Symptoms based endometriosis prediction using machine learning

Visalaxi Sankaravadivel, Sudalaimuthu Thalavaipillai
School of Computing Science, Hindustan Institute of Technology and Science, Chennai, India

ABSTRACT

Endometriosis a painful disorder that stripes the uterus both inside and outside. Endometriosis can be diagnosed by the medical practitioners with the help of traditional scanning procedures. Laparoscopic surgery is the authentic method for identifying the advanced stages of endometriosis. The statistical approach is a state-of-art method for identifying the various stages of endometriosis using laparoscopic images. The paper focuses on a well-known statistical method known as chi-square and correlation coefficients are implemented for identifying the symptoms that are correlated with various stages of endometriosis. Chi-square analysis performs the association between symptoms and stages of endometriosis. With these analysis, an algorithm was proposed known as endometriosis prediction factor algorithm (EPF). The EPF algorithm predicts the presence of endometriosis if the derived value is greater than 1. From the chi-square analysis, it is identified that mild endometriosis is influenced 34% by menstrual flow, minimal endometriosis is influenced 40% by dysmenorrhea, where moderate endometriosis is influenced 31% by tenderness and deep infiltrating endometriosis is influenced 22% by adnexal mass.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Visalaxi Sankaravadivel
School of Computing Science
Hindustan Institute of Technology and Science
OMR, Padur, Chennai, India
Email: sakthi6visa@gmail.com

1. INTRODUCTION

Endometriosis is a diagnosed disease that arises in the 1/15th of women global population. The tissue-like structure is known as the Adnexal mass that covers the uterus, ovary, and Gall bladder, etc. known as endometriosis. Endometriosis is diagnosed through various techniques that include: a) magnetic resonance images, b) transvaginal ultrasound (TVUS), c) laparoscopic images, etc. The most accurate procedure for diagnosing the advanced stages of endometriosis is the Laparoscopic procedure. With the help of those laparoscopic procedures, precise location, as well as area associated with endometriosis, was identified. The symptoms of endometriosis include: a) irregular menstrual cycle, b) adnexal mass, c) tube blockage, d) tenderness, e) dysmenorrhea, f) chronic pelvic pain, etc. There exist several other symptoms that can be identified only by laparoscopic procedures. For instance tube blockage and adnexal mass are visible through laparoscopic procedures.

Statistical analysis plays a vital role in the association. There exist various statistical approaches for analyzing the relationship between variables. The attributes that exist may have a direct or indirect relationship with the variables. Correlation coefficient is a technique used for identifying the type of correlation between variables. There exist: a) positive correlation, b) negative correlation, c) no correlation. Chi-square is yet another statistical technique where it is used for identifying the possibility of a relationship.
between variables. Chi-square observes the difference between actual and expected counts. The hypothesis can be evaluated by performing a chi-square test. Implemented sensitivity analysis to identify the vary in endometriosis fertility index (EFI) score for treating infertility [1], [2]. ENZIAN and rASRM scores were used for finding associations between diseases as well the severity of the disease. Compared normal-weight women with overweight women to lower the risk of endometriosis. Used chronic management plan to relate various signs of endometriosis includes adnexal mass, dyspareunia, retrograde, etc. Kappa statistics were implemented on MRI images for classifying pelvic compartments [3]-[5]. LC-CUSUM was used for correlating laparoscopic findings with TVUS scan [6]. Text mining was implemented on the PubMed database. Based on human gene predicts the depth of endometriosis. Logistic regression was applied on pathology reports to predict the early stages of endometriosis with an accuracy of 90%. Chi-square analyses were performed on women with endometriosis. Using chi-square prolactin levels of women are associated with endometriosis [7]-[9]. Partial least square regression was applied to identify various phenotypes in endometriosis as “ERK1 and 2, AKT, MAPK, and STAT4”. The Cohort study was performed on endometriotic patients with pelvic pain to find the stages of endometriosis [10], [11]. CPP was not associated with the stages of endometriosis. Hormonal oral and subsequent laparoscopy reduces the pain intensity. Multiple logistic regression was used to identify the lesion type, the severity of pain, etc. from the data provided by the American Fertility Society. Psychometrics properties of endometriotic patients were associated with pain [12]-[14]. Endometriosis patients were identified with another important symptom known as Gastrointestinal symptom. These symptoms are evaluated using a p-value [15]. Women affected with endometriosis are recognized by anxiety and depressive syndrome using HAM-A and state-trait method. All possible locations are identified using a transvaginal scan and symptoms are associated using a visual analog scale. Endometriotic symptoms are evaluated using proper supplements [16], [17].

2. ANALYSIS OF INFLUENCING FACTORS BASED ON SYMPTOMS

Endometriosis is a problem that occurs in the female of child-bearing groups. The dataset containing laparoscopic images of 180 patients obtained from [18]. The various stages of endometriosis are identified, they are as; 1) first stage known as minimal endometriosis, 2) second stage known as mild endometriosis, 3) third stage known as moderate endometriosis, 4) fourth stage known as deep infiltrating endometriosis. Different group of people is affected by various stages of endometriosis. The various symptoms of endometriosis are identified [19] as features from the laparoscopic images and analysis. These features are categorized as external and internal factors that influencing endometriosis. They are as; a) menstrual irregularity, b) heavy menstrual flow, c) scant menstrual flow, d) dysmenorrhea, e) dyspareunia, f) chronic pelvic pain, g) tenderness, h) adnexal mass, i) restricted mobility.

2.1. Correlation co-efficient

The steps involved in analyzing the association of endometriosis with various symptoms are illustrated. Correlation coefficient [20] is used to measure how strong a relationship is correlated between two variables [21], [22]. Here the value:

a. 1 shows a robust positive association.
b. 1 shows a robust negative association.c. 0 shows no connection at all.

Pearson correlation coefficient is used to recognize the association between various features. They are represented as:

\[ P_{coefficient} = \frac{\alpha \cdot \beta}{\sqrt{(\alpha - \alpha_1)(\beta - \beta_1)}} \]  

(1)

where; \(\alpha\) represents values of x variable
\(\beta\) : represents values of the y variable
\(\alpha_1\): represents the mean value of the X variable
\(\beta_1\): represents the mean value of y variable
\(P_{coefficient}\) : represents the correlation coefficient

2.2. Chi-square test

This test is an analytical approach. It helps for identifying the association among variables [23], [24]. Here the chi-square is employed to identify how far the symptoms influenced at various stages of endometriosis. The formula for chi-square is calculated as:

---

Symptoms based endometriosis prediction using machine learning (Visalaxi Sankaravadivel)
\[ X = \sum \frac{(\alpha_i - \beta_i)^2}{\beta_i} \]  
\[ \text{(2)} \]

In (2), \( X \) is the chi-square value to be calculated, \( \alpha \) represents the observed value and \( \beta \) represents the expected value to be calculated from the observed value, \( i \) is the total number of observations made. Chi-square is calculated by subtracting all observed values from the expected value and dividing the obtained result from the estimated value. The estimated value is calculated as:

\[ \beta = \frac{\sum (\alpha \cdot p_i)}{\alpha_i} \]  
\[ \text{(3)} \]

In (3), \( p \) represents the symptoms identified in each stage. The estimated value is calculated by multiplying each observed value with the corresponding stages of endometriosis. P-value is calculated, where P-value decides the final result is significant or not. To evaluate the chi-square two more values are calculated by:

a. Calculate the degree of freedom. The degree of freedom value is calculated as:

\[ \text{DoF} = (r - 1) (c - 1) \]  
\[ \text{(4)} \]

In (4), \( r \) represents rows and \( c \) represents columns. The degree of freedom is calculated by subtracting the categories by minus 1.

b. The Alpha value is initialized by the researcher as 5% as either 0.01 or 0.10. The final critical value is calculated as:

\[ CV = \text{Chi} - \text{square. Inv. rt (} \alpha, \text{DoF)} \]  
\[ \text{(5)} \]

In (5) represents the critical value, where \( \alpha \) is the alpha value and DoF is the degree of freedom. From the critical value obtained, it is identified when the critical value is lesser than chi-square then the null hypothesis is false and the alternate hypothesis is true. Chi-square analysis provides a gateway for identifying the influencing factor from the laparoscopic images. The influencing factors are classified as internal influencing factors recognized from laparoscopic Images through chi-square analysis and external influencing factors recognized from the survey. The endometrial influence value is calculated using Endometriosis prediction factor as:

### 2.3. Endometriosis prediction factor algorithm

Step-1: The influencing factors are multiplied with the corresponding weight value

\[ \varepsilon = IF_i \cdot IFW_i \ \forall \ i < -1 \ to \ n \]  
\[ \text{(6)} \]

Where \( IF_i \) represents the influencing factor for each symptom and \( \varepsilon \) represents the internal influencing factor.

Step-2: Endometriosis prediction factor is calculated by summing of obtained influencing factor with the corresponding mean value.

\[ \delta_i = \sum_{i=1}^{n} \varepsilon_i \cdot \mu_i \]  
\[ \text{(7)} \]

Where \( \delta_i \) represents the endometriosis prediction factor (internal factors), \( \mu \) represents meantime.

Step-3: Similarly the external influencing factor is calculated as:

\[ \delta_E = \sum_{i=1}^{n} \lambda_i \cdot \mu_i \]  
\[ \text{(8)} \]

Where \( \delta_E \) represents the endometriosis prediction factor (external factor).

Step 4: 80% of weightage is given to internal factors where 20% of weightage is given to external factors, as a result

\[ \delta = (0.8 \cdot \delta_i) + (0.2 \cdot \delta_E) \]  
\[ \text{(9)} \]

Where \( \delta \) represents the endometriosis prediction factor.

Step -5: If calculated \( \delta \) value is greater than 1 then it’s predicted as endometriosis.
2.4. Metrics for endometriosis prediction factor algorithm

The endometriosis prediction factor (EPF) was analyzed using the following performance metrics. Accuracy: accuracy is calculated as the fraction of the right forecast performed on data.

\[
Accuracy = \frac{tp+tn}{tp+tn+fp+fn}
\]  

(10)

Sensitivity: sensitivity is calculated as the ratio of real positive cases over forecast positive cases.

\[
Sensitivity = \frac{tp}{tp+fn}
\]  

(11)

Specificity: specificity is obtained by calculating the fraction of real negative cases over forecast negative cases.

\[
Specificity = \frac{tn}{tn+fp}
\]  

(12)

Precision: precision is obtained by calculating the proportion of correct positive values with the overall positive values.

\[
Precision = \frac{tp}{tp+fp}
\]  

(13)

3. RESULTS AND DISCUSSION

The statistical test was executed across the symptoms identified and stages of endometriosis. Here the statistical test is performed using pearson correlation coefficient and chi-square execution. From the statistical analysis, various stages of endometriosis are identified across a different groups of endometriotic affected patients from laparoscopic images [25]. A different sets of people are identified with various stages of endometriosis. Accordingly stage 1, endometriosis was identified on 119 patients, stage 2 on 39 patients, stage 3 on 11 patients and stage 4 on 11 patients. Pearson coefficient shows how the symptoms are strongly correlated with stages of endometriosis. A co-efficient close to 1 means that there’s a very strong positive association between the two variables. In our case, the diagonal blue shows very strong associations. The diagonal line is the association of the variables to themselves, so they’ll be 1 was illustrated in Figure 1.

![Figure 1. Heat map for correlating endometriosis](image)

As a result of chi-square execution, from the critical value identified alternate hypothesis was found to be true. i.e. the symptoms identified are directly associated [26] with the various stages of endometriosis. From laparoscopic images, various stages of endometriosis are identified for all 180 patients. Stage 1 of endometriosis represents minimal endometriosis where the symptoms influenced includes menstrual irregularity as 34%, heavy menstrual flow as 8%, Scant menstrual flow as 23%, dyspareunia as 17%, dysmenorrhea as 6%, chronic pelvic pain as 7%, adnexal mass as 8%. Among all symptoms, menstrual irregularity plays a predominant role in the identification of stage 1 endometriosis. All stages are illustrated in Figure 2.
Stage 2 of endometriosis was identified in 39 patients. Stage 2 represent mild endometriosis where the symptoms influenced includes menstrual irregularity as 18%, heavy menstrual flow as 28%, scant menstrual flow as 4%, dyspareunia as 17%, dysmenorrhea as 40%, chronic pelvic pain as 4%, adnexal mass as 1%, tenderness as 5%, chronic pelvic pain as 4%. Among all symptoms, Heavy menstrual flow and dysmenorrhea play a predominant role in influencing stage 2 endometriosis. Stage 3 of endometriosis was identified on 11 patients. Stage 3 represent moderate endometriosis where the symptoms influenced includes menstrual irregularity as 24%, heavy menstrual flow as 8%, scant menstrual flow as 16%, dyspareunia as 4%, dysmenorrhea as 13%, chronic pelvic pain as 2%, adnexal mass as 8%, tenderness as 31%, chronic pelvic pain as 4%, restricted mobility as 16%. Among all symptoms, Tenderness and Menstrual irregularity plays a predominant role in influencing moderate endometriosis. Stage 4 of endometriosis was identified on 11 patients. Stage 4 represent deep infiltrating endometriosis. Along with women’s uterus, ovary, gall bladder are also affected due to deep infiltrating endometriosis. The symptoms influenced include menstrual irregularity as 10%, heavy menstrual flow as 19%, scant menstrual flow as 22%, dyspareunia as 1%, dysmenorrhea as 16%, chronic pelvic pain as 14%, adnexal mass as 2%, tenderness as 13%, chronic pelvic pain as 4%, restricted mobility as 22%. Among all symptoms, restricted mobility, adnexal mass and dysmenorrhea play a predominant role in influencing deep infiltrating endometriosis. The overall symptoms influencing endometriosis are menstrual irregularity as 20%, scant menstrual flow as 22%, dysmenorrhea as 16%, heavy menstrual flow as 19%, chronic pelvic pain as 14%, tenderness as 13% and so on as illustrated.

From the chi-square analysis [27] it was identified critical value is greater. As a result, the alternate hypothesis was found to be true as there is a direct association between symptoms and stages of endometriosis. It was identified that symptoms and stages are dependent on each other. The overall symptoms influencing endometriosis was illustrated in Figure 3. The internal influencing factors are the features identified from the analysis of laparoscopic images: a) adnexal mass where it influences 55%, b) tenderness as it influences 20%, c) tube blockage where it influences 15% and d) retrograde where it influences 10%. Similarly, the external influencing factor is the features obtained from the analysis are as; a) dyspareunia as 35%, b) dysmenorrhea as 20%, c) heavy menstrual flow as 15%, d) menstrual irregularity as 15%, e) chronic pelvic pain as 10%, f) restricted mobility as 5% as illustrated in Figure 4.

The weight value of various symptoms related to internal factors are 0.38,0.32,0.20 and 0.1 for adnexal mass, tenderness, tube blockage, retrograde respectively. Based on the EPF algorithm calculation presence of endometriosis was predicted. The threshold value is set as 1, as a result, if the EPF value is greater than 1 then endometriosis presence is predicted otherwise endometriosis is not present. Based on
those evaluations actual gynecologist prediction from laparoscopic images is compared with the proposed EPF algorithm. As a result, a confusion matrix is constructed. The performance of the EPF algorithm was evaluated using metrics known as: a) accuracy b) sensitivity c) specificity d) precision was illustrated in Figure 5 and the obtained metrics values are defined in Table 1.

Figure 3. Overall symptoms influencing endometriosis

Figure 4. Internal and external symptoms influencing endometriosis

Figure 5. Performance metrics of EPF algorithm

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPF Algorithm</td>
<td>90%</td>
<td>95.50%</td>
<td>78.12%</td>
<td>90.27%</td>
</tr>
</tbody>
</table>

Table 1. Performance metrics of EPF algorithm

Symptoms based endometriosis prediction using machine learning (Visalaxi Sankaravadivel)
4. CONCLUSION
This paper analyzed the various stages of endometriosis using statistical approaches. Chi-square analysis helps in the association of symptoms with various stages of endometriosis. The correlation coefficient derives the value 1. It was proved that there exists strong relation between various symptoms and corresponding stages of endometriosis. The designed EPF algorithm predicts endometriosis using the internal and external influencing symptoms with an accuracy of 90.2%. Chi-square analysis was employed to evaluate the influences of various symptoms of endometriosis with the stages of endometriosis. Adnexal mass and tenderness have the major influencing factor of 55% for the occurrence of advanced endometriosis. Minimal endometriosis was majorly influenced by dysmenorrhea and moderate endometriosis was majorly influenced by tenderness of 31%.

REFERENCES


**BIOGRAPHIES OF AUTHORS**

**Visalaxi Sankaravadivel** is a Research scholar in the School of Computing Sciences, Hindustan Institute of Technology and science, Chennai, India. She has ten years of experience in teaching and research. She has published 10 research article in the field of Endometriosis using Deep Learning in reputed journals and conferences. She is granted a patent on Endometriosis identification. She served as a resource person for numeral FDPs/Training. She is pursuing a research work in endometriosis using deep learning. She is active researcher in the field of medical image analysis. She is a member of many professional societies includes ACM, IEEE, IET.

**Sudalaimuthu Thalavaipillai** is working as a Professor in the School of Computing Science, Hindustan Institute of Technology and Science, Chennai, India. He had completed his PhD degree from Hindustan Institute of Technology and Science, Chennai, India. He is a Certified Ethical Hacker. He has published 50 research article in the reputed International Journals and conferences. He is granted both Indian and Australian patent. He obtained many awards in his career includes Pearson Award under Best Teacher Category, Top innovator Award for patent rights. His research areas include cyber network security, grid and cloud computing, machine learning. He is a life time member of CSI, ACM and IEEE.