Journal of KIIT. Vol. 18, No. 1, pp. 11-17, Jan. 31, 2020. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2020.18.1.11

머신러닝 기반 반려동물 진단을 위한 임상 의사결정 지원 모델의 예비 연구

최우용*, 박래정**

A Preliminary Study on Clinical Decision Support Model for Pet Diagnosis Based on Machine Learning

Wooyong Choi*, Lae-Jeong Park**

이 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원의 수출전략기술개발사업의 지원을 받아 연구되었음. (과제번호 : 317022-03-1-SB010)

요 약

본 논문은 주요 증상과 기본검사 결과로부터 반려동물(개와 고양이)의 진단 범주를 예측하는 머신러닝 기반 반려동물 진단 보조 모델에 관한 연구 결과를 소개한다. 진단 예측 모델을 개발하기 위해서 동물병원 EMR (electronic medical record)의 확진 데이터에서 수의사의 검증을 거쳐 기본혈액검사와 혈액화학검사의 항목 수치 보정, 진단 범주 확정 등 리뷰 과정을 수행하여 변환하였으며, 병원 간 진단명 불일치 문제를 해결하여 최종 진단 범주를 결정하였다. 이후 수집된 데이터의 분석 과정을 통해서, 진단에 영향이 큰 증상을 선택하였으며 총 80개의 증상 코드로 통합하였다. 실용적인 진단 보조를 위해서는 추가적인 검사 정보가 필요하지만, 본 논문에서는 최종적으로 동물병원에서 실시하고 있는 기본적인 검사결과 데이터와 주요 증상 정보를 사용하여 22개의 진단 범주를 예측하는 분류 모델을 개발하였다. 진단 예측 모델은 총 1,296개의 데이터로 검증하였으며, 82%의 예측 성공률을 나타낸다.

Abstract

This paper introduces a preliminary study on a machine learning-based diagnostic support model that predicts diagnostic categories of pets (dogs and cats) on the basis of their major symptoms and basic blood tests. In order to build the veterinary diagnostic support model, EMR (Electronic Medical Record) data collected from eight veterinary clinics were curated under the supervision of veterinarians. The curation consists of defining un-existing standard diagnosis categories, standardizing attribute names of the blood tests, and discretizing their numeric values. Through investigation of the EMR data set, we chose a list of 80 major symptoms that are highly related to the diagnosis. Practical diagnostic support requires additional test information, but in this research, with the curated EMR data set, we built a classification model that predicts likely diagnosis among 22 diagnosis categories based on information about major symptoms and the basic blood test in the veterinary clinics. The classification model was evaluated with the test data of 1,296 EMR data, showing a predictive success rate of 82%.

Keywords
CDSS, machine learning, random forest, EMR data

* (주)메디사피엔스 선임연구원
- ORCID: http://orcid.org/0000-0002-6981-899X
** 강릉원주대학교 전자공학과 교수
- ORCID: http://orcid.org/0000-0002-2672-7270

· Received: Dec. 06, 2019, Revised: Dec. 26, 2019, Accepted: Dec. 29, 2019
· Corresponding Author: Lae-Jeong Park
Gangneung-Wonju National University, 7 Jukheon-gil, Gangneung-si, Gangwon-do, Korea. 35457
Tel.: +82-33-640-2389, Email: ljpark@gwnu.ac.kr

Ⅰ. 서 론

우리나라에서는 가구당 반려동물의 비율이 점차 높아지는 시대에 접어들었으며 이에 따라 반려동물의 치료비용 부담도 증가하고 있다[1]. 일반적으로 증상이 있어 병원을 방문하면, 수의사는 증상을 확인하고 증상과 연관되는 질환 여부를 확인하기 위한 검사를 시행하고 그 결과를 토대로 진단한다. 호소 증상과 검사결과를 리뷰하는 과정에서, 빈도가 낮은 케이스에 대해서 수의사의 임상 경험이 부족하거나 숙련도가 낮은 경우에, 적시에 적합한 검사를 하지 못하거나 불필요한 검사를 수행하게 되어 확진까지의 시간과 비용이 증가할 가능성이 존재한다. 특히, 반려동물의 특성상 증상에 대한 표현과 문진이 용이하지 않으므로 수의사의 초기 판단이 더욱 중요하다. 이러한 증상과 검사결과 리뷰로부터 초기 진단 과정에서의 오류 가능성을 최소화하기 위한 반려동물 진단 보조 시스템의 효용성은 사람 대상 진료 분야에서보다도 훨씬 더 클 수 있다.

머신러닝을 활용한 반려동물의 진단 관련 연구 중 고양이의 장 질환 중 inflammatory bowel 질병과 alimentary lymphoma에 대해 머신러닝 알고리즘을 토대로 분류하여 진단 결정을 보조하는 방법을 소개하였다[2]. 다른 연구에서는 결정 트리를 사용하여 반려동물의 각 다리에 가해지는 체중 정보로부터 반려동물의 이상징후를 진단하는 방법을 제시하였다[3]. 기존 연구와 달리, 저자는 실제 동물병원의 임상 데이터를 토대로 동물병원 현장 (동물병원에 방문하는 대상은 개와 고양이이므로 본 연구의 대상을 개, 고양이로 국한하였음)에서 사용 가능한 진단지원 시스템을 목표로 개발을 진행해왔다.

본 논문에서는 초기 연구 결과를 소개한다. 간략히는, 진단 보조 시스템의 개발을 위해 우선 병원 간 산재된 자료를 수집하였으며, 병원 간 상이한 데이터 포맷과 코드 일치 등의 데이터 포맷 전처리 작업과 증상 및 진단코드를 통합하고 분류하는 작업을 수행하였다. 진단 분류를 위한 모델은 랜덤 포레스트(Random forest) 모델을 기반으로 구축하였다[4][5]. 모델은 반려동물의 개인정보, 증상, 검사결과의 입력에 대해 가능성이 높은 진단명 리스트를 출력한다.

논문의 구성은 다음과 같다. 2장에서 데이터 및 전처리 과정과 모델 구축에 대해 소개한다. 3장에서 분류 모델을 사용한 실험 결과에 대해 상세히 설명하고, 마지막 장에서는 반려동물의 진단 분류 성능에 관한 토론과 한계, 그리고 향후 연구 방향을 소개하고 결론을 맺는다.

Ⅱ. 데이터 전처리 및 모델 구축

실제 동물병원에서 사용할 정도의 진단 정확도를 위해서는 많은 진료 기록 데이터(EMR, Electronic Medical Record)가 필수적이므로, 최대한 여러 동물병원으로부터 많은 양의 데이터를 수집해야 한다. 아쉽게도 진료 기록 데이터의 속성 명칭이 병원별로 상이하며, 무엇보다도 통일된 진단명 범주가 존재하지 않아 동물병원마다 다른 코드를 사용하고 있다. 또한, 병원마다 사용하는 검사장비의 차이로 인해 동일 검사항목에 대한 검사결과 수치의 불일치 문제가 존재한다. 이런 데이터 통합 시 발생하는 문제들을 해결하기 위해 데이터 속성 일치 작업, 일본의 ANICOM code를 활용하여 22개의 진단코드로 수정하는 코드화 과정, 장비별 검사결과 수치를 상호 보정하고 이를 등급화(Discretization)하는 작업을 수행하였다[6]. 그림 1은 수집한 EMR 데이터를 전처리하고 통합하며 이를 사용하는 단계를 도시하였다.

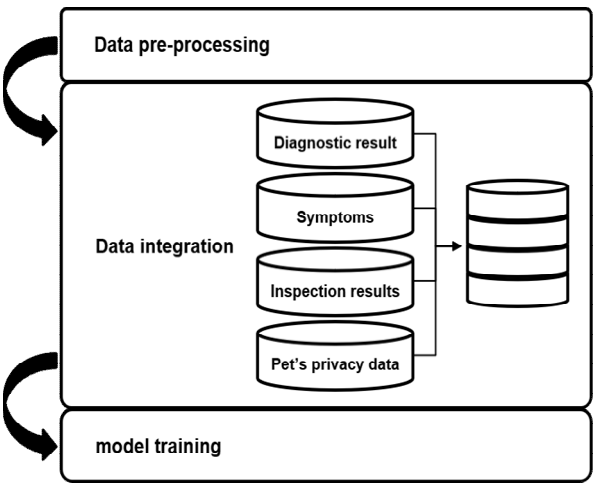


그림 1. 3단계 진행 과정
Fig. 1. Three steps process

첫 단계에서 각 병원의 EMR 데이터에 대해 적절한 전처리를 진행한다. 두 번째 단계에서 전처리한 데이터를 통합하여 최종적으로 반려동물 증상과 검사와 진단 정보를 가진 데이터베이스를 완성한다. 마지막 단계에서는 통합 데이터베이스를 사용하여 랜덤 포레스트 학습을 진행하고 평가한다. 이하 각 단원에서 상세히 설명한다.

2.1 ANICOM 코드

데이터 통합과정에서 중요한 작업 중의 하나가 병원별로 상이한 진단명을 통일하는 것이었다. 아쉽게도 우리나라에는 아직 통합된 반려동물 진단코드가 존재하지 않는다. 이런 이유로, 이미 검증이 완료되고 일본에서 사용 중인 질환 코드인 ANICOM 코드를 사용하여 진단명을 코드화하였다(그림 2 참조).

해당 코드는 질환명을 대분류 및 소분류로 분류하고 있다. 대분류는 16개로 구성되어 있으며, 각 대분류마다 소분류 진단명이 있으며 총 303개로 구성되어 있다. 여기서 대분류 중 이국적인 동물에 해당하는 경우는 제외하여 소분류는 291개로 축소되었다. 대분류는 소화기, 순환기, 호흡기, 비뇨기 등으로 구성된다. 예를 들어 '소화기' 대분류 속에 '위염', '장염', '구토', '설사' 등의 소분류로 세분화된다.

최종 진단 분류는 데이터 수가 적은 8,9,11의 대분류를 제외한 나머지 대분류 12개와 다빈도 질환(291개의 질환 중 반려동물에게 주로 발생하는 질환) 10개를 추가하여 총 22개의 진단 클래스로 하였다. 이는 소분류 차원(291개 클래스)에서의 진단 분류 모델을 구축하기에는 취합된 데이터의 개수가 충분하지 않기 때문이며, 대분류 차원에서의 진단이 갖는 낮은 임상적 효용성을 높이기 위해 소분류 중 상위 10개의 다빈도 질환을 대분류 범주에 추가하였다. 다빈도 질환은 기존 대분류에 속한 질환이기 때문에 중복되지 않도록 해당되는 대분류에서 제거하여 사용하였다.

표 1은 모델 학습에 사용된 22개의 진단 정보에 대한 대분류 12개와 소분류 10개의 ANICOM code를 보여준다.

표 1. 22개의 진단코드 정보
Table 1. Information of 22 diagnostic codes

	Diagnostic codes
Large scale categories	1 2, 3, 4, 5, 6, 7, 10, 12, 13, 14, 15
High frequent diagnosis	2001, 2043, 2046, 2082, 2087, 2092, 2095, 2101, 2133, 2170

2.2 반려동물 EMR 확진 데이터의 전처리 및 통합

본 연구에서는 총 8곳의 동물병원으로부터 데이터를 수집하였다. 반려동물의 증상과 진단 관련한 모든 정보는 기본적으로 EMR 데이터로부터 확보할 수 있다. 하지만 모든 동물병원에서 동일한 증상명, 검사 항목명을 사용하지 않기 때문에 데이터 통합하기 전에 전처리 과정이 필수적이다.

Index	Class	Code	Diagnosis in Japan
1	CARDIOVASCULAR SYSTEM DISORDERS	2001~2016	弁膜症（疑い含む、心雑音+、心不全徴候−）
2	RESPIRATORY SYSTEM DISORDERS	2017~2039	鼻炎/副鼻腔炎/上部気道炎
3	DIGESTIVE SYSTEM DISORDERS	2040~2073	食道炎
4	HEPATOBILIARY and EXOCRINE PANCREATIC DISORDERS	2074~2086	肝炎
5	URINARY TRACT DISORDERS	2087~2099	慢性腎臓病（腎不全含む）
6	REPRODUCTIVE SYSTEM DISORDERS	2100~2118	卵巣の疾患
7	NEUROMUSCULAR DISORDERS	2119~2133	てんかん
8	OPHTHALMOLOGY	2134~2156	結膜炎（結膜浮腫含む）
9	EAR DISEASES	2157~2169	細菌性外耳炎
10	DENTISTRY	2170~2180	歯周病/歯肉炎（乳歯遺残に起因するもの含む）
11	ORTHOPEDICS	2181~2203	椎間板ヘルニア
12	DERMATOLOGY	2204~2229	膿皮症/細菌性皮膚炎
13	HEMATOLOGY	2230~2243	貧血(免疫介在性溶血性）・IMHA
14	ENDOCRINE DISORDERS	2244~2253	糖尿病
15	SYSTEMIC DISORDERS	2254~3000	タマネギ中毒/ネギ中毒
16	EXOTIC ANIMAL DISEASES	3001	スナッフル

그림 2. ANICOM의 질병코드
Fig. 2. ANICOM disease codes

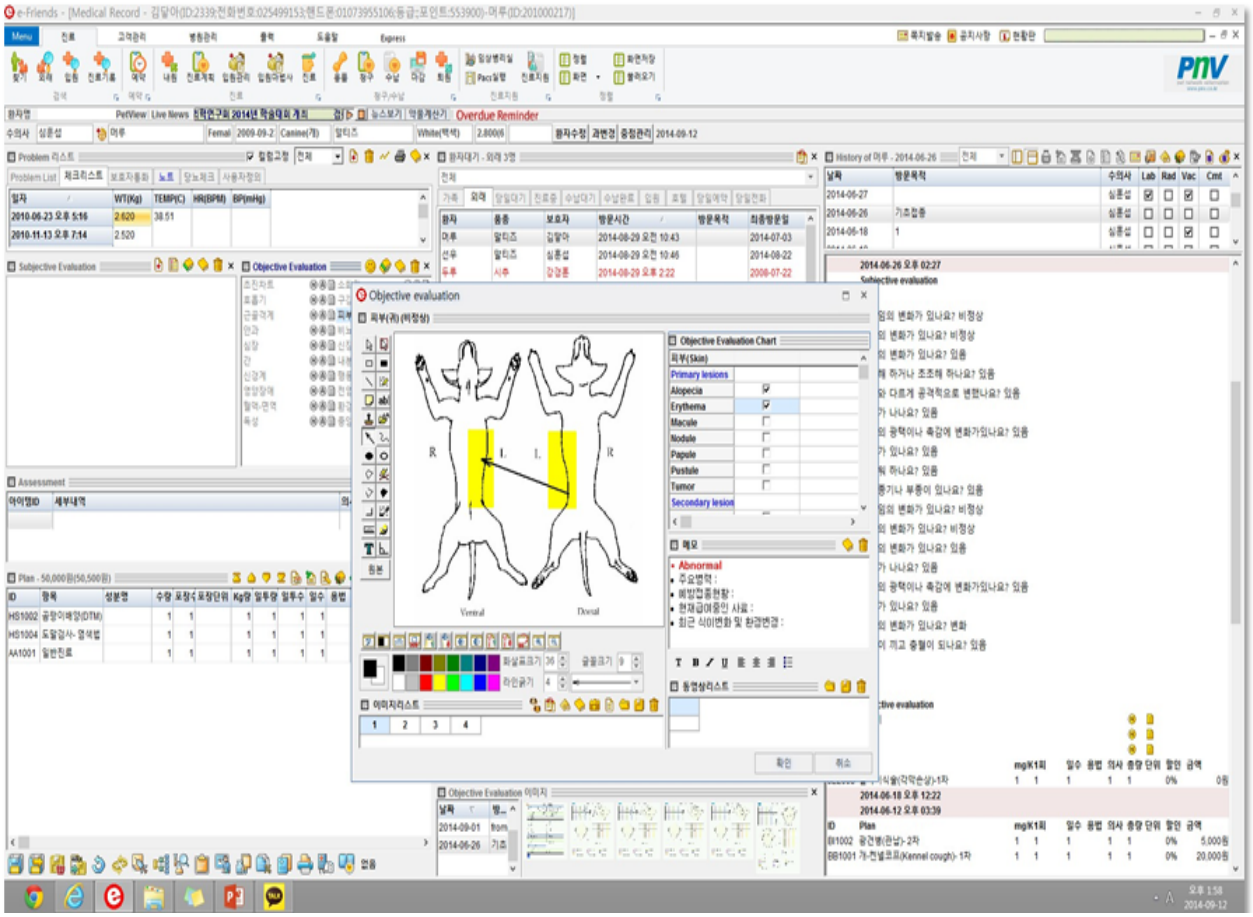


그림 3. 동물병원의 EMR 소프트웨어
Fig. 3. EMR software of animal hospital

그림 3은 본 논문에서 사용된 데이터를 얻기 위해 사용된 EMR 소프트웨어를 보여주고 있다. 8곳의 동물병원으로부터 해당 소프트웨어를 사용하여 수집된 데이터를 사용하였다.

기본적인 반려동물의 품종, 나이, 성별의 정보는 그대로 사용하였다. 전처리 과정은 총 세 단계로서 증상의 코드화, 검사항목의 기준명칭 통합, 검사결과의 등급화로 구성된다.

증상 속성의 경우는 진찰 중에 수의사가 입력한 의견으로서 대부분이 텍스트 형태로 되어 있으므로 범주화(Categorization)하였으며, 진단 속성과 마찬가지로 수의사의 검증을 통해 총 80개의 증상 범주로 표준 코드화하였다. 수집한 데이터의 통합 작업에서의 또 다른 어려운 점은 동물병원에서 사용하고 있는 검사장비(총 장비 개수는 혈액검사 장비 9개, 화학검사 장비 14개)의 속성 (EMR상의 항목명칭, 검사결과 수치 범위 등)이 동일하지 않은 점이다. 동종 검사장비의 검사명칭을 일치하기 위해 수의사의 검증을 받아 64개의 기준검사명칭으로 통합하였다. 또한, 제조사가 달라서 검사 수치의 범위가 상이한 동종 검사장비의 측정 결과 수치를 통합하기 위해 각 검사장비의 정상범위 참조치(Normal range)를 기준으로 재설정하였다. 참조치의 정상인 범위를 기준으로 7등급(Missing value: -1, 미만: 0, 정상: 1~4, 초과: 5)으로 등급화 하였다. 정상 등급에 대해서는 정상범위의 25% 이하: 1, 25% 이상 50% 이하: 2, 50% 이상 75% 이하: 3, 75% 이상: 4로 지정하였다. 8곳의 동물병원에서 수집한 17,458개 EMR 데이터에 대해 위의 전처리 과정을 통해서 총 169개의 속성 (환자의 기본정보: 3개, 증상 관련 속성: 80개, 검사항목 관련 속성: 64개, 진단명: 22개)을 갖는 6,481개의 데이터를 확보하였다. 17,458개 중 약 62%에 해당하는 데이터는 단순 건강검진, 임신, 치석 제거 등 질환이 아닌 경우로 혈액검사와 화학검

사를 통해 확인 불가능한 진단에 해당되어 제외하였다.

2.3 Random Forest 학습

본 논문에서 제시하는 반려동물 진단 분류 모델은 147개의 예측 인자를 기반으로 22개의 진단 클래스 중에서 가능성 높은 진단 정보를 제공한다. 머신러닝의 이론상, 6,500여 개의 데이터 개수에 대하여 147의 입력 차원은 상당히 고차원이다. 또한, 의료 분야라는 특성상 추론 결과에 대한 사용자의 해석 가능성(Interpretability)도 모델의 선택에서 중요한 요소이다. 이러한 점을 고려하여 랜덤 포레스트 모델을 채택하였다. 랜덤 포레스트 모델은, 고차원의 데이터를 다루기 위해서 랜덤성에 의해 트리들이 조금씩 다른 특성을 갖고 각 트리들의 예측들이 비상관화한다[5][7]. 또한, 다양한 예측 인자로 인해 생기는 노이즈 역시 다루기 위해 랜덤화 한다. 랜덤 포레스트는 기존의 분류 회귀 트리의 발전된 형태이다. 분류 회귀 트리는 설명변수 또는 예측 인자의 비선형성과 상호작용을 최대한 활용하여 변수에 대한 영향을 판단하는 기법이다. 설명변수를 중요도 기준에 따라 줄기(Branch)를 만들고 leaf node에서 반응변수에 관한 판단을 내린다[7]-[10].

본 논문에서는 leaf node에서의 높은 probability를 갖는 상위 5개 클래스(진단 분류명)를 사용자에게 제시한다. 147개의 속성을 갖는 입력에 대해 22개 진단 범주 중 1개의 진단만을 제시하는 방식보다는 22개 클래스 중 높은 가능성을 갖는 상위 5개의 진단 범주를 해당 범주의 확률과 함께 제공하는 방식이, 수의사 1인이 진단하는 1차 동물병원 특성상 진단보조 시스템의 효용성이 배가될 수 있다.

본 논문에서는 통합된 EMR 확진 데이터로부터 22개의 다빈도 진단을 예측하기 위해 기준검사명칭 64개, 증상 80개, 기본정보 3개, 총 147개 예측 인자를 사용하여 랜덤 포레스트 모델을 학습한다.

III. 실험 결과 및 토의

3.1 반려동물이 갖는 질환

실험 결과를 소개하기에 앞서 생후 년 수별 반려동물이 갖는 복수의 질환에 대해 먼저 소개한다. 반려동물이 고령일수록 복수의 질환을 갖고 있는 반려동물의 경우가 다수 존재한다. 따라서 본 논문에서는 반려동물의 생후 년 수별에 따른 결과 또한 소개한다.

3.2 학습 환경 & 데이터 세트

본 논문에서는 사용한 데이터 세트는 8곳의 동물병원으로부터 수집하고 전처리 과정을 마친 6,481개의 EMR 데이터이다. 이 중에서 80% (5,185개)는 학습데이터로, 20% (1,296개)는 검증에 사용하였다. 학습데이터 5,185개에는 22개의 진단이 모두 포함되도록 하여 학습을 진행하였다. 랜덤 포레스트 학습 환경은 표 2와 같다.

표 2. 랜덤 포레스트 학습 환경
Table 2. Random forest learning environment

OS	Ubuntu 16.04.5 LTS
Programming language	Python.3.5.6
Python library	Sklearn 0.20.0 Numpy 1.15.2 Pandas 0.23.4 openpyxl 2.5.6

3.3 실험 결과

EMR 확진 데이터의 통합 과정을 통해 새롭게 생성된 포맷의 데이터를 기반으로 랜덤 포레스트 모델을 학습하였다. 모델의 입력은 검사를 진행한 반려동물의 기본정보(품종, 나이, 성별)와 수의사가 확인한 증상을 마지막으로 시행한 64개의 검사의 결과를 사용한다. 모델의 출력은 진단 코드로 출력이 되며 상위 5개 예상 진단명과 코드로 출력한다. 출력되는 코드는 그림 4와 같다.

```
Diagnosis prediction using RF
Result format = [Diagnosis code, probability score]

{'pred': [[2092, 14.7], [5, 9.9], [15, 9.5], [2095, 7.8], [2, 6.2]]}
```

그림 4. 진단 예측 결과
Fig. 4. Prediction result

표 3. 반려동물(개, 고양이)의 진단 분류 예측 정확도 성능
Table 3. Diagnostic classification prediction accuracy performance of pets (dogs, cats)

Class	TOP-3	TOP-4	TOP-5
dogs+cats	68.4%	73.2%	86%
dogs	66.6%	74.2%	82.9%
cats	68.1%	76.9%	85.5%

표 3은 반려동물(개와 고양이)에 대한 진단 분류 예측 성능을 보여주며 개와 고양이에서의 진단 예측 성능의 차이는 크지 않았다. 반려동물의 진단 예측 성능 중 3, 4개의 상위 클래스에 대한 분류 성능은 각각 68.4%, 73.2%로 임상적으로 수의사의 진단을 보조하는데 큰 도움이 되기 어려운 것으로 나타났다 약 30%에 해당되는 케이스에 대해 상위 3~4개의 예측진단에 실제 진단이 포함되지 않기 때문이다. 이는 기본 혈액 검사만으로 진단이 가능한 케이스의 비율의 상한이 있기 때문으로 판단된다. 따라서 분류 성능 향상을 위해서는 데이터의 추가 확보 이외에 기본 혈액 검사 이외에 소변 검사 등의 검사 결과 항목을 추가하는 등의 시도가 필요하다.

이외에도 진단 분류 성능 저하의 원인 중 하나는 고령인 반려동물이 갖는 복수질환임을 확인하였다. 생후 1년 미만인 경우, 생후 1년 이상 5년 미만인 경우, 생후 5년 이상인 경우에 대하여 분류하여 분석하였다. 각각의 경우에 주 질환이 아닌 복수의 질환을 갖고 있는 경우는 표 4와 같다.

표 4. 생후 복수 질환을 갖는 경우의 데이터 비율
Table 4. Rate of data for asymptomatic disease after birth

Class	Number of data(multiple disase case / total)
Less than 1 year after birth	129 / 885 (약 14.5%)
1 to 5 years after birth	299 / 1,631 (약 18.3%)
Over 5 years after birth	1,372 / 3,965 (약 34.6%)

표 5. 반려동물의 생후 년 수에 대한 진단 분류 예측 성능 표
Table 5. Predictive performance of diagnostic classification on the age of pets

Class	Top 3	Top 4	Top 5
Less than 1 year after birth	73.3%	86.7%	91.1%
1 to 5 years after birth	64.6%	73.2%	84.1%
Over 5 years after birth	49.7%	62.2%	72.3%

표 4에서 보듯이, 실제 전처리된 데이터에서 생후 1년 미만의 반려동물에서는 2가지 이상의 복합성 질환의 경우가 14.5%이지만, 생후 1년 이상의 고령 반려동물에서 18.3%, 생후 5년 이상의 고령 반려동물에서는 34.6%의 비율을 차지한다.

표 5에서 볼 수 있듯이, 반려동물의 나이가 고령에 가까울수록 분류 성능이 떨어지는데, 이는 나이에 따른 복수질환 보유 비율과 높은 상관성을 가짐을 나타낸다.

Ⅳ. 결 론

본 논문에서는 반려동물의 EMR 확진 데이터를 전처리 과정을 통해 통합하였으며, 총 6,481개의 EMR 확진 데이터로부터 기본정보(3), 증상(80), 검사항목(64), 진단명(22)의 총 169개의 통합된 형태로 만들었다는 것에도 의미를 두고 있다. 이 통합 방법을 기반으로 일본의 ANICOM과 같이 통합된 정보를 제공할 수 있을 것으로 기대되며, 반려동물의 진단결과를 기반으로 학습된 머신러닝 모델을 수의사의 진단을 보조하는데 적용해볼 수 있을 것이다. 여기에 더해 반려동물 EMR 확진 데이터로부터 다빈도 질환을 포함한 대분류 레벨의 진단 범주를 예측하는 머신러닝 모델 개발 과정을 소개하였다. 사용한 예측 인자는 총 169개의 변수로서 반려동물의 기본정보, 증상 정보, 기본 혈액 검사항목과 결과값, 진단명을 포함하며, 높은 확률을 갖는 상위 5개의 진단을 제시한다. 이러한 이유는 반려동물이 고령일수록 복수의 질환을 갖는 비율이 높아 상위 5개의 진단을 보여주는 것이 수의사의 진단을 보조하는데 의미가 있다고 볼 수 있기 때문이다.

향후 임상적으로 분류 성능을 개선하기 위해서는 추가적인 데이터 사용 이외에 검사항목 추가 등이 필요해 보이며, 이에 대해 연구를 진행 중에 있다.

References

[1] J. Y. Park, "Problems with the current companion animal health system and improvement plans", Journal of Environmental Law and Policy, Vol.

19, pp. 99-130, Sep. 2017.

[2] A. Awaysheh, J. Wilcke, F Elvinger, L Rees., W Fan, and K. L. Zimmerman, "Evaluation of supervised machine-learning algorithms to distinguish between inflammatory bowel disease and alimentary lymphoma in cats", Journal of Veterinary Diagnostic Investigation, Vol. 28, No. 6, pp. 679-687, Oct. 2016.

[3] WIZARD i Co., Ltd. SYSTEM AND METHOD FOR DIAGNOSING COMPANION ANIMAL. Patent No. 10-2018-0045086, 2016.

[4] L. Breiman, "Random forests", Machine learning Vol. 45, No. 1, pp. 5-32, Oct. 2001.

[5] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and Regression Trees", Monterey, CA: Wadsworth, 1984.

[6] ANICOM, List of injury and illness names, https://www.anicom-sompo.co.jp/medical/karte/list_di/ [accessed: Dec. 20,2019]

[7] D. S. Siroky, "Navigating random forests and related advances in algorithmic modeling", Statistics Surveys, Vol. 3, pp. 147-163, Nov. 2009.

[8] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization", Machine Learning, Vol. 40, pp. 139-157, Aug. 2000.

[9] L. Breiman, "Bagging Predictors", Machine Learning, Vol. 24, pp.123-40, Aug. 1996.

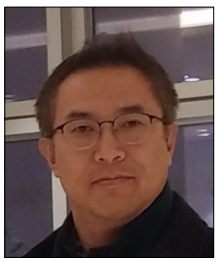
[10] L. Breiman, "Out-of-Bag Estimation", ftp://ftp.stat.berkeley.edu/pub/users/breiman/OOBestimation.ps, 1996.

저자소개

최 우 용 (Wooyong Choi)

2013년 2월 : 수원대학교 물리학과 학사
2013년 3월 ~ 2017 2월 : KIAS Insilico Protein Scieonce 연구원
2018년 6월 ~ 현재 : (주)메디사피엔스 선임연구원
관심분야 : AI/BI

박 래 정 (Lae-Jeong Park)



1991년 2월 : 서울대학교 전기공학과 학사
1997년 8월 : KAIST 전기 및 전자공학과 박사
2000년 3월 ~ 현재 : 강릉원주대학교 전자공학과 교수
2018년 1월 ~ 현재 : (주)메디사피엔스 연구소장
관심분야 : 데이터 과학, 기계학습, 기계학습의 의료/바이오 응용.