A Rule-Based Expert System for Improved Endometriosis Phenotyping

Ananya S. Rajagopalan, BS[1], Shefali Setia-Verma, PhD[1]
[1]University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

Abstract

Endometriosis is a complex chronic condition affecting millions of people around the world, yet our capability to diagnose and treat it is lacking. In order to facilitate more efficient and larger-scale genetic and epidemiological studies of endometriosis, more accurate phenotyping approaches are required. Currently, the baseline mis-classification error from endometriosis ICD codes compared to chart reviewed diagnoses is approximately 10%. The objective of this work was to incorporate additional symptom and risk factors data from a patient's electronic health record in a rule-based expert system to improve the classification accuracy of endometriosis cases and controls. A total of 1480 chart-reviewed endometriosis cases and controls (725 cases, 755 controls) from the Penn Medicine BioBank with corresponding symptom/risk factors are used to evaluate this expert system. The overall accuracy of this expert system is currently 74%, with 71.4% case and 76% control classification accuracy. An ablation analysis is also conducted to determine how different components contribute to the overall decision making process. This work demonstrates the potential of using EHR data to robustly identify endometriosis cases, which will enable more scalable endometriosis research studies in the future.

Code Availability

All code used to implement this expert system is available at `https://github.com/Setia-Verma-Lab/Endo_RuleBased_Phenotyping`.

Background

Endometriosis is a complex chronic, inflammatory, and gynecological disease that affects approximately 190 million people of reproductive age globally[1]. It is characterized by the growth of uterine-like tissue outside of the uterus. Symptoms of debilitating pain, infertility, dysuria, dyschezia, and dyspareunia are often associated with endometriosis, but the condition is difficult to pinpoint to the presence of any universal symptom(s) because of its heterogeneous presentation and poorly understood pathology (driving the current average time to diagnosis for endometriosis of 6-10 years from the onset of symptoms[2]). This has motivated genetic and epidemiological research to better characterize the disease, understand its origins and development, and determine better management strategies.

To lessen this knowledge gap, many AI-based approaches have been applied to endometriosis with the goal of developing non-invasive predictive and diagnostic methods[2,3], identifying relevant genetic, phenotypic and environmental associations with endometriosis[3], and decision-support tools that recommend personalized treatment strategies[4,5]. For the issue of unestablished clinically relevant, distinct phenotypes of endometriosis, researchers have applied unsupervised learning to cluster patients based on similar patterns with the hopes of identifying biomarkers[6,7]. It is notable that none of these applications appear to focus on deductive inference approaches for the purposes of classification for research purpose (instead of automating or improving clinical diagnosis). One study used a rule-based expert system, implemented with CLIPSPy, to diagnose a whole host of obstetrics and gynecological diseases, including endometriosis[8]. However, this was framed as being for the use of patients to identify potential diagnoses for themselves or for medical training purposes.

Any large-scale study using endometriosis ICD codes to identify endometriosis cases versus controls for their research must overcome the challenge of there being an average 10% mis-classification rate among both cases and controls (10% of alleged cases are actually controls, and vice versa), which I will hereby refer to as the baseline mis-classification rate. While the endometriosis ICD code may be a decent proxy, it is not synonymous with using a confirmed, gold-standard diagnosis because the presence of the code may merely indicate a suspected diagnosis with further confirmation (via surgery) as an intended next step. This can negatively affect the analyses conducted in these studies, so it is useful to develop approaches that can reduce this mis-classification and more accurately identify endometriosis cases and controls from patient EHRs. We hypothesize that the incorporation of additional symptoms or

risk factors from EHR data may improve the discriminatory ability from actual versus spurious endometriosis cases. In this project, we explore the application of an EHR-based expert system for this purpose.

Materials and Methods

To create this classification system, I used a subset of 1480 chart-reviewed cases and controls (725 cases, 755 controls) from the Penn Medicine BioBank (PMBB). These chart reviews were conducted by clinician collaborators of Dr. Setia-Verma. I obtained symptoms for these 1480 patients by specifying ICD codes for the relevant phenotypes. The particular set of ICD codes (ICD-9 and ICD-10) I used can be found in the GitHub repository for this project.

I used a mixture of CLIPSPy, with Python embedded, to implement my rule-based expert system. In particular, I used CLIPSPy to define the templates, rules, and facts associated with my algorithm and Python to define functions that were called by some of the rules as well as for the evaluation of my system, where I read input from a CSV file and used that to populate case facts (per patient).

I followed the clinical diagnosis framework outlined in a study by Agarwal et al.[9], which is also shown in **Figure 1**. However, I had to make modifications due to constraints in what I was able to extract using ICD codes. Additionally, I excluded any symptoms pertaining to family history or history of procedures, because of inability to obtain that from EHRs. I also excluded any imaging or physical examination related symptoms because that would not be in the ICD codes. The modified algorithm that I referenced after accounting for these constraints is shown in **Figure 2**.

A modification I made to the algorithm was to count the presence of adenomyosis as a symptom under the diagnosis of endometriosis, as opposed to an additional disease. This was because the broadest ICD code definition for endometriosis includes adenomyosis, making it more involved to extract the codes for non-adenomyosis endometriosis. While acknowledging the complexity and nuances around the clustering of adenomyosis and endometriosis[10], I prioritized simplifying my symptom extraction process. As a result, I decided to keep the code for endometriosis as including adenomyosis (in other words, counting the presence adenomyosis as synonymous to the presence of endometriosis) for my evaluation purposes.

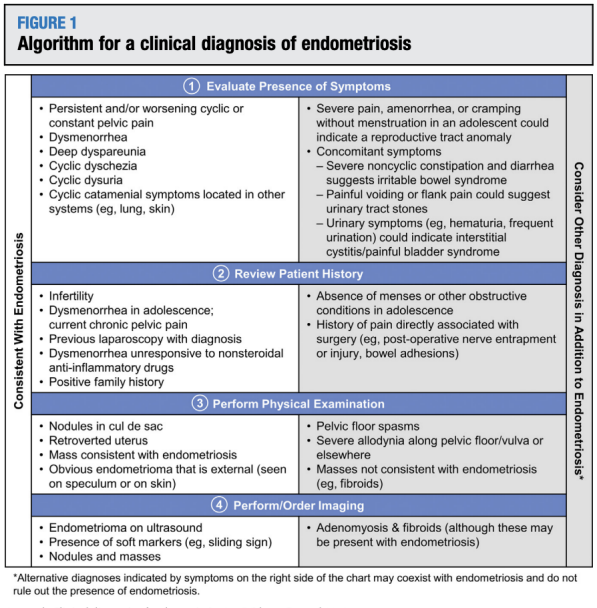

Figure 1: Original clinical diagnosis algorithm[9].

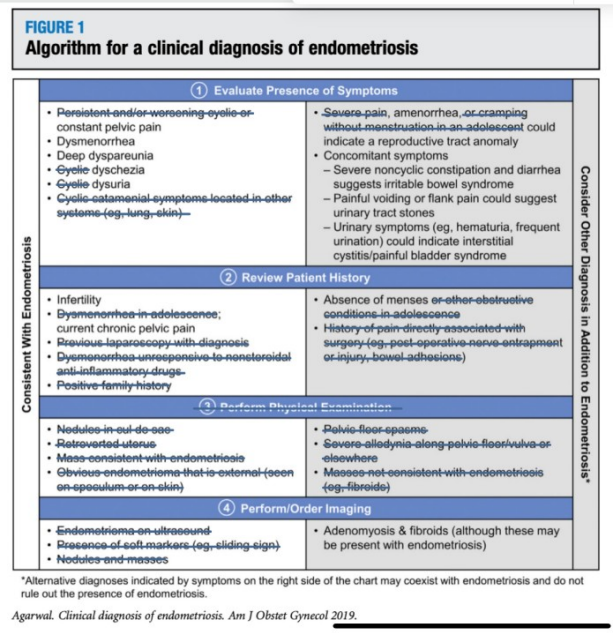

Figure 2: Modified inputs to algorithm, after striking through what ICD codes couldn't cover[9].

Design

0.1 Expert System Inputs: Question Base and Case-Specific Facts

The input to this expert system is simply a list of symptoms; in particular, an indication of whether or not a given patient has each symptom (the system accepts yes, no, or unknown for each symptom). The factors currently incorporated into this expert system are: endometriosis, abdominal pelvic pain, dysmenorrhea, dyspareunia, dyschezia, dysuria, infertility, pelvic perineal pain, amenorrhea, constipation, diarrhea, flank pain, hematuria, frequent urination, irritable bowel syndrome (IBS), interstitial cystitis, and adenomyosis. The expert system uses a question base, implemented (as shown in **Figure 3** and **Figure 4**) as a prompt map in natural language, to solicit case-specific facts from the user—this is the user interface. For the testing data, I have a script to automatically populate case-specific facts from a CSV file.

```
# These are the prompts that will be displayed to the user when requesting various template:slot inputs.
prompt_map = {
    "patient_endo_symptoms:abdominal_pelvic_pain": "Does the patient's EHR contain the abominal pelvic pain ICD code (yes or no
    "patient_endo_symptoms:dysmenorrhea": "Does the patient's EHR contain the dysmenorrhea ICD code (yes or no or unknown)?: ",
    "patient_concomitant_disease_symptoms:interstitial_cystitis" : "Does the patient's EHR contain the interstitial cystitis ICD
    "patient_concomitant_disease_symptoms:ibs" : "Does the patient's EHR contain the irritable bowel syndrome ICD code (yes or
}
```

Figure 3: Code snippet of the prompt map that dynamically obtains patient EHR data from the user. This snippet just shows a sampling of the symptoms that are collected during runtime, the actual code contains all relevant symptoms from the *patient_concomitant_disease_symptoms* and *patient_endo_symptoms* fact templates.

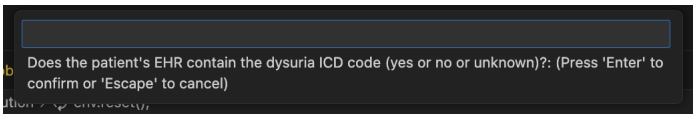

Figure 4: Screenshot of one of the many prompts that pop up during the system's runtime to get case-specific facts from the user. After the user enters a response and hits 'Enter', another question is asked (until all relevant data has been obtained).

0.2 Fact Base

```
# patient concomitant disease symptoms
DEFTEMPLATE_PATIENT_CONCOMITANT_DISEASE_SYMPTOMS = """
(deftemplate patient_concomitant_disease_symptoms
    (slot amenorrhea (type SYMBOL)
        (allowed-symbols yes no unknown))
    (slot constipation (type SYMBOL)
        (allowed-symbols yes no unknown))
    (slot diarrhea (type SYMBOL)
        (allowed-symbols yes no unknown))
    (slot flank_pain (type SYMBOL)
        (allowed-symbols yes no unknown))
    (slot hematuria (type SYMBOL)
        (allowed-symbols yes no unknown))
    (slot frequent_urination (type SYMBOL)
        (allowed-symbols yes no unknown))
    (slot ibs (type SYMBOL)
        (allowed-symbols yes no unknown))
    (slot interstitial_cystitis (type SYMBOL)
        (allowed-symbols yes no unknown))

)
"""
env.build(DEFTEMPLATE_PATIENT_CONCOMITANT_DISEASE_SYMPTOMS)
```

Figure 5: Code snippet for the fact template for *patient_concomitant_disease_symptoms*. An analogous fact template exists called *patient_endo_symptoms*, which has a list of relevant symptoms with allowed symbols being yes, no, or unknown.

```
# Add deffacts that all statuses are unknown initially
DEFFACTS_INITIAL_STATUS = """
(deffacts starting_inclusion_exclusion_facts "Set the initial templates to unknown"
    (endometriosis_inclusion (meets_criteria unknown))
    (concomitant_disease_inclusion (meets_criteria unknown))
    (new_phenotype_status (endometriosis_phenotype_status unknown))
)
"""
env.build(DEFFACTS_INITIAL_STATUS)
```

Figure 6: Code snippet establishing the initial statuses for three fact templates in this knowledge base, set every time the environment is run. The *meets_criteria* slots in the facts *endometriosis_inclusion* and *concomitant_disease_inclusion* will be eventually set to *yes* or *no* by directly evaluating the presence of relevant ICD codes in a patient's EHR. The *new_phenotype_status* fact slot will be set to yes or no depending on the combination of final values from *endometriosis_inclusion* and *concomitant_disease_inclusion*, representing whether a given patient is classified as a case or a control.

0.3 Rule Base

```
DEFRULE_ENDOMETRIOSIS_INCLUSION_CRITERIA_MET = """
(defrule endo-inclusion-criteria-met "Rule to define a person as having sympto
    (patient_endo_symptoms
        (abdominal_pelvic_pain ?v1)
        (dysmenorrhea ?v2)
        (pain_with_sex ?v3)
        (dyschezia ?v4)
        (dysuria ?v5)
        (infertility ?v6)
        (pelvic_perineal_pain ?v7)
        (adenomyosis ?v8)
        (endometriosis ?v9))

    ?f1 <-(endometriosis_inclusion (meets_criteria unknown))

    =>
    (bind ?num_symptoms (count_symptoms ?v1 ?v2 ?v3 ?v4 ?v5 ?v6 ?v7 ?v8 ?v9))
    (if (>= ?num_symptoms 1) then
        (modify ?f1 (meets_criteria yes))
    )
    (if (< ?num_symptoms 1) then
        (modify ?f1 (meets_criteria no))
    )
)
"""
```

Figure 7: Code snippet showing the inclusion rule for endometriosis. This rule fires if at least one symptom in the list is present, by calling a Python function I wrote called *num_symptoms* to count the number of symptoms in a list of arguments that has the slot value "yes". This rule, upon firing, changes the status of the *endometriosis_inclusion* fact to reflect whether or not the threshold is met. There is an analogous rule for determining whether a patient has symptoms consistent with alternative/additional non-Endometriosis diseases (called *concomitant-inclusion-criteria-met*). The threshold of the number of symptoms (in this example, it is set to 1) can be modified to adapt to changing clinical guidelines, and it is a threshold I experimented with to optimize the expert system's performance.

```
DEFRULE_CONCOMITANT_DISEASE_INCLUSION_CRITERIA_NOT_MET = """
(defrule concomitant-inclusion-criteria-not-met "Rule to define a person as havi
    (logical
        (and
            (patient_concomitant_disease_symptoms (amenorrhea ~yes))
            (patient_concomitant_disease_symptoms (constipation ~yes))
            (patient_concomitant_disease_symptoms (diarrhea ~yes))
            (patient_concomitant_disease_symptoms (flank_pain ~yes))
            (patient_concomitant_disease_symptoms (hematuria ~yes))
            (patient_concomitant_disease_symptoms (frequent_urination ~yes))
            (patient_concomitant_disease_symptoms (ibs ~yes))
            (patient_concomitant_disease_symptoms (interstitial_cystitis ~yes))
        )
    )

    ?f1 <-(concomitant_disease_inclusion (meets_criteria unknown))

    =>

    (modify ?f1 (meets_criteria no))
)
"""
env.build(DEFRULE_CONCOMITANT_DISEASE_INCLUSION_CRITERIA_NOT_MET)
```

Figure 8: Code snippet showing the exclusion rule for non-Endometriosis, additional/alternative diseases that a patient might have consistent symptoms with. According to the clinical diagnosis algorithm[9], these are diseases that may be concomitantly occurring in a patient with endometriosis: irritable bowel syndrome, interstitial cystitis, urinary tract stones, or reproductive tract anomaly. This rule fires and determines that a patient does *not* have symptoms consistent with these diseases if none of the listed symptoms are present for a patient, and changes the status of the *concomitant_disease_inclusion* fact to reflect that. An analogous rule exists for determining if a patient has no symptoms consistent with endometriosis, called *endo-inclusion-criteria-not-met*.

```
DEFRULE_ENDOMETRIOSIS_AND_CONCOMITANT_INCLUSION = """
(defrule endo-and-concomitant-inclusion
    (endometriosis_inclusion (meets_criteria yes))
    (concomitant_disease_inclusion (meets_criteria yes))
    ?f1 <-(new_phenotype_status (endometriosis_phenotype_status unknown))
    =>
    (modify ?f1 (endometriosis_phenotype_status yes))
    (println "___________")
    (println "Patient has at least 1 symptom(s) that are consistent with both endometriosis and additional concomitant disease
    (println "___________")
)
"""
env.build(DEFRULE_ENDOMETRIOSIS_AND_CONCOMITANT_INCLUSION)
```

Figure 9: Code snippet of the rule (called *endo-and-concomitant-inclusion*) used to print output to the user when a patient is classified as having symptoms consistent with both endometriosis and concomitant diseases of interest. Three other analogous rules exist for the other combinations of *meets_criteria* slot values for the *endometriosis_inclusion* and *concomitant_disease_inclusion* facts. The *new_phenotype_status* fact is also updated to reflect whether the expert system is ultimately classifying the given patient as an endometriosis case or control.

0.4 Outputs of the Expert System

This system classifies a patient case into one of four categories, which are translated into classes of either an endometriosis case or control: **1)** Patient case exhibits symptoms of only endometriosis (labeled as a case). **2)** Patient case exhibits symptoms of only alternative diseases: interstitial cystitis, IBS, urinary tract stones, or a reproductive tract anomaly (labeled as a control). **3)** Patient case exhibits symptoms of *both* endometriosis and concomitant (interstitial cystitis, IBS, urinary tract stones, or a reproductive tract anomaly) diseases (labeled as a case). **4)** Patient case exhibits symptoms of *neither* endometriosis nor alternative diseases (labeled as a control).

0.5 Integration of Course Topics

Multiple topics covered in class are integrated within this endometriosis phenotyping tool. The **expert system**, a topic discussed extensively during class, employs **deductive reasoning** as its logical framework where general rules (defined by me) are applied to specific case facts (dynamically provided). I chose to represent the knowledge in this expert system as **rules**, given that this a common choice. I used **first order logic implications** to define the rules in this expert system, and inferences were done using forward chaining (which, as covered in class, is based on generalized Modus Ponens to derive new facts from known facts).

To provide a complete expert system tool with a **user interface component**, I have provided two nearly identical versions of the expert system in the GitHub repository, one that has the user interface where a user can enter case facts on their own (dynamically), and one where the case facts are automatically populated by the testing data I use. Other than the user interface component, both versions of the system are identical.

Also, since one hallmark of expert systems is their ability to explain their reasoning (in terms of which facts were fired to arrive at a particular conclusion), this touches on the concept of **explainable AI** that was also covered in class. At the end of each patient case, I have set up my expert system to output all the facts and a high-level justification so that a user can easily interpret the reasoning chain that led to a particular classification. In addition, the ablation analysis is included with the goal of understanding the most important components of the system in its decision-making process, and exploring why certain facts may be more influential than others.

Demonstration

0.6 Expert System Outputs

```
__________
Patient has at least 1 symptom that is consistent only with endometriosis and no symptoms consistent with additional diseases (irritable bowel syndrome
, interstitial cystitis, urinary tract stones, or a reproductive tract anomaly). The purpose of this system is to identify endometriosis cases for larg
e-scale research studies - this is NOT intended to be a formal diagnosis; classification was made based on limited symptoms without lab tests/physical
exam results and further confirmation may be necessary.
__________
(endometriosis_inclusion (meets_criteria yes))
(concomitant_disease_inclusion (meets_criteria no))
(new_phenotype_status (endometriosis_phenotype_status yes))
(patient_endo_symptoms (endometriosis no) (abdominal_pelvic_pain yes) (dysmenorrhea no) (pain_with_sex no) (dyschezia no) (dysuria no) (infertility no)
 (pelvic_perineal_pain no) (adenomyosis no))
(patient_concomitant_disease_symptoms (amenorrhea no) (constipation no) (diarrhea no) (flank_pain no) (hematuria no) (frequent_urination no) (ibs no) (
interstitial_cystitis no))
Total facts: 5
```

Figure 10: Example system output for classification of symptoms only consistent with endometriosis (a case).

```
__________
Patient has at least 1 symptom(s) that are consistent with both endometriosis and additional concomitant disease (irritable bowel syndrome, interstitia
l cystitis, urinary tract stones, or a reproductive tract anomaly). The purpose of this system is to identify endometriosis cases for large-scale resea
rch studies - this is NOT intended to be a formal diagnosis; classification was made based on limited symptoms without lab tests/physical exam results
and further confirmation may be necessary.
__________
(endometriosis_inclusion (meets_criteria yes))
(concomitant_disease_inclusion (meets_criteria yes))
(new_phenotype_status (endometriosis_phenotype_status yes))
(patient_endo_symptoms (endometriosis no) (abdominal_pelvic_pain yes) (dysmenorrhea no) (pain_with_sex no) (dyschezia yes) (dysuria no) (infertility no
) (pelvic_perineal_pain no) (adenomyosis no))
(patient_concomitant_disease_symptoms (amenorrhea no) (constipation yes) (diarrhea no) (flank_pain yes) (hematuria no) (frequent_urination yes) (ibs no
) (interstitial_cystitis no))
Total facts: 5
```

Figure 11: Example system output for classification of symptoms consistent with endometriosis and additional diseases of interest (interstitial cystitis, IBS, urinary tract stones, or a reproductive tract anomaly) (a case).

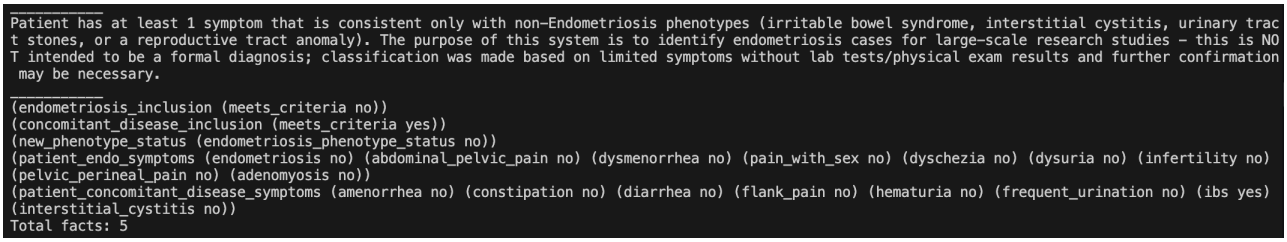

Figure 12: Example system output for classification of symptoms only consistent with non-endometriosis diseases of interest (a control).

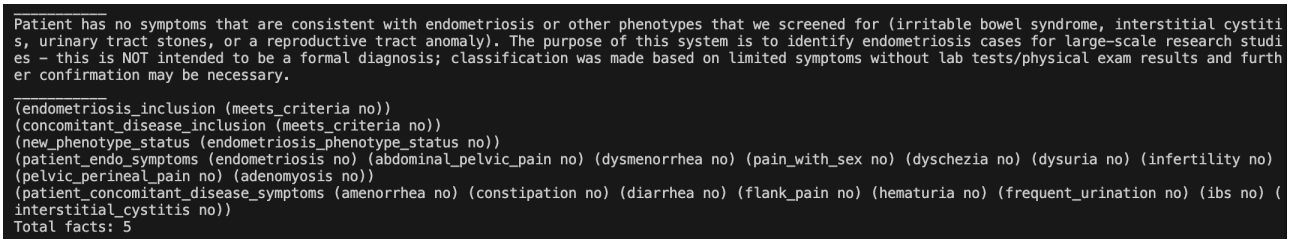

Figure 13: Example system output for classification of symptoms consistent with neither endometriosis or additional diseases of interest (a control).

Here, I incorporated the component of a justifier that can often accompany expert systems, where with each output, I provide a brief explanation of why a particular classification was made and output the list of facts used so that the user can easily infer the decision-making process. The list of facts reveals which symptoms were stored inside which template (of which there are two) so that a user can also see which symptoms being present contributed to the different classifications (of being consistent with endometriosis and/or additional diseases).

Depending on the classification of the system, the output classification is accompanied by a 'disclaimer', underscoring the preliminary and non-comprehensive basis on which the expert system is operating. That is, the expert system is not at all meant to provide an official diagnosis for a patient, as its intended application is for phenotyping as opposed to clinical diagnostic purposes. In addition, the expert system currently does not incorporate laboratory test or physical exam results in its knowledge base, so the incompleteness of the classification is stated.

0.7 Results

Actual Class	**Predicted Class**	
	Case	**Control**
Actual Case	518	207
Actual Control	182	573

Table 1: Confusion Matrix when at least 2 endometriosis symptoms and at least n concomitant symptoms are present. Overall accuracy: 73.7%, Cases accuracy: 71.4%, Controls accuracy: 75.9%.

Actual Class	**Predicted Class**	
	Case	**Control**
Actual Case	698	27
Actual Control	355	400

Table 2: Confusion Matrix when at least 1 endometriosis symptom and at least n concomitant symptoms are present. Overall accuracy: 74.2%, Cases accuracy: 96.3%, Controls accuracy: 47.0%.

The overall classification accuracy of this expert system is around 74%, as seen in **Table 1** and **Table 2**; however, the individual classification accuracies of cases and controls vary depending on the endometriosis rule structure. In **Table 2**, we can see that the case classification accuracy of our system shows an improvement of around 6% compared to the baseline. However, this may be more of an over-fitting effect instead of an improvement in the underlying

discrimination, since the improvement is not maintained across other rule configurations such as in **Table 1**. This does still suggest, though, that the ICD codes additionally considered can help narrow down the cases beyond just the endometriosis ICD code and more accurately identify cases for large-scale studies. However, the overall expert system classification accuracy of 74% compared to the baseline accuracy of 90% must be reconciled, and suggests that additional data are needed to accurately distinguish between cases and controls.

To optimize the performance of this expert system, I experimented with the rules used. In particular, I modified the threshold for the number of symptoms present, that would cause either the endometriosis inclusion rule or the additional diseases rule to fire (e.g., at least 1 symptom present, 2 symptoms present, 3 symptoms, etc.). I found that modifying the symptom number threshold for the endometriosis rules did affect the system's accuracy in classifying endometriosis cases versus controls, but doing so for the concomitant disease rules did not affect accuracy rates. Because of how the cases and controls are defined though, this is expected. To represent the fact that having any number of additional disease symptoms present did not affect the case versus control classification accuracy rates, I captioned the number of concomitant disease symptoms as n in **Table 1**. Interestingly, while increasing the symptom number threshold for the endometriosis rules consistently decreased the overall classification accuracy, it decreased case classification accuracy while increasing control classification accuracy (shown in **Figure 13**). With this in mind, I chose the threshold of 2 endometriosis symptoms being present to best represent the system's typical performance because of the comparable overall accuracy and the similar classification accuracy within endometriosis cases and controls.

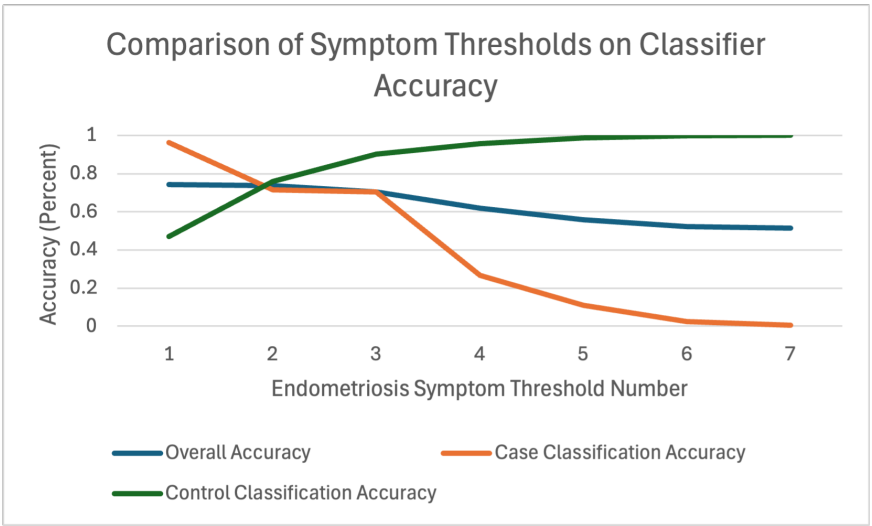

Figure 14: Comparison of symptom number thresholds for endometriosis. Each number x for the symptom threshold corresponds to the rule for endometriosis in the expert being changed to "at least x symptoms must be present" for the rule to fire.

0.8 Ablation Analysis

In order to investigate how influential different components of the expert system are on its decision making process, I manually commented out different facts and observed how a given fact removal affects the system's performance. The following are the corresponding confusion matrices of the system after each fact removal. No other changes were made to the expert system during this comparison process (the minimum number of symptoms for the endometriosis rule to fire was kept at 2, consistently).

Actual Class	Predicted Class	
	Case	Control
Actual Case	383	342
Actual Control	167	588

Table 3: Confusion Matrix when endometriosis is removed. Overall accuracy: 65.6%, Cases accuracy: 52.8%, Controls accuracy: 22.1%.

Actual Class	Predicted Class	
	Case	Control
Actual Case	518	207
Actual Control	182	573

Table 4: Confusion Matrix when Adenomyosis is removed. Overall accuracy: 73.7%, Cases accuracy: 71.4%, Controls accuracy: 75.9%.

	Predicted Class	
	Case	Control
Actual Case	421	304
Actual Control	89	666

Table 5: Confusion Matrix when abdominal pelvic pain is removed. Overall accuracy: 73.4%, Cases accuracy: 58.1%, Controls accuracy: 88.2%.

	Predicted Class	
	Case	Control
Actual Case	502	223
Actual Control	175	580

Table 6: Confusion Matrix when dysmenorrhea is removed. Overall accuracy: 73.1%, Cases accuracy: 69.2%, Controls accuracy: 76.8%.

	Predicted Class	
	Case	Control
Actual Case	516	209
Actual Control	178	577

Table 7: Confusion Matrix when dyspareunia is removed. Overall accuracy: 73.9%, Cases accuracy: 71.1%, Controls accuracy: 76.4%.

	Predicted Class	
	Case	Control
Actual Case	494	231
Actual Control	149	606

Table 8: Confusion Matrix when dyschezia is removed. Overall accuracy: 74.3%, Cases accuracy: 68.1%, Controls accuracy: 80.3%.

	Predicted Class	
	Case	Control
Actual Case	503	222
Actual Control	163	592

Table 9: Confusion Matrix when dysuria is removed. Overall accuracy: 73.9%, Cases accuracy: 69.4%, Controls accuracy: 78.4%.

	Predicted Class	
	Case	Control
Actual Case	497	228
Actual Control	170	585

Table 10: Confusion Matrix when infertility is removed. Overall accuracy: 73.1%, Cases accuracy: 68.6%, Controls accuracy: 77.5%.

	Predicted Class	
	Case	Control
Actual Case	512	213
Actual Control	147	608

Table 11: Confusion Matrix when pelvic perineal pain is removed. Overall accuracy: 75.7%, Cases accuracy: 70.6%, Controls accuracy: 80.5%.

Major takeaways from the ablation analysis:

1. The removal of the endometriosis ICD code significantly decreased the accuracy of the classifier **Table 5**, standing out as a key feature in the expert system's decision making process—this is consistent with what we would expect intuitively.

2. Aside from the endometriosis code, most other symptoms appear to be negligible in their impact on the expert system's overall performance, although:

   (a) The removal of the abdominal pelvic pain fact significantly worsened the accuracy of classifying endometriosis controls (from 71% in **Table 1** to 58% in **Table 5**), indicating its importance in the classification process.

   (b) The removal of the dysmenorrhea, dyschezia, dysuria and infertility facts slightly lowered the expert system case classification accuracies (**Table 6**, **Table 8**, **Table 9**, **Table 10**).

3. For brevity, I chose not to include in this report the second ablation analysis I conducted. The second ablation analysis was identical except the expert system had the endometriosis rule set to at least 1 symptom being present, instead of 2 (the configuration resulting in the performance summarized in **Table 2**). There, the improved case classification accuracy and the decreased controls classification accuracy compared to the baseline performance was maintained across the ablation analysis.

Discussion and Conclusions

0.9 Challenges

A bulk of the challenges I overcame were the inevitable ones that had to do with learning a new way of coding/documenting my work with CLIPSPy and LaTeX/Overleaf, since I was not previously familiar with them. The best I could do here was utilize online documentation[11], but that was (as it always is) a time-intensive process. Particularly with CLIPSPy, I felt myself moving slowly in figuring out the most efficient way to implement things in the way that I wanted; for example, when using a threshold for the number of symptoms present in total instead of combining particular symptoms logically. For a while, I was about to implement a very clunky rule where I wrote out every combination of a given number of symptoms that I wanted as my threshold, but then I realized I could embed a Python function within the CLIPS code. I was also frustrated by the performance of my system when I first implemented it, but then began an incremental process of improvement as I incorporated more data/experimented with the rule structures. I also learned to adjust my mindset and expectations along the way. As I thought critically about what might be behind the expert system's performance, I realized that without the incorporation of even more data—surgical and procedural CPT codes, for example—these results were not unreasonable and not something to be discouraged by.

0.10 Advantages/Disadvantages of this Expert System

As it stands currently, this expert system does not appear to actually improve over the baseline endometriosis ICD codes in accurately identifying true endometriosis cases and controls. This may be in part due to the limitations of ICD code specificity when trying to implement the clinical diagnosis algorithm for endometriosis (e.g., there is no ICD code for *cyclic* dysuria or dyschezia, or information about family history). In addition, PMBB data is limited to patients above the age of 18, so information from their EHR encoding any symptoms specific to adolescence (also part of the clinical diagnosis algorithm[9]) was difficult to obtain. While the classification accuracy of endometriosis cases improved by around 6% (**Table 2**) in a particular configuration of the expert system rules, this seems more likely due to an over-fitting effect as opposed to a truly improved discriminatory capability. This does indicate, however, the potential of this rule-based EHR-based approach to improve endometriosis phenotyping. If additional development of this rule-based expert system can more accurately identify endometriosis cases and controls compared to the baseline, this will be a useful tool in allowing larger-scale genomic and epidemiological studies of endometriosis.

0.11 Future Work

To improve the performance of the expert system, I plan to use CPT codes (encoding medical procedures and services) to determine a more confident classification. This would be implemented by considering whether a patient had a surgery and whether the ICD code for endometriosis was added to their EHR before or after the surgery occurred. If the code was added after the surgery occurred, this suggests that an endometriosis classification can be made (since endometriosis is currently officially diagnosed through surgery). An additional layer to this would be to incorporate unstructured patient notes from the EHR and inspect whether an earlier diagnosis of endometriosis was mentioned. This may be especially relevant in cases where a patient was diagnosed with endometriosis through another provider before entering into the Penn Medicine system (if they moved, for example). While the doctor would know about their endometriosis diagnosis, it might not be in their EHR because of health record duplication errors, or because a current provider would not need to reimburse a past endometriosis diagnosis.

In the future, I would also like to improve the user interface process for obtaining case facts dynamically from the user. In particular, I envision the user being queried for symptoms only as needed instead of all simultaneously at the beginning. This would be only if previously provided symptoms are not present in a patient, so the threshold of symptoms for a certain phenotype is not met. This would make it so that the user need not always provide all the patient symptoms.

References

[1] Krina T. Zondervan, Christian M. Becker, and Stacey A. Missmer. Endometriosis. *New England Journal of Medicine*, 382(13):1244–1256, March 2020.

[2] Anat Goldstein and Shani Cohen. Self-report symptom-based endometriosis prediction using machine learning. *Scientific Reports*, 13(1):5499, April 2023.

[3] Brintha Sivajohan, Mohamed Elgendi, Carlo Menon, Catherine Allaire, Paul Yong, and Mohamed A. Bedaiwy. Clinical use of artificial intelligence in endometriosis: a scoping review. *npj Digital Medicine*, 5(1):109, August 2022.

[4] Brie Dungate, Dwayne R Tucker, Emma Goodwin, and Paul J Yong. Assessing the Utility of artificial intelligence in endometriosis: Promises and pitfalls. *Women's Health*, 20:17455057241248121, January 2024.

[5] Giulia Emily Cetera, Alberto Eugenio Tozzi, Valentina Chiappa, Isabella Castiglioni, Camilla Erminia Maria Merli, and Paolo Vercellini. Artificial Intelligence in the Management of Women with Endometriosis and Adenomyosis: Can Machines Ever Be Worse Than Humans? *Journal of Clinical Medicine*, 13(10):2950, May 2024.

[6] Andrew Horne. Big data: closing the endometriosis knowledge gap. *Healthcare News Scotland Limited*, April 2023.

[7] Iñigo Urteaga, Mollie McKillop, and Noémie Elhadad. Learning endometriosis phenotypes from patient-generated data. *npj Digital Medicine*, 3(1):88, June 2020.

[8] Mohammed F. El-Habibi, Mosa M. M. Megdad, Mohanad H. Al-Qadi, Mohammed J. A. AlQatrawi, Raed Z. Sababa, and Samy S. Abu-Naser. A proposed expert system for obstetrics & gynecology diseases diagnosis. *International Journal of Academic Multidisciplinary Research (IJAMR)*, 6(5):305–321, 2022.

[9] Sanjay K. Agarwal, Charles Chapron, Linda C. Giudice, Marc R. Laufer, Nicholas Leyland, Stacey A. Missmer, Sukhbir S. Singh, and Hugh S. Taylor. Clinical diagnosis of endometriosis: a call to action. *American Journal of Obstetrics and Gynecology*, 220(4):354.e1–354.e12, April 2019.

[10] Jessica Ottolina, Roberta Villanacci, Sara D'Alessandro, Xuemin He, Giorgia Grisafi, Stefano Maria Ferrari, and Massimo Candiani. Endometriosis and Adenomyosis: Modern Concepts of Their Clinical Outcomes, Treatment, and Management. *Journal of Clinical Medicine*, 13(14):3996, July 2024.

[11] CLIPS Reference Manual: Basic Programming Guide (Version 6.4.1), 2023.