

# Development of an Intelligent Support System for Acupuncture and Moxibustion Treatment Based on Machine Learning

YANG Hang\*

WU Ren\*\*

GE Qi-Wei\*\*\*

(Abstract)

Artificial Intelligence (AI) has demonstrated potential in Acupuncture and Moxibustion Treatment (AMT), offering data-driven support for clinical decision-making. A key application is to generate personalized acupoint prescriptions based on patient symptoms, yet challenges remain in data processing, model selection, and clinical applicability. This study aims to address these challenges by developing a machine learning-based AMT support system that provides personalized acupoint prescriptions to assist physicians in clinical decision-making. Specifically, we construct an acupoint prescription database, which serves as a foundation for training AI models. Then, we employ the hierarchical classification model Hierarchical Attention-based Recurrent Neural Network (HARNN) to predict disease categories, syndrome subcategories, and personalized acupoint prescriptions based on patient symptoms. To enhance interpretability and clinical trust, we introduce the Local Interpretable Model-Agnostic Explanations (LIME) to visualize the decision-making process of the model. Building upon this, we design an AI-assisted physician decision-making framework, integrating AI recommendations with physician expertise to ensure that AI-driven AMT aligns with Traditional Chinese Medicine (TCM) principles and clinical standards. Finally, we implement the system using Streamlit, providing an interactive interface for AI-assisted clinical decision-making. Experimental results on 1,000 test samples show that the system achieves an Intersection over Union (IoU) of 0.9165 in acupoint prescription prediction, reflecting reliable performance. This study offers an effective approach for advancing intelligent AMT and a practical tool for validating model reasoning against medical theories, enhancing both credibility and clinical applicability.

*Keywords:* machine learning, acupuncture and moxibustion, acupoint prescription, hierarchical classification

## 1. Introduction

Acupuncture and Moxibustion Treatment (AMT), a fundamental modality within Traditional Chinese Medicine (TCM), has been practiced for thousands of years in the

---

\* The Graduate School of East Asian Studies, Yamaguchi University

\*\* Faculty of Information Science, Shunan University

\*\*\* The Graduate School of East Asian Studies, Yamaguchi University

treatment of various ailments. It is particularly recognized for its efficacy in managing pain, neurological conditions, and chronic diseases, serving as a valuable alternative or complementary therapy in modern healthcare [1]. Despite its long-standing history and widespread application, the clinical practice of acupuncture remains largely experience-driven, resulting in inconsistencies in acupoint prescriptions. The absence of standardized treatment protocols presents challenges for reproducibility and scientific validation, thereby hindering the broader integration of acupuncture into evidence-based medical frameworks [2].

With the advancement of artificial intelligence (AI), new opportunities have emerged for integrating data-driven approaches into AMT. AI technologies, particularly machine learning models, have shown considerable promise in analyzing complex medical data, uncovering hidden patterns, and providing decision support based on large-scale evidence [3]. The application of AI in acupuncture holds the potential to enhance the precision of acupoint prescriptions, improve therapeutic outcomes, and promote the standardization of clinical practices. Moreover, AI-assisted systems can support practitioners by generating recommendations informed by extensive clinical data, contributing to more consistent, accurate, and personalized treatment strategies.

However, the reliability of artificial intelligence in acupuncture treatment still faces multiple challenges, including the limited quality and consistency of clinical data, the difficulty in selecting appropriate machine learning models that can effectively capture the hierarchical and domain-specific characteristics of TCM, and the lack of transparency in machine learning models, particularly complex deep learning models, hinders their adoption in healthcare.

In response to these challenges, we have developed an intelligent AMT support system. This system comprises three key components: (1) Acupoint Prescription Database: An acupoint prescription database has been constructed based on symptoms, disease names, subtypes, and corresponding acupoint prescriptions. It assists physicians in retrieving relevant historical cases and provides training data for machine learning models to support AI-assisted treatment; (2) Hierarchical and Interpretable Machine Learning Model: Based on our analysis of AMT data from a machine learning perspective, we adopted the Hierarchical Attention-based Recurrent Neural Network (HARNN) model [4] for hierarchical classification tasks in this domain. To enhance model transparency and clinical trust, Local Interpretable Model-Agnostic Explanations (LIME) [5] method is integrated to visualize the key symptoms contributing to each prediction; (3) Clinical Decision Support Interface: A web-based platform developed with Streamlit [6] enables clinicians to input patient symptoms and receive AI-generated diagnostic suggestions along with interpretability visualizations.

The rest of this paper is structured as follows: Section 2 reviews related work on machine learning for AMT and hierarchical classification. Section 3 details the design and implementation of our system. Section 4 presents experimental results and case analysis. Section 5 concludes the paper and outlines future work.

## **2. Related work**

### **2.1 Machine learning for AMT**

The rapid advancement of AI has led to significant progress in various aspects of AMT, such as acupoint selection and prescription, identification of treatment techniques, and efficacy prediction [7]. With respect to technical methodologies, existing studies have primarily adopted data mining and unsupervised machine learning techniques. For instance, data mining has been applied to identify effective acupoint combinations, acupoint compatibility patterns, and optimized treatment strategies, thereby enhancing clinical reliability and consistency [8, 9]. Among unsupervised approaches, association rule mining and clustering algorithms are widely used, with the Apriori algorithm being particularly prominent [10, 11]. With respect to research objectives, the majority of existing studies focus on developing acupoint prescriptions tailored to individual diseases. In these studies, machine learning models are typically trained on data from a single condition, such as lumbar disc herniation, facial paralysis, neurogenic dysphagia, urticaria, or insomnia [7].

Differing from existing research, this study adopts supervised machine learning techniques and integrates data from multiple diseases into a unified model. By learning from a broader and more diverse dataset, the model is able to capture both shared treatment patterns across conditions and disease-specific characteristics. This design enhances its ability to generate personalized and accurate acupoint prescriptions based on each patient's symptom profile. Acupoint selection involves establishing a mapping from symptom combinations to acupoint combinations, rather than simple one-to-one associations. In machine learning terms, this task can be formulated as a multi-label classification problem, with symptom combinations as input features and acupoint combinations as output labels. To achieve the prediction of acupoints prescriptions from symptoms using machine learning models, various approaches are evaluated [12], including traditional machine learning methods (such as Support vector machines [13], Decision trees [14]), ensemble learning techniques (such as Extra tree, Boosting [15]), and deep learning models (such as FNN [16], TextCNN [17], TextGCN [18], ML-GCN [19], Seq2seq with attention [20]). However, these methods do not fully leverage the hierarchical structure inherent in AMT data. To overcome these shortcomings, this study adopts hierarchical learning techniques to better utilize the structural characteristics of AMT data.

### **2.2 Hierarchical Classification**

Hierarchical classification is a specialized task in machine learning where labels are structured in a multi-level hierarchy, reflecting parent-child relationships. Unlike flat classification methods, hierarchical approaches capture dependencies between labels, enabling more informed predictions [21].

Koller and Sahami [22] were among the first to address hierarchical text classification (HTC),

highlighting the limitations of flat classifiers in representing hierarchical topics. Later, Silla and Freitas [23] formalized the concept of HTC and proposed a unifying framework applicable across various domains beyond text. More recently, Zangari et al. [21] provided a comprehensive review of the theoretical foundations and applications of hierarchical classification. Building on these foundations, Huang et al. [4] introduced the Hierarchical Attention-based Recurrent Neural Network (HARNN), which automatically assigns document labels level by level in a hierarchy. Given its strong performance in hierarchical classification tasks, we adopt HARNN as the core model for our system’s hierarchical classification module.

### **2.3 Explainable AI in Medical Applications**

While machine learning has great potential to advance AMT, its application in healthcare is often hindered by the limited transparency of many models—particularly deep learning approaches. In the medical field, interpretability is essential not only for earning the trust of practitioners and patients, but also for ensuring the reliability and clinical validity of predictions.

To address this, interpretable machine learning techniques have been developed to provide insights into how models make decisions. Methods such as SHAP (SHapley Additive exPlanations) [24], LIME (Local Interpretable Model-agnostic Explanations) [5], and attention mechanisms [25] improve model transparency by revealing the contribution of input features to specific outputs.

In the context of AMT, interpretability allows for visualization of the diagnostic logic learned from acupoint prescription data. This not only supports clinical understanding and trust but also offers a valuable tool for validating the model’s reasoning against established medical theories, ultimately enhancing the credibility and practical utility of intelligent AMT systems.

## **3. System Design and Implementation**

This section presents the design and development of the proposed intelligent support system for AMT. It begins with an overview of the system architecture, followed by a description of the database construction process. Next, we introduce the hierarchical and interpretable machine learning model that supports multi-level predictions. Finally, we present the technical implementation of the system, including its key components and user interface design.

### **3.1 Overview of the Intelligent Support System**

Fig. 1 illustrates the diagnostic and treatment workflow of AMT using our proposed intelligent AMT support system. This system primarily focuses on optimizing ③, ④, ⑤, ⑨, and ⑩ in the process. We first construct an acupoint prescription database ⑩, which serves as the training data source for a hierarchical and interpretable machine learning model ③. At the beginning of diagnosis and treatment, symptoms are first collected from patients using

relevant medical equipment, based on the TCM diagnostic methods of inspection, auscultation and olfaction, inquiry, and palpation. Next, the symptoms ② obtained from the four diagnostic methods are fed into the trained machine learning model ③, which generates a disease classification, syndrome subtype, personalized acupoint prescription ④, and corresponding LIME-based interpretability analysis ⑤. The doctor subsequently examines relevant historical cases from the database in conjunction with recommendations provided by the model, to determine whether it is appropriate to modify the proposed acupoints ⑥. If necessary, the prescription is adjusted based on clinical expertise; otherwise, the recommended acupoint prescription is directly applied. After receiving treatment, the patient provides feedback on their recovery within a specified period. This feedback is recorded in the database ⑩ to further enhance the model's accuracy and improve future predictions.

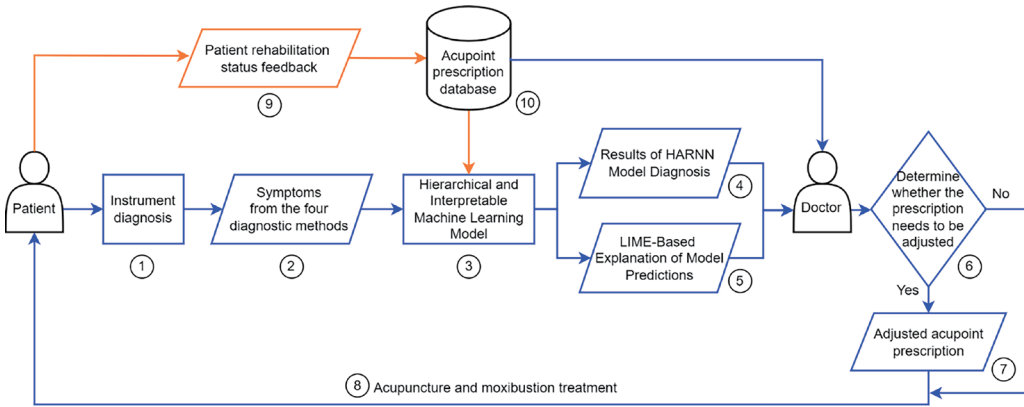


Fig. 1. Diagnosis and Treatment Workflow with the Intelligent AMT Support System. (Source: Created by the authors)

### 3.2 Database construction

AMT, with its thousand years history, has yielded a vast corpus of acupoint prescriptions that serve as the foundation for our database. However, terminological inconsistencies across different data sources require meticulous manual standardization to ensure consistency. To standardize the data, we adopted the symptom names provided in reference [26] as the baseline and supplemented them with new terms whenever the existing vocabulary proved inadequate. In our approach, we broaden the definition of ‘symptom’ to include additional relevant details such as affected body parts, severity, and gender, in order to provide a more accurate representation of the patient’s condition. Moreover, the original textual records were restructured into a format comprising symptoms, disease names, syndrome types, and corresponding acupoint prescriptions. This structured format not only simplifies the analysis of interrelationships among these elements but also enables the generation of diagnostic and treatment recommendations directly from patient symptoms. Currently, our dataset includes 770 distinct symptom terms. As to the names of acupoints, our database includes 410 acupoint

names, consisting of 409 acupoints certified by WHO [27] and the Ashi acupoint. Ashi points refer to areas that lack a specific name or fixed position but are identified by tenderness or other reactions when stimulated [1].

To date, our database has collected 5,000 acupoint prescription cases from books [1] [28][29]. As shown in Table 1, each case consists of the following components: disease case number, disease name, subcategory, symptoms, and acupoints. To meet the requirements of machine learning, we assigned corresponding numerical IDs to each disease name, subcategory, symptom, and acupoint (as shown in Fig. 2). Table 2 presents the statistical information of the database. Each disease case in the database is labeled across three hierarchical levels: The first level is the disease name layer, which includes 159 categories; The second level is the subcategory layer, which consists of 1,576 categories; The third level is the acupoint layer, comprising 410 categories, corresponding to 410 acupoints names. It should be noted that the statistical data presented here reflects the current state of our database, rather than the categorization practices in TCM field.

**Table 1. Structure and examples of our database.**  
(Source: Created by the authors)

Disease case number	Disease name	Subcategory	Symptoms	Acupoints
1	apoplexy	apoplexy (organ strike block pattern)	dizzy, inhibited urination, difficult defecation, curled tongue, retracted genitals, skin color red, thick breath, phlegm rale in the throat, bad breath, fever, agitation	PC6, DU26, LU11, PC9, HT9, LI1, SJ1, SI1, SP1, LR1, KI1, ST45, GB44, BL67
2	apoplexy	apoplexy (organ strike block pattern), upper limb, hemiplegia	dizzy, inhibited urination, difficult defecation, curled tongue, retracted genitals, skin color red, thick breath, phlegm rale in the throat, bad breath, fever, agitation, loss of use of the limbs, upper limb	PC6, DU26, LU11, PC9, HT9, LI1, SJ1, SI1, SP1, LR1, KI1, ST45, GB44, BL67, LI15, LU5, LI4, EX-UE9, SJ5
3	apoplexy	apoplexy (organ strike block pattern), omarthralgia	dizzy, inhibited urination, difficult defecation, curled tongue, retracted genitals, skin color red, thick breath, phlegm rale in the throat, bad breath, fever, agitation, shoulders, shoulder pain and soreness	PC6, DU26, LU11, PC9, HT9, LI1, SJ1, SI1, SP1, LR1, KI1, ST45, GB44, BL67, LI17, LI15, SI9, SI15, SI14, ASHI
...	...	...	...	...
5000	postencephalitis	postencephalitis	aphasia, sequelae period, red or crimson tongue, thin fur, yellow tongue fur, stringlike pulse, rapid pulse	DU15, RN23, DU14, DU13, LU5, PC6, LI4, SI3, GB34, ST36, LR3

Disease number	Disease name	Symptom number	Symptom	Acupoint number	Acupoint
1	Apoplexy	1	Irritable	1	LU1
2	Headache	2	Feel depressed	2	LU2
3	Dizziness	3	Oversensitive	3	LU3
...	...	...	...	...	...
159	Postencephalitis	770	Urinary retention	410	Ashi

Subcategory number	Subcategory
1	apoplexy (organ strike block pattern)
2	apoplexy (organ strike block pattern), upper limb, hemiplegia
3	apoplexy (organ strike block pattern), omarthralgia
...	...
1576	postencephalitis

**Fig. 2. Diseases, symptoms, acupoints, and subcategories with assigned IDs.**  
(Source: Created by the authors)

**Table 2. Statistics of our database.**  
(Source: Created by the authors)

	Statistics
# Disease case	5000
# Hierarchical levels	3
# Categories in level-1	159
# Categories in level-2	1576
# Categories in level-3	410
# Total categories	2145

Among the 5,000 original samples, there were 1,576 subcategories with an average of only 3.2 samples per subcategory. In addition, due to limitations in data sources, some subcategories have a large number of samples, while many have only one sample. This severe data imbalance poses challenges for model training. Thus, in the absence of rapidly expanding metadata, data augmentation becomes essential to address this issue. Our data augmentation strategy is based on the premise that the impact of symptoms on diagnosis and treatment varies by case. We categorized symptoms into three types: primary symptoms (explicitly indicated in the original data), special symptoms (e.g., symptom location and disease severity), and minor symptoms. During augmentation, primary and special symptoms were preserved while some minor symptoms were randomly omitted, without altering the acupoint prescription. This approach expanded the database from 5,000 to 10,138 samples.

### 3.3 Hierarchical and Interpretable Machine Learning Model

As the core of the intelligent AMT support system, we have developed a hierarchical and interpretable machine learning model by integrating a HARNN for classification with LIME for explainability.

The hierarchical classification module utilizes the HARNN model proposed in [4], an established method for hierarchical classification tasks. HARNN generates unified representations for both document texts and hierarchical category structures, models dependencies across different levels, and predicts categories in a top-down manner. In our study, this model is employed to support a three-level diagnostic labeling framework, where the first level corresponds to disease names, the second to disease subtypes, and the third to corresponding acupoint prescriptions. This hierarchical structure enables consistent and structured prediction from general diagnostic categories to specific treatment suggestions. Unless otherwise stated, we follow the configurations outlined in the original paper.

The result interpretation module in this study employs LIME, a model-agnostic and locally interpretable machine learning method [5]. LIME explains the predictions of complex black-box models by approximating them with simple interpretable models, such as linear regression

or decision trees. Its core mechanism involves generating new samples by perturbing the input around a given instance and using the black-box model's predictions on these perturbed samples to train a local interpretable model. In the context of our research, LIME is used to analyze the contribution of individual patient symptoms to each predicted label across the three diagnostic levels. This helps clinicians understand which symptoms most strongly influence the model's decision-making, thereby enhancing interpretability and supporting clinical validation.

### **3.4 System Implementation**

The following details the development of the intelligent acupuncture treatment support system, covering its technology stack, model integration, and user interface design. The system was built using modern, lightweight technologies that enable rapid prototyping and interactive model inference, allowing clinicians to input symptoms and receive acupoint recommendations along with interpretability feedback.

#### **Technology Stack and Model Integration**

The system is implemented in Python using TensorFlow for deep learning and Streamlit to build the web-based interface. At startup, the pre-trained HARNN model is loaded in inference mode. User-input symptoms (selected from predefined sets) are processed in real time to generate predictions, including disease classification, subtype identification, and recommended acupoint prescriptions. To enhance transparency, LIME is incorporated into the pipeline to visualize the key contributing symptoms for each prediction.

#### **User Interface Design**

The user interface is developed with Streamlit and is organized into the following functional sections:

- Symptom Input Panel: Enables users to select or enter symptom descriptions.
- Prediction Display Panel: Shows the predicted disease, syndrome subtype, and acupoint prescription.
- Interpretability Visualization: Presents LIME-generated visualizations that highlight the symptoms most influential in the decision-making process.
- Case Feedback Section: (For future integration) Allows recording of patient rehabilitation feedback.

Example screenshots of the system interface are provided below. Fig. 3 and 4 provide illustrative screenshots of the AMT Intelligent Support System interface. Fig. 3 presents the main interface of the system. A language selection menu is available in the sidebar, offering three language options: Chinese, English, and Japanese. On the main panel, clinicians first input the patient's symptoms, then click the "Generate Recommended Prescription" button to initiate the prediction process. The system returns a three-level diagnostic prediction generated by the HARNN model, along with corresponding LIME-based interpretability visualizations. The navigation bar includes three modules: the Home page, "Enter Treatment Record," and "View

Past Records.” Fig. 4 shows the treatment input page, where clinicians can enter patient information, symptom descriptions, and corresponding prescriptions into the system.



Fig. 3. Main Interface of the AMT Intelligent Support System.  
(Source: Created by the authors)



Fig. 4. Treatment Record Interface of the AMT Intelligent Support System.  
(Source: Created by the authors)

### Notes on Deployment

Currently, the system is deployed and tested in a local environment, suitable for research and demonstration purposes. Future extensions may include cloud-based deployment, database integration, and mobile-friendly adaptations.

## 4. Experiments and Results

This section presents the performance evaluation of the proposed hierarchical and interpretable model, showcases its practical application through two case studies, and concludes with a discussion of its strengths and limitations.

### 4.1 Model Performance

To evaluate the performance of the machine learning model used in our system, 1,000 original records from the database were reserved as the test set, while the remaining 9,138 samples were used for training. The task is structured as a three-level hierarchical multi-label classification problem: predicting the disease category, identifying the disease subcategory, and generating the corresponding acupoint prescription. The model's predictions at each level were compared with the reference labels from the database using standard evaluation metrics, including accuracy, precision, recall, and Intersection over Union (IoU). The calculation formulas are shown in Equations (1) - (4) [30, 31], where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives. The detailed results of 1,000 test data are shown in Table 3 and visualized as a bar chart in Fig. 5.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

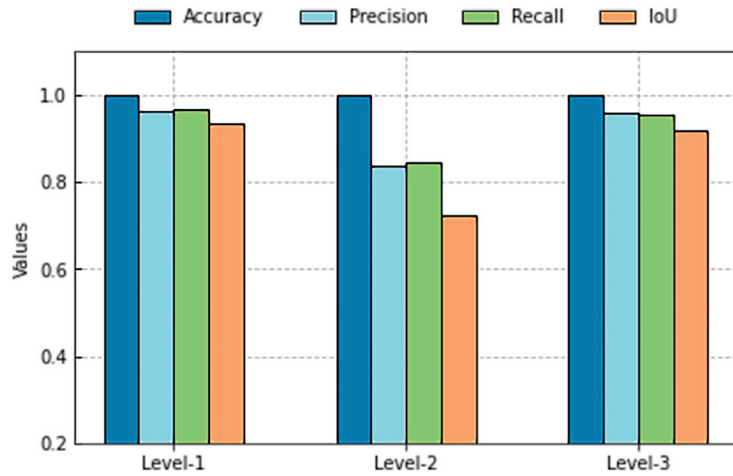
$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (4)$$

**Table 3. System Performance at Each Hierarchical Level Based on 1,000 Original Test Cases.**  
(Source: Created by the authors)

	Accuracy	Precision	Recall	IoU
Level-1	0.9996	0.9622	0.9680	0.9326
Level-2	0.9998	0.8343	0.8460	0.7243
Level-3	0.9984	0.9599	0.9529	0.9165



**Fig. 5. Bar Chart of System Performance at Each Hierarchical Level (1,000 Test Cases).**  
(Source: Created by the authors)

The experimental results show satisfactory performance in predicting both the first-level disease names and the third-level acupoint prescriptions, with an IoU of 0.9165 for acupoint prediction. However, the prediction performance for the second-level disease subcategories is comparatively unsatisfactory, with an IoU of only 0.7243. A plausible explanation is the large number of labels at this level (1,576 in total), which makes it difficult for the limited training data to provide sufficient examples for each label.

## 4.2 Case Study

### 4.2.1 Experimental Setup

To further demonstrate the effectiveness and interpretability of the proposed system, we conducted detailed case studies on two representative test samples. These samples were selected from a held-out test set of 1,000 patient cases that were not seen during training. Each case includes a set of input symptoms and the corresponding predicted acupoint prescription generated by the HARNN model.

The case studies were conducted on a local machine with the following experimental

setup:

- CPU: Intel Core i7-1195G7
- Memory: 8GB RAM
- Operating System: Windows 11
- Programming Language: Python 3.6.13
- Framework: TensorFlow 1.15.0
- Visualization & Interface: Streamlit 1.10.0
- Interpretability Tool: LIME 0.2.0.1

In the following case studies, we present the model’s prediction results alongside interpretability visualizations, highlighting how the model arrives at its recommendations and how these insights can aid physicians in understanding the rationale behind them. These examples demonstrate the system’s potential to support clinical decision-making by providing recommendations that are both accurate and interpretable.

#### 4.2.2 Case 1

This case is selected from the reserved test set to demonstrate the system’s prediction and interpretability output for a real-world clinical record.

- Input Symptoms:  
myosalgia, muscle swelling, muscle soreness, skin numbness, pale red tongue, yellow tongue fur, slimy tongue fur, dry mouth, bitter taste in the mouth, stringlike pulse, slippery pulse
- Target Labels from Database:
  - Level 1 (Disease Category): restless leg syndrome
  - Level 2 (Subtype): restless leg syndrome (damp-heat impeding the channels)
  - Level 3 (Acupoint Prescription): BL56, BL57, SP9, ST40

Fig. 6 illustrates the user interface and system response after entering the symptom descriptions for Case 1. Once the user clicks the “Generate Recommended Prescription” button, the system returns its three-level prediction within approximately one second. The predicted results include the disease name (“restless leg syndrome”), the disease subcategory (“restless leg syndrome: damp-heat impeding the channels”), and the recommended acupoint prescription (“BL56, BL57, SP9, ST40”). Meanwhile, a prompt message—“Generating LIME visualization analysis for you, please wait”—appears on the screen along with an estimated computation time. Under the current system configuration, generating the interpretability results takes approximately 40 minutes. Fig. 7 to 10 present the LIME-based interpretability analysis for each of the three prediction levels.

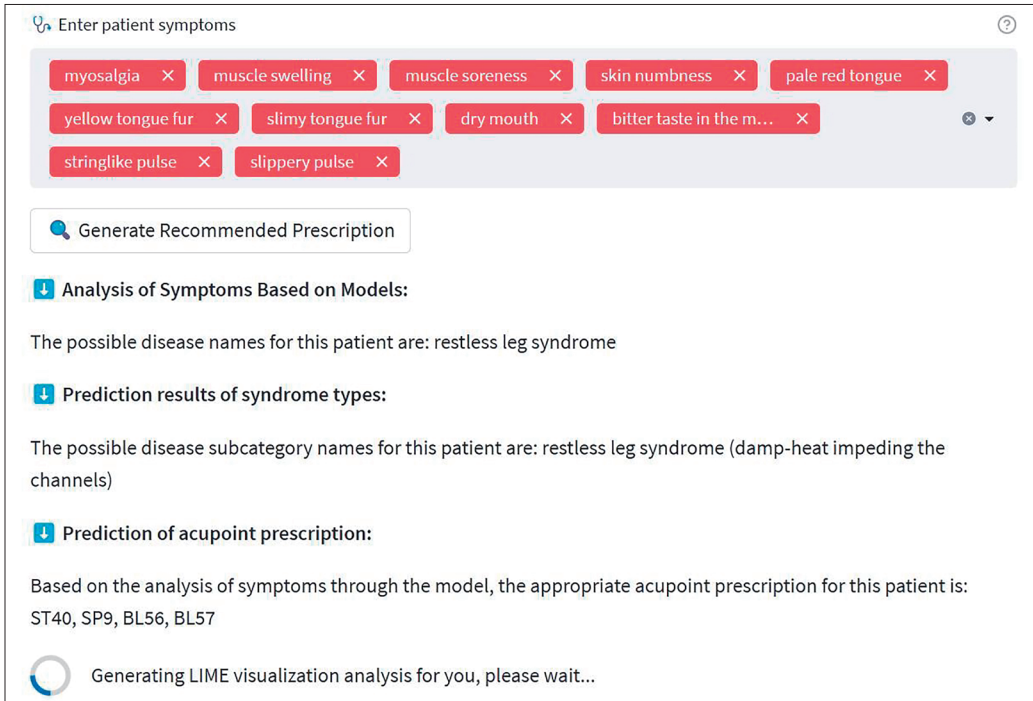


Fig. 6. System Output for Case 1: Hierarchical Predictions Including Disease Name, Subcategory, and Recommended Acupoints. (Source: Created by the authors)

Fig. 7 presents the model’s prediction for the disease name associated with this case. The model identifies “restless leg syndrome” with a high confidence of approximately 99.65%, while also listing other possible disease names along with their respective probabilities. This prediction is consistent with the reference label in the database.

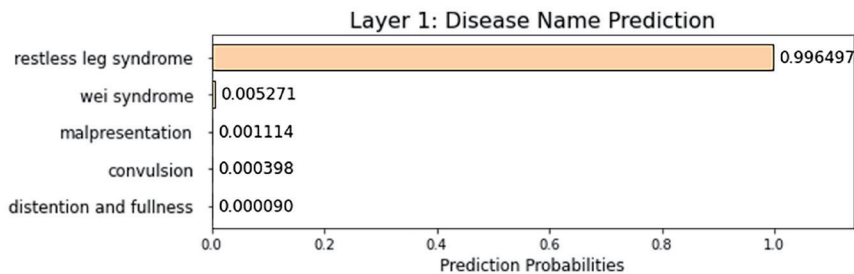


Fig. 7. Top Prediction Probabilities for Disease Names - Case 1 (Level 1). (Source: Created by the authors)

Fig. 8 presents the LIME interpretability analysis of the model’s prediction of “restless leg syndrome” for this case. The vertical axis lists the patient’s symptoms, while the horizontal axis shows each symptom’s contribution score toward the diagnosis. Positive scores indicate

supportive symptoms, with higher values reflecting stronger influence on the model's decision; negative scores suggest symptoms that contradict or reduce confidence in the diagnosis. The model ultimately classified the case as “restless leg syndrome” because the total contribution of supportive symptoms outweighed that of the opposing ones. By comparing the model's reasoning with traditional TCM theories, the interpretability analysis helps judge whether the prediction is reasonable and offers a helpful reference for manual review.

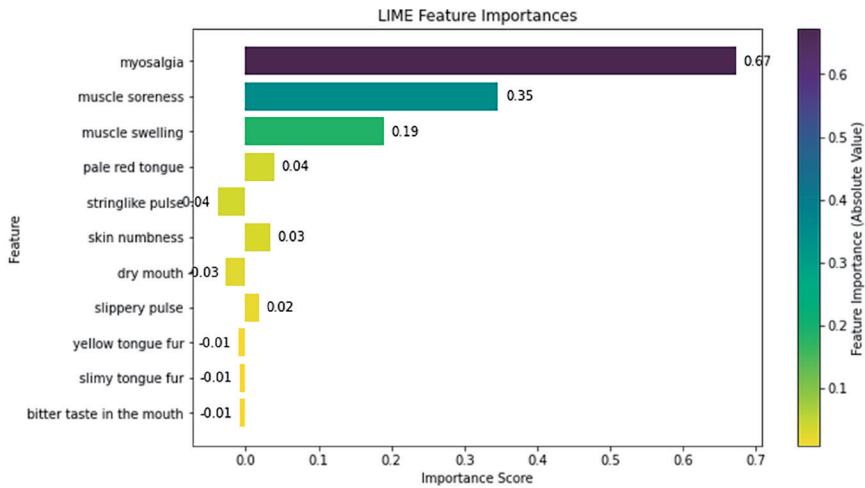


Fig. 8. LIME Explanation for Top-1 Disease Name Prediction - Case 1 (Level 1).  
(Source: Created by the authors)

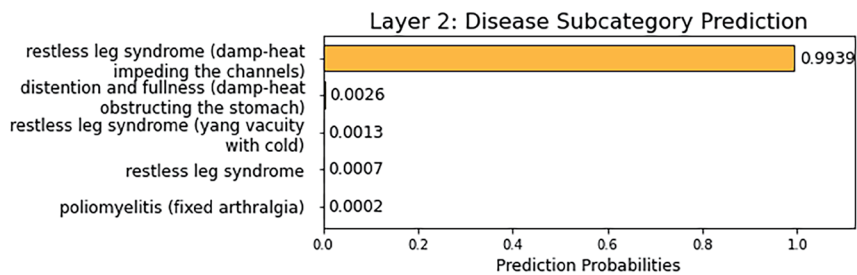


Fig. 9. Top Prediction Probabilities for Disease Subcategories - Case 1 (Level 2).  
(Source: Created by the authors)

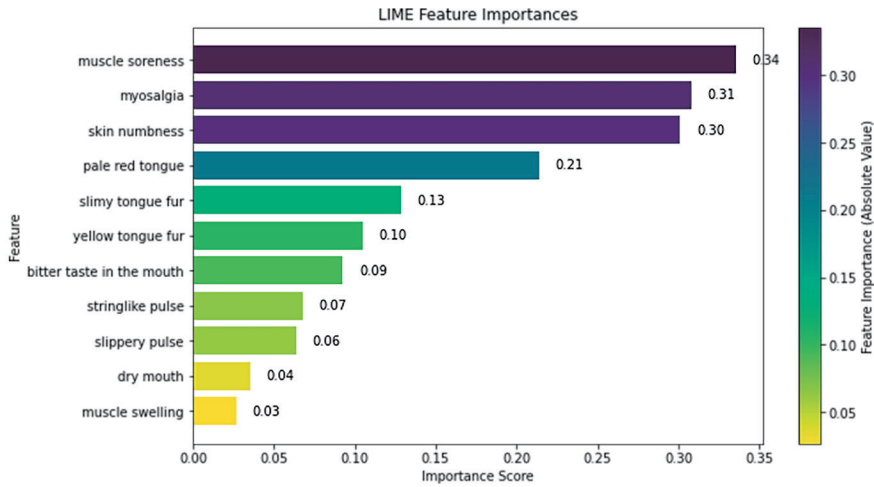


Fig. 10. LIME Explanation for Top-1 Disease Subcategory Prediction - Case 1 (Level 2).  
(Source: Created by the authors)

Fig. 9 and 10 present the LIME-based interpretability results for the second-level prediction. The model predicted the patient’s disease subcategory as “restless leg syndrome (damp-heat impeding the channels)” with a high confidence score of 0.9939, which is consistent with the ground truth. The interpretability visualization further reveals the contribution of individual symptoms to this prediction, highlighting which symptoms played a key role in the model’s diagnostic decision.

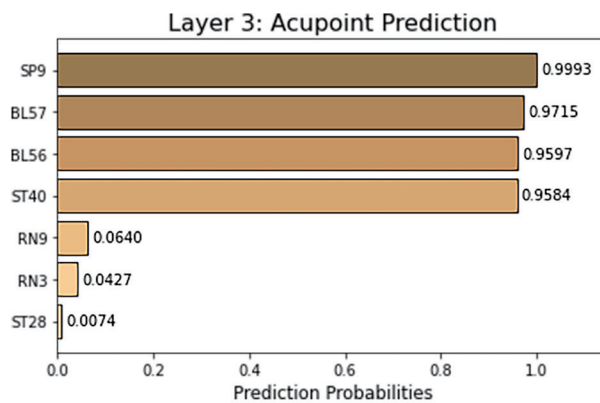


Fig. 11. Top Prediction Probabilities for Acupoint Prescriptions - Case 1 (Level 3).  
(Source: Created by the authors)

Fig. 11 illustrates the predicted probabilities of acupoints recommended by the system for this case. Only acupoints with probabilities greater than 0.5 are considered for use. As shown in the Appendix A, the top four predicted acupoints match the target prescription. A more detailed interpretability analysis of the top 15 predicted acupoints is presented in Appendix A using LIME.

#### 4.2.3 Case 2

A second test case is presented to further assess the system's diagnostic accuracy and interpretability.

- Input Symptoms:  
cough, rapid pulse, fever, red or crimson tongue, thin fur, dry mouth, dry nose, nosebleed, dry throat
- Target Labels from Database:
  - Level 1 (Disease Category): epistaxis
  - Level 2 (Subtype): epistaxis (evil heat invading the lung)
  - Level 3 (Acupoint Prescription): SP4, SP1, SP3, LU6, LI4, LI20, DU23, LU11

Fig. 12 to 17 present the system's diagnostic results, recommended acupoints, and corresponding LIME explanations for Case 2. Detailed interpretability analyses of the top 15 recommended acupoints are provided in Appendix B. These results further demonstrate the system's robust performance across different clinical scenarios.

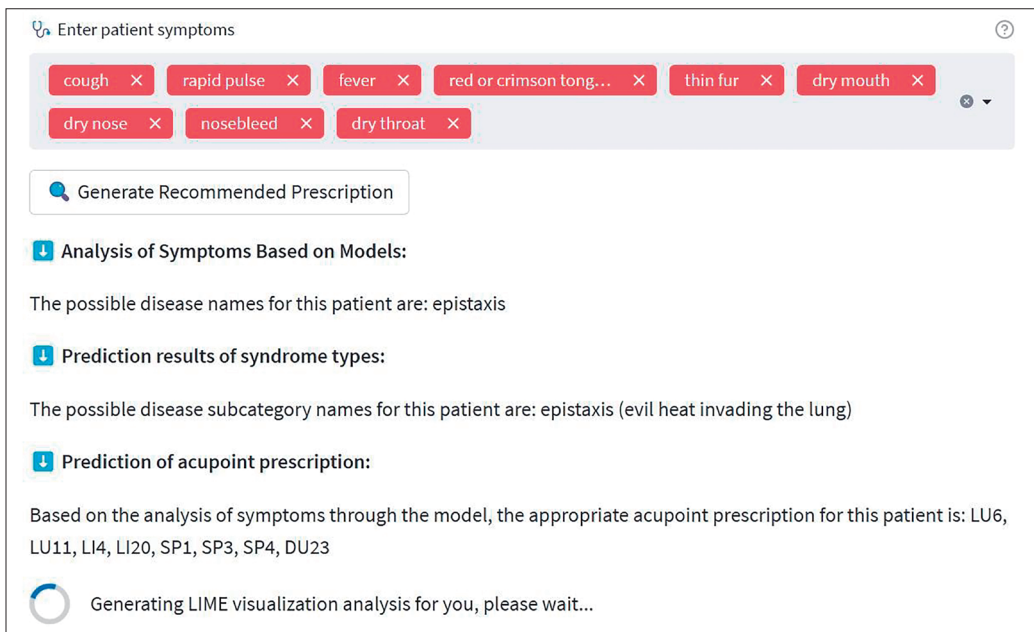


Fig. 12. System Output for Case 2: Hierarchical Predictions Including Disease Name, Subcategory, and Recommended Acupoints.  
(Source: Created by the authors)

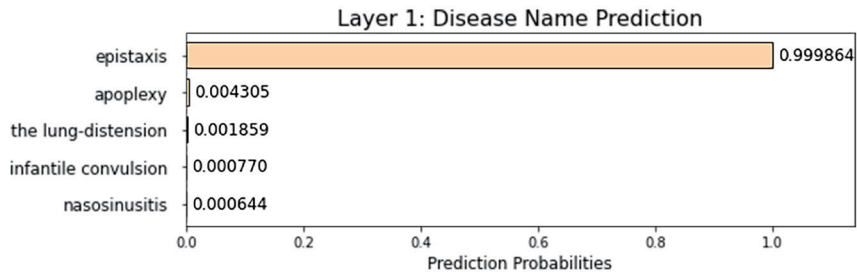


Fig. 13. Top Prediction Probabilities for Disease Names - Case 2 (Level 1).  
(Source: Created by the authors)

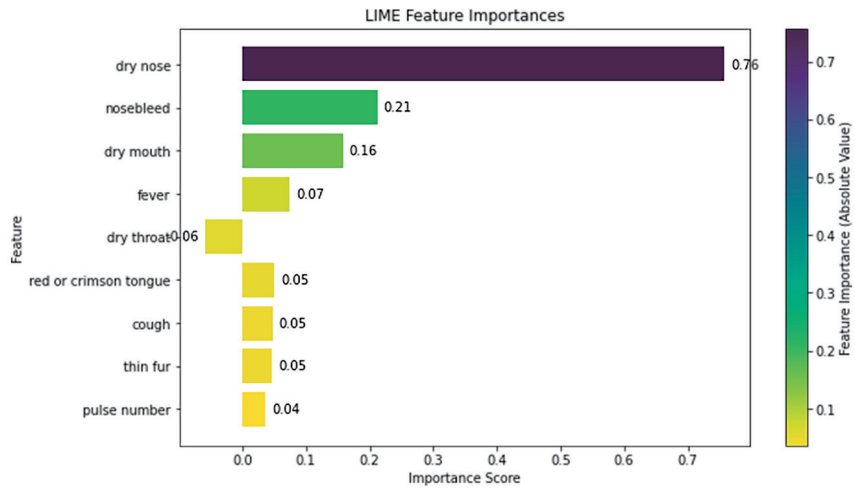


Fig. 14. LIME Explanation for Top-1 Disease Name Prediction - Case 2 (Level 1).  
(Source: Created by the authors)

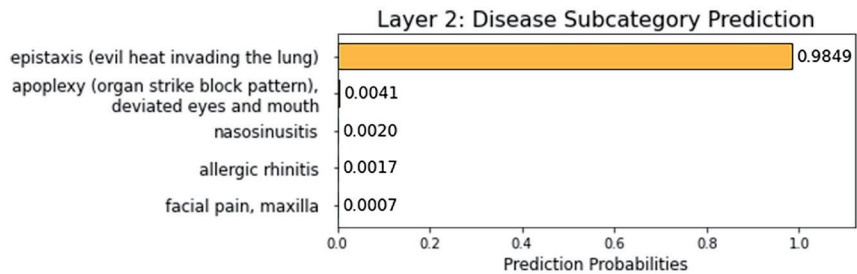


Fig. 15. Top Prediction Probabilities for Disease Subcategories - Case 2 (Level 2).  
(Source: Created by the authors)

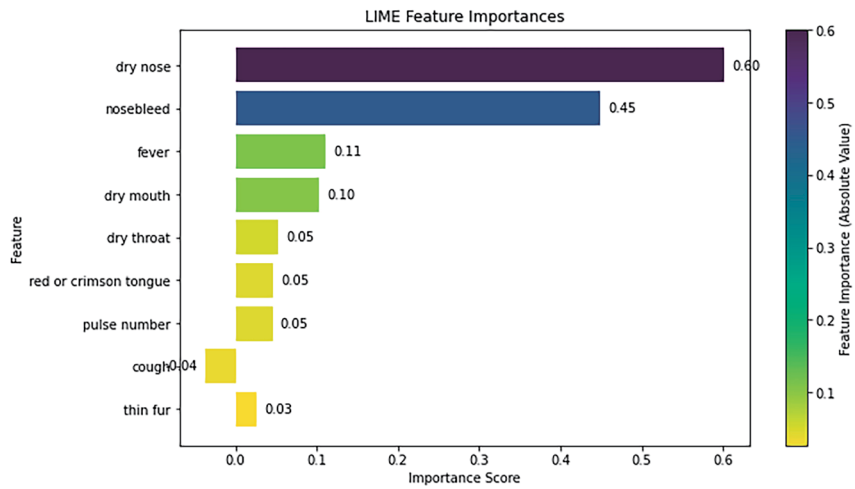


Fig. 16. LIME Explanation for Top-1 Disease Subcategory Prediction - Case 2 (Level 2).  
(Source: Created by the authors)

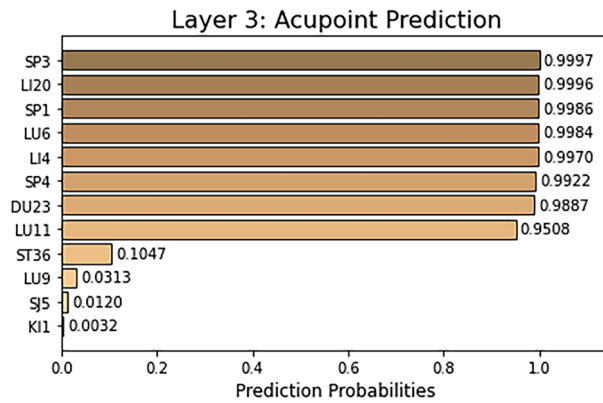


Fig. 17. Top Prediction Probabilities for Acupoint Prescriptions - Case 2 (Level 3).  
(Source: Created by the authors)

### 4.3 Discussion

This section evaluated the performance of the proposed system using 1,000 original test cases and demonstrated its capabilities through two representative examples. The system predicts disease names, subcategories, and corresponding acupoint prescriptions based on patient symptoms. Experimental results show that it performs well in predicting disease names and acupoint prescriptions on the test set. By integrating explainability techniques such as LIME, the system enhances model transparency and supports clinical interpretability. In practice, the system operates within a physician oversight framework and continuously updates its model based on patient feedback, maximizing the advantages of AI while mitigating potential risks.

Despite these promising results, several limitations remain. The prediction performance for subcategories is relatively lower than that for disease names and acupoint prescriptions, indicating areas for improvement. AI models, especially those trained on historical data, may fail to recognize rare or atypical cases, which can lead to missed or incorrect diagnoses. Moreover, although the model demonstrates high predictive accuracy, the actual effectiveness of treatment recommendations remains heavily dependent on the quality and representativeness of the input data, which requires further refinement and continuous evaluation by clinical experts. Finally, in the current experimental environment, generating LIME-based explanations takes over 30 minutes per case, significantly hindering practical deployment. Enhancing the efficiency of interpretability modules is therefore critical for real-time clinical application.

In summary, the proposed system offers a promising step toward intelligent, interpretable support tools for acupuncture and moxibustion treatment planning. With continued validation and technical refinement, it has the potential to assist clinical decision-making, improve efficiency, and promote more consistent and evidence-based therapeutic practices while maintaining the essential role of physician judgment.

## **5. Conclusion**

In this paper, we proposed an intelligent support system for acupuncture and moxibustion treatment (AMT) based on a hierarchical and interpretable machine learning model. Given patient symptoms, the system generates data-driven suggestions for acupoint selection to assist physicians in clinical decision-making. Specifically, the system integrates a Hierarchical Attention-based Recurrent Neural Network (HARNN) to predict disease categories, syndrome subcategories, and to recommend personalized acupoint prescriptions. To enhance interpretability and clinical credibility, the LIME method is employed to visualize the reasoning behind model predictions. Furthermore, an interactive interface is implemented using Streamlit to facilitate real-time usage and improve practical accessibility. Experimental results of 1,000 test cases achieved an IoU of 0.9165 in predicting acupoint prescriptions, demonstrating the effectiveness of the system.

Despite the promising results, the system still faces several limitations. The current dataset is relatively small and lacks diversity, which may affect the robustness and generalizability of the model. In addition, the generation of LIME-based explanations remains time-consuming, hindering real-time clinical application. Future work will focus on expanding the dataset and improving the efficiency of the interpretability module to better support practical deployment.

## Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 23K20390 (Grant-in-Aid for Scientific Research (B)).

## References

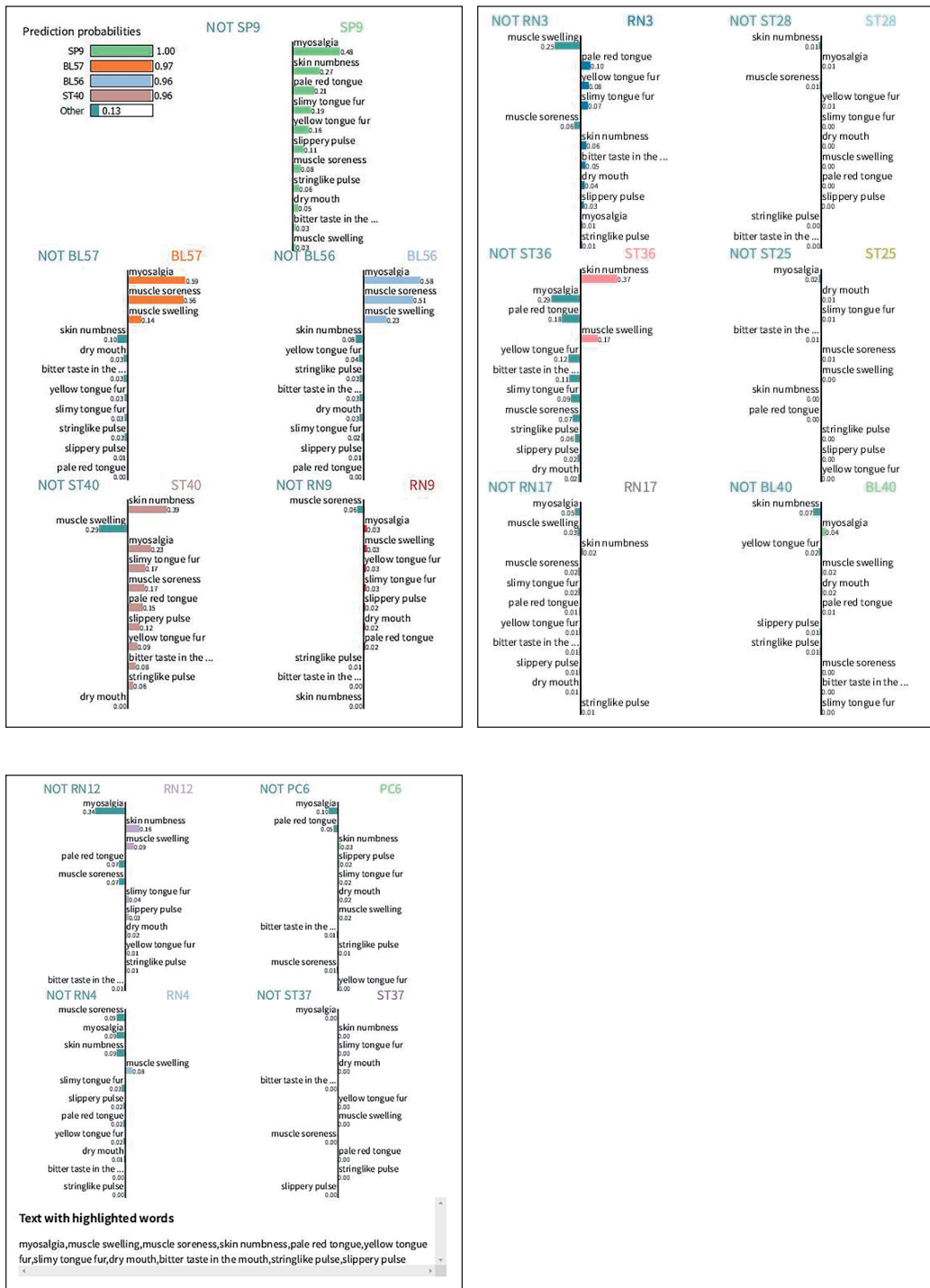
- [1] B.L. Zhang, X.M. Shi, and F.R. Liang, *Theory and Practice of Acupuncture & Moxibustion*, China Press of Traditional Chinese Medicine, 2019 (in Chinese).
- [2] Y.L. Wang, X.M. Shi, T. Efferth, and D. Shang, "Artificial intelligence-directed acupuncture: a review," *Chinese medicine*, vol.17, no.1, pp.1-10, 2022.
- [3] R.C. Deo, "Machine Learning in Medicine," *Circulation*, vol.132, no.20, pp.1920-1930, 2015.
- [4] W. Huang, E. Chen, Q. Liu, Y. Chen, Z. Huang, Y. Liu, and S. Wang, "Hierarchical multi-label text classification: An attention-based recurrent network approach," *Proceedings of the 28th ACM international conference on information and knowledge management*, pp.1051-1060, 2019.
- [5] M.T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp.1135-1144, 2016.
- [6] Streamlit, "Streamlit-A fastest way to build and share data apps," [Online]. Available: <https://streamlit.io>. [Accessed: Apr.15, 2025].
- [7] J. Liang, M.Y. Ming, C.B. Wang, X.L. Lv, Z.R. Sun, and H.N. Yin, "Research progress in the integration of machine learning and the science of acupuncture and moxibustion," *Acupuncture Research*, vol.46, no.6, pp.460-463, 2021.
- [8] W. Li, B. Li, Y. Guo, T. Zhao, X. Pan, and B. Chen, "Feasibility Analysis of Applying Data Mining Technology to the Researches of Major Problems in Acupuncture and Moxibustion," *Acta Chinese Medicine and Pharmacology*, vol.46, no.6, pp.6-9, 2018.
- [9] J. Yu, Y. Jiang, M. Tu, B. Liao, and J. Fang, "Investigating prescriptions and mechanisms of acupuncture for chronic stable angina pectoris: an association rule mining and network analysis study," *Evidence-Based Complementary and Alternative Medicine*, vol.10, pp.1-11, 2020.
- [10] X. Zhong, Y. Jia, D. Li, and X. Zhang, "Classification of acupuncture points based on the Bert model," *Journal of Data Analysis and Information Processing*, vol.9, no.3, pp.123-135, 2021.
- [11] P.H. Lu, J.L. Keng, F.M. Tsai, P.H. Lu, and C.Y. Kuo, "An Apriori Algorithm-Based Association Rule Analysis to Identify Acupoint Combinations for Treating Diabetic Gastroparesis," *Evidence-Based Complementary and Alternative Medicine*, vol.17, pp.1-9, 2021.

- [12] H. Yang, R. Wu, M. Nakata, and Q.W. Ge, "On Application of Machine Learning for Deciding Acupoints in Acupuncture and Moxibustion Treatment," *IEIE Transactions on Smart Processing and Computing*, vol.14, no.2, pp.152-164, 2025.
- [13] C. Cortes, and V. Vapnik, "Support-vector networks," *Machine Learning*, vol.20, pp.273-297, 1995.
- [14] J.R. Quinlan, "Induction of decision trees," *Machine Learning*, vol.1, pp.81-106, 1986.
- [15] M.A. Ganaie, M. Hu, A.K. Malik, M. Tanveer, and P.N. Suganthan, "Ensemble deep learning: A review," *Engineering Applications of Artificial Intelligence*, vol.115, pp.105151, 2022.
- [16] M. Freat, "The Upstart Algorithm: A Method for Constructing and Training Feedforward Neural Networks," *Neural Computation*, vol.2, no.2, pp.198-209, 1990.
- [17] Y. Kim, "Convolutional Neural Networks for Sentence Classification," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp.1746-1751, 2014.
- [18] L. Yao, C.S. Mao, and Y. Luo, "Graph Convolutional Networks for Text Classification," *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, pp.1-8, 2019.
- [19] Z.M. Chen, X.S. Wei, P. Wang, and Y. Guo, "Multi-label image recognition with graph convolutional networks," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.5177-5186, 2019.
- [20] I. Sutskever, O. Vinyals, and Q.V. Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, pp.1-9, 2014.
- [21] A. Zangari, M. Marcuzzo, M. Rizzo, L. Giudice, A. Albarelli, and A. Gasparetto, "Hierarchical Text Classification and Its Foundations: A Review of Current Research," *Electronics*, vol.13, no.7, pp.1-57, 2024.
- [22] D. Koller, and M. Sahami, "Hierarchically Classifying Documents Using Very Few Words," *Proceedings of the Fourteenth International Conference on Machine Learning*, pp.170-178, 1997.
- [23] C.N. Silla, and A. Freitas, "A survey of hierarchical classification across different application domains," *Data mining and knowledge discovery*, vol.22, pp.31-72, 2011.
- [24] S. Lundberg, "A unified approach to interpreting model predictions," *Proceedings of the 31st Conference on Neural Information Processing Systems*, pp.1-10, 2017.
- [25] K. Saito, *Deep Learning from Scratch 2*, O'Reilly Japan, Inc., 2018 (in Japanese).
- [26] Z. Luo, *The Classification Research of the Symptomatic Units of TCM*, M.S. thesis, Shandong University of Traditional Chinese Medicine, 2012 (in Chinese).
- [27] A. Hyodo, *Traditional Chinese medicine meridians and acupoint textbooks*, Shinsei Publishing, 2012 (in Japanese).
- [28] R.M. Yan, *Essence of Yan Runming's 60 years of clinical experience in acupuncture and moxibustion*, China Press of Traditional Chinese Medicine, 2013 (in Chinese).
- [29] S.Z. Gao, and J. Yang, *Therapeutics of Acupuncture and Moxibustion*, China Press of Traditional Chinese Medicine, 2016 (in Chinese).

- [30] M.L. Zhang, and Z.H. Zhou, "A review on multi-label learning algorithms," *IEEE transactions on knowledge and data engineering*, vol.26, no.8, pp.1819-1837, 2013.
- [31] H. Rezatofighi, N. Tsoi, J.Y. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.658-666, 2019.

## Appendix A. LIME Visualization of the Top 15 Predicted Acupoints for Case 1.

(Source: Created by the authors)



## Appendix B. LIME Visualization of the Top 15 Predicted Acupoints for Case 2. (Source: Created by the authors)

