

# 자연어 처리 기반 『傷寒論』 辨病診斷體系 분류를 위한 기계학습 모델 선정

김영남\*

연세대학교 보건과학대학원 의생명과학전공 석사과정

## Selecting Machine Learning Model Based on Natural Language Processing for *Shanghanlun* Diagnostic System Classification

Young-Nam Kim\*

Master degree, Department of Biomedical Life Science, Graduate School of Public Health Science, Yonsei University

**Objective :** The purpose of this study is to explore the most suitable machine learning model algorithm for *Shanghanlun* diagnostic system classification using natural language processing (NLP).

**Methods :** A total of 201 data items were collected from 『*Shanghanlun*』 and 『*Clinical Shanghanlun*』, ‘Taeyangbyeong-gyeolhyung’ and ‘Eumyangyeokchahunobokbyeong’ were excluded to prevent oversampling or undersampling. Data were pretreated using a twitter Korean tokenizer and trained by logistic regression, ridge regression, lasso regression, naive bayes classifier, decision tree, and random forest algorithms. The accuracy of the models were compared.

**Results :** As a result of machine learning, ridge regression and naive Bayes classifier showed an accuracy of 0.843, logistic regression and random forest showed an accuracy of 0.804, and decision tree showed an accuracy of 0.745, while lasso regression showed an accuracy of 0.608.

**Conclusions :** Ridge regression and naive Bayes classifier are suitable NLP machine learning models for the *Shanghanlun* diagnostic system classification.

**Key words :** Artificial intelligence, Machine learning, Natural Language Processing, *Shanghanlun*, Diagnostic system

\* Corresponding author : Young-Nam Kim. Department of Biomedical Life Science, Graduate School of Public Health Science, Yonsei University.

E-mail : worshipper1994@yonsei.ac.kr . Tel : 82-33-330-2145. Fax : 82-33-330-4809.

· Received : 2022/11/30 · Revised · 2022/12/26 · Accepted : 2022/12/27

## 서 론

최근 기계학습 분야의 진보와 함께 자연어 처리(Natural Language Processing, 이하 NLP) 기술이 주목받고 있다. 자연어 처리(NLP)란, 인공지능의 한 분야로서 인간이 사용하는 일상언어를 컴퓨터가 이해할 수 있는 형태로 변환시켜주는 기술이다. 그간 텍스트는 혈압, 혈당 등 수치화된 생의학 데이터와 달리, 언어를 사용하고 기록하는 사람에 따라 매우 다양한 형태를 보이는 비정형화된 데이터이기에 활용의 장벽이 높은 편이었다<sup>1)</sup>. 그러나 기계학습 기술의 발달로 인간의 일상언어를 정형화된 데이터로 변환하여 컴퓨터에 학습시키는 것이 가능해짐에 따라 점차 텍스트 데이터 또한 분석의 대상으로 인식되기 시작하며, 뉴스와 신문 기사 등의 주제 분류(topic modeling), 문서 작성자의 심리와 감정을 파악하는 감성 분석(sentiment analysis) 등에 텍스트 데이터가 활용되기 시작하였다<sup>2)</sup>. 최근 자연어 처리 기술은 무엇보다 의료 영역에서 큰 주목을 받고 있다. 환자가 호소하는 주소증(chief complaint) 등 전자의무기록(EMR)의 텍스트 데이터를 활용하여 제2형 당뇨병 환자의 경증 저혈당 발생을 진단해내는 기계 학습 모델이 개발되었고<sup>3)</sup>, 이 밖에도 자연어 처리를 통하여 외과 수술 후 합병증 환자를 분류하는 모델<sup>4)</sup>과 폐렴 치료를 위한 모델이 개발되었다<sup>5)</sup>. 한편, 한의학 분야에서는 동

의보감을 자연어 처리하여 한의변증진단 분류 모델을 개발한 것<sup>6)</sup>과 사상체질 진단을 위한 기계 학습 모델<sup>7,8)</sup> 및 한의 증상 추천 모델 개발<sup>9)</sup> 등 주로 후세방을 대상으로 진행된 연구가 존재한다. 이러한 텍스트 분석은 특히 정신과 영역에서 활용도가 매우 높다. 정신과의 진단은 수치화된 생의학 데이터가 아닌, 환자가 호소하는 임상 증상과 환자가 이야기하는 서사(narrative) 및 의사의 관찰에 의해 기술된 텍스트 데이터에 의존할 수 밖에 없기에 객관화가 어려웠는데<sup>10)</sup>, 자연어 처리 기술이 발달함에 따라 비로소 진단의 객관화가 가능해진 것이다<sup>11)</sup>. 본 연구자는 이 점에 주목하였다. 김과 이의 연구<sup>12)</sup>에서 밝혔듯, 『傷寒論』의 辨病診斷體系는 근본적으로 서사의학적 체계를 지니고 있다. 이는 환자를 기계로 보지 않고 전인적으로 진단해내는 장점을 갖고 있으나, 서사 의학의 특성상 기존의 생의학 모델과 달리 객관화가 어려웠다. 본 연구를 통해 환자와의 대화를 중요한 수단으로 활용하는 『傷寒論』의 辨病診斷體系를 자연어 처리(NLP) 기술을 활용하여 분류하기 위해 가장 적합한 기계 학습 모델을 선정하고자 한다.

## 방 법

### 1. 상한론 변병진단 임상증례 수집

『상한론, 고문자적 번역과 해석』\*<sup>9)</sup>에

\* 傷寒論 조문들을 甲骨文 등 古文字적 해석을 통하여 번역한 서적이다. 傷寒論 辨病診斷體系의 기초 이론을 담고있는 서적이기에 선택하였다. 宋代 교정의서국의 교정을 거치지 않은 康平本 傷寒論을 판본으로 삼고 있다.

수록된 상한론 조문 중 처방이 명시되어 있는 조문들 및 『임상 상한론, 상한론의 정신 질환 및 난치성질환 적용과 실제』\*<sup>10)</sup>에 수록된 임상증례 중 총 201례를 수집하였고, 이를 Python 3.8의 pandas 패키지가 제공하는 DataFrame 클래스를 사용하여 데이터베이스화 하였다. 데이터의 과대표집(oversampling) 및 과소표집(undersampling)을 방지하기 위하여 임상증례 수가 너무 많은 太陽病 중 結胸篇과, 임상증례 수가 너무 적은 陰陽易差後勞復病은 분석에서 제외하였다. 수집된 데이터의 구성은 [Tab. 1]과 같다.

**Table 1.** Number of Data by *Shanghanlun* Diagnostic System

Class	Num
Taeyangbyeong (TYB)	83
Yangmyeongbyeong (YMB)	34
Soyangbyeong (SYB)	7
Taeumbyeong (TEB)	11
Soeumbyeong (SEB)	43
Gworeumbyeong (GEB)	23
total	201

## 2. 학습 데이터 전처리

자연어는 비정형 데이터이기 때문에, 컴퓨터가 학습할 수 있는 정형데이터 형태로 표준화시키는 전처리 작업이 선행되어야 한다. 텍스트 데이터 전처리를 위해 KoNLPy

패키지가 제공하는 twitter 한글 형태소 분석기를 사용하여 201개의 임상증례들을 토큰화(tokenization)하였다<sup>11)</sup>. 형태소 분석을 수행한 이후에는, 모델의 성능 극대화를 위하여 각 토큰에 품사를 태깅(POS tagging)한 후<sup>7)</sup>, TF-IDF (Term Frequency - Inverse Document Frequency)를 수행하였다. TF-IDF는 토큰이 여러 문서 중 특정 문서 내에서 얼마나 빈도수가 높은지를 평가하기 위한 통계적 기법으로, 텍스트 분류와 문서 유사도 측정 등에 활용된다<sup>12)</sup>.

## 3. 기계 학습 및 성능 평가

전처리된 데이터들을 바탕으로 傷寒論 辨病診斷體系의 기계 학습을 위해 가장 적합한 분류 모델을 선정하기 위하여 텍스트 분류에 주로 활용되는 6가지 모델의 성능을 평가하였다. train\_test\_split로는 디폴트값인 0.25를 사용했다.

1) 로지스틱 회귀분석(Logistic regression) : 로지스틱 회귀분석은 회귀분석에서 과생된 분류 알고리즘으로서, 회귀분석이 예측하고자 하는 값이 연속변수일 때 활용되는 반면 로지스틱 회귀분석은 불연속적인 범주형 변수의 분류에 가장 보편적으로 사용된다. 본 연구에서는 Python 패키지의 사이킷런(sklearn) 1.1 version이 제공하는 LogisticRegression 클래스를 사용하였다. 본 연구는 분류의 수가 6개인 다중분류이기에 이진분류와 달리 별도의 알고리즘을 사

\* 傷寒論 辨病診斷體系를 활용하여 환자들을 진단하고 치료한 증례를 모아놓은 임상 서적이다. 위와 마찬가지로 傷寒論 조문들은 康平本 傷寒論을 판본으로 삼고 있다.

용해야 하지만, 사이킷런은 이를 자동으로 반영해주고 있다. 모델의 성능은 기계 학습 분류 모델 평가에 가장 많이 활용되는 정확도(accuracy)를 지표로 평가하였다<sup>7)</sup>.

2) 릿지 회귀(Ridge regression): 릿지 회귀는 회귀분석에 정규화까지 활용하는 알고리즘으로, 최적화를 위한 목적함수에 정규화 항목을 넣어 특성(feature)에 대한 계수가 지나치게 커지는 것을 방지한다. 본 연구에서는 사이킷런이 제공하는 RidgeClassifier 클래스를 사용하였고, 정확도를 지표로 성능을 평가하였다.

3) 라쏘 회귀(Lasso regression): 라쏘 회귀는 특성의 계수에 대해 정규화를 하는 점에서 릿지 회귀와 유사하나, 정규화 항에서 차이가 있다. 릿지 회귀는 L2 정규화를 쓰는 반면, 라쏘 회귀는 L1 정규화를 사용한다. L1 정규화는 특성을 정규화할 때 특성의 계수가 0에 가까워지면 이를 완전히 0으로 대체한다는 점에서 L2 정규화와 차이가 있다. 이는 자연어 처리에 거의 영향을 미치지 않는 특성의 값은 사용되지 않게 하여, 결과적으로 특성의 수를 줄여 빠른 속도로 정보를 처리하게 된다는 장점이 있다. 라쏘 회귀는 사이킷런의 LogisticRegression 클래스에서 매개변수를 L1 정규화 방식으로 입력하여 활용하였으며, 정확도를 지표로 성능을 평가하였다.

4) 나이브 베이즈(Naive bayes): 나이브 베이즈는 위의 회귀분석 기반 모델들과는

달리 모든 특성들이 서로 독립이라고 가정한 조건부 확률 모델로서, 자연어 처리 분야에서 회귀분석 모델에 비하여 높은 성능을 발휘하는 것으로 알려져있다<sup>14)</sup>. 나이브 베이즈 기계 학습을 위하여 사이킷런이 제공하는 MultinomialNB 클래스를 사용하였고, 모델의 성능은 정확도를 지표로 평가하였다.

5) 디시전 트리(Decision tree): 디시전 트리는 결정나무 모델이라고도 부르며, 종속 변수를 가장 잘 나뉘는 독립변수의 특성을 가지고 데이터를 점진적으로 Yes/No로 가지치기 분류해나가는 모델이다. 일반적으로 높은 정확도를 보이는 등 좋은 성능을 지니지만, 학습 데이터에 과적합 되는 성향이 강하여 보편 모델을 형성하기 어렵다는 단점이 있다. 사이킷런이 제공하는 DecisionTreeClassifier 클래스를 사용하였고, 모델의 성능은 정확도를 지표로 평가하였다.

6) 랜덤 포레스트(Random forest): 랜덤 포레스트는 기본적으로 디시전 트리와 같은 원리로 학습을 진행하나, 데이터를 여러 디시전 트리로 나누어서 학습시킨 후 다시 합치는 앙상블 기법을 사용하는 것이 특징이다. 디시전 트리의 단점인 과적합을 방지할 수 있으며, 일반적으로 디시전 트리에 비하여 성능이 좋다고 알려져있다. 사이킷런이 제공하는 RandomForestClassifier 클래스를 사용하였으며, 모델의 성능은 정확도를 지표로 평가하였다.

**Figure 1.** Naive bayes Classifier Expression (Each of ‘x1,...,xn’ means featu

$$\begin{aligned}
 p(C_k | x_1, \dots, x_n) &\propto p(C_k, x_1, \dots, x_n) \\
 &\propto p(C_k) p(x_1 | C_k) p(x_2 | C_k) p(x_3 | C_k) \dots \\
 &\propto p(C_k) \prod_{i=1}^n p(x_i | C_k).
 \end{aligned}$$

**Table 2.** Pretreated Data by Twitter Korean Tokenizer

Num	Pretreated Data	Classes
1	'이/Noun', '환자/Noun', '는/Josa', '20년/Number', '전/Noun', '마루/Noun', '에서/Josa', '떨어지다/Verb', '외상/Noun', '을/Josa', '입다/Verb', '틀니/Noun', '가/Josa', '부러지다/Verb', '얼굴/Noun', '을/Josa', '다치다/Verb', './Punctuation', '그/Noun', '후/Noun', '겨울/Noun', '이/Josa', '시작/Noun', '되다/Verb', '안면/Noun', '부/Noun', '우측/Noun', '으로/Josa', '칼/Noun', '로/Josa', '자르다/Verb', '한/Determiner', '통증/Noun', '이/Josa', '발생/Noun', '하다/Verb', '시작/Noun', '하다/Verb', './Punctuation', '그/Noun', '이후/Noun', '에도/Josa', '증상/Noun', '이/Josa', '계속/Noun', '되다/Verb', '12년/Number', '전/Noun', '병원/Noun', '에서/Josa', '시술/Noun', '을/Josa', '받다/Verb', '증상/Noun', '은/Josa', '지속/Noun', '되어다/Verb', './Punctuation', '추위지다/Verb', '바람/Noun', '이/Josa', '불면/Noun', '눈동자/Noun', '가/Josa', '쭈시다/Verb', '우측/Noun', '안면/Noun', '부위/Noun', '통증/Noun', '이/Josa', '심해지다/Verb', '경향/Noun', '도/Josa', '있다/Adjective', './Punctuation'	TYB
...	...	...
201	'3년/Number', '간/Foreign', '취업/Noun', '준비/Noun', '를/Josa', '하다/Verb', '정신/Noun', '적/Suffix', '으로/Josa', '스트레스/Noun', '를/Josa', '많이/Adverb', '받다/Verb', '매우/Noun', '몰입/Noun', '하다/Verb', '준비/Noun', '하다/Verb', './Punctuation', '1년/Number', '전/Noun', '부터/Josa', '속/Noun', '쓰리다/Adjective', '명치/Noun', '부위/Noun', '가/Josa', '체/Noun', '한/Josa', '것/Noun', '처럼/Josa', '막히다/Verb', '증상/Noun', '이/Josa', '발병/Noun', '하다/Verb', './Punctuation', '현재/Noun', '취업/Noun', '1/Number', '개월/Noun', '정도/Noun', '되다/Verb', '상황/Noun', '에서/Josa', '직장/Noun', '내/Noun', '에서/Josa', '자신/Noun', '의/Josa', '일/Noun', '을/Josa', '완벽하다/Adjective', '해내다/Verb', '한다는/Modifier', '부담/Noun', '과/Josa', '스트레스/Noun', '를/Josa', '겪다/Verb', '있다/Adjective', '상태/Noun', '이다/Verb', './Punctuation', '소화도/Noun', '자다/Verb', '안/Noun', '되다/Verb', './Punctuation', '취업/Noun', '을/Josa', '준비/Noun', '하다/Verb', '상황/Noun', '에서도/Josa', '공부/Noun', '하다/Verb', '잠/Noun', '을/Josa', '자다/Verb', '자지/Noun', '못/VerbPrefix', '하다/Verb', '하다/Verb', './Punctuation'	GEB

## 결 과

각 모델에 학습시킬 데이터를 twitter 한글 형태소 분석기로 토큰화(tokenization) 및 품사 태깅(POS tagging)하여 전처리한 결과는

[Tab. 2]과 같다. 각 토큰에 태깅된 ‘/Noun’은 명사를, ‘/Verb’는 동사를, ‘/Adjective’는 형용사를, ‘/Josa’는 조사, ‘/Number’는 수사, ‘/Suffix’는 접미사를, ‘/Prefix’는 접두사를, ‘/Conjunction’는 접속사를, ‘/Determiner’

는 어미를, ‘/Modifier’는 수식언을, ‘/Punctuation’은 문장부호를 의미한다.

위와 같이 전처리된 데이터를 바탕으로, ‘로지스틱 회귀분석(Logistic regression)’, ‘릿지 회귀(Ridge regression)’, ‘라쏘 회귀(Lasso regression)’, ‘나이브 베이즈(Naive bayes)’, ‘디시전 트리(Decision tree)’, ‘랜덤 포레스트(Random forest)’의 6가지 알고리즘 모델에 기계학습을 수행하였다. 분류 모델의 성능을 평가하기 위하여 정확도 (accuracy)를 지표로 삼았다. 각 모델에 모든 품사를 학습시켰을 때의 정확도는 로지스틱 회귀가 0.804, 릿지 회귀가 0.824, 라쏘 회귀가 0.510, 나이브 베이즈가 0.843, 디시전 트리가 0.725, 랜덤 포레스트가 0.804를 나타냈으며, 그 결과를 [Tab. 3]로 정리하였다.

**Table 3.** Accuracy of Algorithms Using All POS

Algorithm	Accuracy
Logistic regression	0.804
Ridge regression	0.824
Lasso regression	0.510
Naive bayes	0.843
Decision tree	0.725
Random forest	0.804

반면에 ‘명사(Noun)’와 ‘동사(Verb)’, ‘형용사(Adjective)’만을 학습시켰을 때의 성능은 [Tab. 4]과 같았다. 로지스틱 회귀와 나이브 베이즈, 랜덤 포레스트의 경우 모든 품사를 사용했을 때와 정확도의 차이가 없으나, 릿지 회귀와 라쏘 회귀, 디시전 트리의 경우 ‘명사

(Noun)’와 ‘동사(Verb)’, ‘형용사(Adjective)’만을 학습시켰을 때의 성능이 각각 0.843과 0.608, 0.745로 개선됨을 확인하였다.

**Table 4.** Accuracy of Algorithms Using ‘Noun’, ‘Verb’, ‘Adjective’

Algorithm	Accuracy
Logistic regression	0.804
Ridge regression	0.843
Lasso regression	0.608
Naive bayes	0.843
Decision tree	0.745
Random forest	0.804

추가적으로 인공지능이 어떤 특성 (feature)을 높은 coef\_ 값으로 처리하였는지 확인하고자 상위 10개의 특성들을 [Tab. 5]로 정리해보았다. 사이킷런(sklearn) 1.1부터는 본 연구에서 우수한 성능을 보이는 나이브 베이즈와 릿지 회귀 모델 중, 나이브 베이즈 분류기의 coef\_ 값을 제공하지 않기에 coef\_ 값을 제공하는 릿지 회귀 모델의 상위 10개의 특성을 추출하였다. 릿지 회귀 모델은 ‘과도’, ‘일’, ‘낮’, ‘짜증’, ‘시작’, ‘계지’, ‘증상’, ‘부위’, ‘호전’, ‘있다’를 태양병의 주된 분류 특성으로 판단하였다. 이 중에서 ‘부위’나 ‘있다’의 경우 태양병의 특성을 특별하게 반영한다고 보기 어렵기에 불용어 처리(stop word)를 해준다면 모델의 성능이 조금 더 개선될 수 있을 것으로 보인다. 불용어란 대체로 텍스트에서 높은 출현빈도를 갖지만 텍스트 주제 구현에는 기여도가 매우 낮은 단어들을 말한다. 다시 말해 텍스트

의 주제와는 상관성이 적은 잉여 성분들로서, 제거해준다면 모델의 성능이 개선되게 된다<sup>18)</sup>. 양명병의 경우에는 ‘정확하다’, ‘분명하다’, ‘열등감’, ‘채우다’, ‘것’, ‘만족하다’, ‘확인’, ‘대변’, ‘자신’, ‘반복’을 주된 분류 특성으로 판단하였는데, 마찬가지로 의존명사 ‘것’의 경우 양명병의 특성을 특별히 반영한다고 볼 수 없기에 불용어 처리한다면 성능이 개선될 것이다. 소양병의 경우 ‘눈’, ‘악화’, ‘어지럽다’, ‘거나’, ‘목’, ‘토’, ‘낮’, ‘경우’, ‘입’, ‘설사’를 주된 분류 특성으로 파악하였는데, ‘-(하)거나’는 어미에 해당하는 토큰이기에 학습에서 제외되었어야 함에도 불구하고, twitter 한글 형태소 분석기의 성능 한계로 인하여 고유명사로 판

단한 것으로 보인다. 태음병의 주된 분류 특성은 ‘가스’, ‘복부’, ‘배’, ‘식사’, ‘늦다’, ‘차고’, ‘차다’, ‘아프다’, ‘과도’, ‘앓다’로 나타났다. ‘차고’와 ‘차다’는 같은 어근이지만 twitter 한글 형태소 분류기는 ‘차고’를 명사로, ‘차다’는 형용사로 오판하여 별도의 특성으로 취급된 것으로 보인다. 소음병은 ‘염려’, ‘수면’, ‘방해’, ‘근심’, ‘무기력’, ‘움직임’, ‘줄어들다’, ‘걱정’, ‘사소하다’, ‘건강’이 주된 분류 특성으로 판단되었는데, 다른 분류에 비하여 가장 잘 판별된 것으로 생각된다. 율음병은 ‘몰입’, ‘집중’, ‘구역질’, ‘끓다’, ‘냄새’, ‘대해’, ‘스트레스’, ‘다시’, ‘민감하다’, ‘설사’가 주된 분류 특성으로 나타났다. 이 중 ‘대해’는 불용어로 여겨진다.

**Table 5.** Top 10 Features by Ridge Regression

	<b>TYB</b>	<b>YMB</b>	<b>SYB</b>	<b>TEB</b>	<b>SEB</b>	<b>GEB</b>
1	과도 (overwork)	정확하다 (precise)	눈 (eye)	가스 (gas)	염려 (concern)	몰입 (immerse)
2	일 (work)	분명하다 (clear)	악화 (degeneration)	복부 (abdomen)	수면 (sleep)	집중 (concentration)
3	낮 (day)	열등감 (complex)	어지럽다 (dizzy)	배 (abdomen)	방해 (disturbance)	구역질 (nausea)
4	짜증 (irritation)	채우다 (fill)	거나 (Geona)	식사 (meal)	근심 (anxiety)	끓다 (stop)
5	시작 (start)	것 (Geot)	목 (throat)	늦다 (late)	무기력 (lethargy)	냄새 (smell)
6	계지 (Gyeji)	만족하다 (satisfy)	토 (vomit)	차고 (cold)	움직임 (movement)	대해 (about)
7	증상 (symptom)	확인 (confirmation)	낮 (day)	차다 (cold)	줄어들다 (decrease)	스트레스 (stress)
8	부위 (region)	대변 (stool)	경우 (case)	아프다 (sick)	걱정 (worry)	다시 (again)
9	호전 (improvement)	자신 (self)	입 (mouth)	과도 (overwork)	사소하다 (trivial)	민감하다 (sensitive)
10	있다 (be)	반복 (repeat)	설사 (diarrhea)	앓다 (sit)	건강 (health)	설사 (diarrhea)

## 고 찰

본 논문은 실제 의무기록을 바탕으로 자연어 처리 기술(NLP)을 傷寒論 辨病診斷體系에 적용한 최초의 논문으로서, 傷寒論 辨病診斷體系에 대한 자연어 처리 기술 적용의 가능성을 확인했으며, 향후 관련 연구자들을 위한 연구 모델을 제시하였다는 것에 큰 의의가 있다. 노와 김<sup>9)10)</sup>의 저서에서 201개의 데이터를 수집하여 twitter 한글 형태소 분석기로 전처리 과정을 거친 후 ‘로지스틱 회귀분석(Logistic regression)’, ‘릿지 회귀(Ridge regression)’, ‘라쏘 회귀(Lasso regression)’, ‘나이브 베이즈(Naive bayes)’, ‘디시전 트리(Decision tree)’, ‘랜덤 포레스트(Random forest)’의 6가지 모델을 기반으로 기계학습을 수행하였다. 대체적으로 품사는 모든 품사를 학습시키는 것 보다 명사, 동사, 형용사만을 학습데이터로 사용할 때 성능이 개선되었는데, 그 결과 릿지 회귀 모델과 나이브 베이즈 분류기가 0.843로 가장 높은 정확도(accuracy)를 나타내었고, 로지스틱 회귀분석과 랜덤 포레스트는 0.804의 정확도를, 디시전 트리는 0.745를, 라쏘 회귀는 0.608로 가장 낮은 정확도를 보였다 [Tab. 4]. 따라서 환자들의 임상증례를 기반으로 한 傷寒論 辨病診斷體系의 자연어 처리에는 릿지 회귀와 나이브 베이즈 분류기가 가장 적합한 것으로 보인다. 릿지 회귀 모델의 경우, 후속 연구에서 의존명사 '것', 어미 ‘-(하)거나’ 등을 불용어 처리(stopword) 한다면 성능이 더욱 개선될 것으로 보인다. 또한, 본 연구에서는 어간 추출

(Stemming)을 전처리에 사용하였으나 표제어 추출(Lemmatization)을 사용해 보는 것도 성능개선에 도움이 될 수 있다. 그러나, 보편적인 기계학습 모델을 개발하기에는 494개의 임상증례 데이터를 사용한 백 등의 연구<sup>7)</sup>에 비하여 데이터의 수가 201개로 비교적 적기에 정확도 개선에 한계를 지닌다. 특히, 소양병과 태음병의 경우 데이터가 현저히 적기에 실제 임상현장에서 적용되기에는 주의를 요한다. 또한 이러한 한계로 인하여 많은 데이터가 필요한 딥러닝(deep learning) 모델 개발은 시도해보지 못하였다. 딥러닝 모델은 여러 연구를 통하여 일반적인 기계학습 모델 보다 성능이 뛰어난 모델로 밝혀졌다<sup>1)</sup>. 향후 허락을 구하여 傷寒論 辨病診斷體系를 활용하는 임상의 및 연구자들로부터의 다양한 임상증례들을 활용할 수 있다면, 기계학습뿐 아니라 보다 정확도가 우수하며 실제 임상에서까지 활용가능한 딥러닝 모델에 기반한 연구도 진행해볼 수 있을 것으로 기대된다.

## 결 론

1. 자연어 처리(NLP)에 기반한 기계 학습(machine learning)을 통하여 傷寒論 辨病診斷體系진단에 대한 객관화의 가능성을 확인할 수 있었다.
2. 傷寒論 辨病診斷體系를 자연어 처리하여 분류하기 위한 최적의 알고리즘 모델은 릿지 회귀(Ridge regression)와 나이브 베이

즈(Naive bayes) 분류기이며, 모든 품사를 학습시키는 것 보다는 명사와 동사 및 형용사를 학습 데이터로 활용할 때 성능이 개선됨을 확인할 수 있었다.

3. 기계 학습뿐 아니라 추후 傷寒論 辨病診斷體系에 대한 딥러닝(deep learning) 모델 연구가 수행된다면, 첨단 인공지능을 통하여 한의사가 환자의 진단에 도움을 받는 등 한의학 과학기술의 발전이 기대된다.

## 감사의 글

인공지능에 대해 문의한이었던 한의사에게 Python 코딩의 기초부터 세밀하게 가르쳐주신 김화중 교수님과, 무엇보다 『傷寒論』의 고문자적 해석을 통하여 辨病診斷體系를 밝혀주시고 책으로 내주신 노영범 박사님과 김경일 교수님, 그리고 난치성질환으로 고통받는 환자들을 위하여 傷寒論 辨病診斷體系를 발전시켜오신 모든 선배 연구자 분들께 깊은 존경과 감사의 인사를 드립니다. 함께 공부하는 UOI 공동체에도 항상 감사합니다.

## Reference

1. Sang-Un P. Analysis of the Status of Natural Language Processing Technology Based on Deep Learning. The Journal of Bigdata. 2021;6(1):63-81.

doi.org/10.36498/kbigdt.2021.6.1.63

2. Srivastava A, Sahami M. Text mining : classification, clustering, and applications. CRC Press. 2009.

3. Misra-Hebert AD, Milinovich A, Zajichek A, Ji XG, Hobbs TD, Weng WN. Natural Language Processing Improves Detection of Nonsevere Hypoglycemia in Medical Records Versus Coding Alone in Patients With Type 2 Diabetes but Does Not Improve Prediction of Severe Hypoglycemia Events: An Analysis Using the Electronic Medical Record in a Large Health System. Diabetes Care. 2020;43(8):1937-40. doi.org/10.1055/s-0038-1626725

4. Mellia JA, Basta MN, Toyoda Y, Othman S, Elfanagely O, Morris MP. Natural Language Processing in Surgery A Systematic Review and Meta-analysis. Annals of Surgery. 2021;273(5):900-8. doi.org/10.1097/sla.0000000000004419

5. Jones BE, South BR, Shao Y, Lu CC, Leng J, Sauer BC. Development and Validation of a Natural Language Processing Tool to Identify Patients Treated for Pneumonia across VA Emergency Departments. Applied Clinical Informatics. 2018;9(1):122-8. doi.org/10.1055/s-0038-1626725

6. SeungHyeon L, DongPyo J, KangKyung S. Donguibogam-Based Pattern Diagnosis Using Natural Language Processing and Machine Learning. The Journal of Korean Medicine. 2020;41(3):1-8. doi.org/10.13048/jkm.20021

7. Jae-Hwa L, Hyun-Hak L. Selecting Sasang-Type classification model using machine learning and designing the service flow. Journal of Digital Contents Society.

- 2019;20(2):321-7.  
doi.org/10.9728/dcs.2019.20.2.321
8. Musun P, Minwoo H, Jeongyun L, Chang-Eop K, Young-Kyu K. Research on the Evaluation and Utilization of Constitutional Diagnosis by Korean Doctors using AI-based Evaluation Tool. *J Physiol & Pathol Korean Med.* 2022;36(2):73-8.  
doi.org/10.15188/kjopp.2022.04.36.2.73
  9. Ho J, Sang-Hun L, Sangjun Y. Korean Medicine Symptom Recommendation Based on Textual Description of the Patient's Condition (Digestive disorder version). *Journal of Knowledge Information Technology and Systems.* 2022;15(5):831-43.  
doi.org/10.34163/jkits.2020.15.5.025
  10. Craddock N, Mynors-Wallis L. Psychiatric diagnosis: impersonal, imperfect and important. *The British Journal of Psychiatry.* 2014;204(2):93-5.  
doi.org/10.1192/bjp.bp.113.133090
  11. Pak DH MD, Hwang MG MD, Lee MJ MD, Woo SI MD, Hahn SW MD, Lee YJ MD, Hwang JU MD. Application of Text-Classification Based Machine Learning in Predicting Psychiatric Diagnosis. *Korean J Biol Psychiatry.* 2020;27(1):18-26.  
doi.org/10.22857/kjbp.2020.27.1.003
  12. Jin-A K, Sung-Jun L. Shanghanlun Diagnostic System : Exploring Value as Narrative Medicine by Analyzing Cases. *J of KMediACS.* 2014;6(1):1-25.  
doi.org/10.22891/kmedia.2014.6.1.1
  13. Kim KI, Roh YB. Etymological Shanghanlun, translation and interpretation based on old chinese characters. *BADABOOKS.* 2015.
  14. Roh YB, Kim KI. Clinical Shanghanlun, application and practice of mental and incurable diseases. *BADABOOKS.* 2020.
  15. Park EL, Cho S. KoNLPy: Korean natural language processing in Python. *Proceedings of the 26th Annual Conference on Human and Cognitive Language Technology;*2014 Oct 10-11, Chuncheon, Korea.
  16. Ramos JA. Using TF-IDF to determine word relevance in document queries. *Proceedings of the First Instructional Conference on Machine Learning;*2003 Dec 8-10, Rutgers, NJ, USA.
  17. Sadhasivam J, Kalivaradhan RB. An empirical comparison of supervised learning algorithms and hybrid WDBN algorithm for MOOC courses. *Journal of ambient intelligence and humanized computing.* 2019.  
doi.org/10.1007/s12652-019-01190-9
  18. LUXING H, Jin K, Jing. A study on the lists of common korean stopwords for text mining. *Korean Language Research.* 2022;63(13):1-15.  
doi.org/10.16876/klrc.2022.63.13.1