

## Predicting Thrombophilia using Neural Networks and Decision Trees

A. R. Avdić, N. Z. Đorđević, U. A. Marovac, L. M. Memić, Z. Č. Dolićanin, G. M. Babić

**Abstract:** The occurrence of thrombophilia during pregnancy results from a complex interaction of inherited and acquired factors, followed by an increase in blood coagulation and subsequent placental ischemic conditions. In this paper, a novel method is presented, whose aim is early identification of the risk of developing thrombophilia in pregnancy. The proposed method is based on machine learning algorithms: decision trees and neural networks. The research uses a dataset consisting of demographic, lifestyle, and clinical information from 35 pregnant women (22 healthy and 13 with thrombophilia). The results show the effectiveness of decision trees and neural networks in accurately predicting the risk of developing thrombophilia in pregnancy. The implications of this research are significant for clinical practice and it provides a valuable tool for early identifying women with high risk of thrombophilia in pregnancy that can enable improvement of preventive measures, such as lifestyle modifications and the use of therapeutic prophylaxis. In conclusion, this paper demonstrates the potential of machine learning algorithms for the prediction of thrombophilia in pregnancy. By combining advanced computational techniques with comprehensive datasets, we can enhance our understanding of thrombophilia in pregnancy risk factors and improve patient outcomes through personalized preventive measures.

**Keywords:** neural networks, decision trees, machine learning, thrombophilia in pregnancy, and prediction.

### 1 Introduction

Pregnancy is a condition characterized by increased blood coagulability due to hormonal changes. Increased blood coagulability during pregnancy represents an adaptation mechanism that allows for a reduction in the risk of bleeding during pregnancy and after childbirth, but also represents a high risk of thromboembolic events. The risk of thromboembolic events in pregnant and postpartum women is 5 times higher compare to non-pregnant women. Blood hypercoagulability and hemostasis disorders in pregnancy lead

---

Manuscript received August 27, 2023; accepted November 1, 2023.

A. R. Avdić is with the State University of Novi Pazar, Novi Pazar, Serbia; N. Z. Đorđević is with the State University of Novi Pazar, Novi Pazar, Serbia; U. M. Marovac is with State University of Novi Pazar, Novi Pazar, Serbia; L. M. Memić is with the State University of Novi Pazar, Novi Pazar, Serbia; Z. Č. Dolićanin is with State University of Novi Pazar, Novi Pazar, Serbia; G. M. Babić is with the University of Kragujevac, Faculty of Medical Sciences, Kragujevac, Serbia.

to thrombophilia. Thrombophilia in pregnancy is accompanied by the formation of microthrombi that prevent normal uteroplacental blood flow, which results in the occurrence of ischemic conditions in the placenta that lead to the occurrence of various pathologies (recurrent pregnancy loss, fetal growth restriction, late miscarriages, preeclampsia, and stillbirth). Thrombophilia during pregnancy occurs as a result of the interaction of hereditary and acquired factors. The most common genetic factors inducing thrombophilia in pregnancy are FV Leiden (G1691A), FII (G20210A), MTHFR (C677T) and PAI-1 mutations. Stresses, as well as numerous environmental, demographic, sociopsychological and lifestyle factors are associated with the occurrence of thrombophilia in pregnancy. However, many of these factors are not taken into account when assessing the risk of developing thrombophilia in pregnancy due to their still unknown influence on the etiology of this disease. The identification of certain demographic and sociopsychological risk factors associated with genetic predisposition to thrombophilia in pregnancy would enable an individual assessment of the risk of developing this disease, and therefore more adequate and timely prevention and therapy [13].

Machine learning (ML) algorithms [8, 24] have emerged as powerful tools for analyzing complex clinical data and making accurate predictions. By leveraging ML techniques, researchers have made significant progress in various medical domains, including disease diagnosis, prognosis, and risk prediction. In recent years, the application of ML algorithms, particularly neural networks, has shown promise in thrombophilia prediction. The aim of this research was to develop a new approach to identifying environmental risk factors in pregnant women with thrombophilia using ML algorithms, with a special focus on decision trees and neural networks. The new approach would contribute to improving the accuracy of risk assessment of the occurrence of thrombophilia in pregnancy and to a better understanding of the environmental factors that contribute to its development. In this study, we used a comprehensive dataset consisting of clinical, demographic, and lifestyle data. Feature engineering techniques were used to extract relevant features that capture the essential aspects of thrombophilia risk in pregnancy. These features were then used to train and evaluate different ML models using Weka and Python.

The results of this research have the potential to influence clinical practice significantly. Accurately predicting the risk of developing thrombophilia can enable healthcare professionals to identify high-risk pregnant women early, which leads to the improvement of preventive measures, such as lifestyle modifications and the use of therapeutic prophylaxis. Ultimately, this could contribute to reducing the incidence and severity of thrombotic events in pregnancy, improving pregnancy outcomes and the quality of life of pregnant women with thrombophilia. In the following sections, we will detail the methodology, results, and discussion of our study, presenting the effectiveness of machine learning algorithms in thrombophilia prediction.

## 2 Related Research

Prediction systems [20] using ML algorithms have gained significant attention in the

field of healthcare for various diseases, but we didn't find papers that are related to predicting thrombophilia in pregnancy using ML methods. Researchers have explored the potential of these systems to improve risk assessment, early detection, and personalized interventions. In this section, we review related work on prediction systems for other diseases. Several studies have investigated the use of machine learning algorithms for different disease prediction. In [2, 4, 22], authors gave a review of using ML algorithms for prediction in diseases. Especially, the application of neural networks and deep learning over medical data are described in papers [16, 18, 3, 23, 15, 12, 9, 14, 6, 5, 22]. Machine learning algorithms were used in predicting patient deterioration in vital-sign data [7], and prediction of the waring score of hospital mortality using decision trees [11, 10]. Then the LSTM, diabetic retinopathy, ventricular tachycardia, cardiac arrest, heart failure, and breast cancer [15, 9, 14, 19, 6, 17] were predicted using different ML models. In [21], the boosting model was used to predict transfer to the pediatric care unit. Overall, the literature demonstrates the potential of ML algorithms in predicting thrombophilia in pregnancy.

### 3 Materials and Methods

The studied population consisted of 35 pregnant women who were accommodated at the Gynecology and Obstetrics Clinic of the University Clinical Center Kragujevac for treatment or delivery, and 13 of whom had a diagnosis of thrombophilia. Thrombophilia in pregnant women was diagnosed based on confirmed mutations: FV Leiden (G1691A), FII (G20210A), MTHFR (C677T) and PAI-1. The dataset consists of 35 clinical features which are numeric (method of conception, week of pregnancy, systolic blood pressure, number of erythrocytes, myometrial thickness, fetal femur length, etc.) and 25 nominal demographic (smoking in pregnancy, number of cigarettes before pregnancy, place of residence, family relative (brothers and sisters), number of persons in the household, assessment of a healthy lifestyle, etc.), collected for each participant. There were no missing values. In addition, data on the way emotions are regulated were collected from each pregnant woman using the Affective styles questionnaire. The Ethics Committee of the University Clinical Center Kragujevac approved the study protocol, and all patients gave informed consent before data collection.

The statistics of numeric clinical data are given in Table 1, and statistics of numeric demographic data are given in Table 2, and the statistics of nominal demographic and lifestyle data are given in the Table 3.

The raw data is transformed into a suitable format for ML algorithms. Decision Trees are a popular supervised ML algorithm for classification and regression tasks. They recursively split the input space based on feature values to create a hierarchical structure that represents decision rules. Neural networks, a type of ML algorithm inspired by the structure and functioning of the human brain, were utilized in this study. Specifically, deep learning techniques were employed to leverage the power of neural networks in capturing complex patterns and relationships within the data.

Decision Trees represent a machine learning model used for classification and regres-

Table 1. Statistics of clinical data

Features/Statistics	count	mean	std	min	25%	50%	75%	max
Maternal age	35	29.333	4.321	18	27	30	31.5	39
Week of gestation	35	38.779	1.529	36.06	37.05	39.01	40.025	42
Maternal weight	35	78.675	12.324	57	70	78.65	87	109
Maternal height	35	169.736	5.401	155	168	170	170.45	185
Maternal weight gain during pregnancy	35	14.307	3.701	10	12.215	13	15.71	25
Fetal biparietal diameter (BPD)	35	12.201	14.742	8.76	9.37	9.5	9.9	96.8
Fetal head circumference (HC)	35	33.465	1.201	31.33	32.76	33.33	34.25	37.08
Fetal abdomen circumference (AC)	35	34.103	1.462	30.52	33.2	34.13	35.345	36.29
Fetal femur length (FL)	35	7.553	0.33	6.79	7.375	7.5	7.775	8.3
Estimated fetal weight (EFW)	35	3459.086	322.075	2800	3300	3460	3650	4140
Myometrium thickness	35	0.776	0.18	0.41	0.7	0.79	0.82	1.23
Length of the fetal cerebellum	35	5.439	0.32	4.88	5.27	5.42	5.565	6.59
Placenta thickness	35	4.199	1.02	2.3	3.5	4.12	4.67	7.73
Maternal blood type	35	2.114	1.051	1	1	2	3	4
Maternal Rh factor	35	1.143	0.355	1	1	1	1	2
Maternal hematocrit (Hct)	35	0.344	0.03	0.28	0.33	0.34	0.35	0.43
Maternal systolic blood pressure	35	115.971	10.709	100	110	115	120	140
Maternal diastolic blood pressure	35	73.571	7.374	60	70	72	80	90
Mother's erythrocyte count	35	3.907	0.403	2.98	3.705	3.87	4.015	4.99
Mother's leukocyte count	35	9.366	2.402	3.7	8.35	9.2	10.03	17.3
Maternal hemoglobin concentration	35	113.03	10.273	95	106.5	110.75	117.5	142
Mother's platelet count	35	220.546	61.992	106	193.5	207.75	223.56	463
Maternal iron (Fe) concentration	35	15.459	5.961	4.8	11.77	11.77	21.7	27.2
Maternal creatinine concentration	35	54.699	8.542	36	51	55.31	55.655	84
Maternal urea concentration	35	2.369	0.653	1.3	2.1	2.22	2.62	4
Maternal prothrombin time	35	11.155	0.456	10.1	11.08	11.2	11.2	12.7
Maternal international normalised ratio (INR)	35	0.973	0.026	0.914	0.97	0.97	0.98	1.045
Maternal activated partial thromboplastin time (aPTT)	35	26.473	2.349	22.2	25.18	25.18	28.48	33.6
Fetal heart rate	35	136.743	7.578	121	133	138	139	150
Maternal heart rate	35	87.286	11.372	51	85.5	86	92	118
Umbilical cord diameter	35	1.756	0.258	1.14	1.625	1.78	1.84	2.64
Fetal middle cerebral artery resistance index (MCARI)	35	0.754	0.083	0.56	0.725	0.75	0.81	0.92
Uterine artery resistance index (AURI)	35	0.752	0.166	0.41	0.73	0.79	0.885	0.94
Umbilical artery resistance index (UARI)	35	0.577	0.063	0.44	0.56	0.58	0.605	0.71
Umbilical cord wall thickness	35	0.286	0.083	0.17	0.23	0.28	0.29	0.53
Fetal peak systole	35	52.14	10.793	30.47	43.53	50.78	56.6	80.41
Fetal abdominal artery resistance index (AARI)	35	0.766	0.053	0.62	0.745	0.76	0.78	0.87

sion tasks. The fundamental idea is to decompose decision-making problems into a series of smaller decisions based on data attributes. Each node in the tree represents a test of a specific attribute, and branches from the node reflect different values of that attribute. Classification is performed at the leaves, which represent the final decisions of the model.

CNNs (Convolutional Neural Networks) are particularly effective in processing data with spatial structures, such as images. These networks use convolutional layers to extract local patterns and features from input data. Convolutional layers employ filters that move across input data, and pooling layers are used to reduce dimensionality and extract essential information. CNNs are commonly employed in tasks like image recognition, object detection, and similarity analysis.

RNNs (Recurrent Neural Networks) are designed to work with sequential data where the order of information is crucial. The key characteristic of RNNs is the ability to store previous states and transmit information between current and previous time steps. However, RNNs face challenges in handling long-term dependencies. Variations of RNNs, such as

Table 2. Statistics of numeric demographic and lifestyle data

Features/Statistics	count	mean	std	min	25%	50%	75%	max
Pregnant women	35	0.743	0.98	0	0	0	2	2
Number of persons in household	35	2.571	1.577	1	1	2	3.5	7
Material status satisfaction	35	3.629	0.646	3	3	4	4	5
Number of children	35	0.486	0.658	0	0	0	1	2
Pregnancy in order	35	2.086	1.121	1	1	2	3	4
Spontaneous miscarriages	35	0.571	0.884	0	0	0	1	3
The length of the partner relationship	35	5.426	3.417	1	3	5	7	14
Sisters and brothers	35	1.629	1.536	0	1	1	2	6
Number of cigarettes before pregnancy	35	8.086	7.516	0	0	10	15	20
Number of cigarettes in pregnancy	35	1.686	3.636	0	0	0	0.5	15
Covering up	35	22.686	5.88	9	19	23	27	34
Adaptation	35	24.486	4.553	16	21	25	27.5	34
Toleration	35	17.057	2.92	11	15	16	19	24
Affective style	35	64.229	9.496	43	57	65	71.5	87
Maternal Rh factor	35	1.143	0.355	1	1	1	1	2
Maternal hematocrit (Hct)	35	0.344	0.03	0.28	0.33	0.34	0.35	0.43
Maternal systolic blood pressure	35	115.971	10.709	100	110	115	120	140
Maternal diastolic blood pressure	35	73.571	7.374	60	70	72	80	90
Mother's erythrocyte count	35	3.907	0.403	2.98	3.705	3.87	4.015	4.99
Mother's leukocyte count	35	9.366	2.402	3.7	8.35	9.2	10.03	17.3
Maternal hemoglobin concentration	35	113.03	10.273	95	106.5	110.75	117.5	142
Mother's platelet count	35	220.546	61.992	106	193.5	207.75	223.56	463
Maternal iron (Fe) concentration	35	15.459	5.961	4.8	11.77	11.77	21.7	27.2
Maternal creatinine concentration	35	54.699	8.542	36	51	55.31	55.655	84
Maternal urea concentration	35	2.369	0.653	1.3	2.1	2.22	2.62	4
Maternal prothrombin time	35	11.155	0.456	10.1	11.08	11.2	11.2	12.7
Maternal international normalised ratio (INR)	35	0.973	0.026	0.914	0.97	0.97	0.98	1.045
Maternal activated partial thromboplastin time (aPTT)	35	26.473	2.349	22.2	25.18	25.18	28.48	33.6
Fetal heart rate	35	136.743	7.578	121	133	138	139	150
Maternal heart rate	35	87.286	11.372	51	85.5	86	92	118
Umbilical cord diameter	35	1.756	0.258	1.14	1.625	1.78	1.84	2.64
Fetal middle cerebral artery resistance index (MCARI)	35	0.754	0.083	0.56	0.725	0.75	0.81	0.92
Uterine artery resistance index (AURI)	35	0.752	0.166	0.41	0.73	0.79	0.885	0.94
Umbilical artery resistance index (UARI)	35	0.577	0.063	0.44	0.56	0.58	0.605	0.71
Umbilical cord wall thickness	35	0.286	0.083	0.17	0.23	0.28	0.29	0.53
Fetal peak systole	35	52.14	10.793	30.47	43.53	50.78	56.6	80.41
Fetal abdominal artery resistance index (AARI)	35	0.766	0.053	0.62	0.745	0.76	0.78	0.87

Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have been developed to address these challenges, enabling better management of long-term dependencies.

To evaluate the performance of the model, multiple metrics were utilized, including precision, recall, accuracy, F1-score, and error. Precision measures the proportion of true positives out of all predicted positive instances, recall measures the proportion of true positives identified correctly, accuracy measures the overall correctness of predictions and the F1-score provides a balanced evaluation of precision and recall.

The entire study, including data preprocessing, model development, training, and evaluation, was implemented using Weka and Python programming language. Python libraries such as TensorFlow, Keras, and sci-kit-learn were employed for implementing neural networks, deep learning techniques, and performance metric calculations.

To ensure the reliability and robustness of the model, cross-validation techniques, such as 10-fold cross-validation and Leave-on-out, were employed. Cross-validation (10-fold) involved partitioning the dataset into 10 subsets (10-fold) and performing training and eval-

Table 3. Statistics of nominal demographic data

Attribute	Value	Occurrences	Frequency	Entropy	Median
Marital status	'married'	35	1	0	'married'
Education	'medium'	18	0.514	1.158	'medium'
Education	'high'	16	0.457	1.158	'medium'
Education	'basic'	1	0.029	1.158	'medium'
Education	'yes'	25	0.714	0.863	'yes'
Education	'no'	10	0.286	0.863	'yes'
Type of job	'private job'	13	0.371	1.72	'private job'
Type of job	'state job'	11	0.314	1.72	'private job'
Type of job	'does not have'	10	0.286	1.72	'private job'
Type of job	'self employed'	1	0.029	1.72	'private job'
Place of living	'city'	22	0.629	1.282	'city'
Place of living	'village'	9	0.257	1.282	'city'
Place of living	'suburb'	4	0.114	1.282	'city'
Object of living	'house'	27	0.771	0.776	'house'
Object of living	'apartment'	8	0.229	0.776	'house'
Apartment ownership	'partner'	24	0.686	1.309	'partner'
Apartment ownership	'common'	7	0.2	1.309	'partner'
Apartment ownership	'personally'	2	0.057	1.309	'partner'
Apartment ownership	'subtenant'	2	0.057	1.309	'partner'
Proximity to relatives	'yes'	22	0.629	0.951	'yes'
Proximity to relatives	'no'	13	0.371	0.951	'yes'
Children from marriage	'does not have'	21	0.6	1.121	'does not have'
Children from marriage	'yes'	13	0.371	1.121	'does not have'
Children from marriage	'no'	1	0.029	1.121	'does not have'
Planned pregnancy	'yes'	29	0.829	0.66	'yes'
Planned pregnancy	'no'	6	0.171	0.66	'yes'
Desired pregnancy	'yes'	35	1	0	'yes'
Mode of pregnancy	'naturally'	33	0.943	0.315	'naturally'
Mode of pregnancy	'in vitro fertilization'	2	0.057	0.315	'naturally'
Pregnancy outside of marriage	'no'	34	0.971	0.189	'no'
Pregnancy outside of marriage	'yes'	1	0.029	0.189	'no'
Condom use	'no'	20	0.571	0.985	'no'
Condom use	'yes'	15	0.429	0.985	'no'
Marital status of parents	'married'	30	0.857	0.592	'married'
Marital status of parents	'divorced'	5	0.143	0.592	'married'
Growing up in the family	'own'	35	1	0	'own'
Childhood	'happy'	34	0.971	0.189	'happy'
Childhood	'not happy'	1	0.029	0.189	'happy'
Smoker before pregnancy	'yes'	22	0.629	0.951	'yes'
Smoker before pregnancy	'no'	13	0.371	0.951	'yes'
Smoker during pregnancy	'no'	25	0.714	0.863	'no'
Smoker during pregnancy	'yes'	10	0.286	0.863	'no'
Healthy lifestyles	'stick to it'	23	0.657	1.15	'stick to it'
Healthy lifestyles	'healthy'	10	0.286	1.15	'stick to it'
Healthy lifestyles	'completely healthy'	2	0.057	1.15	'stick to it'
Need for psychologist	'no'	27	0.771	0.776	'no'
Need for psychologist	'yes'	8	0.229	0.776	'no'

uation iteratively, ensuring that each subset served as both the training and testing set at least once. Leave-one-out cross-validation is performed because it is recommended for

small datasets such as ours.

By following these materials and methods, we aimed to develop a predictive model for thrombophilia in pregnancy using neural networks and decision trees. The Python programming language facilitated the implementation of the model, and performance evaluation metrics such as precision, recall, accuracy, F1-score, and error provided comprehensive insights into the model's predictive capabilities.

## 4 Experiment Results

The following ML algorithms based on decision trees were applied to the demographic, lifestyle, and clinical data:

- DecisionTree classification in Python,
- RandomTree classification in Weka.

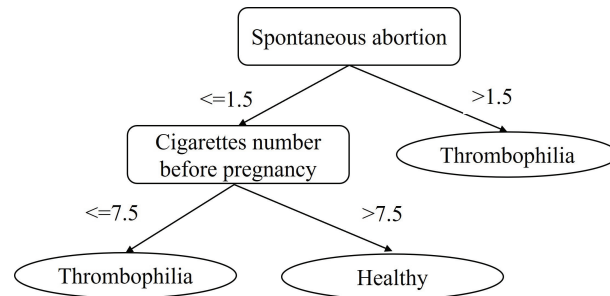


Fig. 1. Demographic and lifestyle data - the decision tree obtained in Python

Since the demographic and lifestyle data contained textual attributes, it was necessary to encode them using the OneHotEncoder package from the sklearn library, for DecisionTree classification. In Weka, it was necessary to use a filter for demographic and clinical data, ie. for the last class, NumericToNominal, to apply the RandomTree classification. The following decision tree rules for demographic and lifestyle data were obtained and shown in Fig. 1 and Fig. 2. The decision trees for clinical data are given in Fig. 3 and Fig. 4.

The classification on the test set was done using a training set (80% - 20% split) (Test), 10-fold cross-validation (CV-10 fold) and Leave One Out (CV-Loo). The results of the classification of the training set in Weka is 100%, and for algorithms implemented in Python, the results are given in tables (Train). Leave One Out cross-validation is performed only in Python. The following results were obtained and shown in Table 4. The results indicate that using cross-validation slightly higher accuracy is obtained using the DecisionTree classifier, and, on such a small data set, the results are good. RandomTree gave better classification results based on the training set. In addition, both trees take smoking during pregnancy and previous spontaneous abortions as key factors for predicting thrombophilia in pregnancy, while RandomTree considers the living environment as an important factor.

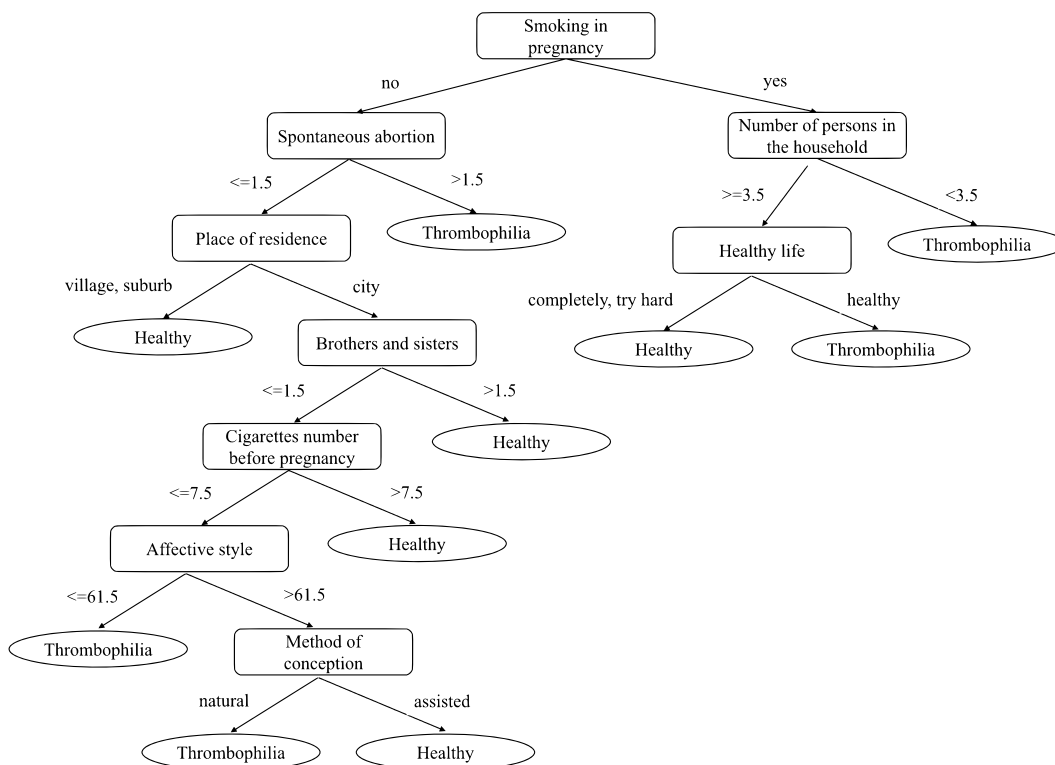


Fig. 2. Demographic and lifestyle data - the decision tree obtained in Weka

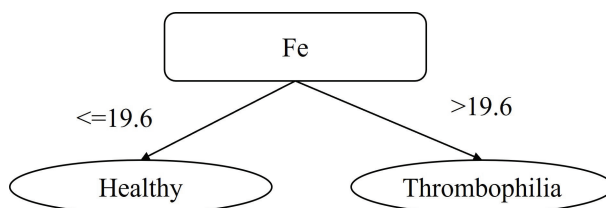


Fig. 3. Clinical data - the decision tree obtained in Python

The methods were applied to clinical data in the same way, and again a simpler tree was obtained for the DecisionTree classifier, where Fe level is crucial, while RandomTree showed that week of pregnancy, systolic blood pressure, myometrial thickness, fetal femur length, and a number of erythrocytes are important. Also, there are very good classification results, over 80%, which means that clinical factors give us a more accurate prediction than demographic ones (Table 5).

Algorithms based on neural networks were applied to demographic and clinical data, namely MultilayerPerception in Weka, then using convolutional and recurrent neural networks in Python. Classification results for demographic and clinical data using the MultilayerPerception algorithm are shown in Table 6, and classification accuracy using CNN



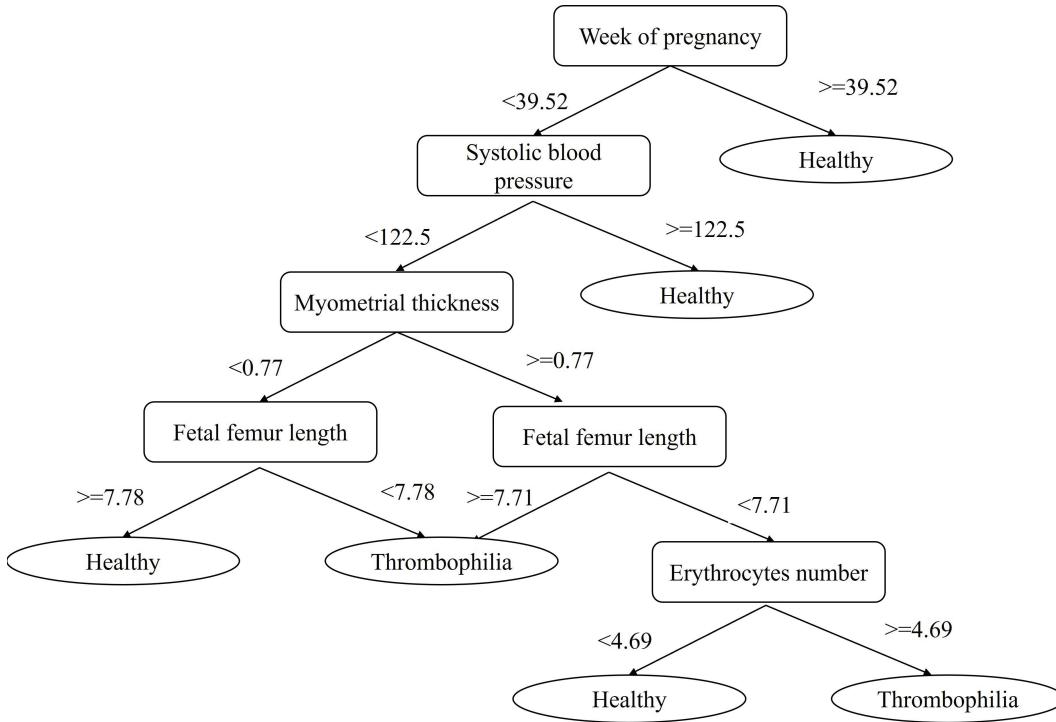


Fig. 4. Clinical data - the decision tree obtained in Weka

Table 4. Classification results on demographic and lifestyle data

Classifier	Accuracy	Precision	Recall	F1-score	Error
<b>DecisionTree (Train)</b>	88.57%	95.00%	86.36%	90.48%	11.43%
<b>DecisionTree (Test)</b>	42.86%	50.00%	25.00%	33.33%	57.14%
<b>RandomTree (Test)</b>	71.42%	71.42%	71.42%	71.42%	28.58%
<b>DecisionTree (CV-10 fold)</b>	71.67%	62.92%	70.83%	62.67%	28.33%
<b>RandomTree (CV-10 fold)</b>	62.86%	62.3%	62.9%	62.5%	37.14%
<b>DecisionTree (CV-Loo)</b>	74.29%	64.29%	69.23%	66.67%	25.11%

and RNN in Table 7.

The results show that good prediction (above 85%) is achieved using MultilayerPerception on clinical data, and also that RNN is more suitable for this data, which is expected having in mind its application. CNN should be used in the future if the dataset will be enlarged using some images (some health recording, for example, ultrasound etc.).

Table 5. Classification results on clinical data

Classifier	Accuracy	Precision	Recall	F1-score	Error
<b>DecisionTree (Train)</b>	97.14%	100.00%	95.45%	97.67%	2.86%
<b>DecisionTree (Test)</b>	85.71%	100.00%	75.00%	85.71%	14.29%
<b>RandomTree (Test)</b>	71.43%	71.43%	71.43%	71.43%	28.57%
<b>DecisionTree (CV - 10 fold)</b>	85.83%	82.08%	85.00%	81.45%	14.17%
<b>RandomTree (CV - 10 fold)</b>	82.86%	82.80%	82.90%	82.50%	17.14%
<b>DecisionTree (CV - Loo)</b>	85.71%	83.33%	76.92%	80.00%	14.29%

Table 6. Classification results on clinical and demographic data using MultilayerPerception

Data	Accuracy	Precision	Recall	F1-score	Error
<b>Clinical (Test)</b>	85.71%	88.10%	85.70%	84.00%	14.29%
<b>Demographic (Test)</b>	71.43%	71.40%	71.40%	71.40%	28.57%
<b>Clinical (CV - 10 fold)</b>	85.71%	87.4%	85.7%	85.9%	14.29%
<b>Demographic (CV - 10 fold)</b>	74.29%	74.8%	74.3%	74.5%	25.71%

Table 7. Classification results on clinical and demographic data using CNN and RNN

Data	Accuracy CNN	Accuracy RNN
<b>Clinical</b>	66.67% ± 14.43%	79.17% ± 23.35%
<b>Demographic</b>	32.5% ± 24.85%	11.67% ± 14.53%

## 5 Conclusion

In this study, we developed a prediction model of thrombophilia in pregnancy using decision trees and neural network techniques. Using a comprehensive data set consisting of clinical, lifestyle, and demographic information, we aimed to improve the accuracy of risk assessment of the occurrence of thrombophilia in pregnancy and to a better understanding of the environmental factors that contribute to its development. Through our experiments, we demonstrated the potential of decision trees and neural networks in predicting of thrombophilia in pregnancy. The model achieved promising results, outperforming the baseline models and showing high precision, recall, accuracy, F1-score and error. These findings highlight the effectiveness of ML algorithms, with over 85% accuracy on clinical data in identifying women at risk of developing thrombophilia during pregnancy. Despite the promising results, there are several areas for future research. Having in mind, that we achieved good results without feature extraction, we can expect better results in the future using FE techniques and with an enlarging dataset. Furthermore, other ML algorithms can be implemented and compared in order to get even better results. In conclusion, this study

demonstrates the potential of machine learning techniques in predicting thrombophilia in pregnancy. The developed model is promising in the precise identification of women at risk of developing thrombophilia during pregnancy, which leads to the development of personalized preventive and therapeutic measures. With further refinement and validation prediction systems may contribute to better risk management and improved health outcomes in pregnant women susceptible to thrombophilia.

## References

- [1] A. AVDIĆ, N. ĐORĐEVIĆ, U. MAROVAC, L. MEMIĆ, Z. DOLIĆANIN, and G. BABIĆ, *Thrombophilia Prediction Using Machine Learning Algorithms*, In 2nd International Conference on Chemo and BioInformatics. University of Kragujevac, Institute for Information Technologies (2023), 140-143.
- [2] E. CHRISTODOULOU, J. MA, G. S. COLLINS, E. W. STEYERBERG, J. Y. VERBAKEL, and B. VAN CALSTER, *A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models*, J. Clin. Epidemiol., vol. 110 (2019), 12–22.
- [3] Y. CHENG, F. WANG, P. ZHANG, and J. HU, *Risk prediction with electronic health records: A deep learning approach*, in Proceedings of the 2016 SIAM International Conference on Data Mining, 2016.
- [4] M. CHEN, Y. HAO, K. HWANG, L. WANG and L. WANG, *Disease prediction by machine learning over big data from healthcare communities*, IEEE Access, vol. 5 (2017), 8869–8879.
- [5] E. CHOI, M. T. BAHADORI, A. SCHUETZ, W. F. STEWART, and J. SUN, *Doctor AI: Predicting clinical events via recurrent neural networks*, JMLR Workshop Conf. Proc., vol. 56 (2016), 301–318.
- [6] E. CHOI, A. SCHUETZ, W. F. STEWART, and J. SUN, *Using recurrent neural network models for early detection of heart failure onset*, J. Am. Med. Inform. Assoc., vol. 24, 2 (2017), 361–370.
- [7] L. CLIFTON, D. A. CLIFTON, P. J. WATKINSON, and L. TARASSENKO, *Identification of patient deterioration in vital-sign data using one-class support vector machines*, in 2011 Federated Conference on Computer Science and Information Systems (FedCSIS), 2011, 125–131.
- [8] Ö. ESKICIOĞLU, E. DOLIĆANIN, A. IŞIK, and K. RIFAI, *Recognition and detection with deep learning methods*, Scientific Publications of the State University of Novi Pazar. Ser. A: Appl. Math. Inform. And Mech., 13, 2 (2021), 105-115.
- [9] V. GULSHAN ET AL., *Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs*, JAMA, vol. 316, 22 (2016), 2402.
- [10] M. HOIKKA, T. SILFVAST, and T. I. ALA-KOKKO, *Does the prehospital National Early Warning Score predict the short-term mortality of unselected emergency patients?*, Scand. J. Trauma Resusc. Emerg. Med., vol. 26, 1 (2018).
- [11] S. W. JARVIS ET AL., *Development and validation of a decision tree early warning score based on routine laboratory test results for the discrimination of hospital mortality in emergency medical admissions*, Resuscitation, vol. 84, 11 (2013), 1494–1499.

- [12] Y. J. KIM, Y.-G. LEE, J. W. KIM, J. J. PARK, B. RYU, and J.-W. HA, *High-risk Prediction from electronic medical records via deep attention networks*, arXiv [cs.LG], 2017.
- [13] M. J. KUPFERMINC, *Thrombophilia and pregnancy*, *Curr Pharm Des.*, vol. 11 (2005), 735-748.
- [14] H. LEE, S. Y. SHIN, M. SEO, G.-B. NAM, and S. JOO, *Prediction of ventricular tachycardia one hour before occurrence using artificial neural networks*, *Sci. Rep.*, vol. 6, 1 (2016), 32390.
- [15] Z. C. LIPTON, D. C. KALE, C. ELKAN, and R. WETZEL, *Learning to diagnose with LSTM recurrent Neural Networks*, arXiv [cs.LG], 2015.
- [16] F. MA, R. CHITTA, J. ZHOU, Q. YOU, T. SUN, and J. GAO, *Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks*, in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017.
- [17] M. MILOSEVIC, D. JANKOVIC, A. MILENKOVIC, and D. STOJANOV, *Early diagnosis and detection of breast cancer*, *Technol. Health Care*, vol. 26, 4 (2018), 729–759.
- [18] P. NGUYEN, T. TRAN, N. WICKRAMASINGHE, and S. VENKATESH, *Deepr: A Convolutional Net for Medical Records*, *IEEE J. Biomed. Health Inform.*, vol. 21, 1 (2017), 22–30.
- [19] M. E. H. ONG ET AL., *Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score*, *Crit. Care*, vol. 16, 3 (2012), R108.
- [20] D. RADOVIĆ, I. FETAHOVIĆ, E. DOLIĆANIN, *The Process of Making, Organizing and Using Thematic Layers in GIS*, *Scientific publications of the State University of Novi Pazar. Ser. A: Appl. Math. Inform. And Mech*, vol. 14, 2 (2022), 89-99.
- [21] J. RUBIN ET AL, *An ensemble boosting model for predicting transfer to the pediatric intensive care unit*, *Int. J. Med. Inform.*, vol. 112 (2018), 15–20.
- [22] F. SHAMOUT, T. ZHU, and D. A. CLIFTON, *Machine learning for clinical outcome prediction*, *IEEE Rev. Biomed. Eng.*, vol. 14 (2021), 116–126.
- [23] B. SHICKEL, P. J. TIGHE, A. BIHORAC, and P. RASHIDI, *Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis*, *IEEE J. Biomed. Health Inform.*, vol. 22, 5 (2018), 1589–1604.
- [24] E. ZOLOTAREVA, *Machine Learning Techniques for Stock Market Trends Identification*, *Scientific publications of the State University of Novi Pazar. Ser. A: Appl. Math. Inform. And Mech.* vol. 11, 1 (2019), 47-54.