# Predictive Analytical Model for Ectopic pregnancy diagnosis: Statistics vs. Machine learning methods.

**Ploywarong Rueangket** ( ploywarong.24@gmail.com )
Phramongkutklao hospital

**Kristsanamon Rittiluechai**
Phramongkutklao hospital

**Assistant Professor Akara Prayote, Ph.D.**
King Mongkut's Institute of Technology North Bangkok, Thailand

**Research Article**

**Keywords:** Machine learning, Neural networks, support vector machines, decision tree, pregnancy of unknown location, ectopic pregnancy

**Posted Date:** April 26th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1580372/v1

## Abstract

# Background:

Ectopic pregnancy is well-known for its serious outcome. Early detection could make the difference between life and death in pregnancy. Although, prompt diagnosis was challenged and numbers of predictive models had been developed, there's still lack of gold standard tool. Our research introduces predictive analytical models using both conventional statistics and machine learning methods based on all three domains of features (clinical factors, serum human chorionic gonadotropin and ultrasound findings).

# Methods:

Retrospective cohort study on 377 pregnancy of unknown location women, training and validating for model for prediction of ectopic pregnancy outcome, using set of 22 features. Analysis performed by three difference machine learning models neural networks (NNs), decision tree (DT) and support vector machines (SVMs) using cross validation technique. In addition, with traditional statistical model, logistic regression (LR). In which we compare the model performance with ROC-AUC and accuracy, PPV and NPV. Finally, new 30 PULs were tested for external validation.

# Result

Comparing of model performance (validation) to predict EP, LR Ranked first, followed by NNs, DT and SVMs with mean ROC-AUC ± SD of 0.904, 0.895, 0.871 and 0.862 respectively. In terms of clinical parameters, we found sensitivity ± SD performed in the same sequence as AUC (LR; 87.97% ± 7.86%, NNs; 85.97% ± 9.34%, SVM; 87.47% ± 8.23% and DT; 85.47% ± 10.37%, orderly. While specificity (± SD) was ranked by DT, then followed by SVMs, LG and NNs, which were 84.08 ± 8.56%, 83.53 ± 9.18%, 82.94 ± 9.84% and 81.76 ± 9.84%, respectively. In testing data, NNs and SVM performed equally best, then DTs and LR at ROC-AUC of 0.784, 0.778 and 0.739, respectively.

# Conclusion

The result demonstrates that both statistics and ML model could be utilized to achieve satisfied predictions for EP. Surprisingly, the highest ranked model was LR in validating test, but in new testing data machine learning, NNs and SVMs could overcome statistics. This could shade a new light for further research of unsolved medical problem with more complexity and bigger database.

## Introduction

Incidence and impact

EP: Ectopic pregnancy occurs when a fertilized egg implants outside the uterine cavity, results from various factors that interrupt the successful migration of conceptus (3). The incidence rate of EP in Thailand and worldwide were 9.3 and 10–20 per 1,000 pregnancies respectively (3–6). Unfortunately, the mortality rate was high comparing with low numbers of incidence (7). UK's Healthcare Safety Investigation Branch (HSIB) has uncovered that failed or delayed in diagnosis were the main concern (8, 9) and declared early diagnosis of EP to be a life or death medical decision (10).

Early diagnosis limitation and benefit

EP usually diagnosed in first trimester pregnancy. Presenting symptoms range from vaginal bleeding, abdominal pain, missed menstruation and faint. Examination findings were abdominal and/or adnexal tenderness, cervical motion tenderness or hypotension. Unfortunately, many studies found that history and physical examination do not reliably. Since, up to 50% of patients has no risk factor (11) and 9% has no pain. Also, normal examination found nearly one-third of cases (12). When pregnancy test was confirmed and early pregnancy complications is suspected, ultrasound examination is commonly used to confirm the location of pregnancy(13). However, only 73.9% of tubal EPs were visualized by initial TVS (14). Recently, serial measurements serum hCG has been shown to improved diagnostic rate(13). Anyhow, the result couldn't differentiate those with EP from normal intrauterine pregnancy or miscarriage precisely enough (15). Consequently, clinicians were misdiagnose more than 40% of EPs on the initial ED visit in former study (16). While many clinical protocols was improved, in addition with modern investigation tools, there was still limitation in diagnosis as in UK national health service (NHS) uncovered 30 missed ectopic pregnancies that leading to "serious harm" in a one year period (2017–2018) (10). Therefore, it's a time-critical condition and become life-threatening when the implantation site ruptured causing immediate bleeding into intra-abdomen and eventually hemorrhagic shock.

Diagnostic model

Attempts in developing model for EP diagnosis was established from variety of utilized domains. Including of

- Clinical model using clinical risk factor classified into high risk group (17) and also risk factors combined with single serum hCG model (18), but the result is inconsistence between specificity and sensitivity.

- Serum level of progesterone model with single cut-offs (AUC 0.725), and later widely known serial hCG, M1, M4 and M6 which present high performance, but the accuracy was lower in different cohort and required at least two follow up exam (0/48h)(19–21)

- Ultrasound score present subclassification patient using indeterminate ultrasound readings(22), although there was unavoidable limitation of ultrasound user expertise and patient's factor confounding.

In addition to the limitations mentioned, there was still no international consensus, neither gold standard tools established to identify early EP. Our aim is to develop model combining all three domains using traditional statistical analysis, and the early method, machine learning.

Machine learning (ML) has dramatically contributed new knowledge in medical field in the last two decades. It's defined the evolution of interdisciplinary sciences between statistics, artificial intelligence and medicine. Ability to conduct complex tasks, automatically finds hidden patterns too complex for humans to find, has its advantage of discovering rules for behavior and adaption to changes in the word, which make ML suitable for predict new EP case (23).

There have been numerous EP studies based on traditional statistical analysis. Although EP is dangerous and difficult to detect, small numbers of studies have applied machine learning into this field. One was used as decision support model for treatment. Another was interestingly studying different ML methods for prediction EP in PUL based on seum hCG and clinical information. To our knowledge, this might be the first combining all diagnostic feature domains using both statistics and ML method.

Our research problem applied to classification technique based on supervised learning method. Three widely used ML method including of decision tree (DT), support vector machine (SVM), neural network (NN) and logistic regression (LR) which is traditional most common used statistical method. Each model process different characteristics of algorithms that suitable for different set of data problems.

Our goal is to compare all four models and search for the best model suitable for the problem.

# Materials And Methods

### Problem definition and formulation

This was a retrospective cohort study, conducting from electronic medical records of pregnant women presenting with first trimester complication symptoms including abdominal pain and abnormal vaginal bleeding at Phramongkutklao Hospital between October 2010 and March 2022. The criteria for inclusion were those who suspected of pregnancy of unknown location (PUL) with medical report of clinical history, physical examination, and ultrasound evaluation. Women were included regardless of report with or without taking serum hCG due to medical judgement at that time. Women whose had definite signs of intrauterine pregnancy or extrauterine pregnancy via ultrasonography at the first visit. The patients presenting clinically suspicious of ruptured EP (clinical instability or sign of intra- abdominal hemorrhage) or show any evidence of intra-uterine gestational content or EP (adnexal mass consisted of fetal pole or fetal heart motion) by ultrasound at first visit were excluded. The study was approved by the Royal Thai Army Medical Department Institutional Review Board, reference number R048h/62_Exp. Patients' identifications were coded before analyzing and discussing.

### Analysis:

For supervised learning analysis, four basic and powerful classification methods were chosen by their unique classifying ability. Despite a variety of method developed, each provide their own characteristics, the method capability and model requirement should be matched.

**Logistic regression (LR)** is a traditional statistical method, invented by the British statistician in 1958(28), dealing with classification problem by using logistic function, whose result always falls between 0 and 1 and the graph of function is S-shaped. Regression method has advantage in its interpretability, which could explain how the model works and more importantly lead to understanding of “why?” this patient was predicted ‘yes’. Although regression coefficients in LR are not easy to interpret and understand as in linear regression, it could interpret whether the relationship is proportional or inversely proportional between each feature (probability) (29, 30).

**Support vector machines (SVMs)** are also used for classification, as an alternative to LR, devised by Soviet statisticians in 1963 and has become feasible with the introduction of kernels and soft margin classifiers in 1990s (31) Advantage over simple regression is that linear or logistic regression uses all the data points in the calculation of the line of best fit, while SVM focus on only the set of points (called “support vector”) closet to the margin. However, in term of interpretability, SVM perform relatively like black box. (30)

**Decision tree (DT)** classify training data by sorting them down the tree from the root to leaf node. Each internal node is a feature and prediction is made at leaf nodes. A leaf is a collection of examples that may not be classified any further(32). It has the ability to train both discrete and continuous values and can be used even when some training data have unknown values(23). However, there are practical issues from learning to determining how deeply to grow the decision tree, handling continuous data, choosing an appropriate feature selection measure, and ling training data with missing attribute values.

Neural networks are ML methods …and competitive in real-world problem, also non-linear data. It’s often used in comparison with LR.

**Neural networks** were named as a simulation of how brain cellular network, that were believed to be in 1950s. NN comprise one or more layers of autonomous computational units or nodes that received input from other nodes (including within the same layer), and send outputs or even feedback to previous input, to present the final output prediction. Although, the earliest NN were used for classification prediction like basic SVMs or linear discriminant analysis, they become more useful in more complex or non-linear task like handwriting or imaging recognition, which competitive in real-world problem, also non-linear data. However, disadvantage found in NN were the longer time for model running compared to the same category of problems by LR, SVM and DT. Secondly, because only numbers of node and layer were identified. Lastly, NNs have no explanatory power of ‘why’ this is predicted.

**Software:** RapidMiner Studio 9.9.003, well-known data analyst tools and specially used for predictive analysis and statistical computing.(33, 34)

### 2. Data gathering

**Study population:** 377 of PUL patients

**Feature (Predictive values):** Three domains of 22 features composed of clinical history (demographic data, history risk factors, clinical manifestations), initial serum hCG levels and ultrasound results. All factors were extracted and selected from literature reviews for the statistical and clinically relevant to our research outcome.

**Data preparation**

**Missing value:** Due to retrospective study, missing input data in inevitable. After reviewing cases, there are approximately 10-20% missing value in all 22 features and were missing at random (mostly unsignificant negative findings or assumed irrelevant past history in those hospital visit)(35). Since, our object is to understand the data for training, not deleting ones, which could bias the classifier performance(36). Missing value imputation method has shown to improve prediction capability. Naïve-Bayes, simple, probability machine learning was applied(37).

**Remove correlated features: To** avoid confusing correlation and causation, features with high or substantial absolute correlation more than 0.95 were removed **(38).**

**Feature selection:** To select the attribute that is most useful for classifying examples, optimization selection using forward/ backward stepwise was applied (n [generation without improvement] = 1).

### 3. Model analysis

**Model training 10-fold cross validation:** To optimize process of training set to estimate their accuracy and to overcome model overfitting, by providing 10-fold (9:1) training and validating data (23). All four models were trained using entire dataset (n=377).

**Performance evaluation:** Comparing between four model using Area Under the Receiver Operating Characteristic Curve (ROC-AUC) (39) report as mean ± SD of cross validation process. Also, report accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for more insights.

### 4. Prediction

**Deployment (external validation):** Apply four models to new separated patient data (n-30), then compare model performance using ROC-AUC.

# Results

| | Characteristics | Total (N = 377) N (%) | EP (N = 200) N (%) | Non EP (N = 177) N (%) | p-value |
|---|---|---|---|---|---|
| Demographic | Age (years) mean ± SD | 30.10 ± 6.22 | 29.48 ± 6.08 | 30.81 ± 6.31 | 0.038 [b] |
| | Age group (years) | | | | 0.302[a] |
| | < 35 (%) | 286 (75.86) | 156 (54.55) | 130 (45.45) | |
| | ≥ 35 (%) | 91 (24.14) | 44 (48.35) | 47 (51.65) | |
| | BMI (kg/m²) | | | | 0.844[a] |
| | < 23 (kg/m²) (%) | 251 (67.29) | 131 (52.19) | 120 (47.81) | |
| | ≥ 23 (kg/m²) (%) | 122 (32.71) | 65 (53.28) | 57 (46.72) | |
| | Parity | | | | 0.142[a] |
| | Nulliparity (%) | 166 (44.03) | 81 (48.80) | 85 (51.20) | |
| | Multiparity (%) | 211 (55.97) | 119 (56.40) | 92 (43.60) | |
| | Gestational age at diagnosis (days) | | 52.65 ± 15.97 | 51.51 ± 16.28 | 0.491 [b] |
| | Underlying disease | | | | 0.098 [a] |
| | No (%) | 321 (85.15) | 176 (54.83) | 145 (45.17) | |
| | Yes (%) | 56 (14.85) | 24 (42.86) | 32 (57.14) | |
| Risk factors | History of pelvic surgery | | | | 0.331 [a] |
| | No (%) | 283 (78.39) | 159 (56.18) | 124 (43.82) | |

SD: Standard Deviation, BMI: body mass index; Kilogram/meter2, PID: pelvic inflammatory disease

[a] chi-square test

[b] t-test

| Characteristics | Total (N = 377) | EP (N = 200) | Non EP (N = 177) | p-value |
|---|---|---|---|---|
| | N (%) | N (%) | N (%) | |
| Yes (%) | 78 (21.61) | 39 (50.00) | 39 (50.00) | |
| **Smoking** | | | | 0.016 [a] |
| No (%) | 353 (98.06) | 192 (54.39) | 161 (45.61) | |
| Yes (%) | 7 (1.94) | 7 (100.00) | 0 | |
| **History of ectopic pregnancy** | | | | 0.982 [a] |
| No (%) | 357 (95.45) | 190 (53.22) | 167 (46.78) | |
| Yes (%) | 17 (4.55) | 9 (52.94) | 8 (47.06) | |
| **History of PID** | | | | 0.032 [a] |
| No (%) | 319 (95.80) | 181 (56.74) | 138 (43.26) | |
| Yes (%) | 14 (4.20) | 12 (85.71) | 2 (14.29) | |
| **Current use of Emergency pill** | | | | 0.032 [a] |
| No (%) | 319 (95.80) | 163 (56.74) | 148 (43.26) | |
| Yes (%) | 39 (4.20) | 33 (84.62) | 6 (15.38) | |
| **Assisted reproductive technology** | | | | 0.936 [a] |
| No (%) | 362 (97.05) | 193 (53.31) | 169 (46.69) | |

SD: Standard Deviation, BMI: body mass index; Kilogram/meter2, PID: pelvic inflammatory disease

[a] chi-square test

[b] t-test

| | Characteristics | Total (N = 377) N (%) | EP (N = 200) N (%) | Non EP (N = 177) N (%) | p-value |
|---|---|---|---|---|---|
| | Yes (%) | 11 (2.95) | 6 (54.55) | 5 (45.45) | |
| **Symptoms** | **Abdominal pain** | | | | < 0.001 [a] |
| | No (%) | 73 (19.41) | 25 (34.25) | 48 (65.75) | |
| | Yes (%) | 303 (80.59) | 174 (57.43) | 129 (42.57) | |
| | **Abnormal vaginal bleeding** | | | | 0.104 [a] |
| | No (%) | 82 (21.75) | 50 (60.98) | 32 (39.02) | |
| | Yes (%) | 295 (78.25) | 150 (50.85) | 145 (49.15) | |
| | **Nausea, vomiting** | | | | 0.556 [a] |
| | No (%) | 238 (84.70) | 138 (57.98) | 100 (42.02) | |
| | Yes (%) | 43 (15.30) | 27 (62.79) | 16 (37.21) | |
| | **Faint** | | | | 0.030 [a] |
| | No (%) | 263 (91.64) | 148 (56.27) | 115 (43.73) | |
| | Yes (%) | 24 (8.36) | 19 (79.17) | 5 (20.83) | |
| **Signs** | **Abdominal tenderness** | | | | < 0.001 [a] |
| | No (%) | 187 (49.60) | 62 (33.16) | 125 (66.84) | |

SD: Standard Deviation, BMI: body mass index; Kilogram/meter2, PID: pelvic inflammatory disease

[a] chi-square test

[b] t-test

| | Characteristics | Total (N = 377) N (%) | EP (N = 200) N (%) | Non EP (N = 177) N (%) | p-value |
|---|---|---|---|---|---|
| | Yes (%) | 190 (50.40) | 138 (72.63) | 52 (27.37) | |
| | **Cervical motion tenderness** | | | | < 0.001 [a] |
| | No (%) | 296 (79.14) | 120 (40.54) | 176 (59.46) | |
| | Yes (%) | 78 (20.86) | 77 (98.72) | 1 (1.28) | |
| **Serum marker** | **Serum β-hCG level at first visit (mIU/ml)** | | | | < 0.001 [a] |
| | < 1,000 (%) | 136 (53.13) | 41 (30.15) | 95 (69.85) | |
| | ≥ 1,000 (%) | 120 (46.88) | 72 (60.00) | 48 (40.00) | |
| **Ultrasound finding** | **Intra-uterine anechoic content** | | | | 0.583 [a] |
| | No (%) | 327 (87.43) | 174 (53.21) | 153 (46.79) | |
| | Yes (%) | 47 (12.57) | 23 (48.94) | 24 (51.06) | |
| | **Endometrial thickness > 14mm** | | | | 0.491 [a] |
| | No (%) | 298 (83.94) | 156 (52.35) | 142 (47.65) | |
| | Yes (%) | 57 (16.06) | 27 (47.37) | 30 (52.63) | |
| | **Adnexal mass of complex echogenicity** | | | | < 0.001 [a] |

SD: Standard Deviation, BMI: body mass index; Kilogram/meter2, PID: pelvic inflammatory disease

[a] chi-square test

[b] t-test

| Characteristics | Total (N = 377) N (%) | EP (N = 200) N (%) | Non EP (N = 177) N (%) | p-value |
|---|---|---|---|---|
| No (%) | 172 (45.74) | 24 (13.95) | 148 (86.05) | |
| Yes (%) | 204 (54.26) | 175 (85.78) | 29 (14.22) | |
| **Free fluid in cul-de-sac** | | | | < 0.001 [a] |
| No (%) | 214 (57.84) | 70 (32.71) | 144 (67.29) | |
| Yes (%) | 156 (42.16) | 125 (80.13) | 31 (19.87) | |
| SD: Standard Deviation, BMI: body mass index; Kilogram/meter2, PID: pelvic inflammatory disease | | | | |
| [a] chi-square test | | | | |
| [b] t-test | | | | |

**Table 1: Descriptive demographic and features of the study population**

In validation cohort, 377 patients whose initial visit were identified as PUL, the final diagnosis was 200 (53%) EPs and 177 (47%) non-EPs. Among the 177 patients with non-EP, 21 (11.8%) of them were threatened abortion, 1 (0.6%) blighted ovum, 1 (0.6%) corpus luteal leakage and the other 154 (87.0%) spontaneous abortion. The mean age was 30.1 years with 55.9% multiparity. Comparison of demographic data presented in Table 1, there were no difference between two groups. The testing cohort, using new separated population with a total of 30, was 12 (40%) EPs and 28 (60%) non-Eps, was also no difference between groups.

| Machine learning model | Factor / Feature selections |
|---|---|
| Logistic regression,<br>Decision tree<br>Support vector machine | Multipara, vaginal bleeding, Cervical tenderness, serum hCG, inhomogeneous adnexal mass in ultrasound |
| Neural network | Multipara, Cervical tenderness, serum hCG, inhomogeneous adnexal mass in ultrasound |
| **Table:2** Features selected in four models | |

In data preparation process, features selected to be in the model were shown in Table 2. We then run the model analysis and present the performance comparison in ROC-AUC in Fig. 3

The average performance ROC-AUC was high in all models (AUC ≥ 0.862, Figs. 3 and 4), also highlights that statistical model (LR) is superior to ML in validating dataset.

Of the four models' performances Fig. 4 shown, LR ranks first, followed by NNs, SVMs and DT with ROC-AUC ± SD of 0.904 ± 0.054, 0.895 ± 0.058, 0.871 ± 0.044 and 0.862 ± 0.032 respectively. For clinical aid, we report sensitivity of mean (± SD) in LR; 87.97% ± 7.86%, NNs; 85.97% ± 9.34%, SVM; 87.47% ± 8.23% and DT; 85.47% ± 10.37%, orderly. While specificity (± SD) was ranked by DT, then followed by SVMs, LG and NNs, which were 84.08 ± 8.56%, 83.53 ± 9.18%, 82.94 ± 9.84% and 81.76 ± 9.84%, respectively.

Figure 5 shown more insight of how DT model predict the outcome, shown that the prediction process is in prioritizing order. As tree grown downwardly, we found that adnexal mass is the highest priority decision nodes, firstly used to classify patient, which indicate it the main classified feature. Followed by cervical tenderness. The second pathway if none of these two features were existed, we found initial serum hCG, multiparity and vaginal bleeding provide additional decisional data.

Table 3
ROC-AUC performance comparison of the four models applied on the internal and external validation datasets.

| Model | Internal validation | External validation |
|---|---|---|
| LR | **0.904** | **0.739** |
| SVMs | **0.871** | **0.784** |
| DT | **0.862** | **0.778** |
| NNs | **0.895** | **0.784** |

We then evaluated the models on new cohort for external validation and found that NNs and SVMs models performed best with ROC-AUC of 0.784, followed by DTs and LR as shown in Table 3.

# Discussion

Our study report incidence of 53% EPs among initially suspected PUL women. Although, others report incidence range from 7 to 31%(40–43), similar rate was observed in one large prospective observational study by Monia et al, 2013 (44) with rate of 43%. This could be explained by spontaneous resolution of EP in PUL that might be failed to diagnose, since the location was never known, while some case might be misclassified as missed abortion. Also, in higher number might resulted from the sensitivity of ultrasound at the initial diagnosis that subjectively difference between cohorts. Through the methodology of data science for model development, there are two main steps result.

**Firstly, model learning and validation.** While, ML believed to empower prediction fields, theoretically by using complex algorithms that enable high accurate model (45). Our result shown that LR has a better predictive ability throughout ROC-AUC, accuracy, sensitivity and NPV. One explanation could be that model was obtained from features, chosen by reviewing and studying in many literatures, resulting in true causation, which unavoidable because medical data reasonably based on fact. Evidently, due to presence of non-random variation (causative variables) in the input variables make logistic regression perform best out of 4 models in internal validation process.

Particularly interesting for researchers is the new feature, nulliparity, for predicting EP. While other study found different findings association (46) and to our knowledge, there was no model using nulliparity as a feature in prediction, yet. Also shown in these four models that cervical tenderness, adnexal mass, and serum hCG were selected by optimized selecting feature process. This could be interpretable by DT model. While ultrasound finding of inhomogeneous adnexal mass are prioritized, following by physical examination of cervical tenderness, then initial level of serum hCG up to 1,000 IU/ml, clinical risk factor of nulliparity and vaginal bleeding, are shown to be useful to classify PUL women in orderly. This was related to evidenced-based observational studies that these factors shown to be correlated with EP with high odds ratio (44, 47). In this research might prove the four factors were not just related to EP but might be the biological plausibility as well.

**Comparing between ML models**, major drawback of ML model especially NNs and SVMs occurs in its training phase, we found that accuracy is highly dependent on the size of input data (48). Although, extraordinary generalization capability of SVM and its discriminative power make SVMs perform better than DT, but practically and theoretically lower than NNs. However, DT has advantages in dealing with training data' s missing value, which could be more useful in practical use.

**For decisional making**, Ideally, we would prefer a diagnostic test that is both 100% sensitivity and 100% specificity. Unfortunately, it was rarely occurs and usually trade-off(49).

In practical use, we would like to focus on two circumstances. on the first patient visit, model focusing on diagnosis EP might be the most important. Since limitations or pitfalls occur in many settings. From lack of obvious clinical presentation or ultrasound findings, to lack of specialist to consult in primary care hospital(50). To decrease rate of misdiagnosis (ruled in), high sensitivity is crucial. Due to the fact that, 11 out of 17 deaths in the first trimester resulted from ruptured ectopic pregnancies ( report by the Confidential Enquiry into Maternal and Child Health 2000–2002)(51), so we concern if positive disease (EP) was not result in positive test(52), leading to inappropriate discharge or inadequate follow-up. Furthermore, to emphasize confidence in the test sensitivity, patients predicted as non-EP result, still need to be counselling for NPV rate, since sensitivity cannot categorize other people as not having the condition when in fact, they do have it. In the second following circumstance, following up EP group or in practice, high risk PUL, serial serum hCG and ultrasound would have been followed as standard protocol to definite cases of ectopic pregnancy and intrauterine pregnancy would be identified by ultrasound. Unfortunately, we found that for counselling for treatment (ruled out), high specificity is more important.

Because people in positive test are very unlikely to be categorized as having a condition if they indeed do not have it and prevent harmful unnecessary treatment for normal pregnancy.

**Secondly, model deployment and testing**. When deploy the models to the new real data, ML prove to be more superior. Unfortunately, due to small sample size, further study is required for more validation. Afterward, there are pros and cons in ML. When more new data becomes available, and more features are explored, ML models tend to be more complex and harder to incorporate all data into a single optimal model and need their ongoing maintenances. While basic statistics don't.

# Conclusion

Introducing machine learning into medical research has become a great challenge, but also an innovative way for disease prediction using its complex algorithm to discover unknown pattern or information inside black box. Advantage in dealing with missing values, selecting the most optimized features, and analyzed non-parametric data, proved to be ground-breaking methodology for clinical usage.

Difference between two main methods is that while LR can only train discrete or categorized input data, ML can do both discrete and continuous. This could open a window for more information gathering to build the model using complex or even unknown irrelevant data e.g., complicated numerical numbers

To our knowledge, while selection of model for scientific problem can markedly influence predictive performance and building complex model in some data might be the only model which is powerful enough to create accurate predictions but might become useless in some question. Since the more complex a model, the harder the results of a prediction are to explain, you might never have the answer for 'why' this says 'yes'. Secondly, while keeping up with changed patient's information world, simple model tends to maintain its performance, but complex one need up-to-date maintenance, so the additional key could be to focus on data nature instead of creating complex models.

Lastly and more importantly, decision for what model is the best might depend on nature of data and the question of 'what is the answer' vs. 'why is this answer'.

# Limitation

Due to low incidence of EP in Thai population, retrospective study was chosen, which provide sufficient power of data for statistics, while obtaining unavoidable missing data. Therefore, we designed our study protocol to avoid limitation using machine learning missing value imputation technique. Secondly, while input data type related to analyzing process, then the prediction performance effect by type of data. In our research, mostly are categorical data instead of continuous, which could limit the performance of NNs and SVMs by its nature.

# Future Work

Until now, healthcare organization had produced and recorded tons of patients' information, which might never been use. Organized electronic collection of data could properly process as a database management system (53). With machine learning model, learning from these data could produce an ultimate benefit in terms of prediction and inventing new insight.

- Optimization or prescriptive provide the best value by of the consequence after prediction. Not only know the future but also know the cost of action or outcome after actions. (know cost of actions)

- create new features that can reveal a more accurate prediction

- more focused business questions. Subsequent data mining processes will benefit from the experiences of previous ones(54)

# References


1. Cox LA. Improving causal determination. Global Epidemiology. 2019;1:100004.

2. Condous G, Okaro E, Khalid A, Bourne T. Do we need to follow up complete miscarriages with serum human chorionic gonadotrophin levels? Bjog. 2005;112(6):827-9.

3. Voedisch AJ, Cahill, Erica P. . Early pregnancy loss and ectopic pregnancy In: Berek JS, editor. Berek & Novak's Gynecology 16 ed. Philadelphia: Wolters Kluwer Health/LippincottWilliams &Wilkins; 2019. p. 1912-59.

4. Liampongsabhuddhi P. Epidemiological Study of Ectopic Pregnancy in Lampang Hospital. [illegible]. 2010;31(1):20-7.

5. Leke RJ, Goyaux N, Matsuda T, Thonneau PF. Ectopic Pregnancy in Africa: A Population-Based Study. Obstetrics & Gynecology. 2004;103(4):692-7.

6. [illegible], et al. The assessment of emergency obstetric care (EmOC) in the lower 5 southern Provinces of Thailand.

7. MARION LL, MEEKS GR. Ectopic Pregnancy: History, Incidence, Epidemiology, and Risk Factors. Clinical Obstetrics and Gynecology. 2012;55(2):376-86.

8. Awoleke JO, Adanikin AI, Awoleke AO. Ruptured tubal pregnancy: predictors of delays in seeking and obtaining care in a Nigerian population. International Journal of Women's Health. 2015;7:141.

9. The diagnosis of ectopic pregnancy: an

independent report [Internet]. 2020. Available from: www.hsib.org.uk/investigations-cases/diagnosis-ectopicpregnancy/final-report.

10. Thornton J. Women are at serious risk of harm from late diagnosis of ectopic pregnancy. Bmj. 2020;368:m924.

11. Portuondo JA, Remacha MJ, Llaguno MR. Ectopic pregnancy early diagnosis limitations. Int J Gynaecol Obstet. 1982;20(5):371-8.

12. Kaplan BC, Dart RG, Moskos M, Kuligowska E, Chun B, Adel Hamid M, et al. Ectopic pregnancy: prospective study with improved diagnostic accuracy. Ann Emerg Med. 1996;28(1):10-7.

13. ACOG Practice Bulletin No. 193: Tubal Ectopic Pregnancy. Obstet Gynecol. 2018;131(3):e91-e103.

14. Kirk E, Daemen A, Papageorghiou AT, Bottomley C, Condous G, De Moor B, Timmerman D, Bourne T. Why are some ectopic pregnancies characterized as pregnancies of unknown location at the initial transvaginal ultrasound examination?. Acta obstetricia et gynecologica Scandinavica. 2008 Nov;87(11):1150-4.

15. Barnhart K, Sammel MD, Chung K, Zhou L, Hummel AC, Guo W. Decline of serum human chorionic gonadotropin and spontaneous complete abortion: defining the normal curve. Obstetrics & Gynecology. 2004 Nov 1;104(5):975-81. .

16. Abbott J, Emmans LS, Lowenstein SR. Ectopic pregnancy: ten common pitfalls in diagnosis. Am J Emerg Med. 1990;8(6):515-22.

17. Buckley RG, King KJ, Disney JD, Gorman JD, Klausen JH. History and physical examination to estimate the risk of ectopic pregnancy: validation of a clinical prediction model. Annals of emergency medicine. 1999;34(5):589-94.

18. Barnhart KT, Sammel MD, Takacs P, Chung K, Morse CB, O'Flynn O'Brien K, et al. Validation of a clinical risk scoring system, based solely on clinical presentation, for the management of pregnancy of unknown location. Fertil Steril. 2013;99(1):193-8.

19. Bobdiwala S, Saso S, Verbakel JY, Al-Memar M, Van Calster B, Timmerman D, et al. Diagnostic protocols for the management of pregnancy of unknown location: a systematic review and meta-analysis. Bjog. 2019;126(2):190-8.

20. Condous G, Okaro E, Khalid A, Timmerman D, Lu C, Zhou Y, et al. The use of a new logistic regression model for predicting the outcome of pregnancies of unknown location. Hum Reprod. 2004;19(8):1900-10.

21. Condous G, Van Calster B, Kirk E, Haider Z, Timmerman D, Van Huffel S, et al. Prediction of ectopic pregnancy in women with a pregnancy of unknown location. Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2007;29(6):680-7.

22. Dart R, Howard K. Subclassification of indeterminate pelvic ultrasonograms: stratifying the risk of ectopic pregnancy. Acad Emerg Med. 1998;5(4):313-9.

23. Mitchell TM. Machine learning. 1997. Burr Ridge, IL: McGraw Hill. 1997;45(37):870-7.

24. Drife J, Lewis G. Why Mothers Die 2000–2002 – The Sixth Report of Confidential Enquiries into Maternal Deaths in the United Kingdom; London, United Kingdom: Royal College of Obstetricians and Gynaecologists; 2004.

25. Bickell NA, Bodian C, Anderson RM et al. Time and risk of ruptured tubal pregnancy. Obstet Gynecol 2004; 104: 789–794.

26. Rollins J. Why we need a methodology for data science. Retrieved February. 2015;1:2017.

27. Foroughi F, Luksch P. Data science methodology for cybersecurity projects. arXiv preprint arXiv:180304219. 2018.

28. Cox DR. The regression analysis of binary sequences. Journal of the Royal Statistical Society: Series B (Methodological). 1958;20(2):215-32.

29. Zekić-Sušac M, Šarlija N, Has A, Bilandžić A. Predicting company growth using logistic regression and neural networks. Croatian operational research review. 2016;7(2):229-48.

30. Nadkarni P. Core Technologies: Machine Learning and Natural Language Processing. 2016. p. 85-114.

31. Vapnik V. Pattern recognition using generalized portrait method. Automation and remote control. 1963;24:774-80.

32. Fan W, editor On the optimality of probability estimation by random decision trees. AAAI; 2004.

33. KDnuggets. RapidMiner Named a Leader in the 2016 Gartner Magic Quadrant for Advanced Analytics Platforms 2016 [Available from: https://www.kdnuggets.com/2016/02/rapidminer-leader-2016-gartner-mq-advanced-analytics-platforms.html.

34. Hemlata PG. Experimental Evaluation of Open Source Data Mining Tools: R, Rapid Miner and KNIME.

35. Van Buuren S. Flexible imputation of missing data: CRC press; 2018.

36. Minhas S, Khanum A, Riaz F, Alvi A, Khan SA, Initiative AsDN, editors. Early alzheimer's disease prediction in machine learning setup: Empirical analysis with missing value computation. International Conference on Intelligent Data Engineering and Automated Learning; 2015: Springer.

37. Luengo J, García S, Herrera F. On the choice of the best imputation methods for missing values considering three groups of classification methods. Knowledge and information systems. 2012;32(1):77-108.

38. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med. 2018;18(3):91-3.

39. Fawcett T. An introduction to ROC analysis. Pattern recognition letters. 2006;27(8):861-74.

40. Condous G, Kirk E, Van Calster B, Van Huffel S, Timmerman D, Bourne T. Failing pregnancies of unknown location: a prospective evaluation of the human chorionic gonadotrophin ratio. Bjog. 2006;113(5):521-7.

41. Hahlin M, Thorburn J, Bryman I. The expectant management of early pregnancies of uncertain site. Hum Reprod. 1995;10(5):1223-7.

42. Banerjee S, Aslam N, Woelfer B, Lawrence A, Elson J, Jurkovic D. Expectant management of early pregnancies of unknown location: a prospective evaluation of methods to predict spontaneous resolution of pregnancy. Bjog. 2001;108(2):158-63.

43. Banerjee S, Aslam N, Zosmer N, Woelfer B, Jurkovic D. The expectant management of women with early pregnancy of unknown location. Ultrasound Obstet Gynecol. 1999;14(4):231-6.

44. Malek-mellouli M, Oumara M, Ben Amara F, Zouch O, Neji K, Reziga H. Prediction of ectopic pregnancy in early pregnancy of unknown location. Tunis Med. 2013;91(1):27-32.

45. Ishibuchi H, Nojima Y. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. International Journal of Approximate Reasoning. 2007;44(1):4-31.

46. Li C, Zhao WH, Zhu Q, Cao SJ, Ping H, Xi X, et al. Risk factors for ectopic pregnancy: a multi-center case-control study. BMC Pregnancy Childbirth. 2015;15:187.

47. Mol BW, Hajenius PJ, Engelsbel S, Ankum WM, Van der Veen F, Hemrika DJ, et al. Serum human chorionic gonadotropin measurement in the diagnosis of ectopic pregnancy when transvaginal sonography is inconclusive. Fertility and sterility. 1998;70(5):972-81.

48. Cervantes J, Lamont FG, López-Chau A, Mazahua LR, Ruíz JS. Data selection based on decision tree for SVM classification on large data sets. Applied Soft Computing. 2015;37:787-98.

49. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Front Public Health. 2017;5:307.

50. Nzaumvila DK, Govender I, Ogunbanjo GA. An audit of the management of ectopic pregnancies in a district hospital, Gauteng, South Africa. Afr J Prim Health Care Fam Med. 2018;10(1):e1-e8.

51. Lewis G, Drife J. Confidential enquiry into maternal and child health. Why mothers die 2000-2002. Sixth report of the Confidential Enquiries into Maternal Deaths in the United Kingdom. 2004.

52. Alexander LK, Lopes B, Ricchetti-Masterson K, Yeatts KB. Assessment of Diagnostic and Screening Tests.

53. Hammer M. Reengineering work: Don't automate, obliterate. Harvard business review. 1990;68(4):104-12.

54. Linoff GS, Berry MJ. Data mining techniques: for marketing, sales, and customer relationship management: John Wiley & Sons; 2011.

## Declarations

Competing interests: The authors declare no competing interests.

## Figures

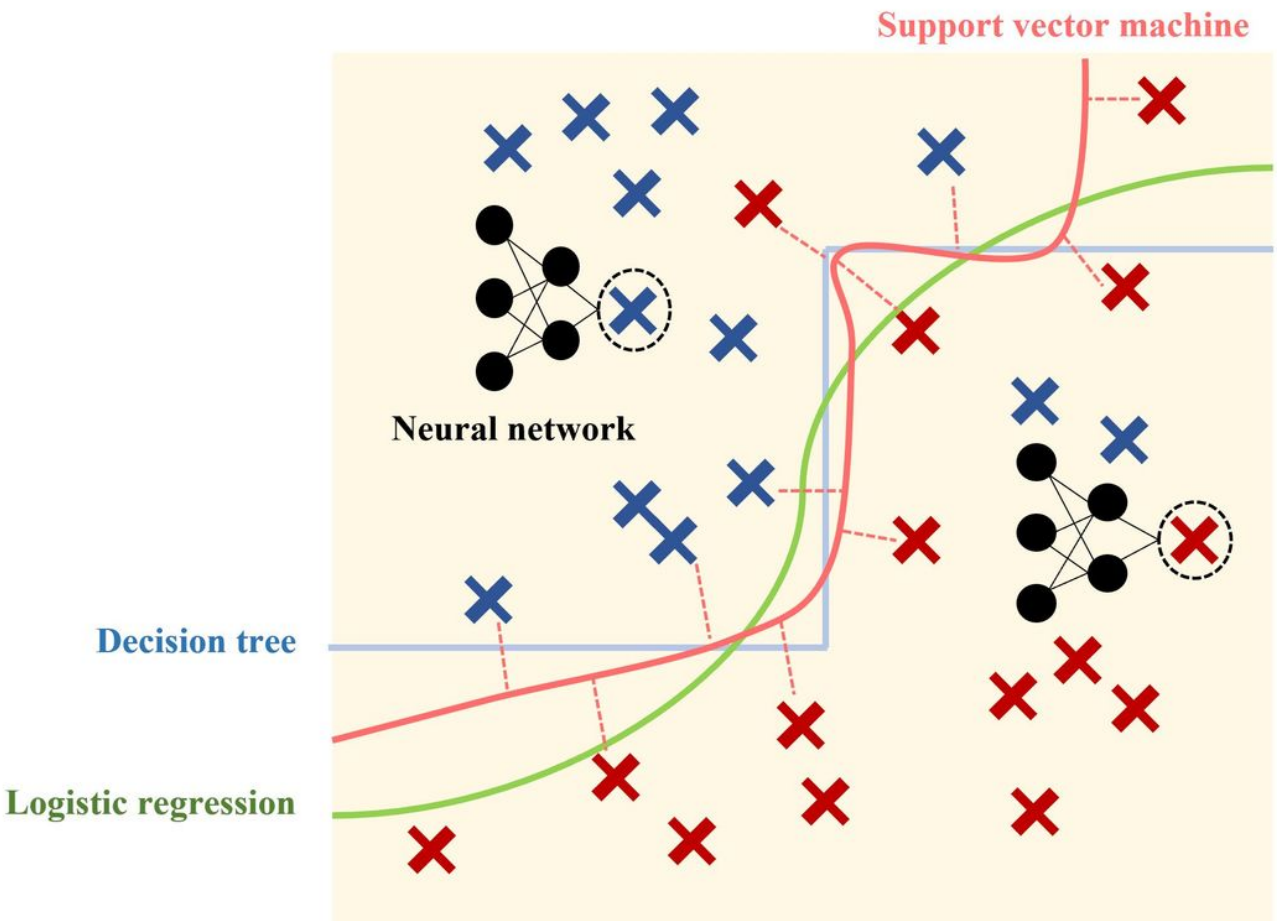
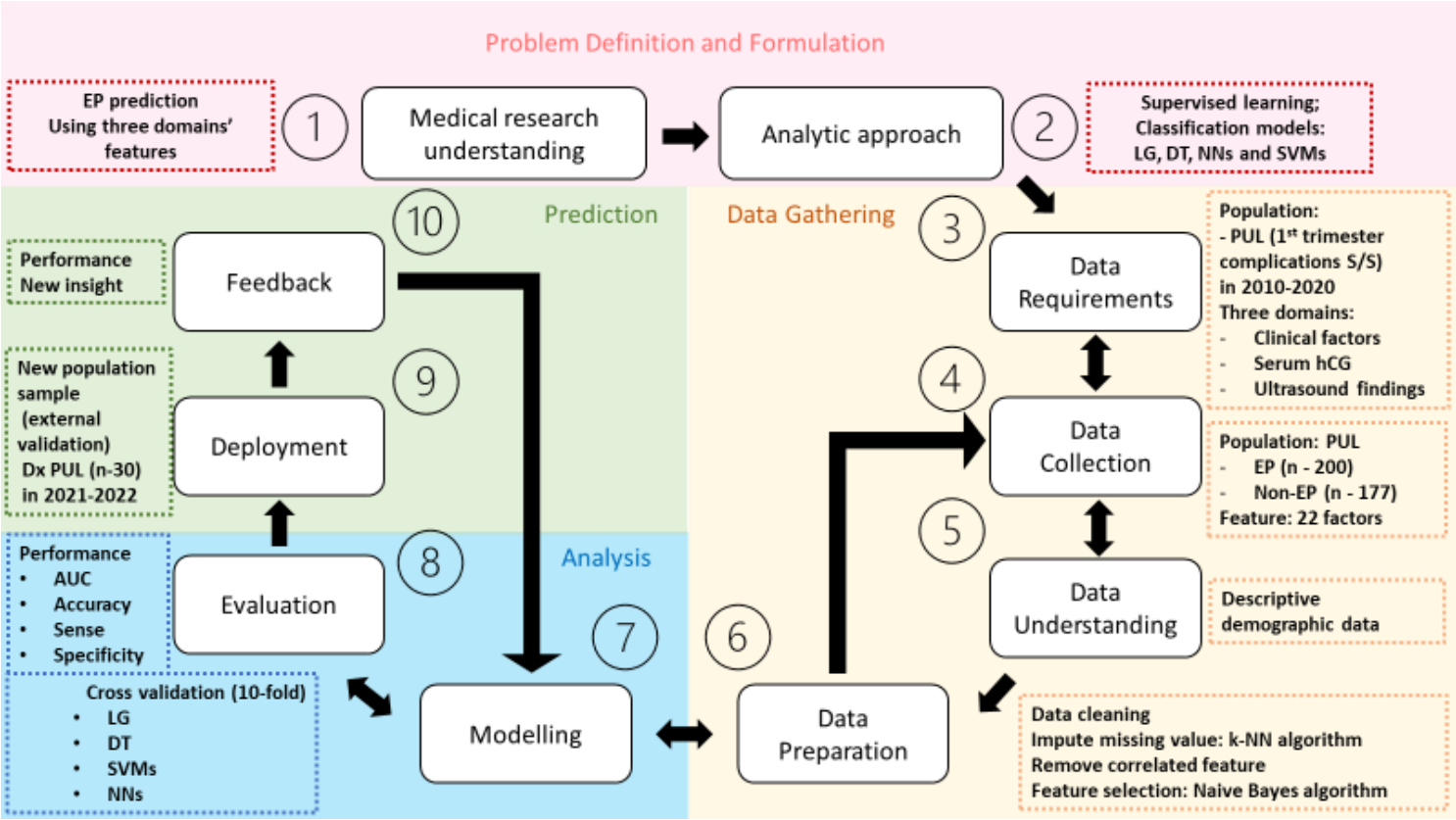


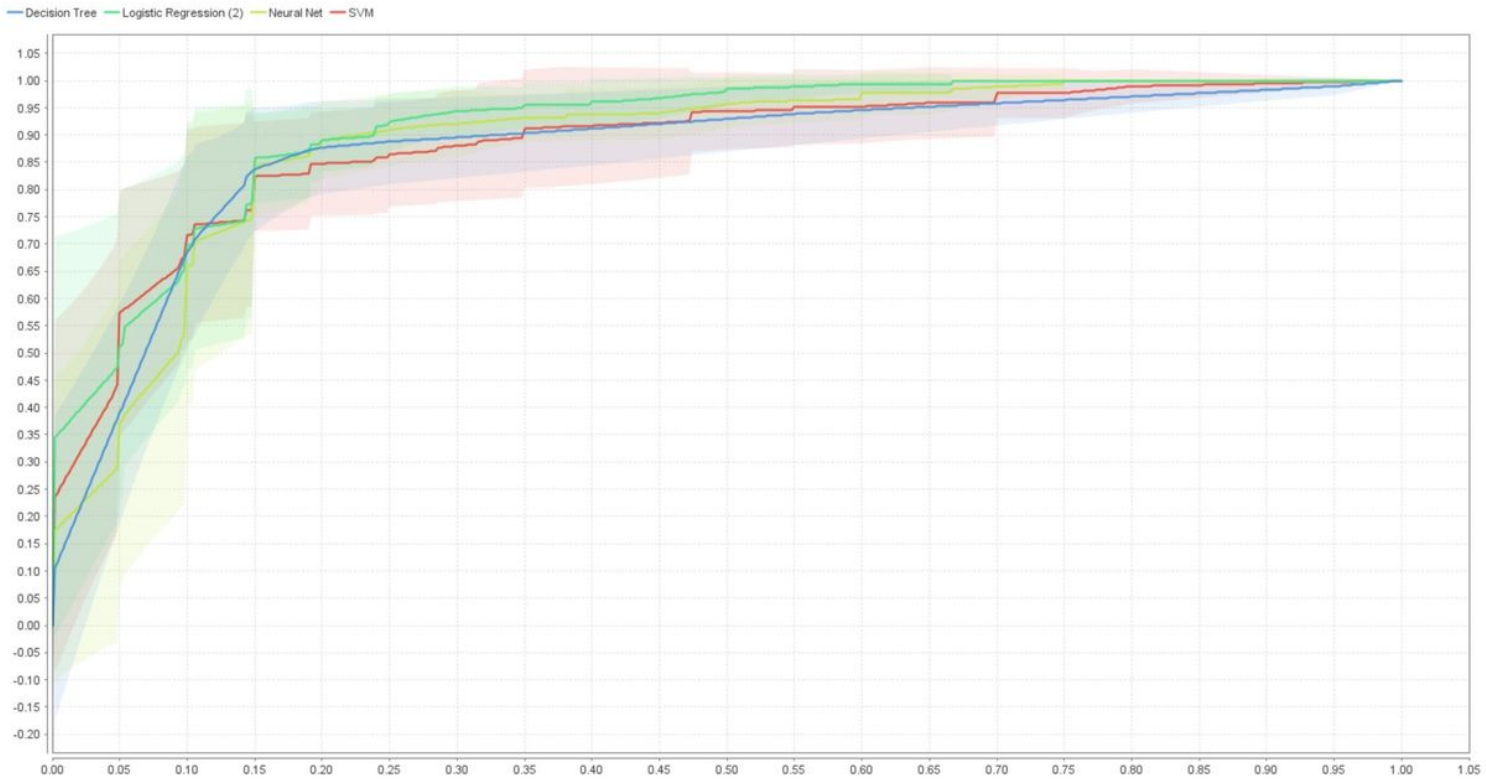

**Figure 1**

Conceptual overview of four predictive models




## Figure 2

Study flow diagram




## Figure 3

AUC-ROC performances

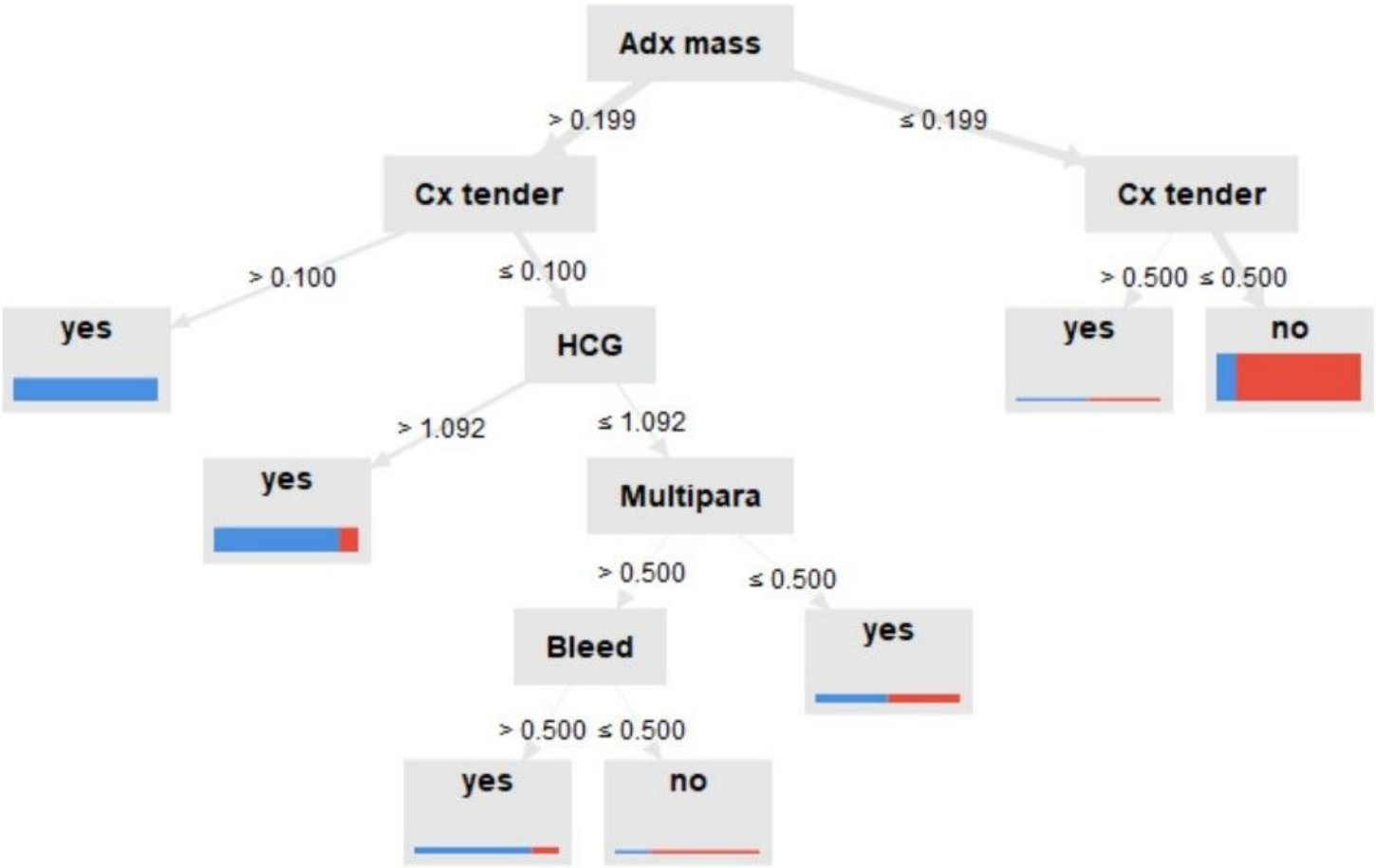
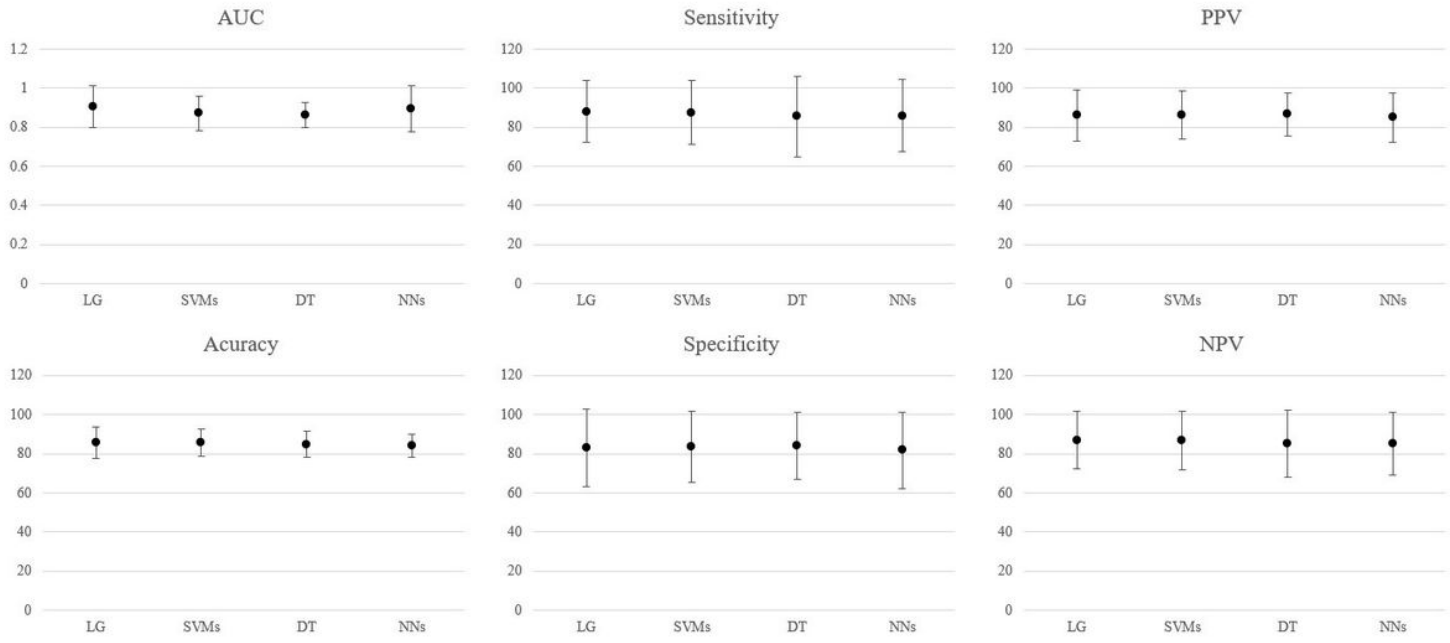



**Figure 4**

Decision tree model




## Figure 5

Predictive performance of the four models using cross validation (Internal validation)