Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

Hybrid Neuro-Fuzzy Estimation of Cold-Urticaria Threshold and Severity from Multi-Sensor Provocative Testing

Dr. Adarsha Sagar H V
Assistant Professor

Viamrsh Buyyani

Tanmay Agarwal

Yeruva Taniya Paul

Dept. of Computer Science and Engineering
B. M. S. College of Engineering, Bengaluru, India

Abstract—Danilychev et al. built TempPro, a 140 mm, 8-sensor applicator that measures the temperature threshold (TTV) of cold and heat urticaria, but it recovers the threshold with a fixed linear interpolation and reports no learning component or classification metrics. We ask whether soft computing can extract more from the same measurements. Because no public per-patient provocation dataset exists, we generate a simulated cohort of 2,000 sessions calibrated to the base paper's device physics and to published cohort statistics, and compare a first-order Takagi-Sugeno ANFIS (a neuro-fuzzy hybrid) with the paper's method, logistic/linear regression, a hand-written fuzzy inference system and an MLP. Over 5 by 3 repeated cross-validation, ANFIS raises severity ROC-AUC from 0.877 to 0.910, recall at 90 percent specificity from 0.677 to 0.773, and cuts threshold MAE from 0.752 to 0.465 degrees Celsius. It beats the MLP and the plain fuzzy system on most metrics while remaining a readable rule base. A genetic-algorithm-tuned ANFIS gives the same accuracy (no significant difference), and neither hybrid beats well-tuned logistic or linear regression. All results are from simulated data and show feasibility, not clinical validity.

Keywords—urticaria, provocative testing, ANFIS, neuro-fuzzy, genetic algorithm, hybrid soft computing, ROC-AUC

I. INTRODUCTION

Physical urticaria is diagnosed by provocation: a cold or warm stimulus is applied and the clinician looks for a wheal. The classical ice-cube test is qualitative, yet the temperature at which the wheal first appears (the threshold, TTV) is clinically informative about severity and treatment response [1]. TempPro [1] addresses this with a linear applicator carrying eight temperature sensors every 20 mm, a monotone profile from Tmin to Tmax inside 0–45 °C, and an estimated accuracy of ±0.5 °C, against about ±2 °C for the only existing device, TempTest [2].

The base paper is a hardware report. It contains no dataset, no classifier and no ROC or recall figures, and the threshold is read by linear interpolation between sensors at the blister edge. Real sessions add imprecision that such a rule ignores: poor skin contact and warm rooms shift the skin-side temperature away from the applicator reading, sensors drift outside the +10…+27 °C ambient window, and the blister edge is hard to call. Imprecise, uncertain inputs of this kind are what fuzzy logic and neuro-fuzzy systems are designed for. This paper therefore asks a narrow question: on the same measurements, does a hybrid neuro-fuzzy model improve threshold estimation and severity classification over (i) the paper's rule and (ii) non-hybrid alternatives?

Contributions: (1) a reproducible, physics-calibrated simulated cohort for TempPro-style sessions; (2) a numpy ANFIS with hybrid learning (least squares plus gradient descent) and a GA-tuned variant (GA-ANFIS); (3) a symmetric comparison against the base method and four non-hybrid models with matched hyper-parameter tuning, repeated cross-validation, paired significance tests and a noise-robustness study; (4) an explicit account of where the hybrid does not win.

II. BASE PAPER AND RELATED WORK

TempPro [1] measures the applicator profile independently at the skin contact and recovers the threshold from the blister position Lx along the 140 mm applicator. More than half of cold-urticaria patients in the authors' sample had a TTV of +12 to +18 °C. Clinical cohorts split typical cold urticaria into a low-threshold (≤17 °C) and high-threshold (>17 °C) group, with the high group showing more asthma and more antihistamine dose escalation [3]; a Thai cohort reports a mean critical temperature of 21 °C [4]; and a Spanish series links higher thresholds and shorter critical stimulation times (CSTT) to more severe disease, though not significantly [5]. These sources fix the clinical structure we simulate. Neuro-fuzzy systems combine the interpretable rule form of fuzzy inference with data-driven tuning of neural networks [6].

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

III. SIMULATED COHORT

We found no public per-patient dataset of provocation temperatures (web searches for open urticaria data returned only published summary statistics and an image dataset), so we simulate sessions. Nothing here is clinical data. The generator (simulate_cohort.py) encodes:

- Patients. 40% high-threshold (TTV~N(21.5, 3.5)) and 60% low-threshold (N(14.5, 2.2)), clipped to 8–28 °C, giving mean TTV 17.4 °C with 55% of patients in the 12–18 °C range (the base paper reports more than 50%).
- Device. Eight sensors at 0, 20, …, 140 mm on a monotone non-linear profile Tmin→Tmax with exponent $\gamma \in [0.8, 1.5]$. Reading noise σ=0.15 °C, a per-session offset, and an ambient drift that grows beyond 27 °C.
- Physiology. The skin side is warmer than the applicator face by a factor that grows with poor contact and room temperature, so the true threshold sits above the reading at the blister edge. The edge position carries σ=2 mm reading error, with 10% hard-to-call outliers (σ=10 mm).
- Clinical context. Critical stimulation time (CSTT), asthma, itch score and a latent severity combining a higher TTV, a shorter CSTT, their interaction and asthma, with noise. About 36% of sessions are 'severe'.

Observed features per session (19): the eight sensor readings, Tmin and Tmax set-points, blister position Lx, the paper's interpolated estimate T_lin, ambient temperature and humidity, estimated contact quality, itch, exposure time, CSTT and asthma. Targets: Task A, severe vs mild (ROC-AUC, recall, precision, F1); Task B, TTV in °C (MAE, RMSE, share within ±0.5 and ±2 °C). The ground-truth TTV is defined at the skin, not at the applicator, which is what the paper's rule cannot see. This is a deliberate design choice and a limitation, discussed in Section VII.

IV. METHODS

A. Non-Hybrid Baselines

Paper rule: T_lin, the linear interpolation between adjacent sensors at the blister edge. For severity the score is T_lin and the decision threshold is 17 °C, the clinical group boundary. Logistic and linear regression use all 19 standardised features. Plain fuzzy FIS: a hand-written Sugeno system with fixed membership functions and fixed consequents (no learning). For TTV, nine rules on contact quality and ambient temperature add a correction to T_lin; for severity, six rules on T_lin and CSTT give a risk in [0, 1] (high TTV with short CSTT is the worst quadrant). MLP: a one- or two-hidden-layer perceptron (scikit-learn, early stopping).

B. Hybrid: ANFIS

We use a first-order Takagi–Sugeno ANFIS [6]. Rule k has Gaussian membership functions on five selected inputs (T_lin, CSTT, itch, contact and ambient for severity; T_lin, contact, ambient, itch and humidity for TTV) and a linear consequent over all 19 inputs. Rule centres are initialised by k-means. Each epoch runs the hybrid-learning pair: the consequents are solved in closed form by ridge least squares, then the membership centres and widths are updated by Adam gradient steps on the squared error, keeping the best epoch (80 epochs). For severity the target is the 0/1 label and the output is used as the ranking score. This is a hybrid in the soft-computing sense: fuzzy rule structure (interpretable) trained by neural-style learning.

GA-ANFIS keeps the same rule structure but searches the membership centres and widths with a genetic algorithm instead of gradient descent: population 24, 30 generations, tournament selection (size 3), blend crossover, Gaussian mutation (rate 0.15) and elitism of 2. Fitness is the validation MSE (25% hold-out of the training fold) after solving the consequents by ridge least squares; the best chromosome is refit on the whole training fold. This is a second hybrid (fuzzy + evolutionary) and tests whether a gradient-free global search finds better premises.

C. Protocol

Evaluation cohort: 2,000 sessions (seed 1), 5-fold stratified cross-validation repeated 3 times (15 folds). Hyper-parameters of every learned model (C for logistic regression; hidden size and L2 for the MLP; number of rules and ridge for ANFIS and GA-ANFIS) were chosen once by 3-fold CV on a separate tuning cohort (seed 7, 1,500 sessions), so the evaluation folds never influence tuning. Both ANFIS variants ended with 2 rules in both tasks. We report ROC-AUC, recall at the model's default threshold, recall at 90% specificity (a like-for-like operating point), specificity, precision, F1, and TTV error. ANFIS is compared with each other model by a paired Wilcoxon signed-rank test over the 15 folds. Robustness: models trained at nominal noise are tested on fresh cohorts (2,000 sessions each) with sensor, drift, contact and blister-edge noise multiplied by 1.5, 2 and 3.

V. RESULTS

A. Task A — Severity Classification

TABLE I. SEVERITY CLASSIFICATION, MEAN OVER 15 FOLDS (SD ≤ 0.044; R@90 = RECALL AT 90% SPECIFICITY)

Model	AUC	Recall	R@90	Spec.	Prec.	F1
Paper rule	0.877	0.784	0.677	0.813	0.700	0.739
Logistic	0.916	0.747	0.779	0.915	0.830	0.786
Plain FIS	0.905	0.697	0.744	0.924	0.836	0.760
MLP	0.896	0.823	0.732	0.822	0.721	0.768
ANFIS	**0.910**	**0.707**	**0.773**	**0.930**	**0.849**	**0.771**
GA-ANFIS	0.910	0.729	0.762	0.917	0.831	0.776

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

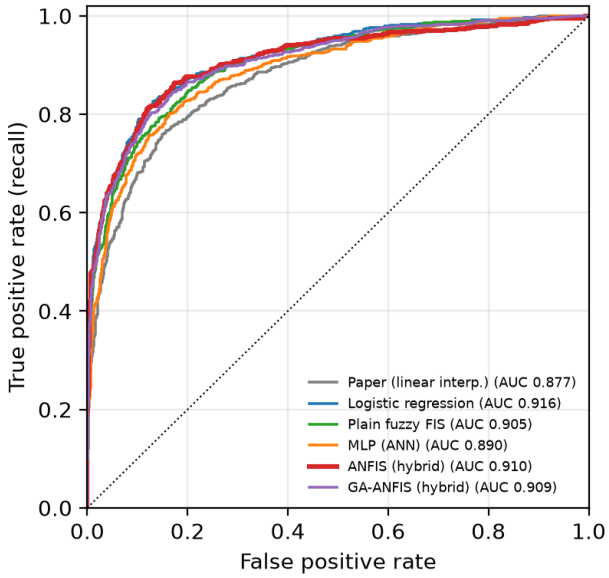


Fig. 1. Pooled out-of-fold ROC curves for severe vs mild (repeat 1).

ANFIS improves ROC-AUC over the paper's rule by +0.033 ($p<0.001$) and over the MLP by +0.014 ($p<0.001$). Against the plain fuzzy system the AUC gain is +0.005 and not significant ($p=0.12$); against logistic regression ANFIS is slightly worse (-0.006, $p=0.002$). At a matched 90% specificity, ANFIS recovers 0.773 of severe cases versus 0.677 for the paper's rule (+0.095, $p<0.001$), 0.732 for the MLP ($p=0.003$) and 0.744 for the plain FIS ($p=0.03$); the difference from logistic regression is not significant ($p=0.15$).

GA-ANFIS is statistically indistinguishable from gradient-trained ANFIS: AUC +0.0005 ($p=0.93$), recall at 90% specificity -0.011 ($p=0.22$). Like ANFIS it is slightly behind logistic regression on AUC (-0.006, $p<0.001$).

Recall at each model's default threshold needs care. The paper's rule (0.784) and the MLP (0.823) reach higher recall than ANFIS (0.707) only by flagging more patients as severe: their specificity is 0.813 and 0.822 against 0.930 for ANFIS, and ANFIS has the highest precision (0.849). The operating point differs, not the ranking quality, which is why we also report recall at fixed specificity.

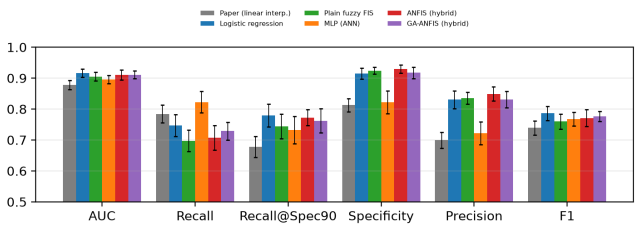


Fig. 2. Classification metrics by model (mean±SD over 15 folds).

B. Task B — Threshold Temperature

TABLE II. TTV ERROR (°C) AND SHARE WITHIN TOLERANCE, MEAN OVER 15 FOLDS (SD ≤ 0.053)

Model	MAE	RMSE	≤0.5 °C	≤2 °C
Paper rule	0.752	0.989	0.413	0.964
Linear reg.	0.469	0.681	0.657	0.981
Plain FIS	0.547	0.802	0.588	0.975
MLP	0.551	0.748	0.561	0.980
ANFIS	**0.465**	**0.668**	**0.666**	**0.981**
GA-ANFIS	0.465	0.673	0.667	0.981

ANFIS cuts MAE from 0.752 to 0.465 °C and raises the share of sessions within ±0.5 °C from 41% to 67%. It also beats the MLP (-0.086 °C, $p<0.001$) and the plain FIS (-0.082 °C, $p<0.001$). GA-ANFIS matches it to within 0.001 °C ($p=0.72$). ANFIS ties linear regression (-0.0037 °C, $p=0.52$): in this simulation the contact and ambient bias is nearly additive, so a linear model captures it. The paper's rule falls short of its ±0.5 °C claim here because the simulation includes contact, ambient and blister-edge effects that the claim presumably excludes; that is a property of our simulator, not evidence against the device.

C. Robustness to Noise

TABLE III. ROC-AUC (SEVERITY) AND TTV MAE (°C) UNDER NOISE MULTIPLIERS 1×, 2×, 3×

Model	AUC 1×	2×	3×	MAE 1×	2×	3×
Paper rule	0.877	0.879	0.876	0.74	0.97	1.33
Logistic	0.911	0.919	0.913	0.46	0.77	1.10
Plain FIS	0.900	0.907	0.904	0.54	0.99	1.39
MLP	0.900	0.904	0.899	0.51	0.82	1.14
ANFIS	**0.910**	**0.917**	**0.893**	**0.46**	**0.77**	**1.09**
GA-ANFIS	0.909	0.911	0.903	0.45	0.76	1.08

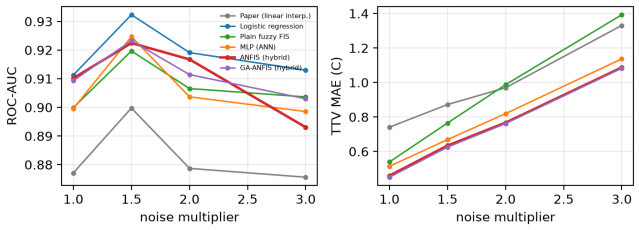


Fig. 3. Degradation with noise (each level is a fresh cohort, so AUC curves are not strictly monotone).

Every model degrades as noise grows, and the ordering mostly holds. At 3× noise ANFIS keeps an AUC of 0.893 versus 0.876 for the paper's rule and TTV MAE of 1.09 versus 1.33 °C. (Table III lists linear regression's TTV error beside logistic regression's AUC.) GA-ANFIS is similar (AUC 0.903, MAE 1.08 °C at 3×). For severity, logistic regression and the plain FIS are slightly more stable than ANFIS (0.913 and 0.904 at 3×).

VI. DISCUSSION

The consistent gain is over the paper's rule: any model that uses contact, ambient and clinical context beats a rule that

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

uses the reading alone. The hybrid-specific evidence is narrower. ANFIS outperforms both non-hybrid soft-computing models on TTV error and beats the MLP on AUC and recall at fixed specificity, at a fraction of the training time (about 0.03 s per fit against 0.06–0.45 s for the MLP). The GA variant shows the premise-search method is not the bottleneck: a gradient-free global search reaches the same accuracy but needs about five times longer per fit (0.14 s). ANFIS ties the hand-written FIS on AUC, so for severity the learned membership functions add little over expert rules here. Its practical advantage is interpretability: with only two rules, each rule's premise centres and linear consequent can be read directly, which a clinician cannot do with an MLP.

The hybrid does not beat logistic regression on severity AUC or linear regression on TTV error. A reader should not take from this paper that neuro-fuzzy dominates; the data are simulated with a mostly additive bias and a smooth severity link, which favours simple models. We expect the hybrid to matter more when real rule-like interactions or heavy-tailed imprecision are present, but that is untested.

VII. LIMITATIONS

- All data are simulated. The generator encodes our assumptions about skin-versus-applicator bias and the severity link; results measure how models handle that structure, not clinical performance. Calibration used only published summary statistics.
- The skin-side bias that the learned models correct is an assumption of the simulator; the base paper measures the applicator at the contact point, so real bias may be smaller.
- Ground-truth TTV is known by construction, which real sessions never provide. The severity label is a noisy function of latent variables, not a clinical outcome from patients.
- Heat-form urticaria, operator range errors and multi-visit data were not simulated.
- A real dataset from the authors of [1] or a clinic, with ethics approval, is needed before any claim of clinical benefit.

VIII. CONCLUSION

On a physics-calibrated simulation of TempPro sessions, an ANFIS hybrid raised severity ROC-AUC from 0.877 to 0.910, recall at 90% specificity from 0.677 to 0.773, and cut threshold MAE from 0.75 to 0.47 °C relative to the base paper's interpolation, outperforming a plain MLP and a hand-written fuzzy system on most metrics while staying an inspectable two-rule model. It did not outperform tuned logistic or linear regression. Next steps are validating on real provocation records, adding the heat form and operator-error sessions. A genetic-algorithm search of the membership functions matched, but did not improve on, gradient-trained ANFIS.

REFERENCES

[1] M. Danilychev, G. Mansurov, V. Kershner et al., “Determination of the temperature threshold value during provocative testing of patients with temperature urticaria,” in Proc. 2024 IEEE 9th All-Russian Microwave Conf. (RMC), pp. 277–280, doi:10.1109/RMC62880.2024.10846820.

[2] Courage + Khazaka electronic GmbH, “TempTest.” [Online]. Available: www.courage-khazaka.de

[3] M. S. Lebedkina et al., “Analysis of patients with cold urticaria in the context of typical and atypical forms,” Russ. J. Allergy, vol. 21, no. 1, pp. 98–111, 2024, doi:10.36691/RJA16910.

[4] K. Kulthanan, P. Tuchinda, L. Chularojanamontri and R. Kiratiwongwan, “Cold urticaria: clinical features and natural course in a tropical country,” Allergy Asthma Immunol. Res., vol. 11, no. 4, pp. 538–547, 2019, doi:10.4168/aair.2019.11.4.538.

[5] M. E. Martinez-Escala, L. Curto-Barredo, L. Carnero, R. M. Pujol and A. M. Giménez-Arnau, “Temperature thresholds in assessment of the clinical course of acquired cold contact urticaria: a prospective observational one-year study,” Acta Derm. Venereol., vol. 95, no. 3, 2015, doi:10.2340/00015555-1918.

[6] J.-S. R. Jang, “ANFIS: adaptive-network-based fuzzy inference system,” IEEE Trans. Syst., Man, Cybern., vol. 23, no. 3, pp. 665–685, 1993.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)