Journal of RADIATION INDUSTRY
Vol. 18, No. 3, pp. 173-176, 2024
https://doi.org/10.23042/radin.2024.18.3.173

Original article

머신러닝을 이용한 다낭성 난소 증후군 분류 모델 평가

조소영[1] · 예수영[2,*]

[1]부산대학교병원 산부인과, [2]부산가톨릭대학교 보건과학대학 방사선학과

Evaluation of Polycystic Ovary Syndrome Classification Model Using Machine Learning

So-Young Jo[1] and Soo-Young Ye[2,*]

[1]Department of Obstetrics and Gynecology, Pusan National University Hospital, 179, Gudeok-ro, Seo-gu, Busan 49241, Republic of Korea
[2]Department of Radiological Science, Catholic University of Pusan, 74, Oryundae-ro, Geumjeong-gu, Busan 46252, Republic of Korea

ABSTRACT In this paper, general characteristics, blood tests, and ultrasound examination results were used to classify the presence of polycystic ovary syndrome (PCOS). The classification algorithms used were SVM (Support Vector Machine) and k-NN (k-Nearest Neighbors). Out of a total of 300 data samples, 210 were used as training data and 90 as test data. The results showed that SVM achieved higher accuracy compared to k-NN, confirming its greater utility in diagnosing the presence of PCOS. Future research is expected to improve classification performance by incorporating various additional indicators and securing more data. Additionally, it is expected to serve as a foundational resource for predicting and classifying other diseases.

Key words: Polycystic ovary syndrome (PCOS), Machine learning, Support Vector Machine (SVM), k-Nearest Neighbor (k-NN)

1. 서 론

다낭성난소증후군(PCOS; polycystic ovary syndrome)은 만성 무배란과 고안드로겐혈증을 특징으로 가임여성의 5~10%에서 발생하는 흔한 내분비계 질환이다[1]. 주 증상으로 무배란과 안드로겐 과다 이외에 다모증, 비만, 남성형 탈모, 여드름 등 나타난다. 현재까지 PCOS의 통일된 진단 기준이 없으며 정확한 병인이 밝혀져 있지 않다.

1990년 National Institutes of Heath (NIH) conference는 다른 원인을 배제한 임상적 고안드로겐혈증이 있으며 만성적 무배란이 있는 경우 다낭성 난소 증후군으로 진단할 수 있다고 한다[2]. 2003년 European Society of Human Reproduction and Embryology (ESHRE)에서는 만성적인 무배란, 임상적 고안드로겐혈증, 그리고 다낭성 난소의 초음파 소견 중 2가지가 만족하는 경우를 다낭성 난소 증후군으로 진단할 것을 권유하고 있다[3]. 하지만 PCOS에 대한 정확한 진단 기준이 마련되어 있지 않다. 이는 상태에 대한 보편적인 정의가 없고 증상이 여성에 따라 다르기 때문이다. 또한 PCOS 관련 증상은 반드시 PCOS 환자에게 나타나는 것은 아니고, 다른 내분비 질환이나 비만 및 갑상선 기능 저하증과 관련이 있을 수 있다[4].

최근 의료 분야에서는 컴퓨터를 활용하여 학습할 수 있는 인공지능(AI) 기법들이 많이 사용되고 있다. 특히 질병의 발견, 예측, 분류 등의 임상에서 빠르게 적용되고 있으며, 그 정확도도 매우 우수하다. 인공지능(AI) 기법 중 머신러닝은 인간의 신경계를 모방한 인공신경망 (Artificial Neural Network)를 기반하여 설계된 개념으로, 목적은 데이터의 구조를 이해하고 데이터를 활용할 수 있는 모델에 적합하게 이용하는 것이다[5]. PCOS와 같이 정확한 진단의 기준이 없는 질환에 머신러닝을 이용한다면, 진단의 보조적인 역할로서 임상 데이터를 작성하고 모델을 구축하여 질병 진단에서 정확성을 유지하면서 효율성을 보일 수 있다.

본 연구에서는 다낭성 난소 증후군 환자의 데이터를 활용하여, 각 알고리즘 모델의 정확도를 평가한 후, 알고리즘 간의 성능을

ISSN 1976-2402 http://www.ksri.kr/

Copyright© 2024 by Korean Society of Radiation Industry

*Corresponding author: Soo Young Ye
Tel: +82-51-510-0586, E-mail: syye@cup.ac.kr

Received 22 July 2024 Revised 3 September 2024 Accepted 3 September 2024

비교하고자 한다. 이를 통해 머신러닝을 사용한 다낭성 난소 증후군 진단의 가장 적합한 모델을 제시하고자 한다.

2. 재료 및 방법

2.1. Subjects

본 연구는 GitHub의 오픈 데이터를 활용하여 환자의 기본 정보를 수집하였으며, 150명의 PCOS 환자와 150명의 대조군을 대상으로 연구하였다. PCOS의 진단은 2003 ESHRE Rotterdam consensus를 통해 설정된 'PCOS revised diagnostic criteria'에 따라 만성적인 무배란, 임상적 고안드로겐혈증, 그리고 다낭성 난소의 초음파 소견 중 2가지가 만족하는 경우이며, 대조군은 PCOS 진단을 받지 않은 여성으로 하였다.

2.2. Parameters

신체적 변화 및 대사 작용에 따른 일반적 특성, 혈액 검사, 초음파 검사의 데이터를 활용하여 연구하였다.

2.2.1. 일반적 특성

일반적 특성에서는 신체적 변화 및 대사 작용에 따른 나이(Age), 체지방률(BMI), 생리주기(Menstrual cycle), 생리 기간(menstrual cycle length), 임신(pregnant), 임신 중절 횟수(abortion), 체중 증가(weight gain), 모발 성장(hair growth), 탈모(hair loss), 여드름(pimples), 즉석식품(fast food), 운동(exercise) 여부를 활용하였다.

Table 1. Classification of the incidence factors of PCOS.

	Generating factor	Principle
Generating	Age BMI Pregnant Abortion Weight gain Hair growth Hair loss Pimples Fastfood Exercise Menstrual cycle length Menstrual cycle	Physical changes and metabolism
A blood test	HCGI,II FSH LH RBS PRG TSH VtiD3 AMH	Hormone level derivation
Sonography	Average of Follicle size Number of follicles Endometrium Thickness	Uterus, measuring follicular changes

2.2.2. 혈액 검사

혈액 검사 데이터는 초기 난포기, 즉 생리 시작 10일 이내에 적어도 8시간 이상 금식 후 채혈하였다. 임신 여부 확인 호르몬 I, II (HCG I, II), 난포자극/황체형성 호르몬(FSH/LH), 혈당 농도(RBS), 여성 호르몬(PRG), 갑상선 자극 호르몬(TSH), 황체 형성 호르몬(LH), 난포 자극 호르몬(FSH), 젖 분비 호르몬(PRL), 비타민 D3 (Vit D3), 난소기능과 생식기능 판단검사(AMH)를 활용하였다. 호르몬 수치 데이터를 바탕으로 각 호르몬의 정상 수치 범위를 비교하여 이상 여부를 확인하였다. 모든 데이터를 통계적으로 분석하여 PCOS 환자 그룹과 대조군 간의 차이를 평가하였다.

2.2.3. 초음파 검사

초음파 검사 데이터는 난포의 총수를 난포 개수로 하였으며, 난소의 부피는 길이(length)×넓이(width)×두께(thickness)×0.523으로 계산하였다. 좌우 양측 난소의 부피와 난포의 개수를 측정하였고 각각 평균값을 계산하였다. 난포의 양쪽 개수(Number of follicles), 난포 평균 크기(Average Follicle size), 자궁 내막 두께(Endometrium Thickness)를 활용하였다. 자궁 내막 두께는 자궁 내막의 가장 두꺼운 부분에 대한 데이터이다. 난소낭종 등의 병변이 있는 경우는 대상자에서 제외하였다.

2.3. Machine Learning Models

본 연구에서는 머신러닝 프로그램으로 MATLAB R2023b 64비트를 사용하여 SVM (Support Vector Machine)과 k-NN (K-Nearest Neighbors) 모델을 구현하였다.

머신러닝은 인공지능의 한 분야로서 경험을 통해 주어진 데이터를 기반으로 패턴을 학습하고 결과를 예측하는 알고리즘 기법을 통칭한다[6,7]. 머신러닝은 학습 방식에 따라 3가지로 나누어지며 이는 지도학습(Supervised learning), 비지도학습(Unsupervised learning), 강화학습(Reinforcement learning)이다. 지도학습은 입력 데이터와 출력 데이터를 모두 제공하여 기계를 학습시키는 방법을 말하며, 비 지도학습은 데이터 자체에서 어떠한 패턴을 찾아내는 학습 방법을 의미한다. 강화학습은 여러 시행착오를 거쳐서 얻은 데이터를 기반으로 모델을 지속해서 개선하는 방식을 말한다[8]. 본 연구에서는 기계학습 중에서 지도학습 분류 모형 중 가장 기본적인 SVM과 K-NN을 이용하였다.

2.3.1. SVM (Support Vector Machine)

SVM은 데이터를 분류하기 위해 고차원 공간에서 최적의 초평면을 생성하는 머신러닝 알고리즘이다. 각 클래스 간의 최대 경계를 보장하는 초평면을 찾아내며, 새로운 객체가 어느 클래스에 속하는지 결정한다. SVM은 고차원 데이터에서도 효율적으로 동작

하며, 정규화 매개변수를 통해 과적합을 방지하고 분류 및 회귀 문제에 모두 적용할 수 있다.

2.3.2. k-NN (k-Nearest Neighbors)

k-NN은 새로운 객체를 분류할 때 가장 가까운 k개의 이웃 데이터 포인트를 참조하는 비지도 학습 알고리즘이다. 이 알고리즘은 투표 시스템을 사용하여 미분류된 객체가 어느 클래스에 속하는지를 결정한다. k-NN은 구현하기 쉽고 간단하여 빠른 학습이 가능하며 수치 기반 데이터 분류 작업에서 성능이 우수하다.

2.3.3. 데이터 분할

머신러닝 모델의 훈련 및 평가를 위해 데이터세트는 훈련용(training set)과 테스트용(testing set)으로 나누었다. 데이터 세트는 일반적으로 훈련용 70%, 테스트용 30% 비율로 나누어지며, 본 연구에서도 이 비율을 적용하였다.

2.3.4. 성능 평가 방법

머신러닝 모델의 성능을 측정하고 평가하기 위해 혼동 행렬(Confusion matrix)을 사용하였다. 혼동 행렬은 예측된 값과 실제 값의 발생 빈도를 나타낸 것을 말한다[9]. 혼동 행렬을 바탕으로 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F-1 score를 이용하여 예측 성능을 평가한다. 혼동 행렬은 참으로 예측하고 참인 경우 TP (True positive), 거짓으로 예측하고 거짓인 경우 TN (True negative), 참으로 예측하였으나 거짓인 경우 FP (False positive), 거짓으로 예측하였으나 참인 경우 FN (False negative)이라 한다[10,11].

2.3.4.1. 정확도(Accuracy) 평가

정확도란 예측값과 실제 값의 비교 시 정확한 정도를 나타내며, 가장 직관적인 모델의 성능 평가지표이다.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

2.3.4.2. 정밀도(Precision) 평가

정밀도란 모델이 Positive라고 예측한 값 중에서 실제 Positive의 비율을 나타낸 것이다.

$$Percision = \frac{TP}{TP + FP}$$

2.3.4.3. 재현율(Recall) 평가

재현율이란 예측값과 실제 값과의 비교 시 True인 것 중에서 모델이 Positive라고 예측한 것의 비율을 나타낸 것이다.

$$Recall\ (Sensitivility) = \frac{TP}{TP + FN}$$

2.3.4.4. F-1 score 평가

F-1 score란 정밀도와 재현율의 조화평균이다.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

이러한 성능 지표를 사용하여 SVM과 k-NN 모델의 성능을 비교 분석하였다.

3. 결 과

본 논문에서는 PCOS 환자와 대조군의 데이터 세트를 이용하여 SVM과 k-NN 알고리즘을 이용한 분류 및 정확도 평가를 하였다. 교차검증은 5회를 시행하였다.

3.1. 모델 성능 평가

트레이닝을 완료한 SVM, k-NN 알고리즘을 이용하여 90개의 데이터로 테스트를 시행하였다.

SVM 알고리즘의 경우, 훈련용 데이터 98.1%, 검증 데이터 94.4%의 정확도를 나타내었다. k-NN 알고리즘의 경우, 훈련용 데이터 98.1%, 검증 데이터 91.1%의 정확도를 나타내었다.

3.2. 참양성률 결과

SVM 알고리즘의 경우, 참양성률(TRP)은 PCOS 95.6%, 대조군 93.3%를 나타내었다. 거짓음성률(FNR)의 경우 PCOS 4.4%, 대조군 6.7%를 나타내었다. k-NN 알고리즘의 경우, 참양성률(TRP)은 PCOS 93.3%, 대조군 88.9%를 나타내었다. 거짓음성률(FNR)의 경우 PCOS 6.7%, 대조군 11.1%를 나타내었다.

Table 2. The result of training and test set.

	Training	Test
SVM	98.1%	94.4%
k-NN	98.1%	91.1%

Table 3. The result of true positive rate and false negative rate.

	TPR (PCOS)	TPR (comparison)	FNR (PCOS)	FNR (comparison)
SVM	95.6%	93.3%	4.4%	6.7%
k-NN	93.3%	88.9%	6.7%	11.1%

Table 4. The result of classification models.

	Accuracy	Precision	Recall	F-1 score
SVM	94.4%	93.4%	95.5%	94.4%
k-NN	94.4%	89.3%	93.3%	91.2%

3.3. 분류모델 결과

혼동 행렬을 바탕으로 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F-1 score를 계산하였다.

SVM 알고리즘의 경우, 정확도 94.4%, 정밀도 93.4%, 재현율 95.5%, F-1 score 94.4%를 나타내었다. k-NN 알고리즘의 경우, 정확도 94.4%, 정밀도 89.3%, 재현율 93.3%, F-1 score 91.2%를 나타내었다.

4. 고 찰

다낭성 난소 증후군은 고안드로겐 혈증과 희발 월경을 특징으로 하는 질환으로, 내분비계 및 대사계에 다양한 임상 양상을 나타낸다. 이러한 다낭성 난소 증후군은 현재까지 근본적인 원인이 없으며, 통일된 진단 기준이 마련되어 있지 않다.

본 연구는 PCOS 환자 150명 대조군 150명을 대상으로 시행하였다. MATLAB의 분류 학습기를 이용하여 300명의 검사자 데이터값 중 210명의 데이터를 가지고 각 알고리즘을 트레이닝한 후, 나머지 90명의 데이터를 활용하여 본 테스트를 시행하였다. 그 결과 트레이닝을 진행한 SVM 값은 98.1%가 나왔고, k-NN 값은 98.1%의 정확도를 나타냈으며, 테스트를 진행한 SVM 94.4%, k-NN은 91.1%의 정확도를 나타내었다. 이를 통하여 SVM이 다낭성 난소 증후군 질환 분류에 있어 더 정확함을 알 수 있었다. 본 연구의 제한점은 연구에 활용된 데이터의 수가 300명이며, 다낭성 난소 증후군 발생 인자의 일부분을 선택하여 연구를 진행하였기에 결과의 신뢰값이 다소 부족하였다. 이러한 점을 보완하고, 높은 수준의 정확성을 얻기 위해서 더 많은 인자 및 데이터의 수를 확보해야 한다. 본 논문을 기초 자료로 활용하여 알고리즘을 이용한다면, 다낭성 난소 증후군 질환 유무 진단에 높은 정확성을 가질 수 있다. 나아가, 다른 질환의 예측 및 분류 부분에서 많은 도움이 될 수 있을 것으로 기대된다.

5. 결 론

본 논문에는 다낭성 난소 증후군 질환 유무를 분류하기 위해 일반적 특성, 혈액 검사, 초음파 검사 결과를 사용하였다. 분류에 사용된 알고리즘 모델은 SVM과 k-NN을 이용하였다. 그 결과, SVM이 k-NN에 비하여 높은 정확도를 나타내어 다낭성 난소 증후군 질환 유무의 진단에 유용함을 확인하였다. 향후 연구에서 다양한 지표를 추가하고 더 많은 데이터를 확보한다면 질환 분류 성능을 더 높일 수 있을 것으로 기대할 수 있다. 또한, 타 질환의 예측 및 분류를 해결하는 데 있어 기초 자료로 활용될 것으로 기대된다.

Acknowledgment

본 연구는 2024년 부산가톨릭대학교 교내학술연구과제로 수행되었다.

참고문헌

1. Diamanti-Kandarakis E, Kouli CR, Bergiele AT, Filandra FA, Tsianateli TC, Spina GG, Zapandi ED, and Bartzis MI. 1999. A survey of the polycystic ovary syndrome in the Greek island of Lesbos: hormonal metabolic profile. *J. Clin. Endocrinol. Metab.* **84**:4006-4011.
2. Knochenhauer ES, Key TJ, Kahsar-Miller M, Waggoner W, Boots LR, and Azziz R. 1998. Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. *J. Clin. Endocrinol. Metab.* **83**:3078-3082.
3. Zawadski JK and Dunaif A. 1992. Diagnostic criteria for polycystic ovary syndrome; towards a rational approach. *In*: Dunaif A, Givens JR and Haseltine F (ed.) Polycystic Ovary Syndrome. Blackwell Scientific, Boston. pp. 377-384.
4. Hdaib. Dana, Almajali. Noor, Alquran. Hiam, Mustafa. Wan Azani, Al-Azzawi. Waleed, and Alkhayyat. Ahmed. 2022. Detection of Polycystic ovary syndrome using machine learning algorithms. Engineering Technology and its Applications (IICETA), 2022 5th International Conference on.: 532-536 May.
5. LeCun Y, Boser B, Denker JS, and Henderson D. 1989. Backpropagation applied to handwritten zip code recognition. *Neural Comput*. **1**(4):541-551.
6. Tei, C. 1995. New non-invasive index for combined systolic and diastolic ventricular function. *Journal of Cardiology* **26**(2):135-136.
7. Yassin AM, Abdelrazek GA, Soliman RA, Elkhashab KA, and Zaky SH. 2018. Role of Tissue Doppler Tei Index in Evaluating Myocardial Performance after Coronary Revascularization. *Journal of Clinical and Experimental Cardiology* **9**(6). http://dx.doi.org/10.4172/2155-9880.1000590
8. Harada K, Tamura M, Toyono M, Oyama K, and Takada G. 2001. Assessment of global left ventricular function by tissue Doppler imaging. *The American Journal of Cardiology* **88**(8):927-932. http://dx.doi.org/10.1016/S0002-9149(01)01912-9
9. https://scikit-learn.org/stable/modules/svm.html
10. Song KD, Kim MC, and Do SH. 2019. The Latest Trends in the Use of Deep Learning in Radiology Illustrated Through the Stages of Deep Learning Algorithm Development. *Korean Journal of Radiology* **80**(2):202-212. http://dx.doi.org/10.3348/jksr.2019.80.2.202
11. Warrens J. 2010. Cohen's kappa can always be increased and decreased by combining categories. *Statistical Methodology* **7**(6):673-677. http://dx.doi.org/10.1016/j.stamet.2010.05.003