Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Computable Boundary of TCM Pattern Differentiation: An External Empirical Study Based on a Four-Axis Quantitative Framework

Guanfeng Yang

jjaarr@163.com

Foshan Yesheng Health Consulting Co., Ltd. https://orcid.org/0009-0009-4933-390X

Research Article

Keywords: pattern differentiation, quantitative model, ill-posed problem, conditional entropy, falsifiability, TCM therapeutic methods

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11877524/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Objective: To reconstruct pattern differentiation (bianzheng lunzhi) from an experiential narrative into a falsifiable decision model between a state side and a formula side, and to measure the boundary of the model's capability on real external clinical cases.

Methods: Using the four quantitative axes of the Same-Coordinate Food-Nourishment Theory (SCFNT) — Cold-Heat (H), Ascent-Descent (Q), Deficiency-Excess (D), and Meridian Tropism — to characterize both states and formulas, we build profiles of 246 classical formulas and 158 herbs. On 128 external clinical cases we evaluate against a random baseline, paired bootstrap, Monte Carlo tests, and inter-rater agreement (Cohen's kappa), and use conditional entropy to measure the non-injectivity of the formula layer.

Results: Exact formula selection showed no significant difference from the random baseline (overlap 0.103 vs 0.096, ratio 1.07, confidence interval containing 0, name hit 1/128). The treatment-method layer was marginally significantly better than random (sign test p=0.0479). For official therapeutic categories, the external 69 cases yielded 34.8% vs 17.4% (2.00-fold, p=0.0003), while the internal 58 cases yielded 39.7% vs 19.5% (2.03-fold, p=0.0003). Cross-model agreement among three general-purpose AI models from different vendors gave kappa = 0.774, 0.846, 0.842. The conditional entropy of the exact formula given the direction was 2.404 bits.

Conclusion: Under the conditions of a global domain, text-only input, unique reconstruction, and no feedback, exact prescribing is a mathematically ill-posed task, whereas the direction and category layers can be reliably rule-based. The model's proper role is a reliable co-pilot for direction, not a prescription machine.

1 Introduction

Two seemingly opposed positions have long existed within Chinese medicine. The first may be called the formula-pattern correspondence school, which holds that stable correspondences exist between specific patterns and specific formulas, emphasizing reproducible rules. The second may be called the "medicine is an art (yi)" school, which stresses that clinical practice is infinitely variable; as the History of the Later Han, Biography of Guo Yu, puts it, "medicine is an art: it can be understood but cannot be fully put into words." This two-thousand-year debate has never had an empirical resolution on a single dataset.

This paper does not presuppose either side to be correct. Instead, it recasts the debate into testable questions: if states and formulas can be characterized by the same set of quantitative coordinates, to what degree can the mappings from observable symptoms to direction and then to an exact prescription each be determined? The contribution is not to claim a model is the strongest, but to use real external data to draw the dividing line between the determinable and the indeterminable, and to provide mathematical and information-theoretic foundations for that line. Modern Chinese medicine information systems commonly overstate their capabilities, and third-party benchmarks show that their pattern-

differentiation subtasks still lag behind human performance; against this background, clarifying the computable boundary is more valuable than reporting yet another high score.

2 Theoretical Framework: The SCFNT Same-Coordinate System

2.1 The Four Quantitative Axes

The state side and the formula side share the same coordinates so the two can be compared within one space. The four axes are: the Cold-Heat axis H, characterizing the direction and intensity of yin-yang cold and heat; the Ascent-Descent axis Q, characterizing the tilt of qi movement; the Deficiency-Excess axis D, characterizing the relative strength of healthy qi and pathogenic factors; and the Meridian Tropism axis, characterizing the target via the twelve meridians. A formula profile is obtained by dose-weighted summation of the qi, flavor, and meridian tropism of its constituent herbs, while the state side obtains its axis values by encoding symptoms.

2.2 The Baohe Degree and Correcting Bias with Bias

Let the bias intensity of a herb or formula be its weighted cold-heat amplitude. The Baohe degree C characterizes the intensity of internal cold-heat counterbalance within an entity; C is a non-negative scalar in [0,1]. The core relation of correcting bias with bias is written K = -S: the correction vector K points opposite to the state bias vector S — heat what is cold and cool what is hot, the formula using its own bias to correct the state's bias. This relation recasts the therapeutic principle of the Suwen, Treatise on the Most Essential Truths [11], into a computable symbolic rule. It also makes same-pattern-different-formula analyzable: the direction is determined by the signs of the axis values, whereas the specific point (which formula) is not unique.

3 Data and Methods

3.1 Formula Library and Formula Profiles

The formula side is based on 246 classical formulas and 158 herbs. Each herb is annotated with qi, flavor, meridian tropism, and a bias level. Each formula is dose-weighted over its constituent herbs to compute its profile on the four axes, and a composition-level effect vector characterizes its direction. The distance between a state and a formula combines axis differences, meridian coverage, and the angle between effect vectors, serving as the ranking basis.

3.2 Encoding Engine and External Clinical Corpus

The encoding engine maps free-text symptoms onto the axes: text cleaning first removes boilerplate, dosage, and preparation instructions; a negation engine then handles negations and their scope (none, not, not yet); finally a frozen synonym table expands coverage. The external corpus comes from four public case sources, totaling 128 cases, all outside the data used for rule writing and parameter selection, to test transferability.

3.3 Evaluation Criteria

To avoid exaggerating capability with a single hit rate, this study uses four criteria: a random baseline constructed by in-library random sampling; paired bootstrap to estimate the confidence interval of the difference; a Monte Carlo permutation test to judge significance; and Cohen's kappa to measure agreement among independent judges. Conditional entropy measures the remaining uncertainty of the exact formula once the direction is known. All key rules and data were frozen before validation (v1.0), and thresholds were not revised after the results emerged.

4 Results

4.1 Internal Two-Way Validation and the Encoding Bottleneck

In internal two-way validation across 58 patterns, multiple rounds of encoding audits showed that errors were almost entirely concentrated at the symptom-to-axis-value encoding end rather than the ranking algorithm; experiments adjusting amplitudes and weights failed to improve the hit rate, showing that more precise does not mean more correct, with only the Deficiency-Excess axis write-back holding. This indicates the bottleneck is first a problem of input information and encoding.

4.2 External Exact Formula Selection: No Significant Difference from Random

Across the 128 external cases, exact formula selection overlap was 0.103 versus a random baseline of 0.096, ratio 1.07, paired bootstrap CI [-0.014, 0.029] containing 0, with an exact name hit of only 1/128. The strong version of formula-pattern correspondence — a symptom text uniquely determining a specific formula — is falsified here; exact formula selection is not transferable, and this study no longer claims prescribing capability. In the single name-hit case (case 23, Huanglian Ejiao Tang), the name matched exactly but the herb overlap was 0.833 rather than 1: the actual prescription in that case added Shengdihuang to the five herbs of Huanglian Ejiao Tang, giving six herbs, whereas the in-library classical formula has five; the intersection is 5 and the union 6, so the Jaccard is 5/6. This reflects that real clinical prescriptions commonly involve modifications, rather than a computational contradiction. A question-bank corroboration also supports this: on licensing-examination questions whose options are similar formulas, option-level performance was close to random, mutually corroborating the 128 cases.

4.3 The Treatment-Method Layer: Marginally Significantly Better than Random

When the target was moved up from a specific formula to the treatment direction, results improved markedly. The pooled treatment-method hit across two sources exceeded random by an absolute 0.117, CI [0.040, 0.197] excluding 0, sign test p = 0.0479, which is marginally significant and should not be

overstated; in the first source, the 39 cases were 1.56 times random, with a CI excluding 0. The weak version of formula-pattern correspondence at the direction layer is supported.

4.4 Official Therapeutic Categories: About Twice Random

Under the official therapeutic categories of GB/T 16751.3 [10], the external 69 cases gave a category hit of 34.8% versus random 17.4%, ratio 2.00, p = 0.0003 — the most robust external positive result of this study. As an internal benchmark, the original 58 textbook patterns gave 39.7% vs 19.5%, ratio 2.03, p = 0.0003, CI [0.083, 0.321]. Note the different scopes: the 58 cases are in-library textbook pattern-formula pairs, while the 69 cases are external. The treatment-method layer lands on finite grid points, consistent with the finite classification tradition of Cheng Guopeng's Yixue Xinwu (Medical Understanding), which grouped methods into the eight categories of sweating, emesis, downward-draining, harmonizing, interior-warming, heat-clearing, resolving, and supplementing [13].

4.5 Cross-Model Blind Review (Three AI Models)

Under the condition that a second human TCM physician could not be obtained, three general-purpose AI models from different vendors, independent of one another (denoted AI-A, AI-B, AI-C), blindly classified the 72 treatment categories, serving as a cross-model agreement check rather than human expert reliability: exact agreement occurred in 39 items (54.2%), and at least two agreed in 63 (87.5%); pairwise Cohen's kappa [5] values were 0.774, 0.846, 0.842, indicating high but not identical agreement. The three AI models' rerun category hits were 1.64–1.93 times random, while the present rule was 1.94 times; the conclusion does not flip.

4.6 Coverage Diagnosis and Three-Layer Modification

After cleaning, there were 2935 real clinical phrases, with only 30.6% hits and 69.4% misses; among the 1142 unmatched atomic phrases, 967 occurred only once, showing an extreme long tail. After the three modifications of text cleaning, the negation engine, and synonym expansion, the category hit was finally 37.5%, 1.87 times the random 20.1%, p = 0.0003; the preset success lines of an 8-percentage-point category gain and a 15-percentage-point coverage gain were both missed, confirming the direction-layer ceiling at about 37–38%.

Evaluation layer	Our model	Random baseline	Conclusion
Exact formula (128 cases)	0.103	0.096	no significant difference
Treatment methods (pooled 58)	—	—	p = 0.0479, marginally significant
Official category (internal 58)	39.7%	19.5%	2.03-fold, p = 0.0003
Official category (external 69)	34.8%	17.4%	2.00-fold, p = 0.0003
Three-model cross-model kappa	0.77–0.85	—	high agreement
Conditional entropy H(Y|Z)	2.404 bits	—	formula non-injectivity

5 Formalization of the Ill-Posedness of the Formula Layer

The empirical gap above is not mere engineering deficiency but corresponds to the mathematical structure of the task itself, captured by five propositions.

Proposition 1

(non-injective inverse problem). The forward mapping state/prescription -> symptoms is many-to-one; the inverse mapping fails the uniqueness requirement of Hadamard well-posedness [1] and is ill-posed, so only regularization [2] yields a stable approximation.

Proposition 2

(Bayes irreducible error). When the conditional distribution P(Y|X) has several plausible values and its conditional entropy is positive, the Bayes-optimal predictor still has an irreducible error [9], independent of the specific model.

Proposition 3

(data processing inequality). Variables determining the prescription — constitution, season, locality, disease course, and response to medication — are absent from the symptom text, and no post-processing can create information not present in the input [3]; the only way out is to expand the input.

Proposition 4

(finite-coverage upper bound). The description space grows with combinations of vocabulary and sentence length, while a finite rule table's coverage of an infinite combinatorial space is structurally limited; events with frequency below about 1/N cannot be reliably learned from N samples [6][7], and the 967 single-occurrence terms are its fingerprint [8].

Proposition 5

(construct-validity ceiling). When any single judge is used as the reference and that standard is given by subjective judgment, the model's observable accuracy cannot exceed the agreement among judges; the three-model kappa is only 0.77–0.85, and judge disagreement constitutes the inherent noise of the reference answer.

Empirically, the 128 exact formulas have 61 values and marginal entropy 5.464 bits, while the conditional entropy H(Y|Z) given the direction category is 2.404 bits, with perplexity 2^2.404 about 5.3, roughly equivalent to 5.3 candidate formulas on average. This empirical distribution contains both multiple valid formulas and individual physician habits, so it is used here as upper-bound evidence for non-injectivity rather than a precise causal measure. This provides direct quantitative evidence that the formula layer cannot be uniquely reconstructed.

6 Two-Region Picture and Falsifiable Predictions

Knowledge layer	Region A: global, text-only, unique reconstruction, no feedback	Region B: specialty, full information, distribution output, closed-loop feedback
Principle layer	derivable	derivable
Method layer	partially derivable (~ 1.9x random, marginally significant)	can advance to the labeling-noise limit
Formula layer	provably impossible (conditional entropy > 0)	boundedly attainable, continuously improvable

Two falsifiable predictions follow. First, the method-layer hit rate should rise with input completeness (adding tongue, pulse, and the four examinations). Second, under text-only, single-shot, no-feedback conditions, no model change should make exact formula hits significantly above random [4].

7 Discussion

On the two-thousand-year debate, this study gives an empirical verdict: the strong version of formula-pattern correspondence (symptoms uniquely determining a specific formula) is falsified by the 128 cases, while the weak version (symptoms determining direction and category) receives stable support at about twice random. This is consistent with the traditional division that principles and methods can be rule-based while formulas cannot be exhausted by finite rules; the Lingshu Jing (Spiritual Pivot) says the coarse practitioner guards the form while the superior practitioner guards the spirit [12], pointing precisely to the hard-to-rule part of clinical choice at the formula layer. It also corroborates the fact that all deployed systems use specialty domains, humans in the loop, and structured input.

This study has clear limitations. First, the sample size is limited and the external corpus from public cases carries selection bias. Second, the two competing hypotheses — the domain is inherently hard and the current model is too crude — predict the same phenomenon on the present data, and this study

cannot exclude the latter; a four-examination full-information experiment is needed to distinguish them. Third, the impossibility claim is strictly bound to four qualifiers — global, text-only, unique reconstruction, no feedback — and missing any one constitutes exaggeration; what this paper proves is the impossibility of guaranteed correctness, not the impossibility of being useful, and a narrow specialty domain with closed-loop feedback can keep improving. Fourth, this study does not invoke Godel, the halting problem, chaos, or quantum theory, which are irrelevant to a finite medical domain, in order to avoid forced analogies.

8 Conclusion

Under the same quantitative coordinates, the state side and the formula side can be rigorously compared. The data show that the direction and category layers can be reliably rule-based, while exact prescribing under global, text-only, unique-reconstruction, no-feedback conditions is a mathematically ill-posed task. The model's proper role is a reliable direction co-pilot, not a prescription machine; future work should reduce conditional entropy by expanding input and treatment feedback rather than changing algorithms on the same input.

Declarations

Data Availability

The per-case evaluation data, formula and herb profiles, formula composition, and entropy measures supporting the conclusions are provided as a single multi-sheet Excel file accompanying the paper. It contains the predicted and true prescriptions for the 128 external cases, per-case category judgments for the 69 external cases, the four-axis profiles and herb composition of the 246 formulas (from which herb-set Jaccard overlap can be independently recomputed), the quantitative features of the 158 herbs, the statistics for each evaluation layer, and blind-review aggregates, with a README sheet explaining fields, sources, and reproduction. From the per-case data a reader can independently recompute the 128-case and 69-case hit rates, random baselines, paired bootstrap confidence intervals, information entropies, and blind-review summary metrics. The per-case raw blind labels, phrase-level coverage working files, and original scraped web text are not redistributed because they were not fully retained or are subject to third-party copyright, and can be provided on request together with the source URLs and scripts. The data are offered under CC BY 4.0 and will be archived together with the paper on the preprint platform; the persistent identifier will be given in this statement after publication.

References

1. Hadamard J (1902) Sur les problemes aux derivees partielles et leur signification physique. Princet Univ Bull, : 49–52
2. Tikhonov AN, Arsenin VY (1977) Solutions of Ill-Posed Problems. Winston, Washington

3. Cover TM, Thomas JA (2006) Elements of Information Theory, 2nd edn. Wiley, Hoboken
4. Wolpert DH (1996) The lack of a priori distinctions between learning algorithms. Neural Comput 8(7):1341–1390
5. Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20(1):37–46
6. Shalev-Shwartz S, Ben-David S (2014) Understanding Machine Learning. Cambridge University Press, Cambridge
7. Vapnik VN (1998) Statistical Learning Theory. Wiley, New York
8. Zipf GK (1949) Human Behavior and the Principle of Least Effort. Addison-Wesley, Cambridge, MA
9. Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning, 2nd edn. Springer, New York
10. GB/T 16751 (2023) 3–2023 Clinical diagnosis and treatment terminology of traditional Chinese medicine — Part 3: Therapeutic methods. Standards Press of China, Beijing
11. Huangdi Neijing Suwen (Yellow Emperor's Inner Canon: Basic Questions). Beijing: People's Medical Publishing House, standard edition
12. Lingshu Jing (Spiritual Pivot). Beijing: People's Medical Publishing House, standard edition
13. Cheng Guopeng Yixue Xinwu (Medical Understanding). Beijing: People's Medical Publishing House, standard edition

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SCFNTdataset.xlsx
- Appendix.docx