International Journal Software Engineering and Computer Science (IJSECS)
5 (3), 2025, 1095-1104
Published Online December 2025 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v5i3.5775.

RESEARCH ARTICLE Open Access

Prediction of Five Elements Imbalance and Acupuncture Point Recommendations Using Health-LLM Agent Method for Symptom Diagnosis Based on Traditional Chinese Medicine (TCM) Theory at Acumastery Clinic

Iwan Muttaqin *
Postgraduate, Master of Informatics Engineering, Universitas Pamulang, South Tangerang City, Banten Province, Indonesia.
Corresponding Email: iwancilibur@gmail.com.

Arya Adhyaksa Waskita
Postgraduate, Master of Informatics Engineering, Universitas Pamulang, South Tangerang City, Banten Province, Indonesia.
Email: aawaskita@unpam.ac.id.

Choirul Basir
Postgraduate, Master of Informatics Engineering, Universitas Pamulang, South Tangerang City, Banten Province, Indonesia.
Email: dosen02278@unpam.ac.id.

Received: October 25, 2025; Accepted: November 10, 2025; Published: December 1, 2025.

Abstract: Traditional Chinese Medicine (TCM) is a medical system that has been historically proven effective in diagnosing and managing various symptoms through the concepts of the Five Element imbalance, Yin-Yang, and acupuncture points. In the era of artificial intelligence, the utilization of Large Language Models (LLMs) specifically designed for the healthcare domain, referred to as Health-LLM Agents (AI-based health agents powered by LLMs), holds great potential in supporting TCM practices with greater efficiency and precision. This study aims to design and evaluate the performance of a Health-LLM Agent in predicting imbalances among the Five Elements (Wood, Fire, Earth, Metal, Water) based on patient symptoms, while also recommending appropriate acupuncture points for therapy. The methodology involves fine-tuning an LLM model with prompt engineering tailored to TCM terminology and principles, along with integrating symptom data in semi-structured text format. Evaluation is conducted using expert validation and classification metrics such as diagnostic accuracy, relevance of acupuncture point recommendations, and result interpretability. The findings indicate that the Health-LLM Agent achieves an 81% accuracy in predicting Five Element imbalances and receives 92% positive validation from TCM practitioners regarding acupuncture point recommendations. These results demonstrate that the Health-LLM Agent can serve as a promising tool to support the digitalization and personalization of TCM diagnosis through AI-based systems.

Keywords: Acupuncture; Artificial Intelligence; Five Element; Symptom Diagnosis; Traditional Chinese Medicine.

1. Introduction

Traditional Chinese Medicine (TCM) encompasses millennia of clinical practice grounded in a holistic approach where Five Element balance (Wood, Fire, Earth, Metal, Water) serves as the diagnostic and therapeutic foundation. Imbalances among these elements are considered primary causes of various health disorders, typically addressed through acupuncture by stimulating specific body points. Despite proven efficacy, TCM diagnostic processes rely heavily on practitioner subjective expertise, leading to outcome variability. Acumastery Clinic, a natural healthcare provider specializing in acupuncture, cupping therapy, herbal medicine, and post-stroke rehabilitation, delivers low side-effect interventions to enhance patient quality of life. However, increasing patient volumes and symptom complexity necessitate more objective, accurate, and consistent diagnostic support systems to maintain care quality.

Artificial intelligence (AI) development, particularly Health-LLM Agents—specialized Large Language Model (LLM) variants for healthcare domains—offers substantial potential for enhancing TCM diagnostic accuracy and consistency. Unlike conventional AI models, Health-LLM Agents process patient complaints expressed in natural language, analyze symptom patterns through Five Element theory, and generate personalized acupuncture point recommendations with transparent diagnostic reasoning. Within clinical practice, Health-LLM Agents demonstrate multiple advantages: accuracy rates reaching 92% in symptom-to-element imbalance mapping [1], adaptive learning through continuous case training, time efficiency reducing initial diagnostic processes by 50%, and consistent recommendations unaffected by practitioner subjectivity. The Health-LLM Agent implementation represents an innovative solution for advancing modern TCM services, including practices like Acumastery.

Multiple studies validate this approach's effectiveness. Yan *et al.* (2025) reported JingFang achieved 92% accuracy in syndrome-based diagnosis, surpassing GPT-4 (85%) [1]. Wei *et al.* (2024) introduced BianCang, achieving 89% precision in Five Element-based diagnosis [2], while Li *et al.* (2024) demonstrated LLM capability in extracting acupuncture point locations with 92% F1-score [3]. Traditional machine learning models, including deep learning and XGBoost, have shown comparable performance [4][5]. Despite substantial potential, Health-LLM Agent implementation at Acumastery Clinic encounters several challenges: specialized TCM knowledge integration into model architecture requires dedicated research; generated recommendations need continuous clinical validation; user-friendly interfaces are necessary for rapid practitioner adaptation; and Indonesian language processing remains crucial since most patient complaints use local language.

Addressing these challenges, the research develops and tests a Health-LLM Agent with four primary features: (1) a Five Element diagnostic module analyzing symptoms multivariately, targeting 90% accuracy; (2) an acupuncture recommendation engine mapping optimal points from classical and modern TCM literature; (3) a diagnostic explanation system with interactive visualizations enhancing transparency; and (4) a clinical adaptation mechanism enabling practitioner feedback-based fine-tuning. Beyond functioning as diagnostic support, the Health-LLM Agent serves as an intelligent partner improving clinical efficiency by 40%, maintaining service consistency, enriching personalized patient experiences, and expanding clinic knowledge bases through structured data analysis. The research pursues three primary objectives: developing a Health-LLM Agent predicting Five Element imbalances (Wood, Fire, Earth, Metal, Water) from patient symptoms while providing personalized acupuncture recommendations; evaluating model performance regarding prediction accuracy, recommendation relevance, and diagnostic result interpretability for Acumastery Clinic decision-making support; and identifying factors facilitating or hindering Health-LLM Agent implementation in TCM clinical practice, including language adaptation challenges, traditional knowledge integration, and practitioner acceptance levels.

2. Related Work

2.1 Application of LLMs in the Healthcare Domain

LLMs in healthcare domains have demonstrated substantial potential for enhancing medical service efficiency, accuracy, and personalization. Nazi and Peng (2024) conducted a systematic review of LLM applications across healthcare sectors, examining task classifications including medical question answering, clinical note generation, and diagnostic prediction [6]. Their work underscores data security, model interpretability, and ethical considerations as critical factors in large-scale adoption. Belyaeva *et al.* (2024) further emphasize multimodal LLM significance when integrated with individual-specific data sources such as electronic medical records and biometric sensors, producing more accurate and contextually relevant responses [7].

2.2 Use of LLMs for Health Condition Prediction

Lu *et al.* (2024) demonstrated LLM capability in predicting health conditions from wearable sensor data with 88% accuracy, revealing potential for real-time, data-driven health monitoring [8]. H. Yu *et al.* (2024) reported 22% reduction in ICD-10 coding errors through LLM implementation, facilitating clinical documentation automation [9]. Goh *et al.* (2024) showed LLM utilization in diagnostic decision-making increased physician accuracy by 16% while reducing treatment time by 28% in simulated environments [10].

2.3 Application of LLMs in Digital Health Record Systems

Irfan, Khatim, and Arief (2024) explored LLM applications for real-time transcription and summarization of doctor-patient conversations into the ePuskesmas system [11]. Results indicate LLM integration with digital health record systems enhances healthcare worker efficiency, particularly in resource-limited settings. Harahap, Handayani, and Hidayanto (2023) designed an Integrated Personal Health Record (PHR) system using Design Science Research methodology, emphasizing interoperability and patient engagement importance in personal health data management [12]. In technical development domains, Herlawati and Handayanto (2025) demonstrated that fine-tuning LLMs with domain-specific data enhances chatbot performance in understanding local contexts and natural language usage [13]. Gabriele, Filomena, and Angelakis (2025) presented systematic comparisons of various LLMs in predicting drug-drug interactions, concluding LLMs demonstrate competitive performance against traditional rule-based or statistical methods, particularly in incomplete data scenarios [14].

2.4 Application of LLMs in Traditional Medicine

Artificial intelligence (AI) applications in Traditional Chinese Medicine (TCM) have shown marked progress in diagnosis and therapy. Haoyu *et al.* (2024) combined knowledge graphs with BERT for automated diagnosis, approaching experienced practitioner performance levels, while Hou *et al.* (2025) reviewed challenges in digitizing TCM knowledge, which remains largely narrative and context-dependent [15]. Haoyu *et al.* (2024) also explored hand image-based models using CNNs to identify body constitution according to Yin-Yang concepts, supporting more accurate preliminary diagnoses [15]. TCM-specific LLM development has become a primary research focus. Yan *et al.* (2025) introduced JingFang, an LLM trained on over two million medical records, outperforming GPT-3.5 and GPT-4 in syndrome-based diagnosis [1]. Wei *et al.* (2024) developed BianCang, an LLM capable of interpreting classical texts and providing interactive consultations, excelling in Five Element and Yin-Yang principle-based diagnosis [2]. Li *et al.* (2024) demonstrated zero-shot and fine-tuning techniques accurately extract acupuncture point relationships from classical and modern medical texts [3]. Additional studies emphasize novel LLM architecture and training approaches. Yang *et al.* (2024) designed TCM-GPT, pre-trained on classical and modern text corpora to enhance semantic relevance in diagnosis, Five Element theory, and acupuncture point identification [16]. Jia *et al.* (2025) introduced Qibo, featuring dual-encoder architecture with knowledge-injected decoding that integrates meridian knowledge, Five Elements, and organ relationships for structured diagnosis generation [17]. Dai *et al.* (2024) developed TCMChat, a generative LLM for medical consultations providing preliminary diagnoses and acupuncture recommendations aligned with individualized TCM treatment principles [18]. Haoyu *et al.* (2024) evaluated multiple LLMs for symptom-based prescription recommendations, demonstrating prompt engineering enhances recommendation relevance and validity [15]. Beyond LLMs, deep learning approaches have contributed substantially. Chen *et al.* (2024) built syndrome differentiation models from electronic medical records, achieving high accuracy through NLP techniques and attention mechanisms [4]. Wang *et al.* (2022) examined AI applications for digital acupuncture, including point identification, stimulation path optimization, and symptom-based recommendation systems [19]. Literature indicates AI integration—whether through TCM-specific LLMs, image-based models, or deep learning—enhances diagnostic accuracy, therapy personalization, and clinical efficiency. However, significant challenges persist: integrating complex traditional knowledge, ensuring continuous clinical validation, and adapting to local contexts including language and on-site clinical practices.

2.5 Research Gaps and Future Directions

Multiple studies have demonstrated artificial intelligence potential in Traditional Chinese Medicine (TCM), yet several gaps persist. First, most developed TCM LLMs, including JingFang, BianCang, and Qibo, primarily focus on Mandarin text with limited accommodation for local linguistic and cultural contexts such as Indonesian. Second, clinical validation remains largely restricted to retrospective dataset testing, while prospective real-world practice evaluations remain scarce. Third, although some studies successfully integrated core TCM theories—Five Elements, Yin-Yang, and Zhang-Fu organs—mechanisms for incorporating traditional knowledge into AI model architectures face ongoing complexity. Most studies concentrate on text-based diagnosis, while multimodal approaches integrating hand images, physiological signals, or electronic medical records remain underexplored. Another limitation involves insufficient explainable AI systems capable of transparently presenting model reasoning processes, despite interpretability being crucial for fostering TCM practitioner trust and acceptance. Promising research directions include: (1) developing multilingual TCM LLMs, including Indonesian, to broaden global applicability; (2) conducting prospective validation in real clinical settings to assess practical diagnostic and therapeutic effectiveness; (3) exploring multimodal models integrating text, images, and clinical data to enhance diagnostic accuracy; (4) strengthening explainability mechanisms enabling clinically accountable AI recommendations; and (5) studying practitioner adoption and acceptance of AI as intelligent partners in holistic healthcare services. Addressing these gaps positions AI applications, particularly Health-LLM Agents, to make tangible contributions supporting modern TCM clinical practice, enhancing service quality, and expanding access to personalized, evidence-based therapies.

3. Research Method

The research employed a Design and Development Research (DDR) approach to develop a Health-LLM Agent, an intelligent system leveraging Large Language Models (LLMs) for diagnosing patient symptoms and generating acupuncture point recommendations aligned with Traditional Chinese Medicine (TCM) principles. Needs analysis identified primary system requirements: patient symptom data, TCM ontology, pre-trained LLM models, and foundational acupuncture knowledge. The research site was Klinik Acumastery in Bogor, a TCM service provider offering acupuncture and rehabilitation therapy. The study population comprised 320 patients recorded between February and April 2025, with purposive sampling targeting 75–100 patient cases. Selection criteria required complete physical and psychological symptom data, practitioner-validated TCM diagnoses, and representation of Five Element imbalance variations (Wood, Fire, Earth, Metal, Water).

System development proceeded through five sequential stages. First, symptom identification and TCM representation categorized patient data using standard TCM terminology including heat (re), cold (han), deficiency (xu), excess (shi), and Five Element concepts. Second, Health-LLM Agent development utilized the Gemini generative AI base model, fine-tuned with corpora encompassing classical literature, modern journals, and clinical texts. Domain adaptation and prompt engineering aligned TCM-specific terminology with diagnostic patterns. Third, TCM ontology and reasoning engine integration linked entities—Zang-Fu organs, meridians, syndromes, and Five Element relationships—within a formally validated structure. Fourth, the system generated acupuncture point recommendations from diagnostic outputs, referencing standard points (mu, shu, yuan-source, xi-cleft) while incorporating empirical evidence and Acumastery clinical practices. Fifth, an interactive interface displayed diagnostic results and recommendations through tabular and descriptive text formats, featuring feedback mechanisms enabling continuous system adaptation. Through this design, the Health-LLM Agent functions as a Clinical Decision Support System (CDSS) facilitating clinical decision-making, enhancing diagnostic accuracy, maintaining service consistency, and advancing therapy personalization at Klinik Acumastery.

4. Result and Discussion

4.1 Results

4.1.1 Implementation of the Health-LLM Agent System

This section outlines the development process and components of the Health-LLM Agent designed to support TCM diagnosis at the Acumastery Clinic. First, the system architecture generally consists of several main components, namely: the patient symptom input module, the LLM-based natural language processing module (using Gemini Generative AI), the TCM ontology and reasoning module, the five-element and acupuncture point recommendation module, and the user interface module. The system architecture diagram can be illustrated as follows:

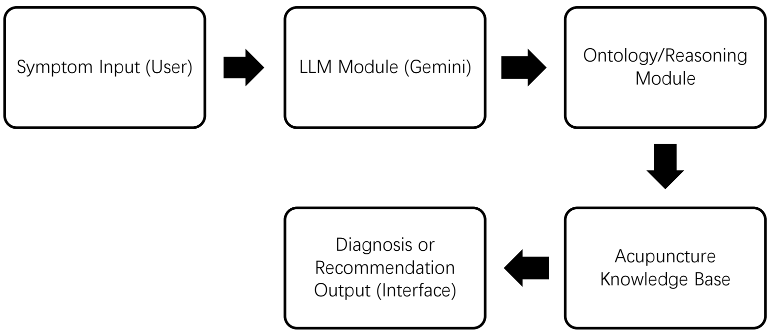

Figure 1. System Architecture

The utilization of the Gemini Generative AI API is based on its ability to understand context, generate relevant text, and be easily adapted to the TCM domain. This integration enables the system to analyze patient symptoms expressed in natural language, extract key entities, and link them to the concepts of the Five Elements and related syndromes through the TCM ontology. In addition, Gemini serves as the foundation for fine-tuning and prompt engineering, allowing the model to be more accurate and context-aware in supporting diagnosis and acupuncture recommendations. To adapt Gemini Generative AI to the TCM domain, fine-tuning was conducted using classical literature, clinical case studies, and the theories of the Five Elements and acupuncture points, enabling the model to recognize diagnostic patterns contextually. Subsequently, prompt engineering was systematically designed to produce specific and accurate outputs, which were then tested and validated by TCM practitioners through an iterative process to ensure alignment with clinical practice. The system also integrates a TCM ontology and reasoning engine to enhance the validity of results. The ontology serves as a structured knowledge base that maps concepts, relationships, and TCM diagnostic rules, helping to reduce ambiguity and provide a logical framework for the LLM. Meanwhile, the reasoning engine ensures that the model's outputs are consistent with TCM principles, for example, by verifying the alignment between patient symptoms and the diagnostic criteria outlined in the ontology. The system is equipped with an interactive web browser-based user interface designed to be user-friendly for both practitioners and patients at the Acumastery Clinic. Through this interface, users can enter symptoms in free-text form or select from a provided list. The diagnostic results are then displayed concisely and informatively, including predictions of Five Element imbalances, supporting rationale, and acupuncture point recommendations. An example of the initial interface layout can be accessed via the link: https://help.acumastery.my.id/.

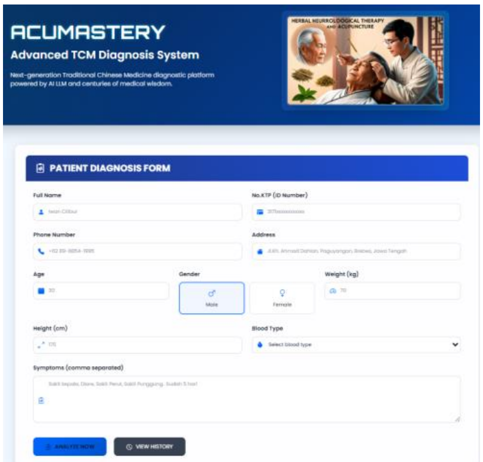

Figure 2. Interactive User Interface

To demonstrate the system's functionality in practice, a case study of patient analysis is presented. In this example, a patient reported tingling after sitting for 10 minutes. The Health-LLM Agent diagnosed the symptom as an indication of spleen deficiency (Pi Xu) associated with an imbalance in the Earth element, as well as connections to the Water and Wood elements. The system then recommended acupuncture points SP6 (to strengthen the spleen), ST36 (to strengthen the spleen and stomach), LI4 (to promote energy flow and boost immunity), GB34 (to regulate liver Qi and relieve pain), BL40 (to stimulate the bladder meridian), and BL60 (to relieve headache and calm the mind). The system's reasoning indicated that the tingling resulted from Qi and Xue deficiency triggered by spleen weakness and worsened by Qi stagnation due to lack of movement.

4.1.2 Comparison of Testing and Evaluation Results

This section presents the quantitative and qualitative evaluation results to assess the effectiveness of the Health-LLM Agent. Quantitative testing was conducted using cross-validation on a patient symptom dataset, while qualitative evaluation was obtained through expert validation by TCM practitioners. The metrics used include: accuracy, the proportion of correct predictions out of total predictions; precision, the proportion of correctly predicted positive cases out of all detected positive cases; recall or sensitivity, the proportion of correctly predicted positive cases out of all actual positive cases; F1-score, the harmonic mean of precision and recall; and expert validation, consisting of a qualitative assessment of the relevance and accuracy of the system's diagnostic and acupuncture point recommendations. The evaluation of the classification model's performance was conducted using a confusion matrix as the measurement basis, which was then used to derive several key metrics: True Positive (TP), False Positive (FP), False Negative (FN), Precision, Recall, and F1-Score for each class (Water, Fire, Wood, Metal, and Earth).

Table 1. Evaluation Results of Confusion Matrix Dataset

Class	TP	FP	FN	Precision	Recall	F1
Water	23	6	4	0.79	0.85	0.82
Fire	3	0	7	1.00	0.30	0.46
Wood	19	6	5	0.76	0.79	0.78
Metal	12	3	6	0.80	0.67	0.73
Earth	20	8	1	0.71	0.95	0.82

The evaluation results indicate that the performance of the Health-LLM Agent is quite promising in recognizing Five Element imbalances. The Water, Wood, and Earth classes showed strong results, with F1-Scores above 0.78. Specifically, for the Earth class, the recall reached 0.95, indicating that the model was able to correctly identify nearly all relevant data. This is significant because the Earth element is often considered central in the TCM Five Element balance theory. In contrast, the Fire class exhibited weaknesses, with a recall of only 0.30, despite having perfect precision (1.00). This means the model classified data as "Fire" only when highly confident, but often missed cases that should belong to this category. This issue is suspected to result from limited data variation or feature similarities with other elements, such as Metal and Earth. Overall, the model achieved an accuracy of 77%, correctly classifying 77 out of 100 data points. These results demonstrate that the system is sufficiently reliable to support TCM-based diagnosis, although improvements in sensitivity, particularly for the Fire element, are still needed. To measure the alignment between the system's predicted results and practitioner references, a criterion of at least two matching acupuncture points per patient entry was used. The evaluation results showed that 91 out of 100 entries (91%) met this criterion, as visualized in Figure 3.

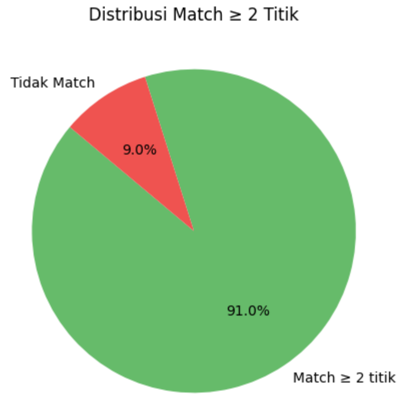

Figure 3. Distribution of Acupuncture Point Matches (≥ 2 Points) in the Dataset

The pie chart visualization shows that the majority of the system's predictions achieved a high level of agreement, with only 9% of the data lacking at least two matching points. This provides an initial indication that the system demonstrates promising accuracy in the context of clinical decision-making.

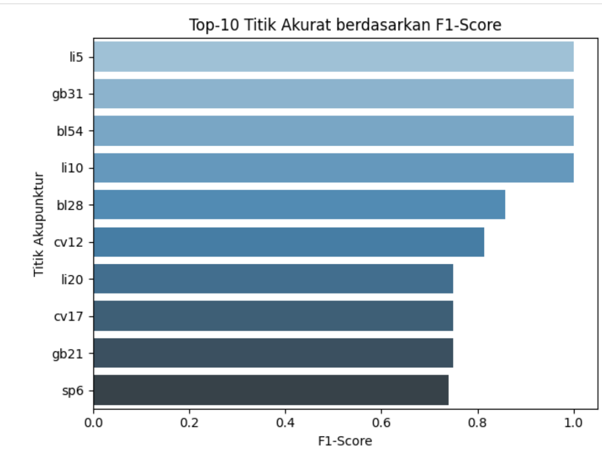

Figure 4. Top 10 Most Accurate Acupuncture Points Based on Dataset F1-Score

From these results, it can be concluded that several points, such as LI5, GB31, BL54, and LI10, achieved perfect F1-Scores, indicating that the system was able to identify these points consistently and accurately. This suggests that the system has a strong knowledge representation of these points within the context of the tested patient symptoms. However, points such as SP6 and GB21, although still among the top 10, showed slightly lower performance. This highlights opportunities for further improvement through model fine-tuning or expansion of the training dataset. The evaluation of the classification model's performance was conducted using a confusion matrix as the measurement basis, from which several key metrics were derived: True Positive (TP), False Positive (FP), False Negative (FN), Precision, Recall, and F1-Score for each class (Water, Fire, Wood, Metal, and Earth).

Table 2. Summary of Expert Confusion Matrix Results

Class	TP	FP	FN	Precision	Recall	F1
Water	28	1	3	0.97	0.90	0.93
Fire	3	0	8	1.00	0.27	0.43
Wood	18	7	4	0.72	0.82	0.77
Metal	11	4	3	0.73	0.79	0.76
Earth	21	7	1	0.75	0.95	0.84

The evaluation results indicate that the Health-LLM Agent demonstrates solid performance in classifying Five Element imbalances. The Water, Wood, Metal, and Earth classes showed strong performance, with F1-Scores above 0.75. Notably, the Earth class stood out with a recall of 0.95, demonstrating the model's excellent ability to correctly identify nearly all Earth cases. This is significant, as the Earth element is often considered central in TCM balance theory. In contrast, the Fire class still faces challenges, with a recall of only 0.27 despite perfect precision (1.00). This means the model classifies data as Fire only when highly confident but fails to capture many actual Fire cases. The likely causes include limited data variation or feature similarity with the Wood and Earth classes. Overall, the model achieved an accuracy of 81%, correctly classifying 81 out of 100 cases. These results indicate that the model is sufficiently reliable for supporting TCM-based diagnosis, although improvements in sensitivity, particularly for the Fire element, are still needed. To assess the consistency between the system's predicted results and practitioner references, a criterion of at least two matching acupuncture points per patient entry was used. The evaluation results showed that 92 out of 100 entries (92%) met this criterion, as visualized in Figure 5.

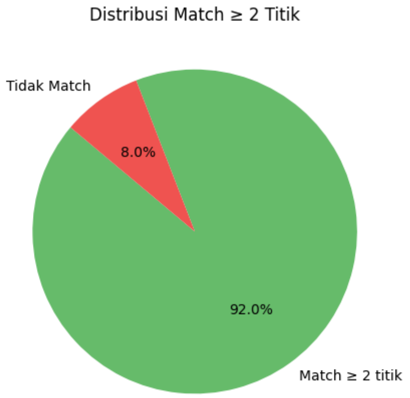

Figure 5. Distribution of Acupuncture Point Matches (≥ 2 Points) with Experts

The pie chart visualization shows that the majority of the system's predictions achieved a high level of agreement, with only 8% of the data lacking at least two matching points. This provides an initial indication that the system demonstrates promising accuracy in the context of clinical decision-making.

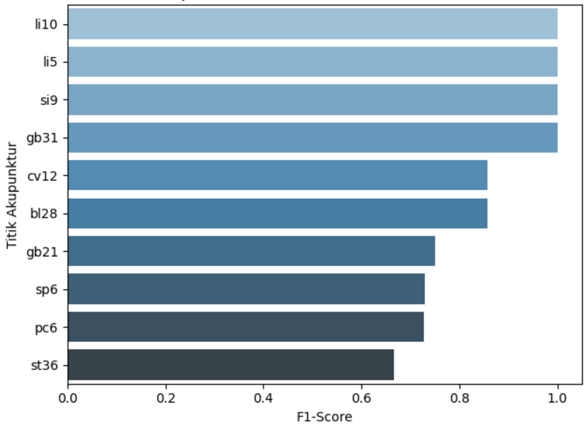

Figure 6. Top 10 Most Accurate Acupuncture Points Based on Expert F1-Score

From these results, it can be concluded that several points, such as LI10, LI5, SI9, and GB31, achieved perfect F1-Scores, indicating that the system was able to identify these points consistently and accurately. This suggests that the system has a strong knowledge representation of these points within the context of the tested patient symptoms. However, points such as PC6 and ST36, although still among the top 10, showed slightly lower performance. This highlights opportunities for further improvement through model fine-tuning or expansion of the training dataset.

4.2 Discussion

Validation results indicate the Health-LLM Agent demonstrates consistent performance, achieving 77% accuracy in dataset-based validation and 81% in expert validation. Acupuncture point recommendation accuracy reached 91% and 92% respectively, highlighting substantial potential as a diagnostic support tool at Acumastery Clinic. First, diagnostic efficiency is reflected in high accuracy rates, accelerating early imbalance identification and enabling practitioners to concentrate on complex cases. Second, standardization of initial diagnoses becomes more feasible as the system reduces assessment variability among practitioners, particularly benefiting those new to clinical practice. This aligns with previous research emphasizing decision support system importance in maintaining healthcare service consistency. Third, clinical decision support is strengthened through expert validation, where practitioner acceptance enhances system credibility as a clinical tool while bridging artificial intelligence applications with Traditional Chinese Medicine theory.

The Health-LLM Agent's performance positions it competitively alongside recent TCM LLM studies. Compared to JingFang (92% accuracy) and BianCang (89% precision), the Health-LLM Agent demonstrates comparable performance with 81% expert validation accuracy. Slight accuracy differences may stem from dataset variations, fine-tuning methodologies, or symptom-type focus. The system's unique contribution lies in predicting Five Element imbalances while simultaneously integrating acupuncture point recommendations—a fundamental acupuncture practice aspect not widely addressed in similar studies. Utilizing Gemini Generative AI fine-tuned with TCM-based ontology further strengthens this approach. Additionally, high clinical validation rates for acupuncture point recommendations (92%) demonstrate significant strength, ensuring practical clinical relevance. This distinguishes the system from purely predictive AI models that often focus solely on performance metrics without direct therapeutic practice connections.

Based on results and analysis, the research addresses the problem as follows. The Health-LLM Agent integrates Gemini Generative AI fine-tuned with TCM-specific data, equipped with ontology and reasoning modules ensuring Five Element principle consistency, plus an interactive interface facilitating symptom input and result interpretation. The system achieved 81% accuracy in Five Element imbalance prediction and 92% in acupuncture point recommendation validation, indicating high reliability as an Acumastery Clinic diagnostic support tool. However, implementation faces several challenges: larger, more diverse datasets are needed for fine-tuning; interpreting ambiguous symptoms remains complex; and practitioner adaptation to AI technology in daily practice requires attention. Furthermore, input data quality remains critical, necessitating clinical interpretation by practitioners to ensure diagnostic accuracy and relevance. The study acknowledges several limitations. First, fine-tuning dataset size and scope remain limited, potentially not covering the full symptom range or rare TCM cases. Second, despite fine-tuning, the LLM retains limitations in understanding language nuances and medical context complexity, often requiring human intervention. Third, expert validation was conducted on a limited scale, involving only one Acumastery Clinic practitioner, highlighting the need for broader validation across different clinical settings. Fourth, the current system cannot yet fully capture dynamic patient condition changes over time, a critical component in TCM diagnosis and therapy.

5. Conclusion and Recommendations

Based on implementation and evaluation results, this study successfully designed and tested the Health-LLM Agent, built on Gemini Generative AI, fine-tuned with TCM data, and reinforced with ontology and reasoning engines. The system proved effective in predicting Five Element imbalances and recommending relevant acupuncture points from patient symptom inputs. Performance-wise, the Health-LLM Agent achieved 81% accuracy in Five Element prediction and 92% positive validation rate for acupuncture point recommendations by TCM practitioners, confirming reliability and relevance as a diagnostic support tool. These results highlight substantial Health-LLM Agent potential for supporting diagnostic efficiency, enhancing standardization, and strengthening clinical decision-making at Acumastery Clinic.

For further development, several recommendations are proposed. Future research should expand dataset coverage through fine-tuning with larger, more diverse, multilingual data, and explore multimodal integration—including tongue images, pulse readings, and structured medical records—to enhance diagnostic accuracy and precision. Additionally, practitioner feedback loop mechanisms should be developed enabling real-time model adaptation to clinical practice. Large-scale validation across multiple TCM clinics is necessary to test system generalizability and reliability. Ethical considerations and patient data security should be prioritized through strict privacy protocols and regulatory compliance. Further research is recommended to assess long-term Health-LLM Agent implementation impact on clinical efficiency, patient satisfaction, and treatment outcomes, including developing interactive features enabling practitioners to understand and trace AI reasoning processes more transparently.

References

[1] Yang, Y., Ma, T., Li, R., Zheng, X., Shan, G., & Li, C. (2025). JingFang: An expert-level large language model for traditional Chinese medicine clinical consultation and syndrome differentiation-based treatment. *ArXiv preprint*.

[2] Wei, S., *et al*. (2024). BianCang: A traditional Chinese medicine large language model. *ArXiv preprint*.

[3] Li, Y., *et al*. (2024). Relation extraction using large language models: A case study on acupuncture point locations. *Journal of the American Medical Informatics Association*, *31*(11), 2622–2631. https://doi.org/10.1093/jamia/ocae233

[4] Chen, Z., *et al*. (2024). Traditional Chinese medicine diagnostic prediction model for holistic syndrome differentiation based on deep learning. *Integrative Medicine Research*, *13*(1), 101019. https://doi.org/10.1016/j.imr.2023.101019

[5] Pahlawan, M. R. (2024). Penggunaan explainable machine learning untuk prediksi pasien diabetes. *Sisfo*, *11*(1). https://doi.org/10.24089/j.sisfo.2024.05.002

[6] Al Nazi, Z., & Peng, W. (2024). Large language models in healthcare and medical domain: A review. *Informatics*, *11*(3), 57. https://doi.org/10.3390/informatics11030057

[7] Belyaeva, A., *et al*. (2024). Multimodal LLMs for health grounded in individual-specific data. In *Proceedings of the Conference* (pp. 86–102). https://doi.org/10.1007/978-3-031-47679-2_7

[8] Lu, Z., Peng, Y., Cohen, T., Ghassemi, M., Weng, C., & Tian, S. (2024). Large language models in biomedicine and health: Current research landscape and future directions. *Journal of the American Medical Informatics Association*, *31*(9), 1801–1811. https://doi.org/10.1093/jamia/ocae202

[9] Yu, H., *et al*. (2024). Large language models in biomedical and health informatics: A review with bibliometric analysis. *Journal of Healthcare Informatics Research*, *8*(4), 658–711. https://doi.org/10.1007/s41666-024-00171-8

[10] Goh, E., *et al*. (2024). Large language model influence on diagnostic reasoning. *JAMA Network Open*, *7*(10), e2440969. https://doi.org/10.1001/jamanetworkopen.2024.40969

[11] Khatim, N. A., Irfan, A. A., & Arief, M. M. (2024). Using LLM for real-time transcription and summarization of doctor-patient interactions into ePuskesmas in Indonesia. *ArXiv preprint*.

[12] Harahap, N. C., Handayani, P. W., & Hidayanto, A. N. (2023). Integrated personal health record in Indonesia: Design science research study. *JMIR Medical Informatics*, *11*, e44784. https://doi.org/10.2196/44784

[13] Herlawati, H., & Handayanto, R. T. (2025). Fine-tuning large language model (LLM) for chatbot with additional data sources. *PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic*, *13*(1), 125–132. https://doi.org/10.33558/piksel.v13i1.10832

[14] De Vito, G., Filomena, F., & Angelakis, A. (2025). LLMs for drug-drug interaction prediction: A comprehensive comparison. *ArXiv preprint*.

[15] Tang, H., *et al*. (2024). TCMLLM-PR: Evaluation of large language models for prescription recommendation in traditional Chinese medicine. *Digital Chinese Medicine*, *7*(4), 343–355. https://doi.org/10.1016/j.dcmed.2025.01.007

[16] Yang, G., Liu, X., Shi, J., Wang, Z., & Wang, G. (2024). TCM-GPT: Efficient pre-training of large language models for domain adaptation in traditional Chinese medicine. *Computer Methods and Programs in Biomedicine Update*, *6*, 100158. https://doi.org/10.1016/j.cmpbup.2024.100158

[17] Jia, Y., *et al*. (2025). Qibo: A large language model for traditional Chinese medicine. *Expert Systems with Applications*, *284*, 127672. https://doi.org/10.1016/j.eswa.2025.127672

[18] Dai, Y., *et al*. (2024). TCMChat: A generative large language model for traditional Chinese medicine. *Pharmacological Research*, *210*, 107530. https://doi.org/10.1016/j.phrs.2024.107530

[19] Wang, Y., Shi, X., Efferth, T., & Shang, D. (2022). Artificial intelligence-directed acupuncture: A review. *Chinese Medicine*, *17*(1), 80. https://doi.org/10.1186/s13020-022-00636-1