder Inflammation

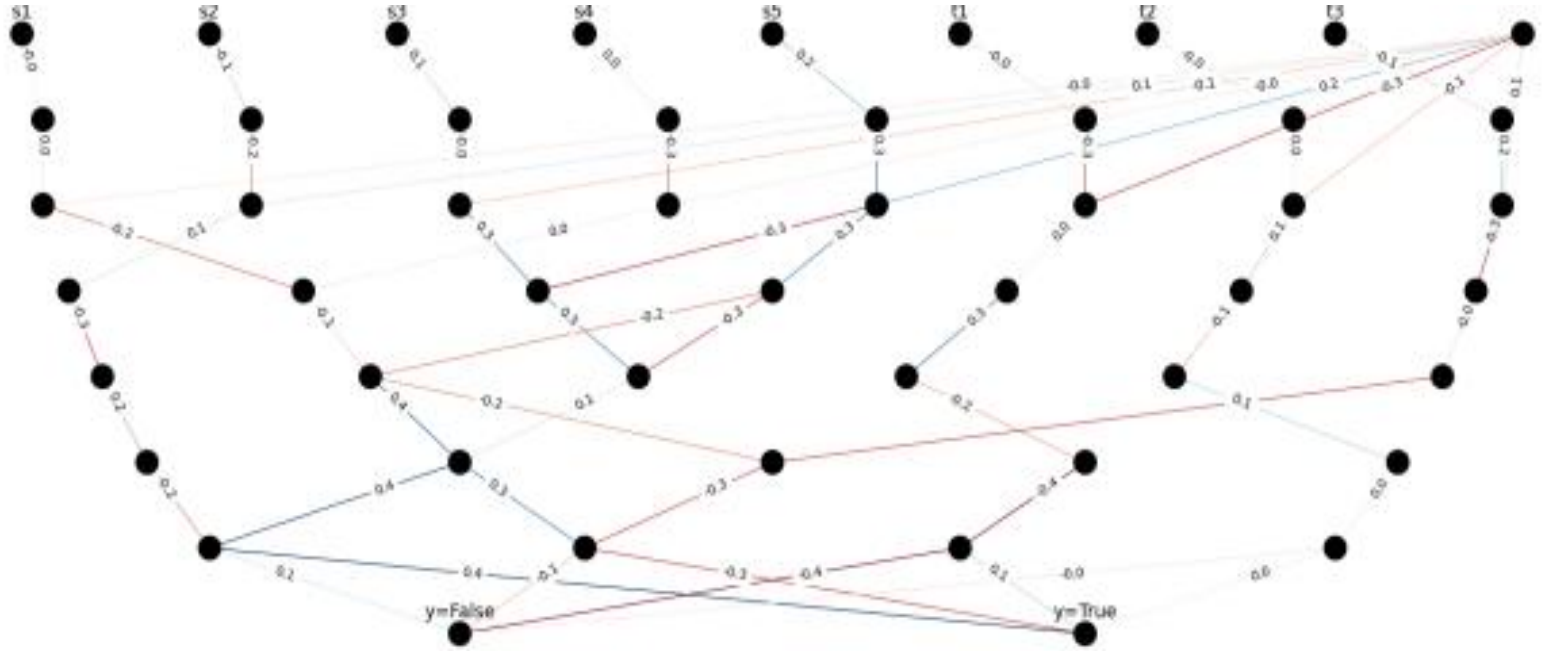


Figure 8: Neural network for Model 7 (Ordinal scale, accuracy to select best concepts, ReLU activation and target attribute Bladder Inflammation).

Second Binarization Result for diagnosing Nephritis

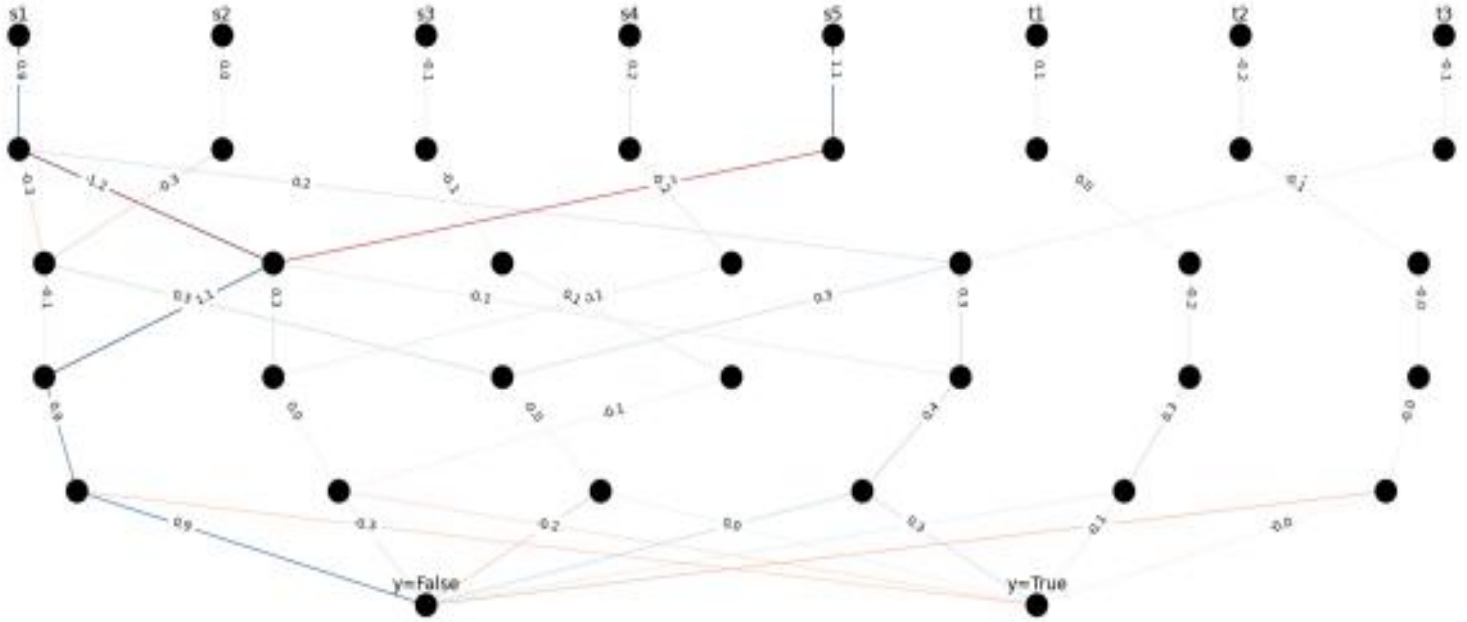


Figure 9: Neural network for Model 8 (Ordinal scale, accuracy to select best concepts, ReLU activation and target attribute Nephritis).

Using different Activation function

The default nonlinearity in the network is ReLU. Let's use our first and best model, i.e. using F1 score to select best concepts, to test Leaky ReLU and hyperbolic tangent.

Leaky ReLU with Dicromatic Scale Binarization Result for diagnosing Bladder Inflammation

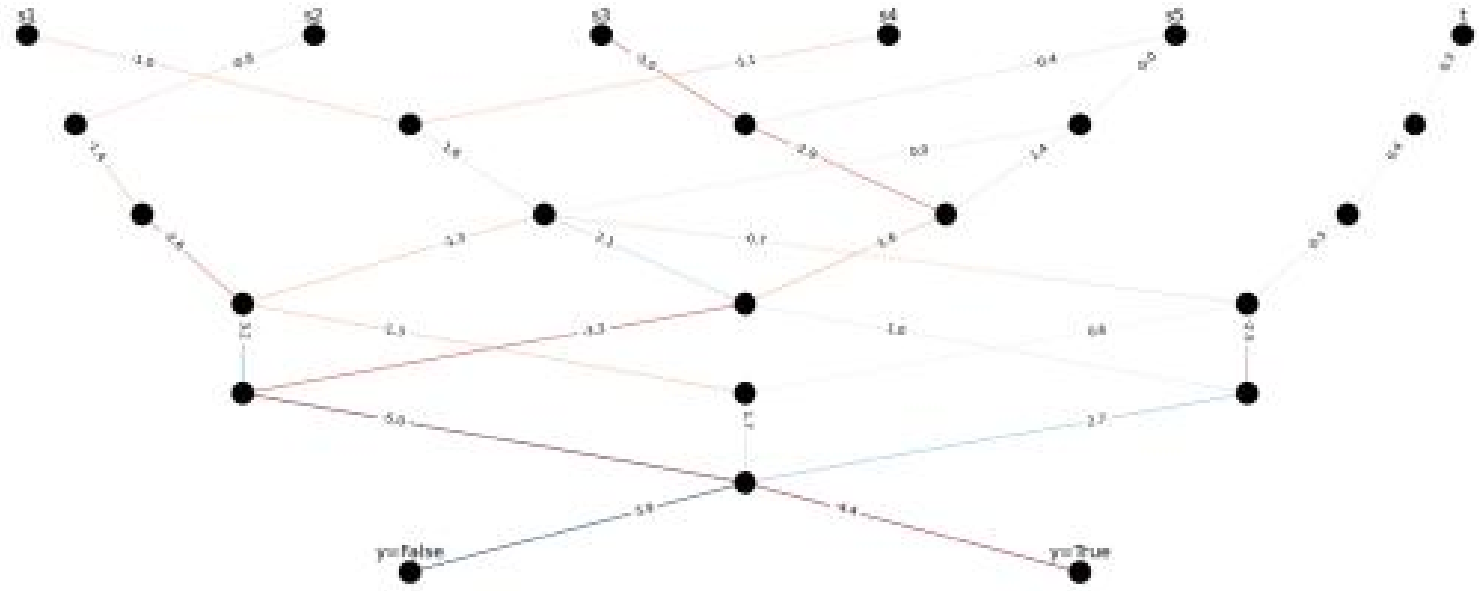


Figure 10: Neural network for Model 9 (Dichotomic scale, F1-score to select best concepts, Leaky ReLU activation and target attribute Bladder Inflammation).

Leaky ReLU with Dicromatic Scale Binarization Result for diagnosing Nephritis

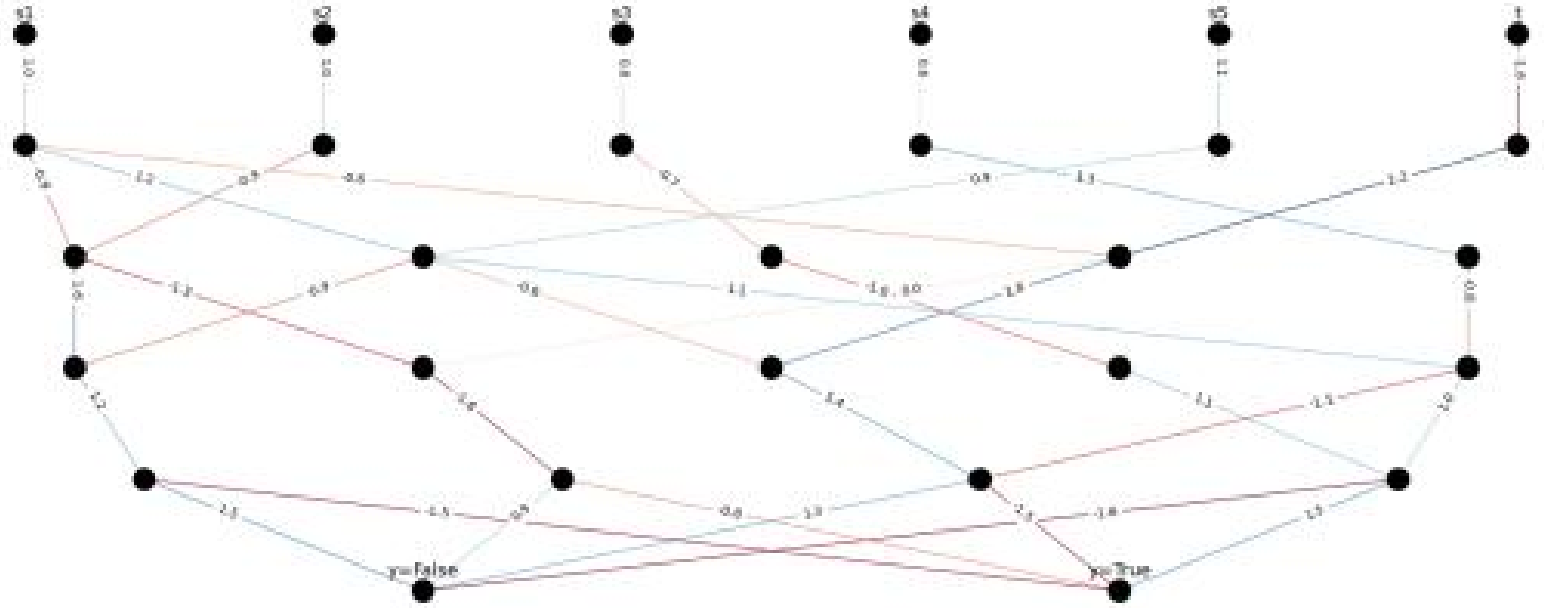


Figure 11: Neural network for Model 10 (Dichotomic scale, F1-score to select best concepts, Leaky ReLU activation and target attribute Nephritis).

Hyperbolic Tangent with Dicromatic Scale Binarization Result for diagnosing Bladder Inflammation

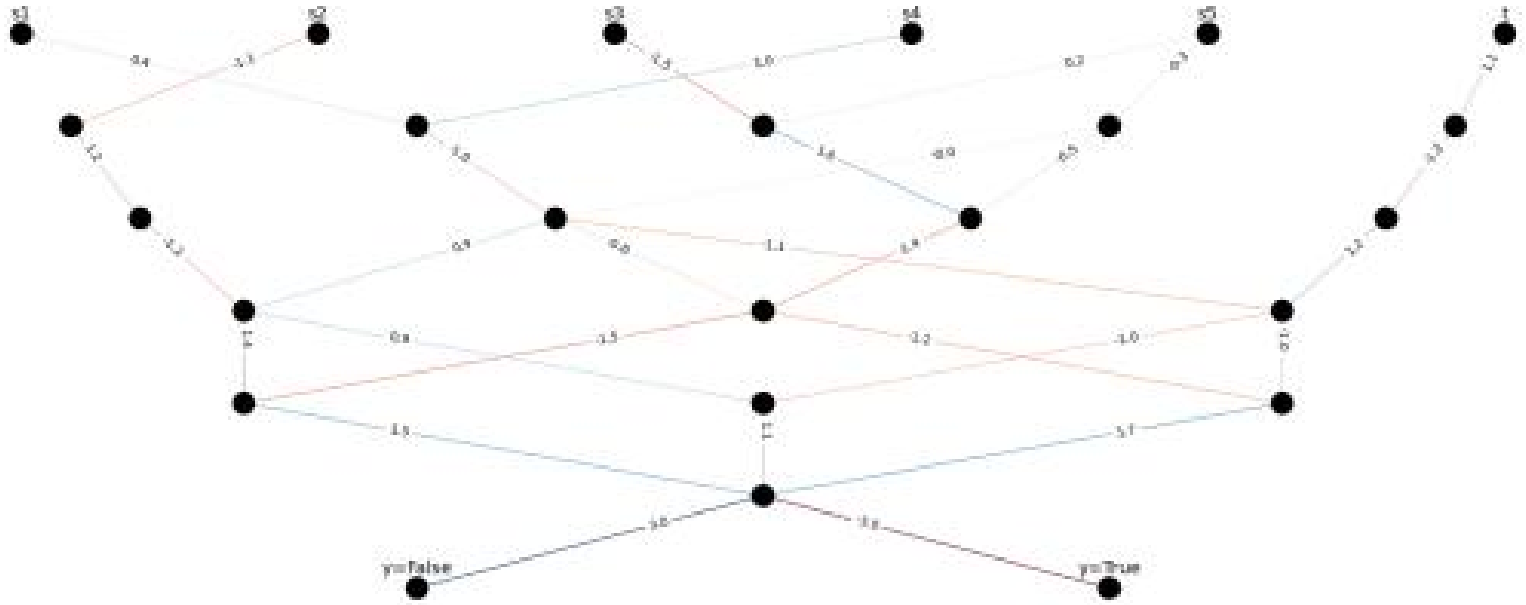


Figure 12: Neural network for Model 11 (Dichotomic scale, F1-score to select best concepts, hyperbolic tangent activation and target attribute Bladder Inflammation).

Hyperbolic Tangent with Dicromatic Scale Binarization Result for diagnosing Nephritis

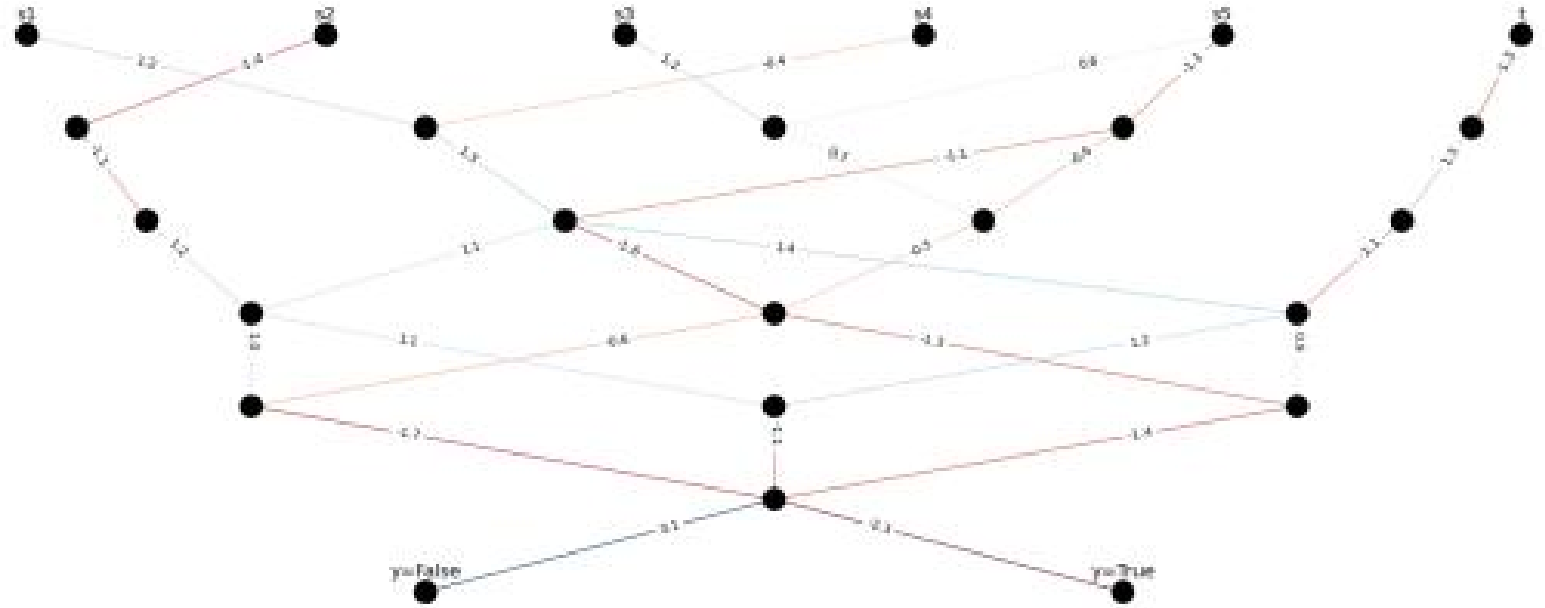


Figure 13: Neural network for Model 12 (Dichotomic scale, F1-score to select best concepts, hyperbolic tangent activation and target attribute Nephritis).

Leaky ReLU with Ordinal Scale Binarization Result for diagnosing Bladder Inflammation

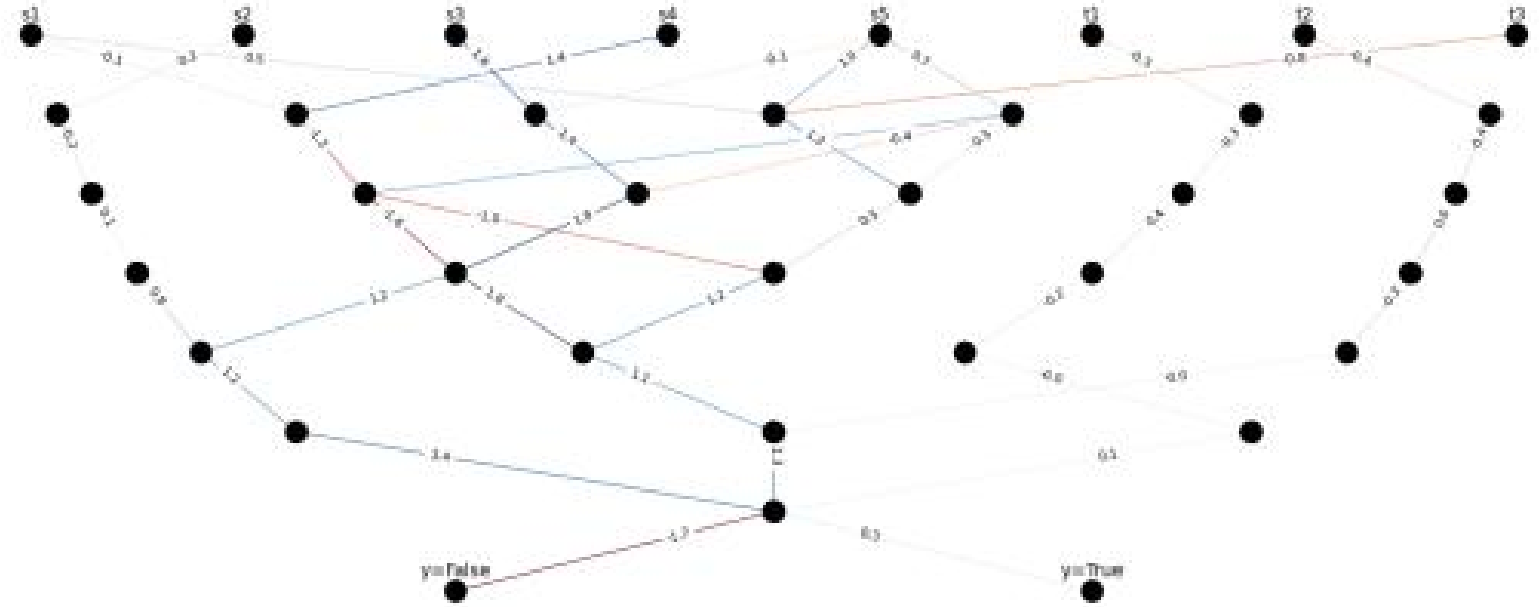


Figure 14: Neural network for Model 13 (Ordinal scale, F1-score to select best concepts, Leaky ReLU activation and target attribute Bladder Inflammation).

Leaky ReLU with Ordinal Scale Binarization Result for diagnosing Nephritis

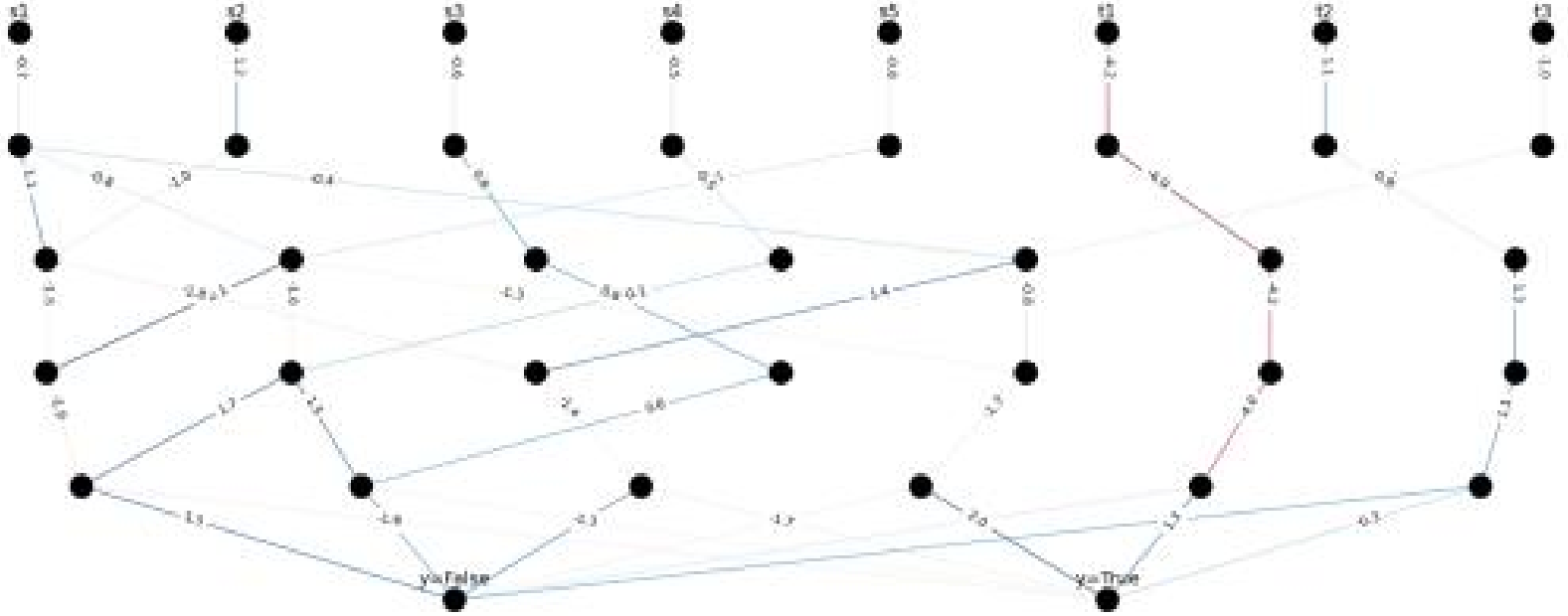


Figure 15: Neural network for Model 14 (Ordinal scale, F1-score to select best concepts, Leaky ReLU activation and target attribute Nephritis).

Hyperbolic Tangent with Ordinal Scale Binarization Result for diagnosing Bladder Inflammation

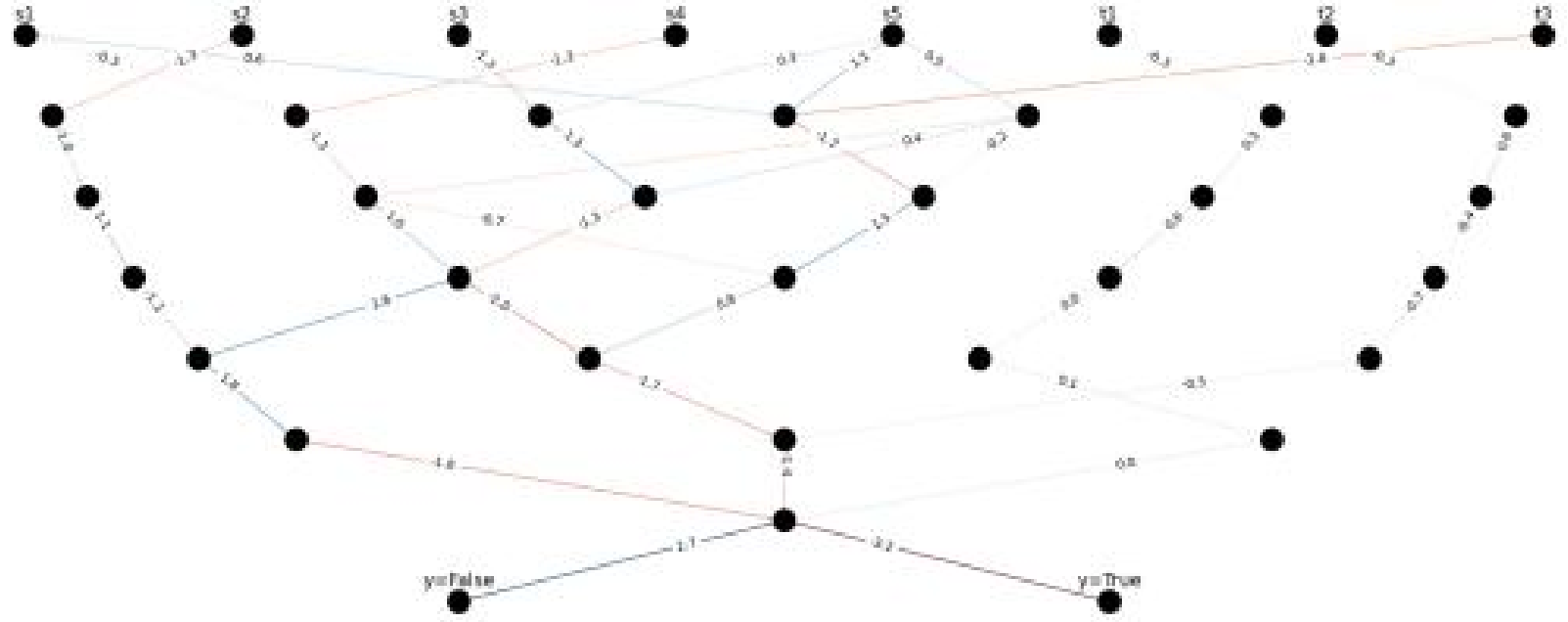


Figure 16: Neural network for Model 15 (Ordinal scale, F1-score to select best concepts, hyperbolic tangent activation and target attribute Bladder Inflammation).

Hyperbolic Tangent with Ordinal Scale Binarization Result for diagnosing Nephritis

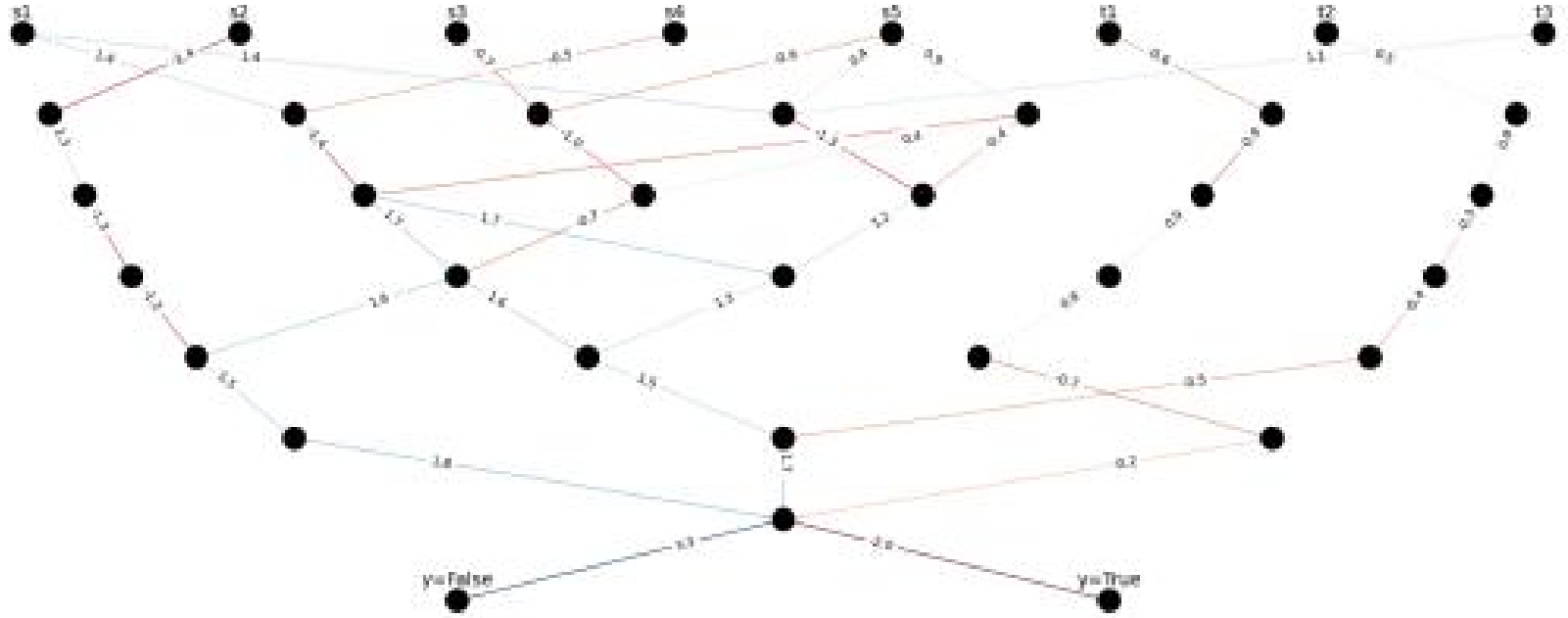


Figure 17: Neural network for Model 16 (Ordinal scale, F1-score to select best concepts, hyperbolic tangent activation and target attribute Nephritis).

Neural FCA Performance

Model	Target attribute	Binarization Type	Best Concept Metric	Activation Function	F1 Score
1	Bladder Inflammation	Binary	F1 Score	ReLU	0.93
2	Nephritis	Binary	F1 Score	ReLU	0.86
3	Bladder Inflammation	Ordinal	F1 Score	ReLU	0.78
4	Nephritis	Ordinal	F1 Score	ReLU	0.875
5	Bladder Inflammation	Binary	Accuracy	ReLU	0.57
6	Nephritis	Binary	Accuracy	ReLU	0.875
7	Bladder Inflammation	Ordinal	Accuracy	ReLU	0
8	Nephritis	Ordinal	Accuracy	ReLU	0.875
9	Bladder Inflammation	Binary	F1 Score	Leaky ReLU	1
10	Nephritis	Binary	F1 Score	Leaky ReLU	1
11	Bladder Inflammation	Binary	F1 Score	Hyperbolic Tangent	1
12	Nephritis	Binary	F1 Score	Hyperbolic Tangent	1
13	Bladder Inflammation	Ordinal	F1 Score	Leaky ReLU	1
14	Nephritis	Ordinal	F1 Score	Leaky ReLU	1
15	Bladder Inflammation	Ordinal	F1 Score	Hyperbolic Tangent	1
16	Nephritis	Ordinal	F1 Score	Hyperbolic Tangent	1

Table 1: F1 Score for each model using different target attributes, binarization types, activation functions, and concept selection method.

Insights from Weighted Neural Networks

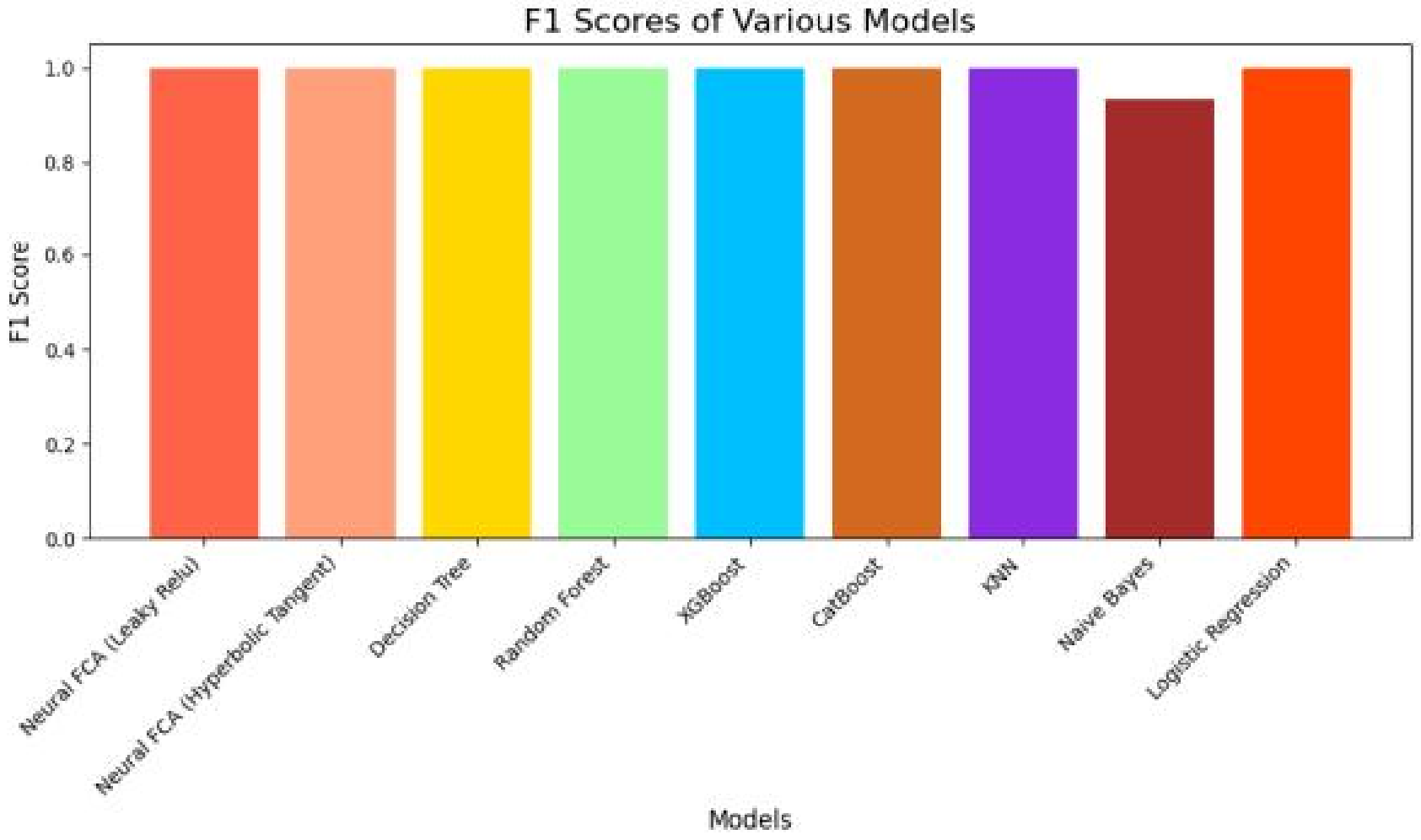
Visualization of the neural network’s weighted edges provided valuable insights into feature importance:

1. Temperature: Had a strong influence on both targets, indicating its critical role in diagnosing fever-related conditions.

2. Lumbar Pain: Was particularly significant for predicting Nephritis, highlighting its relevance to kidney-related symptoms.

3. Urine Pushing and Micturition Pains: Were crucial for Bladder Inflammation prediction.

Comparison with State-of-the-Art Approaches



Comparison with State-of-the-Art Approaches

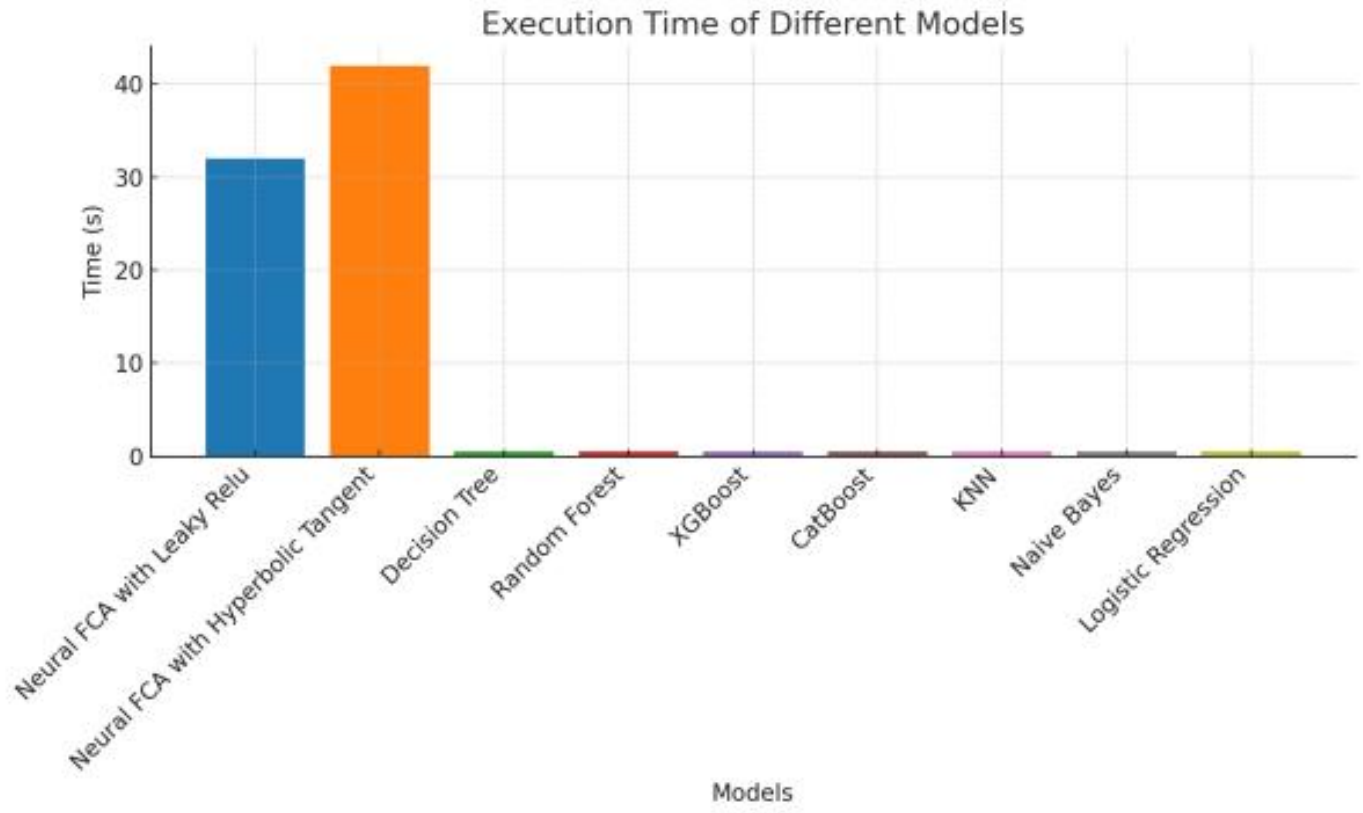


Figure 19: Time performance of Neural FCA against State-of-the-Art Approaches).

Conclusion

Neural FCA provided excellent interpretability through concept lattices and visualizations, which offered a deeper understanding of the relationships between attributes and target outcomes. The results underscore the importance of the activation function choice, with Leaky ReLU and Tanh significantly improving performance over ReLU.

However, the method’s computational complexity limits scalability for larger datasets. Further optimizations, such as more efficient lattice construction or pruning of irrelevant concepts, could enhance performance.

References

- UCI Acute Inflammations Dataset (https://archive.ics.uci.edu/dataset/184/acute+inflammations)’
- Human Body Temperature Binarization (https://en.wikipedia.org/wiki/Human_body_temperature)
- World Health Organization (WHO) Definition of Fever (https://www.who.int/data/gho/indicator-metadata-registry/imr-details/180)