LLM4Sweat: A Trustworthy Large Language Model for Hyperhidrosis Support

Wenjie Lin, Jin Wei-Kocsis

School of Applied and Creative Computing, Purdue University
{lin1790, kocsis0}@purdue.edu

Abstract

While large language models (LLMs) have shown promise in healthcare, their application for rare medical conditions is still hindered by scarce and unreliable datasets for fine-tuning. Hyperhidrosis, a disorder causing excessive sweating beyond physiological needs, is one such rare disorder, affecting 2–3% of the population and significantly impacting both physical comfort and psychosocial well-being. To date, no work has tailored LLMs to advance the diagnosis or care of hyperhidrosis. To address this gap, we present LLM4Sweat, an open-source and domain-specific LLM framework for trustworthy and empathetic hyperhidrosis support. The system follows a three-stage pipeline. In the data augmentation stage, a frontier LLM generates medically plausible synthetic vignettes from curated open-source data to create a diverse and balanced question-answer dataset. In the fine-tuning stage, an open-source foundation model is fine-tuned on the dataset to provide diagnosis, personalized treatment recommendations, and empathetic psychological support. In the inference and expert evaluation stage, clinical and psychological specialists assess accuracy, appropriateness, and empathy, with validated responses iteratively enriching the dataset. Experiments show that LLM4Sweat outperforms baselines and delivers the first open-source LLM framework for hyperhidrosis, offering a generalizable approach for other rare diseases with similar data and trustworthiness challenges.

Introduction

Hyperhidrosis, characterized by excessive sweating beyond thermoregulatory needs, affects approximately 2–3% of the population and can significantly impair both physical comfort and psychosocial well-being (Strutton et al. 2004; Solish, Benohanian, and Kowalski 2007). Current therapeutic protocols span from potent topical antiperspirants to advanced procedural interventions such as botulinum toxin injections, microwave thermolysis (miraDry), iontophoresis, and endoscopic thoracic sympathectomy (Hong, Lupin, and O'Shaughnessy 2012; Gregoriou et al. 2019; Martínez-Hernández et al. 2024). Despite the availability of these treatments, reliable and personalized medical support for diagnosis, treatment selection, and psychosocial management remains limited and unregulated (Parashar, Adlam, and Potts 2023). Notably, the majority of cases represent primary hyperhidrosis, which is inherited and characterized by focal symptoms and has been linked to several genetic loci (Henning, Pedersen, and Jemec 2019).

Beyond clinical progress, parallel efforts in engineering have sought to improve diagnosis and daily management. Examples include polymer-based insole designs using additive manufacturing for plantar hyperhidrosis management (Camargo et al. 2024), rectangular patch antenna electrolyte sensors for sweat composition analysis (Sakhawat, Islam, and Reza 2024), and a wearable system for continuous sweat level monitoring (Bellapukonda, Mohan, and Sahu 2023). Also, daily management devices are proposed, alongside systematic engineering reviews of diagnostic and intervention technologies (Lin and Fang 2022), illustrating the potential for integrating engineering innovations into hyperhidrosis care. However, they remain largely decoupled from intelligent reasoning systems that can deliver holistic diagnosis, treatment guidance, and psychosocial support.

LLMs have shown promising capabilities in a range of healthcare applications, including diagnosis support (Singhal et al. 2023), treatment planning (Zhao et al. 2025), and patient counseling (Qiu and Lan 2024). However, their application adapted to hyperhidrosis and similar rare disorders remains largely absent. Unlike common conditions with abundant datasets, hyperhidrosis suffers from scarce high-quality data and fragmented, unreliable online information. These gaps hinder the trustworthy use of LLMs in clinical care and underscore the need for data-centric frameworks that can turn limited resources into reliable medical knowledge.

To address these limitations, we present LLM4Sweat, the first open-source LLM framework designed for trustworthy hyperhidrosis support. It overcomes the aforementioned challenges through a three-stage pipeline. Our work makes three key contributions: (1) we construct the hyperhidrosis question-answer dataset from open-source data and generate medically plausible synthetic vignettes by a frontier LLM; (2) we introduce a closed-loop fine-tuning and expert-in-the-loop evaluation pipeline that adapts open-source lightweight LLM to perform three integrated tasks, provide diagnosis, recommend treatment, and psychological support, within a unified framework; and (3) experiments show that our framework outperforms baselines and demonstrates strong

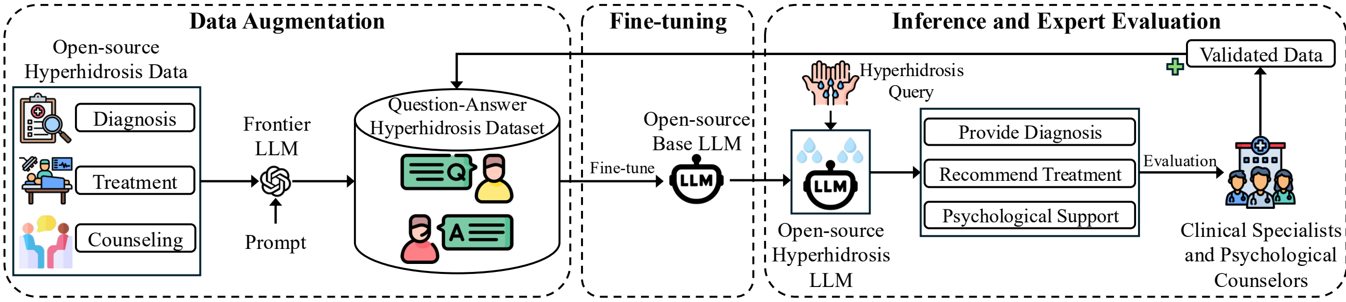

Figure 1: Overview of our proposed LLM4Sweat framework.

potential for clinical deployment. Beyond addressing a critical gap in LLM adaptation for hyperhidrosis, it also offers a transferable solution to tackle data scarcity and trustworthiness challenges across other rare medical conditions.

Related Work

Clinical Diagnosis and Treatment of Hyperhidrosis

The diagnosis of hyperhidrosis is primarily clinical, relying on patient history and physical examination, though these assessments are often subjective in practice (Henning et al. 2021). Once identified, treatment typically follows a stepwise approach, beginning with topical aluminum chloride hexahydrate and advancing to oral anticholinergics, iontophoresis, botulinum toxin injections, or device-based options such as miraDry (Hong, Lupin, and O'Shaughnessy 2012; Gregoriou et al. 2019). For severe and refractory cases, surgical intervention with endoscopic thoracic sympathectomy may be considered, although it carries a significant risk such as compensatory sweating (Martínez-Hernández et al. 2024).

Engineering Solutions for Hyperhidrosis

Recent work has highlighted engineering-driven diagnostic and monitoring innovations. Sakhawat et al. designed an antenna-based electrolyte sensor using Whatman filter paper to detect sodium and chloride concentrations in sweat for palmoplantar hyperhidrosis patients (Sakhawat, Islam, and Reza 2024). 3D-printed thermoplastic polyurethane insoles have been developed to manage plantar hyperhidrosis by improving comfort and reducing moisture retention (Camargo et al. 2024). Hyun et al. leveraged machine learning models to classify hyperhidrosis types and predict compensatory hyperhidrosis level after surgery with high accuracy (Hyun et al. 2023). Bellapukonda et al. applied logistic regression to predict sweat levels with high accuracy based on environmental, physiological, and demographic variables, (Bellapukonda, Mohan, and Sahu 2023). Lin and Fang (Lin and Fang 2022) provided a comprehensive review of current and potential hyperhidrosis interventions from both medical and engineering perspectives, proposing smart wearable devices for daily management.

Synthetic Data in Healthcare

Synthetic data generation has emerged as a solution for medical domains with limited annotated datasets (Peng et al. 2023). GAN-based models have also been used to expand rare disease datasets (Kazeminia et al. 2020). Diffusion models have been successfully applied to dermatological imaging tasks, producing high-fidelity synthetic skin lesion data to improve classifier robustness (Farooq et al. 2024). For text-based healthcare tasks, LLM-driven augmentation has improved performance in low-resource clinical NLP applications (Barr et al. 2025).

LLMs in Medical Applications

Recent work has shown that state-of-the-art LLMs such as GPT and Llama achieve competitive performance in clinical reasoning, treatment recommendation, and patient communication (Singhal et al. 2023). LLMs have also been used for empathetic patient support (Alanezi 2024). Psychological distress is also a key challenge in rare conditions such as hyperhidrosis (Solish, Benohanian, and Kowalski 2007), underscoring the need for both accurate medical guidance and empathetic support.

Building on these advances, our framework leverages a frontier LLM to generate medically plausible vignettes from open-source hyperhidrosis-related data, converting scarce and noisy real-world resources into a structured and balanced dataset. This enriched dataset enables reliable fine-tuning, equipping the model to provide trustworthy diagnostic reasoning, personalized treatment recommendations, and empathetic counseling. Furthermore, in the inference and expert evaluation stage, outputs by the fine-tuned model from users' queries are reviewed by clinical and psychological specialists, creating a feedback loop that further enhances accuracy, appropriateness, and trustworthiness.

Methodology

Figure 1 provides an overview of the LLM4Sweat framework, which comprises closed-loop three stages: data augmentation, fine-tuning, and inference with expert-in-the-loop evaluation.

Model	Diagnosis				Treatment				Overall			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
Llama-3.2-1B	0.425	0.595	0.434	0.425	0.425	0.340	0.380	0.340	0.425	0.543	0.421	0.403
Llama-3.2-1B (**LLM4Sweat**) **w/o Expert Eval**	**0.900**	**0.944**	**0.865**	**0.887**	**0.725**	**0.736**	**0.770**	**0.735**	**0.813**	**0.830**	**0.816**	**0.807**
Llama-3.2-1B (**LLM4Sweat**)	**0.925**	**0.942**	**0.909**	**0.918**	**0.825**	**0.848**	**0.842**	**0.830**	**0.875**	**0.899**	**0.880**	**0.879**
Llama-3.2-3B	0.475	0.798	0.491	0.569	0.700	0.917	0.683	0.753	0.588	0.856	0.587	0.670
Llama-3.2-3B (**LLM4Sweat**) **w/o Expert Eval**	**0.825**	**0.821**	**0.828**	**0.819**	**0.825**	**0.902**	**0.822**	**0.856**	**0.825**	**0.861**	**0.827**	**0.839**
Llama-3.2-3B (**LLM4Sweat**)	**0.875**	**0.883**	**0.850**	**0.861**	**0.925**	**0.941**	**0.917**	**0.923**	**0.900**	**0.902**	**0.898**	**0.900**

Table 1: Comparison of our LLM4Sweat framework against baseline models for hyperhidrosis diagnosis and treatment task performance. We report accuracy (Acc), precision (Prec), recall (Rec), and F1 scores (F1) across models.

Data Augmentation Stage

We begin with a self-curated set of open-source hyperhidrosis-related data

$$\mathcal{D}_{\text{real}} = \{(q_i, a_i, T_i^j)\}_{i=1}^{N} \quad (1)$$

where q_i is a patient query, a_i is the corresponding answer, and T_i^j denotes the task label with $j \in \{1, 2, 3\}$ corresponding to three tasks *diagnosis*, *treatment*, and *counseling*, respectively. $\mathcal{D}_{\text{real}}$ is for testing the fine-tuned models.

To address the low-resource nature of hyperhidrosis-specific data, we use a frontier LLM f_{θ_0} with carefully designed prompts to generate a vigenette synthetic dataset:

$$\mathcal{D}_{\text{syn}} = \{(q'_k, a'_k, T_k^j)\}_{k=1}^{M} \sim f_{\theta_0} \quad (2)$$

where q'_k denotes synthetic queries in vignettes and a'_k represents the synthetic answers. They form synthetic question–answer pairs for one of the defined task categories T_k^j. Besides, we ensure balanced generation so that the three task types T^1, T^2, and T^3 are proportionally represented in $\mathcal{D}_{\text{syn}}$.

Fine-Tuning Stage

We fine-tune an open-source foundation model g_θ to the hyperhidrosis domain via supervised fine-tuning on $\mathcal{D}_{\text{syn}}$. The training objective maximizes the conditional likelihood of the correct answer given the question and task:

$$\theta^* = \arg\max_{\theta} \sum_{(q,a,T^j)\in\mathcal{D}_{\text{syn}}} \log p_\theta(a \mid q, T^j)$$

Inference and Expert Evaluation Stage

At inference time, a user query q^{user} is assigned a task label $T^{j,\text{pred}}$ based on the query context. The fine-tuned model g_{θ^*} generates a response:

$$\hat{a} = g_{\theta^*}(q^{\text{user}}, T^{j,\text{pred}}),$$

All generated outputs are validated by clinical specialists and psychological counselors for accuracy, appropriateness, and empathy. Validated output $\hat{a}_v$ will be fed into the synthetic dataset $\mathcal{D}_{\text{syn}}$ created in the data augmentation stage to further fine-tune g_{θ^*}. This stage creates a dynamic training loop where the fine-tuned model is iteratively refined with validated expert feedback, improving both factual correctness and trustworthiness over successive cycles.

Experiments

We evaluated and demonstrated the performance of the proposed LLM4Sweat framework across diagnosis and treatment tasks under low-resource conditions. The workflow is present in Figure 2. All models operated in a controlled evaluation environment, with prompt formats standardized across tasks to ensure fairness, and inference executed on a server with one NVIDIA A100 GPU to support efficient training, testing, and fair comparison.

Datasets

Testing Dataset We curated a balanced evaluation benchmark of 80 multiple-choice questions (40 diagnosis, 40 treatment) from open-access hyperhidrosis resources, including International Hyperhidrosis Society (IHHS) (International Hyperhidrosis Society 2025), National Health Service (National Health Service 2025), Mayo Clinic (Mayo Clinic 2025a,b), DermNet (DermNet 2025), and MedlinePlus (MedlinePlus 2025). Questions cover both the diagnosis of hyperhidrosis and evidence-based treatment strategies, ensuring coverage of both clinical reasoning and therapeutic guidance.

Synthetic Training Dataset To address data scarcity, we generated 180 synthetic vignettes (90 diagnosis, 90 treatment) using a frontier LLM (GPT-5) based on the real testing dataset. All generated cases were prompted and filtered

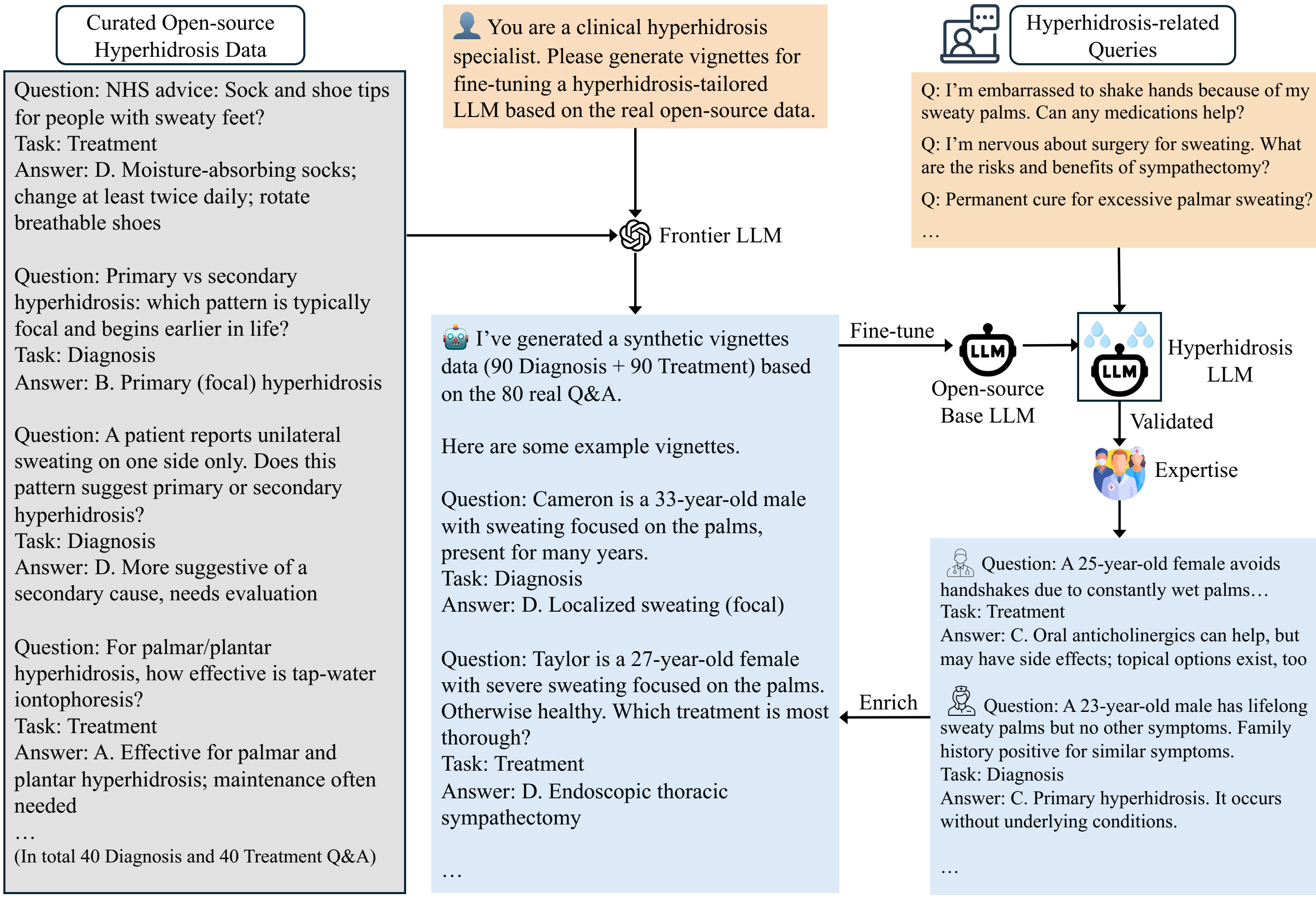

Figure 2: A demonstration of LLM4Sweat workflow.

for medical plausibility. This synthetic training dataset is for fine-tuning the open-source base LLMs before the inference and expert evaluation stage.

Dynamic Training Dataset with Validated Clinical Data In addition to the static synthetic dataset, we incorporated iterative expert feedback cycles into the framework. In real-world applications, after inferences for real users' queries about hyperhidrosis, the model outputs are reviewed by domain experts, and validated responses are fed back into the dataset. In this experiment, we used the state-of-the-art LLM to represent specialists for practice. This process created a dynamic training corpus that improved alignment with clinical expertise and further fine-tuned the base model. For one cycle inference and expert evaluation, we got 40 outputs and corresponding 40 validated responses (20 for diagnosis and 20 for treatment) from specialists to enrich the synthetic training set, which can be used to fine-tune base models.

Baselines and Our Models

We evaluated our framework against open-source lightweight foundation models that had not been adapted to the hyperhidrosis domain: Llama-3.2-1B and Llama-3.2-3B. They served as baseline comparisons, reflecting the performance of general-purpose LLMs when directly applied to rare medical conditions without domain-specific fine-tuning.

In contrast, our proposed LLM4Sweat framework first fine-tuned these same base models on the dynamic synthetic training dataset composed of synthetic hyperhidrosis-related vignettes and also validated expert feedback. This enabled the models to learn domain-specific reasoning while maintaining computational efficiency.

Performance Metrics

For both diagnosis and treatment tasks, we evaluated models using four standard metrics: accuracy, precision, recall, and F1 scores for both diagnosis and treatment tasks. Accuracy provides an overall measure of correctness across all predictions, while Precision captures the proportion of model outputs that are clinically correct among all predicted positives. Recall quantifies the ability to identify true positives, a critical factor in medical contexts where missing a correct diagnosis or treatment option may have serious implications. The F1 score balances Precision and Recall, offering a more stable indicator of performance.

These metrics collectively allowed us to assess not only the correctness of model predictions but also their clinical robustness and reliability. In particular, Recall and F1 are especially important for hyperhidrosis-related tasks, where overlooking appropriate treatment guidance or misclassifying diagnosis categories could compromise patient care. By reporting all four measures, we provided a comprehensive evaluation of model performance and its potential for trustworthy clinical deployment.

Hyperparameter Settings

We trained the base models using LoRA-based supervised fine-tuning implemented in the Hugging Face PEFT library. This allowed us to efficiently adapt lightweight foundation models while keeping the number of trainable parameters small, making the framework more scalable for future deployment. We present other key hyperparameters here. The learning rate was tuned over $\{5 \times 10^{-6}, 5 \times 10^{-5}, 2 \times 10^{-4}, 1 \times 10^{-3}\}$, and training epochs were in the range of $\{1, 3, 5\}$ across models. A LoRA dropout rate of 0.05 was applied to improve generalization. For all the inference, we used the same temperature $= 0.7$ and top-$p = 0.9$ across all models.

These fine-tuning configurations and hyperparameter settings provided a balance between training stability, efficiency, and performance, enabling fair comparison between baselines and fine-tuned models under the LLM4Sweat framework.

Results

Comparison to Baselines

Table 1 presents the performance of LLM4Sweat compared with raw baseline models on both diagnosis and treatment tasks. Baseline foundation models do not perform well under hyperhidrosis-specific evaluation, confirming that general-purpose LLMs struggle in rare, data-scarce medical domains. The smaller model, Llama-3.2-1B, achieves only 0.425 overall accuracy, with balanced but uniformly low performance in diagnosis (0.425) and treatment (0.425). The Llama-3.2-3B baseline shows modest improvements, particularly in treatment (0.700), but overall performance still remains insufficient (0.588). These results demonstrate that, without targeted fine-tuning, both LLMs fail to provide accurate medical decision support. Beyond accuracy, the baseline models also exhibit instability across precision, recall, and F1. For Llama-3.2-1B, precision remains moderate (0.543 overall), but both recall (0.421) and F1 (0.403) are very low, reflecting frequent failures to capture true positives and poor balance between sensitivity and specificity. Llama-3.2-3B achieves stronger precision (0.856 overall), particularly in treatment tasks (0.917), yet recall (0.587) and F1 (0.670) remain insufficient, indicating that while it can produce correct answers in some cases, it struggles to consistently identify them. These results also confirm that they fail to provide stable and clinically trustworthy performance across all evaluation metrics.

LLM4Sweat substantially outperforms baseline models across both tasks. After fine-tuning on the generated synthetic dataset, the 1B model achieves 0.925 diagnosis accuracy and 0.825 treatment accuracy, yielding an overall score of 0.875, more than doubling baseline accuracy. The 3B model improves from 0.588 to 0.900 overall accuracy, with diagnosis accuracy rising from 0.475 to 0.875 and treatment from 0.700 to 0.925. Importantly, gains extend beyond accuracy: both precision and recall increase, leading to F1 scores above 0.85 across all subtasks. This suggests that LLM4Sweat not only improves correctness but also achieves a better balance between sensitivity (capturing true positives) and specificity (avoiding false positives), both critical for clinical safety and trustworthiness.

Taken together, these findings validate the effectiveness of our data-centric fine-tuning strategy for our framework. By augmenting scarce real-world data with medically plausible synthetic cases and fine-tuning LLMs on the enriched data from expert evaluation, LLM4Sweat successfully transforms weak baseline models into high-performing, clinically relevant assistants. Last but not least, the results illustrate that even modestly sized LLMs, when fine-tuned carefully, can achieve competitive performance on specialized medical tasks, making them practical for deployment in resource-constrained environments.

Ablation Studies

To investigate the contribution of individual components, we first compare unadapted baseline models against their fine-tuned counterparts. As shown in Table 1, for the 1B model, fine-tuning doubles performance, lifting overall accuracy to 0.813 and substantially improving both diagnosis (0.900 vs. 0.425) and treatment (0.725 vs. 0.425). Similarly, the 3B baseline achieves only 0.588 overall accuracy, with diagnosis accuracy at 0.475 and treatment at 0.700. After fine-tuning, the 3B model reaches 0.825 overall, 0.825 in diagnosis, and 0.825 in treatment, demonstrating that synthetic data augmentation and domain adaptation are the critical drivers of performance gains.

To further assess the contribution of expert-in-the-loop refinement, we compare fine-tuned models without and with expert evaluation. The results show that expert evaluation contributes measurable and consistent improvements across both model sizes. On the 1B model, expert refinement raises treatment accuracy from 0.725 to 0.825 and overall accuracy from 0.813 to 0.875, while maintaining high diagnosis accuracy (0.925). On the 3B model, diagnosis improves from 0.825 to 0.875, treatment from 0.825 to 0.925, and overall accuracy from 0.825 to 0.900. In both cases, F1 scores increase in parallel, indicating that expert validation enhances not just raw accuracy but also the stability and consistency of medical predictions.

These findings highlight two key insights of LLM4Sweat. First, fine-tuning with enriched synthetic data is an important driver of performance gains, elevating baseline models from near-random to clinically relevant levels. Second, expert-in-the-loop refinement provides the final layer of trustworthiness, ensuring that outputs are aligned with medical plausibility and clinical expectations. This mirrors real-world deployment needs: data-driven adaptation provides a strong foundation, but structured expertise remains essential for scaling applications.

Discussions and Limitations

The results highlight the effectiveness of the LLM4Sweat framework in addressing the challenges of hyperhidrosis support, a domain traditionally limited by scarce and noisy data. By combining generative augmentation with domain-specific fine-tuning and expert evaluation, the system achieved substantial performance improvements across multiple open-source base LLMs, with two Llama3.2 models fine-tuned reaching overall accuracy and strong gains in both diagnosis and treatment. Smaller variants, while starting from weaker baselines, benefited markedly from fine-tuning and further improved through inference-and-expert-evaluation cycles. These findings confirm that both data-centric (augmentation, fine-tuning) and process-centric (structured inference, expert validation) stages are essential to adapt general-purpose LLMs into trustworthy domain-specific medical assistants.

Despite these promising outcomes, several limitations remain. First, the current evaluation relies on 80 curated multiple-choice questions, which, though carefully balanced, may not capture the full complexity of real-world patient interactions, including multi-modal symptoms, longitudinal histories, and emotional nuance. Second, expert evaluation was conducted in limited cycles; broader validation across diverse clinicians and patient populations is necessary to ensure robustness and fairness. Finally, deployment in clinical settings involves non-trivial challenges beyond technical accuracy: integration with electronic health records, regulatory approval, and alignment with ethical standards of care. Addressing these challenges will be essential to transition LLM4Sweat from a research-stage prototype into a clinically deployable decision support and patient guidance system, and the framework offers a blueprint for extending this approach to other rare and underrepresented medical conditions.

Path to Deployment

As demonstrated in Figure 2, the LLM4Sweat workflow shows a clear pathway toward deployment in real-world applications. The following aspects are essential for a trustworthy and scalable deployment.

- **Expert Evaluation:** The immediate step centers on rigorous expert-in-the-loop evaluation. In this work, it is done by the frontier LLM. In the future, the system will undergo structured evaluations with dermatologists, hyperhidrosis experts, and psychologists specializing in hyperhidrosis to assess diagnostic reliability, treatment appropriateness, and empathetic counseling quality. These evaluations will validate accuracy and guide iterative refinement of the fine-tuned model and data augmentation pipeline, ensuring that LLM4Sweat evolves into a clinically credible tool aligned with real-world reasoning and patient needs.
- **Progressive Integration:** Building on this foundation, LLM4Sweat can be gradually introduced into practice through both clinician-facing and patient-facing applications. For physicians, it may serve as a clinical decision support system, surfacing AI-generated suggestions alongside patient records while preserving clinician authority in decision-making. For patients, secure mobile or web interfaces could provide educational content, coping strategies, and personalized lifestyle guidance between consultations, and eventually extend into treatment recommendations under medical supervision. Deployment will follow established standard approval, medical constraints, and regulated certification.
- **Generalization Ability:** Beyond hyperhidrosis, the proposed methodology is transferable to other rare medical conditions, offering a scalable way to deploy trustworthy LLM-based assistants in underserved areas of medicine.

Conclusions

We present LLM4Sweat, the first open-source LLM framework tailored for hyperhidrosis. By combining generative data augmentation, domain-specific fine-tuning, and expert-in-the-loop refinement, our approach addresses the challenges of data scarcity and trustworthiness. Experimental results across LLMs demonstrate that LLM4Sweat substantially outperforms raw baselines, achieving high accuracy in both diagnosis and treatment recommendation while improving stability and sensitivity, highlighting the necessity of both data-driven adaptation and expertise to ensure high and reliable performance.

Looking ahead, LLM4Sweat offers a generalizable and deployable method for applying LLMs in other rare medical conditions where curated data are insufficient but clinical need is high. Future work will also extend to richer patient cases, longitudinal scenarios, and multi-modal studies.

References

Alanezi, F. 2024. Assessing the effectiveness of ChatGPT in delivering mental health support: a qualitative study. *Journal of multidisciplinary healthcare*, 461–471.

Barr, A. A.; Quan, J.; Guo, E.; and Sezgin, E. 2025. Large language models generating synthetic clinical datasets: a feasibility and comparative analysis with real-world perioperative data. *Frontiers in Artificial Intelligence*, 8: 1533508.

Bellapukonda, P.; Mohan, R. N. V. J.; and Sahu, B. 2023. Predicting Sweat Levels to Detect Hyperhidrosis: A Logistic Regression Approach. In *2023 14th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–8. IEEE.

Camargo, F. P.; Gomes, D. R.; Alves, L. R.; and Corrêa, C. E. 2024. Development of polymer insole in additive manufacturing for people with plantar hyperhidrosis. *Disciplinarum Scientia. Série: Naturais e Tecnológicas*, 25(2): 15–26.

DermNet. 2025. Hyperhidrosis. https://dermnetnz.org/topics/hyperhidrosis/. Accessed: 2025-08.

Farooq, M. A.; Yao, W.; Schukat, M.; Little, M. A.; and Corcoran, P. 2024. Derm-t2im: Harnessing synthetic skin lesion data via stable diffusion models for enhanced skin disease classification using vit and cnn. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1–5. IEEE.

Gregoriou, S.; Sidiropoulou, P.; Kontochristopoulos, G.; and Rigopoulos, D. 2019. Management strategies of palmar hyperhidrosis: challenges and solutions. *Clinical, cosmetic and investigational dermatology*, 733–744.

Henning, M.; Pedersen, O.; and Jemec, G. 2019. Genetic disposition to primary hyperhidrosis: a review of literature. *Archives of dermatological research*, 311(10): 735–740.

Henning, M. A.; Thorlacius, L.; Ibler, K. S.; and Jemec, G. B. 2021. How to diagnose and measure primary hyperhidrosis: a systematic review of the literature. *Clinical Autonomic Research*, 31(4): 511–528.

Hong, C.-H. H.; Lupin, M.; and O'Shaughnessy, K. F. 2012. Clinical evaluation of a microwave device for treating axillary hyperhidrosis. *Dermatologic Surgery*, 38(5): 728–735.

Hyun, K. Y.; Kim, J. J.; Im, K. S.; Lee, B. S.; and Kim, Y. J. 2023. Machine learning analysis of primary hyperhidrosis for classification of hyperhidrosis type and prediction of compensatory hyperhidrosis. *Journal of Thoracic Disease*, 15(9): 4808–4817.

International Hyperhidrosis Society. 2025. Hyperhidrosis. https://www.sweathelp.org/. Accessed: 2025-08.

Kazeminia, S.; Baur, C.; Kuijper, A.; van Ginneken, B.; Navab, N.; Albarqouni, S.; and Mukhopadhyay, A. 2020. GANs for medical image analysis. *Artificial Intelligence in Medicine*, 109: 101938.

Lin, W.; and Fang, Y. 2022. Primary Hyperhidrosis: A Systematic Review of Current Status and Potential Interventions. In *2022 6th International Conference on Universal Village (UV)*, 1–8. IEEE.

Martínez-Hernández, N. J.; Estors-Guerrero, M.; Galbis-Caravajal, J. M.; Hervás-Marín, D.; and Roig-Bataller, A. 2024. Endoscopic thoracic sympathectomy for primary hyperhidrosis: an over a decade-long follow-up on efficacy, impact, and patient satisfaction. *Journal of Thoracic Disease*, 16(12): 8292.

Mayo Clinic. 2025a. Hyperhidrosis Diagnosis and Treatment. https://www.mayoclinic.org/diseases-conditions/hyperhidrosis/diagnosis-treatment/drc-20367173/. Accessed: 2025-08.

Mayo Clinic. 2025b. Hyperhidrosis Overview. https://www.mayoclinic.org/diseases-conditions/hyperhidrosis/symptoms-causes/syc-20367152/. Accessed: 2025-08.

MedlinePlus. 2025. Hyperhidrosis. https://medlineplus.gov/ency/article/007259.htm#:~:text=Hyperhidrosis%20is%20a%20medical%20condition,when%20they%20are%20at%20rest./. Accessed: 2025-08.

National Health Service. 2025. Hyperhidrosis. https://www.nhs.uk/conditions/excessive-sweating-hyperhidrosis/. Accessed: 2025-08.

Parashar, K.; Adlam, T.; and Potts, G. 2023. The impact of hyperhidrosis on quality of life: a review of the literature. *American Journal of Clinical Dermatology*, 24(2): 187–198.

Peng, C.; Yang, X.; Chen, A.; Smith, K. E.; PourNejatian, N.; Costa, A. B.; Martin, C.; Flores, M. G.; Zhang, Y.; Magoc, T.; et al. 2023. A study of generative large language model for medical research and healthcare. *NPJ digital medicine*, 6(1): 210.

Qiu, H.; and Lan, Z. 2024. Interactive agents: Simulating counselor-client psychological counseling via role-playing llm-to-llm interactions. *arXiv preprint arXiv:2408.15787*.

Sakhawat, R.; Islam, M. F.; and Reza, M. T. 2024. Design of a rectangular patch antenna-based electrolyte sensor for palmar hyperhidrosis patients. In *2024 27th International Conference on Computer and Information Technology (ICCIT)*, 3343–3347. IEEE.

Singhal, K.; et al. 2023. Large language models encode clinical knowledge. *Nature*, 620: 172–180.

Solish, N.; Benohanian, A.; and Kowalski, J. W. 2007. Impact of hyperhidrosis on quality of life and its assessment. *Dermatologic Clinics*, 25(4): 447–458.

Strutton, D. R.; Kowalski, J. W.; Glaser, D. A.; and Stang, P. E. 2004. US prevalence of hyperhidrosis and impact on individuals with axillary hyperhidrosis: results from a national survey. *Journal of the American Academy of Dermatology*, 51(2): 241–248.

Zhao, L.; Bai, J.; Bian, Z.; Chen, Q.; Li, Y.; Li, G.; He, M.; Yao, H.; and Zhang, Z. 2025. Autonomous Multi-Modal LLM Agents for Treatment Planning in Focused Ultrasound Ablation Surgery. *arXiv preprint arXiv:2505.21418*.