Modern Literature Research on Children Sweat Syndrome based on TF-IDF

Hongli Luo, Jiangwei Yu*, Zhihui Du, Shiyu Tan

Guizhou University of Traditional Chinese Medicine, Guiyang 550025, China

*Corresponding author: Jiangwei Yu, 691105gy@163.com

Abstract: *Objective:* To explore the characteristics of intelligent syndrome diagnosis and intelligent syndrome differentiation of pediatric sweat syndrome by retrospectively studying the traditional Chinese medicine (TCM)-based diagnosis and rule of drug use on pediatric sweat syndrome in China Academic Journal Full-text Database (CNKI), Wanfang Database (WF) and VIP database. *Methods:* Modern literature on pediatric sweat syndrome was collected with Endnote software, and the collected data were input into The Microsoft Excel 2010 table. TF-IDF relative entropy quantitative analysis was used to extract the intelligent diagnosis and intelligent medication rule of pediatric sweat syndrome in modern literature. *Results:* TF-IDF was used to obtain the syndrome's type-symptom fuzzy matrix and syndrome's type-TCM fuzzy matrix of *qi-yin* deficiency and *qi*-deficiency. *Conclusion:* The intelligent syndrome diagnosis and intelligent syndrome differentiation law of modern data mining in the treatment of children sweat syndrome accord with the basic theory of TCM, and can provide clinical reference to a certain extent.

Keywords: Children sweat syndrome; TF-IDF; Medication rule; Data mining

1. Introdcution

Pediatric sweating syndrome refers to the syndrome of excessive sweating of the whole body or some parts of a child when he or she is in the quiet state in the daily living environment; this syndrome is mostly seen in children aged 2 to 6 years [1]. Long-term sweating treatment is improper or not timely, easy to consume *yang qi* of body fluid, and affect children's growth as well as physical and mental health. At present, anticholinergic therapy, antiperspirants, ion electroosmosis therapy and even surgery are mainly used in modern medical treatment, but these methods have severe side effects and low efficiency, and are limited in pediatric clinical application [2]. Chinese medicine has remarkable curative effect in the treatment of children's sweat syndrome, but Chinese medicine pays attention to syndrome differentiation and treatment, so it has certain clinical guiding significance to explore the law of dialectical diagnosis of children's syndrome and the rule of medication for the prevention and treatment of children's sweat syndrome. Based on this, we conducted data mining on the syndrome manifestations and rules of medication on pediatric sweat syndrome, in order to provide some reference for the study of pediatric sweat syndrome.

2. Data and methods
2.1. Data and retrieval
Through the subject-based search method, data in Chinese literature were extracted since database construction until September 1, 2021. The search terms include "card" or "night sweats" or "khan," "sweat"

and/or "children," and "children." The literature must contain the full text pertaining to "syndrome" and "Chinese medicine." A total of 4,689 items were retrieved from China Academic Journal Full-text Database (CNKI), Wanfang (WF) Database and VIP Database. A total of 2,792 items were obtained by deleting 1,897 duplicated items using Endnote. After reading the title, abstract and full text, the data needed for this paper was deleted and selected. Finally, 69 literatures were deleted and selected, and 1,271 pieces of clinical data were obtained.

2.2. Inclusion and exclusion criteria
2.2.1. Inclusion criteria
(1) Literature about clinical studies on children sweating syndrome diagnosis, excluding vitamin D deficiency-related rickets and tuberculosis, rheumatism, hyperthyroidism and other diseases caused by sweating, heat, warm clothing, eating, fear, sports and other physiological sweating;
(2) Literature on the treatment of sweating syndrome in children with traditional Chinese medicine (TCM); and
(3) Complete literatures of syndrome manifestations and TCM composition.

2.2.2. Exclusion criteria
(1) Theoretical studies, animal experiments, pharmacokinetics, review literature and literatures with similar or duplicate contents;
(2) Literature of incomplete syndromes or TCM composition;
(3) Master's and doctoral dissertation;
(4) Literature that cannot be obtained; and
(5) Literature on acupuncture, plaster, pediatric massage, western medicine and western medicine treatment and other literature.

2.3. Creating a Database
A unified "modern literature survey form of pediatric sweat syndrome" was developed, which includes 9 items, such as literature title, author, patient gender, age, sample size, clinical manifestation, syndrome differentiation of TCM, composition of TCM, etc., and was input into Excel database.

2.4. Data normalization
2.4.1. Data normalization
Reference standards of this study are as follows: Zhu Wenfeng's Syndrome Element Differentiation in June 2008, the first edition, "13th Five-Year Plan" textbook "Diagnostics of Traditional Chinese Medicine," "Chinese Pharmacopoeia," "13th Five-Year Plan" textbook "Traditional Chinese Medicine."

"Hand, foot and heart heat" was unified as "hand, foot and heart burn," and "poor sleep, easy to wake up, like turning over" was unified as "sleep is not real." The processing of drugs as well as raw and used attributes were not calculated to retain their essence. For example, "honey, stir-fried, raw, calcined, fried" and so on were deleted; "honey astragalus" was written as "astragalus" and "*zhi baibu*" was written as "*baibu*."

2.5. Data mining
All standardized data were subject to frequency analysis, TD-IDF analysis and complex network analysis.

3. Results

Through retrieval, deletion and selection, 69 literatures were selected and a total of 1,271 medical records were obtained, among which the age of subjects ranged from 0 to 18 years.

3.1. TF-IDF relative entropy quantization results

TF-IDF is a text mining method whose formula is $W_{ij}=tf_{ij} \times idf_j=tf_{ij} \times \log(N/n_j)$, wherein tf_{ij} is term frequency (TF), which refers to the frequency of occurrence of feature. Term t_j in document d_i. idf_j is inverse document frequency (IDF); N represents the total number of documents, and nj represents the number of documents of the feature item Tj in the document [3]. IDF can be understood as the cross entropy (that is, relative entropy) of the probability distribution of keywords under a specific condition.

TF-IDF can be applied to analyze the rules of TCM drug use based on syndrome differentiation, where i represents a syndrome type and j represents a TCM characteristic. w_{ij} refers to the relative entropy quantification of TCM j in syndrome type i, and TF_{ij} refers to the probability or frequency of TCM j appearing in syndrome type i. Ω represents the whole or specific disease domain, n_j represents the frequency of occurrence of TCM j in the whole or specific disease domain [4]. Similarly, we can obtain the relative entropy quantification result of TCM-symptom. Similarly, TF-IDF can be applied in TCM syndrome differentiation, where i represents a certain syndrome type and j represents a certain symptom feature. w_{ij} refers to the relative entropy quantification of symptom j in syndrome type i, and TF_{ij} refers to the probability or frequency of symptom J appearing in syndrome type I. ω represents the whole or specific disease region, and nj represents the frequency of symptom J occurring in the whole or specific disease region [5].

From 1,271 medical record, there are 152 syndrome manifestations (symptoms) and 135 TCM. Total frequency of TCM is 9,643. This analysis selected the first two syndrome name frequency of *qi* and *yin* deficiency and *qi* deficiency. TF-IDF calculation formula was used to calculate with base 2.

The relative entropy of each TCM and the relative entropy of symptoms are shown in **Figure 1**, **Figure 2**, and **Figure 3**. The larger TF-IDF value of a TCM is, the more it has attribute category significance for the corresponding syndrome type. According to the results, the TF-IDF size of the TCM corresponding to each syndrome type is basically consistent with that of TCM belonging to this syndrome type. Similarly, the larger the TF-IDF value of a symptom is, the more diagnostic significance it has for the corresponding syndrome type. TF-IDF of the symptoms corresponding to each syndrome type is roughly consistent with the TCM diagnosis of the syndrome type. Similarly, for the corresponding TCM, the larger the TF-IDF value of a (syndrome manifestation) symptom is, the more therapeutic significance the TCM has for the (syndrome manifestation) symptom. As can be seen from **Figure 3**, the TF-IDF size of the corresponding TCM symptom is roughly consistent with the efficacy and indications of the medicine.

4. Discussion

"Plain ask · *Yin* and *Yang* on another" recorded that "*yang* combines *with* yin to become sweat." Sweat comes from the evaporation of body fluid by *yang qi*. "Simple Question · *Yin* and *Yang* Should be like the Fifth" recorded that "when *yang* is outside, so is *yin*; when *yin* is inside, so is *yang*." *Yin* and *yang* coordination, nutrition harmony, body fluid solidification, and conversely, if *yin* and *yang* viscera are affected by *qi* and blood disorders, nutrition will become discordant [1]. Since ancient times, TCM has great advantages in the treatment of children sweat syndrome, and the curative effect is remarkable, and the syndrome differentiation and syndrome differentiation of TCM medication is more complex; therefore, the study of syndrome differentiation and syndrome differentiation of children sweat syndrome medication rule has a certain significance to guide clinical and academic discussion.

Deficiency of qi and Yin - Relative entropy of Traditional Chinese medicine

traditional Chinese medicine	total frequency	Frequency of this type	TF	IDF	TF*IDF
Ophiopogon japonicus	337	249	0.626	1.915	1.198
Radix pseudostellariae	280	215	0.54	2.182	1.179
Schisandra	576	366	0.92	1.142	1.05
ginseng	132	91	0.229	3.267	0.747
Herba agrimoniae	68	68	0.171	4.224	0.722
Paeonia lactiflora	167	96	0.241	2.928	0.706
keel	380	157	0.394	1.742	0.687
Oyster	730	307	0.771	0.8	0.617
Cassia twig	102	65	0.163	3.639	0.594
shriveled dried peach	86	60	0.151	3.885	0.586
Ephedra root	262	92	0.231	2.278	0.527
windproof	703	233	0.585	0.854	0.5
Astragalus	766	234	0.588	0.731	0.43
Turtle shell	30	30	0.075	5.405	0.407
colla	30	30	0.075	5.405	0.407
Tortoise plastron	31	30	0.075	5.358	0.404
Floating wheat	317	70	0.176	2.003	0.352
Dried Rehmannia	62	31	0.078	4.358	0.339
Codonopsis	270	48	0.121	2.235	0.27
Atractylodes	823	169	0.424623	0.627	0.266239
plum	60	22	0.055276	4.40486	0.243485
Glutinous rice root	80	22	0.055276	3.98982	0.220543
jujube	93	23	0.057789	3.77259	0.218014
Licorice	440	31	0.077889	1.53039	0.119201

Figure 1. Deficiency of *qi* and *yin* relative entropy of TCM

Qi deficiency defends the surface not solid - Chinese medicine relative entropy

Frequency of this type	total frequency	Frequency of this type	TF	IDF	TF*IDF
windproof	703	232	0.996	0.854	0.851
Astragalus	766	233	1	0.731	0.731
Codonopsis	270	73	0.313	2.235	0.7
Oyster	730	198	0.85	0.8	0.68
Schisandra	576	132	0.567	1.142	0.647
Ephedra root	262	64	0.275	2.278	0.626
Atractylodes	823	232	0.996	0.627	0.624
Floating wheat	317	67	0.288	2.003	0.576
Poria	336	39	0.167	1.919	0.321
Licorice	440	40	0.172	1.53	0.263
lotus seeds	1	1	0.004	10.31	0.044
Bellflower	3	1	0.004	8.727	0.037
white lentils	14	1	0.004	6.504	0.028
Amomum	19	1	0.004	6.064	0.026
Scrophularia	33	1	0.004	5.267	0.023
Coix Seed	34	1	0.004	5.224	0.022
ginseng	132	1	0.004	3.267	0.014
Ophiopogon japonicus	337	1	0.004	1.915	0.008
keel	380	1	0.004	1.742	0.007

Figure 2. *Qi* deficiency defends the non-solid surface-Chinese medicine relative entropy

Qi and Yin deficiency-symptom Relative entropy

symptom	The total frequency	Frequency of this type	TF	IDF	TF*IDF
fear of cold	288	288	0.724	2.142	1.55
Thirsty	421	348	0.874	1.594	1.394
thin and fast pulse	423	289	0.726	1.587	1.153
exhausted	358	229	0.575	1.828	1.052
Red tongue	596	349	0.877	1.093	0.958
Thin and white tongue coating	524	288	0.724	1.278	0.925
thin and fast or weak pulse	82	82	0.206	3.954	0.815
Little or no coating on the tongue	88	82	0.206	3.852	0.794
Sweating more after activity	679	336	0.844	0.904	0.764
Cold limbs	60	60	0.151	4.405	0.664
The tongue dry	60	60	0.151	4.405	0.664
Dry lips	60	60	0.151	4.405	0.664
fatigue	606	228	0.573	1.069	0.612
insomnia	92	60	0.151	3.788	0.571
The mood irritability	115	60	0.151	3.466	0.523
Loss of appetite	119	60	0.151	3.417	0.515
Hand, foot and heart is hot	263	60	0.151	2.273	0.343
spontaneous sweating	1033	397	0.997	0.299	0.298
night sweats	1048	337	0.847	0.278	0.236
The tongue coating is thin	129	26	0.065	3.301	0.216
The whole body sweat	200	26	0.065	2.668	0.174
easy to catch cold	156	22	0.055	3.026	0.167
Pale tongue	526	48	0.121	1.273	0.154
thin and weak pulse	318	26	0.065	1.999	0.131
complexion lack of luster	352	22	0.055	1.852	0.102

Qi deficiency cannot guarding the surface-symptom Relative entropy

symptom	The total frequency	Frequency of this type	TF	IDF	TF*IDF
Cold limbs	134	131	0.562	3.246	1.825
Shortness of breath	139	130	0.558	3.193	1.781
The whole body sweat	200	130	0.558	2.668	1.489
Poor health	216	130	0.558	2.557	1.427
complexion lack of luster	352	170	0.73	1.852	1.351
The shoulder sweating	62	62	0.266	4.358	1.16
Back sweat	126	65	0.279	3.334	0.93
Sweating more after activity	679	233	1	0.904	0.904
Don't want to eat	427	130	0.558	1.574	0.878
sweating on the head	352	102	0.438	1.852	0.811
fatigue	606	170	0.73	1.069	0.78
Sweaty neck and back	58	35	0.15	4.454	0.669
Chest sweating	79	38	0.163	4.008	0.654
Easy to sweat	106	39	0.167	3.584	0.6
Easy to sweat	156	39	0.167	3.026	0.507
easy to catch cold	358	40	0.172	1.828	0.314
spontaneous sweating	1033	193	0.828	0.299	0.248
night sweats	1048	194	0.833	0.278	0.232
complexion yellow	5	1	0.004	7.99	0.034
Sweaty shoulder and back	86	2	0.009	3.885	0.033
thin body	124	2	0.009	3.358	0.029
Poor sleep	14	1	0.004	6.504	0.028
deep and weak pulse	51	1	0.004	4.639	0.02
loose stool	66	1	0.004	4.267	0.018
Red tongue	596	1	0.004	1.093	0.005

Figure 3. *Qi* and *yin* deficiency-symptom relative entropy and *qi* deficiency cannot guard the surface-relative entropy

Text classification is an indispensable technology in natural language processing, data mining, data search and other fields. As a method of feature extraction in text classification, TF-IDF has a broad application field and can be applied to the field of TCM, such as the diagnosis of TCM syndromes, the classification of TCM effects, and the classification of TCM indications. TF-IDF embodies the idea that the more frequently a word appears in a particular document, the stronger it is in distinguishing the content attributes of the document (TF); the more widely a word appears in a document, the lower its IDF. This classification idea is consistent with the syndrome diagnosis of TCM, that is, the higher the frequency of a symptom in a particular syndrome, the stronger it is in distinguishing the syndrome (TF); the wider the scope of a symptom in all syndromes, the lower its ability to distinguish a syndrome (IDF) [1]. Similarly, the classification of TCM's indications is consistent with the classification of TCM's effects. This is the theoretical basis for quantifying TCM syndrome, syndrome differentiation and indication of TCM with TF-IDF relative entropy.

As can be seen from **Figure 1** and **Figure 2**, the treatment of *qi* and *yin* deficiency in the sweating syndrome of children mainly includes Radix Ophiopogonis, radix pseudostellariae, schisandra fruit, ginseng, stormgrass, paeonia lactiflora, keel, oyster, cassia twig and dried peach. This is consistent with the basic theory of TCM. Sweat for heart fluid, excessive sweating can damage *yin*, lung body for gas, lung *qi yang*, and gas with *jin* off. Sweating too much will consume lung gas. *Maidong gan*, cold, moist, stomach,

lung, heart, can nourish *yin sheng jin run zao*. Schisandrae collects lung *qi*, generates *jin* and collects sweat, both of which play the effect of generating *jin* and collecting sweat. *Taizi* ginseng is beneficial to *qi* and spleen, *shengjin runfei*, can treat *qi yin* deficiency, self-sweating, etc. Oyster and keel are often compatible with each other, and each must be used to suppress latent and astringent substances. They can collect fluid and exudate fluid, nourish *yin* and absorb *yang*. The combination of oyster and keel can enhance the functions of solidifying surface to collect sweat and calm nerves. In the treatment of *qi* deficiency and non-solid appearance, there are mainly parsnip, astragalus membranaceus, Codonopsis, oyster, schisandra fruit, ephedra root, atractylodes, floating wheat, Poria cocos, licorice root and lotus seed. Various medicines are commonly used to nourish *qi* and solidify the surface, and antiperspirate [6] can treat the deficiency of the surface and self-perspiration.

As shown in **Figure 3**, in the pediatric khan card, *qi* and *yin* deficiency symptoms mainly include chills, thirsty, pulse count, god weakness, red tongue, moss thin white, pulse thin or count, moss or spend less peeling, especially after sweat out activities, cold limbs, tongue dry, dry lips, fatigue, insomnia, upset, loss of appetite, hand, foot and heart burn, etc., the two virtual diagnosis evidence are consistent with *qi* and *yin*. The symptoms of *qi* deficiency and unstable surface mainly include cold limbs, shortness of breath, sweating all over the body, weak body, little complexion, sweating on head, neck, back shoulder and chest, especially after strenuous activity, lethargy, fatigue, easy sweating, easy cold, fatigue, spontaneous sweating, night sweating, yellow complexion, emaciation, irregular sleep, weak pulse, loose stools, red tongue, etc. This is basically consistent with the syndrome diagnosis regardless of *qi* deficiency.

5. Conclusion

In summary, with TF-IDF relative entropy, construction of symptom and syndrome, symptoms and TCM fuzzy matrix, the fuzzy relationship synthetic calculation, using the computer simulation of TCM dialectical reasoning process, process of TCM drugs, largely in line with Chinese medicine theory knowledge, showed that TF-IDF relative entropy is used to simulate syndromes differentiation and syndrome differentiation of TCM drugs are effective, This provides a new idea for the study of TCM intelligent syndrome differentiation and intelligent drug use. Obtained by complex network analysis, the treatment of infantile card at the heart of the khan, drug formula for treating infantile khan card has a certain reference value. Thus, we should pay attention to evidence-based medicine of TCM. This paper delineates the drugs for TCM syndrome differentiation and core drugs, The findings of this study still need to be validated by further experimental and clinical studies. In short, the application of relevant computer technology in the field of TCM conforms to the relevant theory of TCM, which can provide a certain reference for clinical treatment and research, thereby contributing to the modernization of TCM to a certain extent.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Wang L, Xiong L, Yu Y, Wen Z, 2021, Wen Zhongyu's Experience in Treating Children Sweat Syndrome. Journal of Practical Traditional Chinese Medicine, 37(03): 495–496.

[2] Xu H, 2017, Research Overview of Traditional Chinese Medicine in the Treatment of Children Sweat Syndrome. Inner Mongolia Journal of Traditional Chinese Medicine, 36(06): 138+177.

[3] Shi C, Xu C, Yang X, 2009, Review of TFIDF Algorithm. Computer Applications, 29(S1): 167–170+180.

[4] Yu J, Yu Q, Zhang T, et al., 2015, Quantitative Study of TCM Syndromes Based on TF-IDF Relative Entropy. World Science and Technology - Modernization of Traditional Chinese Medicine, 17(10): 1986–1991.

[5] Yu J, Yu Q, Zhang T, et al., 2018, Application of Fuzzy Diagnosis Matrix Based on TF-IDF Quantization in Traditional Chinese Medicine Diagnosis. Shi Zhen Traditional Chinese Medicine, 29(07): 1784–1785.

[6] Zhang B, 2002, Cheng Convenient Reading. China Traditional Chinese Medicine Publishing House, Beijing, 9.