Machine learning approaches to the determinants of women's vasomotor symptoms using general hospital data

Ki-Jin Ryu

Korea University College of Medicine and School of Medicine

Kyong Wook Yi

Korea University College of Medicine and School of Medicine

Yong Jin Kim

Korea University College of Medicine and School of Medicine

Jung Ho Shin

Korea University College of Medicine and School of Medicine

Jun Young Hur

Korea University College of Medicine and School of Medicine

Tak Kim

Korea University College of Medicine and School of Medicine

Jong Bae Seo

Mokpo National University

Kwang-Sig Lee

Korea University College of Medicine and School of Medicine

Hyuntae Park (✉ cyberpelvis@gmail.com)

Korea University College of Medicine and School of Medicine <https://orcid.org/0000-0002-3364-4860>

Research article

Keywords: vasomotor symptoms, hot flushes, menopause age, thyroid stimulating hormone, monocyte, cancer antigen

Posted Date: September 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-74144/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

To analyze the determinants of women's vasomotor symptoms (VMS) using machine learning.

Methods

Data came from Korea University Anam Hospital in Seoul, Korea, with 3298 women, aged 40–80 years, who attended their general health check from January 2010 to December 2012. Five machine learning methods were applied and compared for the prediction of VMS, measured by a Menopause Rating Scale. Variable importance, the effect of a variable on model performance, was used for identifying major determinants of VMS.

Results

In terms of the mean squared error, the random forest (0.9326) was much better than linear regression (12.4856) and artificial neural networks with one, two and three hidden layers (1.5576, 1.5184 and 1.5833, respectively). Based on variable importance from the random forest, the most important determinants of VMS were age, menopause age, thyroid stimulating hormone, monocyte and triglyceride, as well as gamma glutamyl transferase, blood urea nitrogen, cancer antigen 19 – 9, C-reactive protein and low-density-lipoprotein cholesterol. Indeed, the following determinants ranked within the top 20 in terms of variable importance: cancer antigen 125, total cholesterol, insulin, free thyroxine, forced vital capacity, alanine aminotransferase, forced expired volume in one second, height, homeostatic model assessment for insulin resistance and carcinoembryonic antigen.

Conclusions

Machine learning provides an invaluable decision support system for the prediction of VMS. For preventing VMS, preventive measures would be needed regarding the thyroid function, the lipid profile, the liver function, inflammation markers, insulin resistance, the monocyte, cancer antigens and the lung function.

Background

Vasomotor symptoms, referring to hot flashes and sweating, are major symptoms of peri-menopausal and post-menopausal women and a main cause of their hospital visit [1]. Once it was considered to be a temporary symptom but now it is reported to exert a lasting effect on their quality of life [2]. A decline in hormone concentration is considered to be one of its risk factors, disrupting brain neurotransmission and hypothalamic thermoregulation [3]. Other candidates for its risk factors include age, body mass index,

race, smoking and depressive symptoms [4–6]. However, these results were not consistent and more research is to be done in this direction. Based on increasing evidence, moreover, vasomotor symptoms are expected to associate with major chronic diseases such as cognitive impairment, cardiovascular disease, diabetes mellitus and sleep disorder [7–9]. More research is needed to analyze women's vasomotor symptoms and their major determinants.

This study is the first attempt to demonstrate that machine learning provides an invaluable decision support system to predict women's vasomotor symptoms and analyze their determinants. Existing literature did not address which variables are more important for the prediction of vasomotor symptoms. This was because previous studies were based on an unrealistic assumption of *ceteris paribus*, "all the other variables staying constant". Machine learning methods such as the random forest are free from such an unrealistic assumption of "all the other variables staying constant" [10, 11]. As addressed in the section of Results below, the performance of the random forest is much better and much more stable than that of logistic regression, a popular statistical approach in conventional studies. Moreover, data in this study are larger than those in previous studies – 3298 women and 104 independent variables on the thyroid function, the lipid profile, the liver function, inflammation markers, insulin resistance, the monocyte, cancer antigens, the lung function and other information.

Materials And Methods

Participants and Variables

Data came from Korea University Anam Hospital in Seoul, Korea, with 3298 women, aged 40–80 years, who attended their health check from January 2010 to December 2012. The dependent variable was vasomotor symptoms measured by a Menopause Rating Scale item on hot flashes and sweating (MRS-1). In the early 1990s the MRS was introduced as a standardized scale to measure women's aging symptoms and health-related quality of life. The MRS consists of 11 items such as vasomotor symptoms (MRS-1), heart discomfort (MRS-2), sleep problems (MRS-3), and each item has a scoring system from 0 (no symptoms) to 4 (very severe symptoms) [12] And 104 independent variables were included in this study such as age (years), menopause age (years), thyroid stimulating hormone ($\mu\text{U/mL}$), monocyte (%) and triglyceride (mg/dL) (See Tables 1 and 2 in Results for detail).

Table 1
Descriptive statistics on continuous variables

Continuous variables	Values (mean $\pm$ SD)
MRS 1 (vasomotor symptoms)	0.81 $\pm$ 0.99
Age (years)	51.54 $\pm$ 7.81
Age at menopause (years)	49.24 $\pm$ 4.70
Years since menopause (years)	6.85 $\pm$ 6.78
Age at menarche (years)	14.88 $\pm$ 2.20
Age at marriage (years)	25.41 $\pm$ 3.90
Abortion, spontaneous (frequency)	0.34 $\pm$ 0.76
Abortion, artificial (frequency)	1.10 $\pm$ 1.29
Menopausal hormone replacement therapy (years)	0.49 $\pm$ 1.73
Height (cm)	157.37 $\pm$ 5.24
Body weight (kg)	57.15 $\pm$ 7.79
Body mass index (kg/m ²)	23.04 $\pm$ 3.17
Pulse (beats per minute)	65.11 $\pm$ 9.43
Blood pressure, systolic (mmHg)	106.88 $\pm$ 15.26
Blood pressure, diastolic (mmHg)	64.88 $\pm$ 10.81
Waist/hip ratio	0.88 $\pm$ 0.05
Hip circumference (cm)	94.11 $\pm$ 6.02
Neck circumference (cm)	31.59 $\pm$ 2.89
Waist circumference (cm)	82.46 $\pm$ 13.76
Thigh circumference (cm)	48.05 $\pm$ 11.41
Percent body fat (%)	31.20 $\pm$ 9.87
Skeletal muscle mass (kg)	20.50 $\pm$ 4.96
Body fat mass (kg)	18.06 $\pm$ 6.12
Lean body mass (kg)	39.45 $\pm$ 10.37
Basal metabolic rate (kcal)	1215.48 $\pm$ 89.60
Abbreviations: MRS, menopause rating scale; BMD, bone mineral density; HOMA-IR, Homeostatic model assessment for insulin resistance; FVC, Forced vital capacity; FEV1, Forced expiratory volume in one second; PEF, peak expiratory flow rate	

Continuous variables	Values (mean $\pm$ SD)
InBody score	72.92 $\pm$ 3.81
Edema index (extracellular water/total body water)	0.35 $\pm$ 0.02
Visceral fat area (cm ²)	79.11 $\pm$ 20.31
Physical exercise grade	1.97 $\pm$ 1.04
L1 spine BMD (g/cm ²)	0.84 $\pm$ 0.14
L1 spine T-score	-1.26 $\pm$ 2.36
L2 spine BMD (g/cm ²)	0.93 $\pm$ 1.71
L2 spine T-score	-0.71 $\pm$ 1.24
L3 spine BMD (g/cm ²)	0.95 $\pm$ 0.33
L3 spine T-score	-0.74 $\pm$ 1.29
L4 spine BMD (g/cm ²)	0.95 $\pm$ 0.17
L4 spine T-score	-0.97 $\pm$ 1.41
Lumbar spine total BMD (g/cm ²)	1.76 $\pm$ 27.98
Lumbar spine total T-score	-0.81 $\pm$ 1.21
Femur neck BMD (g/cm ²)	1.29 $\pm$ 19.47
Femur neck T-score	-0.97 $\pm$ 1.02
Trochanter BMD (g/cm ²)	0.87 $\pm$ 11.35
Trochanter T-score	-0.11 $\pm$ 0.87
Intertrochanter BMD (g/cm ²)	1.03 $\pm$ 0.14
Intertrochanter T-score	0.20 $\pm$ 0.97
Hip total BMD (g/cm ²)	0.85 $\pm$ 0.11
Hip total T-score	-0.02 $\pm$ 0.96
Ward's BMD (g/cm ²)	0.55 $\pm$ 0.14
Ward's T-score	-1.13 $\pm$ 1.25

Abbreviations: MRS, menopause rating scale; BMD, bone mineral density; HOMA-IR, Homeostatic model assessment for insulin resistance; FVC, Forced vital capacity; FEV1, Forced expiratory volume in one second; PEF, peak expiratory flow rate

Continuous variables	Values (mean $\pm$ SD)
HOMA-IR index	1.83 $\pm$ 1.07
FVC (liters)	2.94 $\pm$ 0.45
FVC (%)	97.04 $\pm$ 11.45
FEV1 (liters)	2.45 $\pm$ 0.39
FEV1 (%)	107.02 $\pm$ 13.56
PEFR (L/sec)	5.54 $\pm$ 1.13
PEFR (%)	97.23 $\pm$ 18.21
FEV1/FVC (%)	83.16 $\pm$ 5.46
Hemoglobin (g/dL)	13.06 $\pm$ 1.09
Hematocrit (%)	38.61 $\pm$ 2.97
Erythrocyte sedimentation rate (mm/hour)	10.34 $\pm$ 8.05
White blood cell count ($\times 10^3$ /mL)	5.46 $\pm$ 1.98
Neutrophils (%)	53.88 $\pm$ 9.12
Lymphocytes (%)	36.67 $\pm$ 8.44
Monocytes (%)	6.47 $\pm$ 1.71
Eosinophils (%)	2.40 $\pm$ 2.08
Basophils (%)	0.56 $\pm$ 0.37
Total protein (g/dL)	7.10 $\pm$ 0.40
Albumin (g/dL)	4.24 $\pm$ 0.20
C-reactive protein (mg/L)	1.20 $\pm$ 3.73
Total cholesterol (mg/dL)	193.40 $\pm$ 33.73
High-density lipoprotein cholesterol (mg/dL)	57.08 $\pm$ 13.59
Low-density lipoprotein cholesterol (mg/dL)	95.89 $\pm$ 20.28
Triglyceride (mg/dL)	104.19 $\pm$ 60.16
Aspartate aminotransferase (IU/L)	21.93 $\pm$ 10.14
Alanine aminotransferase (IU/L)	19.36 $\pm$ 15.0
Abbreviations: MRS, menopause rating scale; BMD, bone mineral density; HOMA-IR, Homeostatic model assessment for insulin resistance; FVC, Forced vital capacity; FEV1, Forced expiratory volume in one second; PEFR, peak expiratory flow rate	

Continuous variables	Values (mean $\pm$ SD)
Alkaline phosphatase (IU/L)	50.53 $\pm$ 16.52
Amylase (U/L)	50.21 $\pm$ 16.45
Gamma glutamyl transferase (IU/L)	23.25 $\pm$ 25.52
Fasting glucose level (mg/dL)	93.46 $\pm$ 16.19
Blood urea nitrogen (mg/dL)	12.44 $\pm$ 3.63
Creatinine (mg/dL)	0.75 $\pm$ 0.12
Hemoglobin A1c (%)	5.71 $\pm$ 0.62
Rheumatoid factor (IU/mL)	7.16 $\pm$ 19.04
Free thyroxine (ng/dL)	1.26 $\pm$ 0.19
Thyroid stimulating hormone (mU/L)	2.97 $\pm$ 3.15
Serum fasting insulin level (uIU/mL)	7.72 $\pm$ 3.72
Alpha fetoprotein (ng/mL)	2.53 $\pm$ 1.50
Carcinoembryonic antigen (ng/mL)	1.13 $\pm$ 0.93
Cancer antigen 19 - 9 (U/mL)	14.73 $\pm$ 22.38
Cancer antigen 125 (U/mL)	7.50 $\pm$ 7.66
Abbreviations: MRS, menopause rating scale; BMD, bone mineral density; HOMA-IR, Homeostatic model assessment for insulin resistance; FVC, Forced vital capacity; FEV1, Forced expiratory volume in one second; PEFr, peak expiratory flow rate	

Table 2
Descriptive statistics on categorical variables

Categorical variables	Values (Percentage)
Hypertension	15.6%
Diabetes mellitus	4.5%
Thyroid disease	8.8%
Dyslipidemia	1.1%
Metabolic syndrome (AHA/NHLBI)	17.1%
Metabolic syndrome (IDF)	16.3%
Parity	
0	3.3%
1	10.5%
2	70.9%
>3	15.3%
Alcohol intake	41.3%
Smoking	
Never	93.9%
Yes	3.4%
Quit	2.7%
Atrophic change on Papanicolaou test	21.9%
Non-alcoholic fatty liver disease	34%
Positive for hepatitis B virus surface antigen	2.8%
Positive for anti-hepatitis B virus surface antibody	83.7%
Positive for anti-hepatitis B core antibody	43.6%
Abbreviations: AHA/NHLBI, American Heart Association/National Heart, Lung and Blood Institute; IDF, International Diabetes Federation.	

Analysis

Five machine learning approaches were used for predicting vasomotor symptoms, the dependent variable of this study: linear regression, random forest and artificial neural networks (ANNs) with one, two and three hidden layers [13]. Based on linear regression, a linear line is found in a way that it minimizes, among an infinite number of linear lines, the sum of the squares of errors (Errors are gaps between actual

and predicted values of the dependent variable). A decision tree consists of (1) internal nodes (each meaning a test on an independent variable), (2) branches (each denoting an outcome of the test) and (3) terminal nodes (each representing a value of the dependent variable). A random forest creates many training sets, trains many decision trees and makes a prediction with a majority vote ("bootstrap aggregation"). An ANN includes one input layer, one, two or three hidden layers, and one output layer. Neurons in a previous layer combine with "weights" in the next layer (Here, the weights are numerical values showing how much effects neurons in a previous layer have on neurons in the next layer). This operation is done in the order of weights in a layer next to the input layer, its following layer, and so on. This process is called the feedforward algorithm. Then, these weights are adjusted based on how much contribution they made to the errors of the ANN (Here, errors are gaps between actual and predicted values of the dependent variable). This operation is done in the order of weights in the output layer, its previous layer, and so on. This process is called the backpropagation algorithm. These algorithms are repeated until a certain standard is achieved for the accurate prediction of the dependent variable [13].

Data on 3298 participants were divided into training and validation sets with a 75:25 ratio. The models were built (or trained) based on the training set with 2474 observations then the models trained were validated based on the validation set with 824 observations. The mean squared error (MSE), the average of the squares of errors among 824 observations, was introduced as a criterion for validating the models trained. Here, errors are gaps between actual and predicted values of the dependent variable, vasomotor symptoms. Variable importance from the random forest, a mean-impurity gap between a complete model and a model excluding a certain variable, was adopted for identifying major determinants of vasomotor symptoms (mean impurity, or the degree of data being mixed at a node on average, is proportional to the MSE). The greater "mean-impurity increase" is defined as the greater variable importance. R-Studio was employed for the analysis on April 2020.

Ethics statement

This retrospective study complied with the tenets of the Helsinki Declaration and was approved by the Institutional Review Board (IRB) of Korea University Anam Hospital on 20 Jan 2020 (2020AN0031). Informed consent was waived by the IRB.

Results

Descriptive statistics for continuous and categorical variables in this study are summarized in Tables 1 and 2, respectively. The mean score of MRS 1 (vasomotor symptoms), age and menopause age of 3298 participants were 0.8093, 52 (years) and 49 (years), respectively. Their mean thyroid stimulating hormone, monocyte and triglyceride were 2.98 ($\mu\text{U/mL}$), 6.47 (%) and 104.20 (mg/dL), respectively. The MSEs of the five machine learning models are shown in Table 3. A single 75:25 split of the training and validation sets would reduce the validity and generalizability of the results. For this reason, the random split and the statistical analysis were repeated 3 times and their average MSE was calculated for each of the five statistical methods, i.e., linear regression, random forest and ANNs with one, two and three hidden layers. The random forest was the best model for predicting vasomotor symptoms. Its average MSE (0.9326)

was much smaller than those of linear regression and the ANNs (12.4856, 1.5576, 1.5184 and 1.5833, respectively). Linear regression was the worst model and it registered a great variation in terms of MSE, from 1.0119 (Run 2) to 29.5104 (Run 3).

Table 3
Model performance: mean squared error

Model	Run 1	Run 2	Run 3	Average
Linear Regression	6.9343	1.0119	29.5104	12.4856
Random Forest	0.9180	0.9351	0.9448	0.9326
Artificial Neural Network 1 Layer	1.5894	1.5616	1.5218	1.5576
Artificial Neural Network 2 Layers	1.4014	1.4321	1.7217	1.5184
Artificial Neural Network 3 Layers	1.5652	1.3787	1.8060	1.5833

Based on variable importance from the random forest, the most important determinants of vasomotor symptoms were age, menopause age, thyroid stimulating hormone, monocyte and triglyceride, as well as gamma glutamyl transferase, blood urea nitrogen, cancer antigen 19 – 9, C-reactive protein and low-density-lipoprotein cholesterol (Run 1 in Table 4, Fig. 1). Indeed, the following determinants ranked within the top 20 in terms of variable importance: cancer antigen 125, total cholesterol, insulin, free thyroxine, forced vital capacity, alanine aminotransferase, forced expired volume in one second, height, homeostatic model assessment for insulin resistance and carcinoembryonic antigen. The findings of linear regression present useful information about the effect of a major determinant on vasomotor symptoms. For example, vasomotor symptoms will decrease by 0.03 if menopause age increases by 1 year. Likewise, vasomotor symptoms will increase by 0.01 if thyroid stimulating hormone increases by 1 unit ($\mu\text{U/mL}$). It is to be noted, however, that the results of linear regression are based on an unrealistic assumption of *ceteris paribus*, “all the other variables staying constant”. In this context, the findings of linear regression are to be considered as just supplementary information to the variable importance from the random forest.

Table 4
Results of the random forest and linear regression

Variable	Random Forest		Linear Regression	
	VI-Value	Rank	Coef	P-Value
Age	89.33	1	0.004263	0.68
Age at menopause	48.20	2	-0.032260	0.00
Thyroid stimulating hormone	47.03	3	0.013600	0.03
Monocyte	46.11	4	-0.015550	0.83
Triglyceride	45.03	5	0.000068	0.89
Gamma glutamyl transferase	44.18	6	0.000925	0.34
Blood urea nitrogen	44.13	7	0.002123	0.72
Cancer antigen 19 – 9	42.54	8	0.001309	0.52
C-reactive protein	39.60	9	0.005826	0.36
Low-density lipoprotein cholesterol	37.02	10	-0.005414	0.24
Cancer antigen 125	36.86	11	-0.000274	0.92
Total cholesterol	36.40	12	0.005096	0.09
Serum fasting insulin level	34.62	13	0.010670	0.69
Free thyroxine	34.61	14	0.139400	0.21
Forced vital capacity	33.99	15	0.303000	0.68
Alanine aminotransferase	33.66	16	0.002964	0.21
Forced expiratory volume in one second	33.24	17	-0.228300	0.78
Height	33.14	18	0.025580	0.45
Homeostatic model assessment for insulin resistance	32.83	19	-0.006738	0.95
Carcinoembryonic antigen	32.83	20	0.036780	0.10
Hematocrit	31.48	21	-0.024060	0.26
Lymphocyte	31.40	22	-0.038240	0.58
Thigh circumference	31.28	23	-0.001295	0.41
White blood cell count	31.16	24	0.004046	0.81
Aspartate aminotransferase	30.82	25	-0.001866	0.61
Body mass index	30.81	26	0.008785	0.93

Variable	Random Forest		Linear Regression	
	VI-Value	Rank	Coef	P-Value
Amylase	30.77	27	-0.000919	0.47
Alpha fetoprotein	29.68	28	0.007071	0.61
Erythrocyte sedimentation rate	29.65	29	0.002115	0.52
Peak expiratory flow rate	29.64	30	-0.576400	0.03
Heart rate	29.46	31	-0.003075	0.21
Neutrophil	29.36	32	-0.037360	0.59
Fasting glucose level	29.05	33	0.001390	0.59
Edema index (extracellular water/total body water)	28.97	34	-0.681500	0.45
High-density lipoprotein cholesterol	28.60	35	-0.005087	0.09
Eosinophil	28.52	36	-0.053860	0.46
Alkaline phosphatase	28.37	37	0.000088	0.95
Hemoglobin	27.78	38	0.116100	0.05
Creatinine	27.45	39	-0.416500	0.03
Visceral fat area	26.11	40	-0.001338	0.67

Discussion

The results of this study are consistent with previous findings on the associations of vasomotor symptoms with age [4], the lipid profile [14], the liver function [9], inflammation markers [15] and insulin resistance [16, 17]. This study provides the following additional information for existing literature as well. Firstly, menopause age was the second most important determinant of vasomotor symptoms and their linkage was found to be negative in this study. Little study has been done on this topic and only a few independent suggestions have been made on a negative association between menopause age and cardiovascular disease [18] and a positive linkage between cardiovascular disease and vasomotor symptoms [19, 20]. In a similar context, one would expect a negative relationship between menopause age and vasomotor symptoms, and this study supports this expectation given the high ranking of menopause age from the random forest in this study (2nd). More effort should be made for identifying their underlying mechanism. Secondly, thyroid stimulating hormone was a top 3 determinant of vasomotor symptoms and their association was reported to be positive for the participants aged 40–80 years in this study. According to the previous results, however, their linkage was (1) neutral for pre- or peri-menopausal women aged 42–52 years [21] and (2) negative for euthyroid menopausal women aged 49–59 years [22]. Age and thyroid status would be important mediating variables here and more

examination is to be done on a relationship among thyroid stimulating hormone, vasomotor symptoms and their mediating variables.

Thirdly, the ranking of the monocyte was fourth among the most important determinants of vasomotor symptoms in this study. It was reported that the ratio of monocyte to high-density-lipoprotein cholesterol has positive linkages with the risks of coronary artery disease [23], metabolic syndrome [24] and polycystic ovary syndrome [25]. One possible explanation for these results is that the higher monocyte ratio is associated with systematic inflammation [25–27], which leads to the higher risks of the diseases above. This explanation can be extended to an association between the monocyte and vasomotor symptoms. More effort is to be made in this direction and this study would be a good starting point. Fourthly, it was unexpected that cancer antigen 19 – 9, cancer antigen 125 and carcinoembryonic antigen were among top 20 determinants of vasomotor symptoms in this study. There were a couple of studies on the effect of hormone replacement therapy on these cancer antigens among post-menopausal women [28, 29]. But no examination has been done and more investigation is needed on a direct relationship between vasomotor symptoms and these cancer antigens. In a similar vein, forced vital capacity and forced expiratory volume in one second ranked fifteenth and seventeenth among the most important determinants of vasomotor symptoms in this study, respectively. There were a few studies on the impact of hormone replacement therapy on these lung-function indicators among post-menopausal women [30, 31]. However, no literature is available and more research is needed on a direct relationship between vasomotor symptoms and these lung-function indicators.

This study is the first machine-learning study to predict women's vasomotor symptoms and analyze their determinants. Based on the results of this study, the random forest could address which variables are more important for the prediction of vasomotor symptoms, while its performance was much better and much more stable than that of logistic regression, a popular statistical approach in conventional studies. In addition, data in this study are larger than those in previous studies – 3298 women and 104 independent variables. The findings of this study demonstrate that machine learning provides an invaluable decision support system for the prediction of vasomotor symptoms. However, this study had some limitations. Firstly, this study adopted a cross-sectional design. Expanding data with a longitudinal design is expected to improve the accuracy of machine learning significantly. Secondly, this study did not consider possible mediating effects among variables. Thirdly, data came from a single center in this study. Expanding data with a multi-center design will be a good topic for future research.

Conclusions

In conclusion, for preventing women's vasomotor symptoms, preventive measures would be needed regarding the thyroid function, the lipid profile, the liver function, inflammation markers, insulin resistance, the monocyte, cancer antigens and the lung function. Machine learning provides an invaluable decision support system to predict women's vasomotor symptoms and analyze their determinants.

Abbreviations

ANN: Artificial Neural Network; IRB:Institutional Review Board; MRS:Menopause Rating Scale; MSE:Mean Squared Error

Declarations

Acknowledgements: Nil

Authors' Contributions

Ki-Jin Ryu: I declare that I participated in the acquisition and interpretation of data and drafting of the manuscript, and that I have seen and approved the final version. I have no conflicts of interest.

Kyong Wook Yi: I declare that I participated in the acquisition of data and that I have seen and approved the final version. I have no conflicts of interest.

Yong Jin Kim: I declare that I participated in the acquisition of data and that I have seen and approved the final version. I have no conflicts of interest.

Jung Ho Shin: I declare that I participated in the acquisition of data and that I have seen and approved the final version. I have no conflicts of interest.

Jun Young Hur: I declare that I participated in the acquisition of data and that I have seen and approved the final version. I have no conflicts of interest.

Tak Kim: I declare that I participated in the acquisition of data and that I have seen and approved the final version. I have no conflicts of interest.

Jong Bae Seo: I declare that I participated in drafting of the manuscript and that I have seen and approved the final version. I have no conflicts of interest.

Kwang-Sig Lee: I declare that I participated in the design of the study, acquisition and interpretation of data, and drafting of the manuscript, and that I have seen and approved the final version. I have no conflicts of interest.

Hyuntae Park: I declare that I participated in the design of the study, acquisition and interpretation of data, and drafting of the manuscript, and that I have seen and approved the final version. I have no conflicts of interest.

Funding

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-0-01405) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), and the National

Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1C1C1013579).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Data will be made available on request

Consent for publication

Not applicable

Competing interests

none.

References

1. Boulet MJ, Oddens BJ, Leher P, Vemer HM, Visser A. Climacteric and menopause in seven south-east Asian countries. *Maturitas*. 2008;61:34–53.
2. Avis NE, Crawford SL, Greendale G, Bromberger JT, Everson-Rose SA, Gold EB, et al. Duration of menopausal vasomotor symptoms over the menopause transition. *JAMA Intern Med*. 2015;175:531–9.
3. Stearns V, Ullmer L, López JF, Smith Y, Isaacs C, Hayes D. Hot flushes. *Lancet*. 2002;360:1851–61.
4. Gallicchio L, Miller SR, Kiefer J, Greene T, Zacur HA, Flaws JA. Risk factors for hot flashes among women undergoing the menopausal transition: baseline results from the Midlife Women's Health Study. *Menopause*. 2015;22:1098–107.
5. Whiteman MK, Staropoli CA, Benedict JC, Borgeest C, Flaws JA. Risk factors for hot flashes in midlife women. *J Womens Health (Larchmt)*. 2003;12:459–72.
6. Gold EB, Crawford SL, Shelton JF, Tepper PG, Crandall CJ, Greendale GA, et al. Longitudinal analysis of changes in weight and waist circumference in relation to incident vasomotor symptoms: the Study of Women's Health Across the Nation (SWAN). *Menopause*. 2017;24:9–26.
7. Sturdee DW, Hunter MS, Maki PM, Gupta P, Sassarini J, Stevenson JC, et al. The menopausal hot flush: a review. *Climacteric*. 2017;20:296–305.
8. Biglia N, Cagnacci A, Gambacciani M, Lello S, Maffei S, Nappi RE. Vasomotor symptoms in menopause: a biomarker of cardiovascular disease risk and other chronic diseases? *Climacteric*. 2017;20:306–12.

9. Ryu KJ, Park H, Kim YJ, Yi KW, Shin JH, Hur JY, et al. Moderate to severe vasomotor symptoms are risk factors for non-alcoholic fatty liver disease in postmenopausal women. *Maturitas*. 2018;117:22–8.
10. Lee KS, Ahn KH. Artificial Neural Network Analysis of Spontaneous Preterm Labor and Birth and Its Major Determinants. *J Korean Med Sci*. 2019;34:e128.
11. Lee KS, Park KW. Social determinants of the association among cerebrovascular disease, hearing loss and cognitive impairment in a middle-aged or older population: Recurrent neural network analysis of the Korean Longitudinal Study of Aging (2014–2016). *Geriatr Gerontol Int*. 2019;19:711–6.
12. Heinemann K, Ruebig A, Potthoff P, Schneider HP, Strelow F, Heinemann LA, et al. The Menopause Rating Scale (MRS) scale: a methodological review. *Health Qual Life Outcomes*. 2004;2:45.
13. Han J, Kamber M. Data mining: concepts and techniques. 2nd ed. San Francisco: Morgan Kaufmann; 2006.
14. Thurston RC, El Khoudary SR, Sutton-Tyrrell K, Crandall CJ, Gold EB, Sternfeld B, et al. Vasomotor symptoms and lipid profiles in women transitioning through menopause. *Obstet Gynecol*. 2012;119:753–61.
15. Thurston RC, El Khoudary SR, Sutton-Tyrrell K, Crandall CJ, Gold E, Sternfeld B, et al. Are vasomotor symptoms associated with alterations in hemostatic and inflammatory markers? Findings from the Study of Women's Health Across the Nation. *Menopause*. 2011;18:1044–51.
16. Thurston RC, El Khoudary SR, Sutton-Tyrrell K, Crandall CJ, Sternfeld B, Joffe H, et al. Vasomotor symptoms and insulin resistance in the study of women's health across the nation. *J Clin Endocrinol Metab*. 2012;97:3487–94.
17. Kwon DH, Lee JH, Ryu KJ, Park HT, Kim T. Vasomotor symptoms and the homeostatic model assessment of insulin-resistance in Korean postmenopausal women. *Obstet Gynecol Sci*. 2016;59:45–9.
18. Zhu D, Chung HF, Dobson AJ, Pandeya N, Giles GG, Bruinsma F, et al. Age at natural menopause and risk of incident cardiovascular disease: a pooled analysis of individual patient data. *Lancet Public Health*. 2019;4:e553-e64.
19. Thurston RC. Vasomotor symptoms: natural history, physiology, and links with cardiovascular health. *Climacteric*. 2018;21:96–100.
20. Muka T, Oliver-Williams C, Colpani V, Kunutsor S, Chowdhury S, Chowdhury R, et al. Association of Vasomotor and Other Menopausal Symptoms with Risk of Cardiovascular Disease: A Systematic Review and Meta-Analysis. *PLoS One*. 2016;11:e0157417.
21. Sowers M, Luborsky J, Perdue C, Araujo KL, Goldman MB, Harlow SD. Thyroid stimulating hormone (TSH) concentrations and menopausal status in women at the mid-life: SWAN. *Clin Endocrinol (Oxf)*. 2003;58:340–7.
22. Slopian R, Owecki M, Slopian A, Bala G, Meczekalski B. Climacteric symptoms are related to thyroid status in euthyroid menopausal women. *J Endocrinol Invest*. 2020;43:75–80.

23. Cetin MS, Ozcan Cetin EH, Kalender E, Aydin S, Topaloglu S, Kisacik HL, et al. Monocyte to HDL Cholesterol Ratio Predicts Coronary Artery Disease Severity and Future Major Cardiovascular Adverse Events in Acute Coronary Syndrome. *Heart Lung Circ.* 2016;25:1077–86.
24. Vahit D, Akboga MK, Samet Y, Hüseyin E. Assessment of monocyte to high density lipoprotein cholesterol ratio and lymphocyte-to-monocyte ratio in patients with metabolic syndrome. *Biomark Med.* 2017;11:535–40.
25. Usta A, Avci E, Bulbul CB, Kadi H, Adali E. The monocyte counts to HDL cholesterol ratio in obese and lean patients with polycystic ovary syndrome. *Reprod Biol Endocrinol.* 2018;16:34.
26. Linker R, Gold R, Luhder F. Function of neurotrophic factors beyond the nervous system: inflammation and autoimmune demyelination. *Crit Rev Immunol.* 2009;29:43–68.
27. Canpolat U, Çetin EH, Cetin S, Aydin S, Akboga MK, Yayla C, et al. Association of Monocyte-to-HDL Cholesterol Ratio with Slow Coronary Flow is Linked to Systemic Inflammation. *Clin Appl Thromb Hemost.* 2016;22:476–82.
28. Cecchi E, Lapi F, Vannacci A, Banchelli G, Mazzei T, Mugelli A. Increased levels of CA 125 and CA 19.9 serum tumour markers following cyclic combined hormone replacement therapy. *J Clin Pharm Ther.* 2009;34:129–32.
29. Cengiz B, Atabekoglu C, Cetinkaya E, Cengiz SD. Effect of hormone replacement therapy on serum levels of tumor markers in healthy postmenopausal women. *Maturitas.* 2003;46:301–6.
30. Carlson CL, Cushman M, Enright PL, Cauley JA, Newman AB. Hormone replacement therapy is associated with higher FEV1 in elderly women. *Am J Respir Crit Care Med.* 2001;163:423–8.
31. Cevrioglu AS, Fidan F, Unlu M, Yilmazer M, Orman A, Fenkci IV, et al. The effects of hormone therapy on pulmonary function tests in postmenopausal women. *Maturitas.* 2004;49:221–7.

Figures

Figure 1

Top 20 variables from random forest variable importance