

데이터 마이닝을 이용한 제주 양식 넙치(*Paralichthys olivaceus*)의 스쿠티카증 발생 패턴 분석

김해란 · 정성주¹ · 김성현² · 박정선³ · 정희택³ · 한순희^{3*}

전남대학교 디지털컨버전스협동과정, ¹전남대학교 수산생명의학과, ²수산질병진단전문연구소(피쉬케어), ³전남대학교 멀티미디어전공

Data Mining for Scuticociliatosis Outbreak Patterns in Cultured Olive Flounder *Paralichthys olivaceus* in Jeju, Korea

Hae-Ran Kim, Sung-Ju Jung¹, Sung-Hyun Kim², Jeong-Seon Park³, Hee-Taek Ceong³ and Soon-Hee Han^{3*}

Department of Digital Convergence, Chonnam National University, Yeosu 59626, Korea

¹Department of Aqualife Medicine, Chonnam National University, Yeosu 59626, Korea

²Fishcare Laboratory, Jeju 63629, Korea

³Department of Multimedia, Chonnam National University, Yeosu 59626, Korea

In the aquaculture industry, few studies are analyzing big data for intrinsic meaning. Fishcare Laboratory (www.fishcare.kr) diagnostic data from 2016-2018 was analyzed for scuticociliatosis (caused by *Miamiensis avidus*) outbreak patterns in cultured olive flounder *Paralichthys olivaceus* in Jeju, Korea. The scuticociliatosis monthly occurrence ratio is reported in the summary table after preparing and filtering the basic dataset model. Nonparametric test results suggest differences in the water temperature, body length, and weight between groups with and without scuticociliatosis. Data distribution visualization revealed that shorter body length and lighter weight increased the occurrence of scuticociliatosis. The association rule mining technique was applied to determine the primary clinical signs of mixed scuticociliatosis and bacterial infections. Venn diagrams were used to report clinical signs and suggest commonalities. These results may help diagnose and treat fish and provide a decision-making reference.

Keywords: Olive flounder, *Paralichthys olivaceus*, Scuticociliatosis, Association rule, Data mining

서론

제주도 내 넙치 양식장에서는 다양한 질병이 발생하고 있고, 최근에는 수온과 상관없이 연중 다양한 병원체의 혼합 감염 형태의 질병이 발생하고 있다. 신종 질병에 대한 대비책 마련과 신속한 진단 치료제 개발을 위해서는 국내외 주요 병원체의 특성, 유전정보, 역학적 정보 등 관련 정보가 체계적으로 수집되어 빅데이터로 관리될 필요성이 대두되고 있다(Kim et al., 2006; Cho et al., 2008; Kim et al., 2010; Cho et al., 2019). 보건 의료 분야에서는 의료 빅데이터에 근거하여 연관규칙 마이닝 기법을 이용한 질병과 증상과의 관계 추론 연구(Tai and Chiu, 2009; Choi et al., 2010; Park et al., 2016), 질병 간의 연관성(Lee et al., 2010; Sabthami et al., 2016; Kim, 2018)을 파악하는 연구

등이 활발히 진행되고 있지만 수산 양식 분야에서는 데이터 기반의 연구가 미비한 실정이다. 수산질병진단전문연구소 피쉬케어(www.fishcare.kr)는 제주지역 양식 어류 진단을 위해 사육수온, 어류의 체장, 체중, 임상증상, 소견, 병원체 진단, 현미경 관찰 정보들을 상세히 기록하여 2015년부터 축적해 오고 있었다. 본 연구의 시발점은 이 축적된 질병진단 데이터로부터 내재되어 있는 의미 있는 패턴을 발견하고자 하는 것이었다. 제주지역 양식 넙치의 여러 질병 중에서 스쿠티카증이 가장 문제되는 질병 중 하나로 보고되고 있다(Jin et al., 2003; Jin et al., 2015). 원인체인 스쿠티카충은 현미경 관찰로 바로 확인이 가능하며 질병의 증상이 다른 세균과 바이러스성 질병과 확연히 구분되는 장점이 있어 새로운 분석법을 적용하고 분석한 결과가 타당한지를 판단하기에 적절한 병원체로 생각되어 스쿠티카증을 질

*Corresponding author: Tel: +82. 61. 659. 7432 Fax: +82. 61. 659. 7449

E-mail address: shhan@chonnam.ac.kr



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

<https://doi.org/10.5657/KFAS.2020.0740>

Korean J Fish Aquat Sci 53(5), 740-751, October 2020

Received 18 June 2020; Revised 5 August 2020; Accepted 25 September 2020

저자 직위: 김해란(대학원생), 정성주(교수), 김성현(대표이사), 박정선(교수), 정희택(교수), 한순희(교수)

병 패턴 분석 대상으로 선정하였다.

본 연구는 2016-2018년 동안 수천 마리의 넙치에 대한 환경·생물학적 인자(사육수온, 체장, 체중), 100가지의 임상증상, 진단정보등을 이용하여 스쿠티카증 질병 발생 패턴을 발견하고 특히, 세균성 질병과 스쿠티카증 혼합 감염 시 주요 임상증상 및 스쿠티카증의 공통 임상증상을 밝히고자 하였다.

재료 및 방법

분석 데이터

본 연구에서는 수산질병진단전문연구소 피쉬케어(Jeju, Korea)에서 워드 형태로 해당 기간(2015년 10월-2018년 12월) 동안 축적한 자료를 이용하였다. 이는 제주지역 다수의 양식장을 방문하여 정기 질병모니터링을 실시한 결과로 대부분은 대량 폐사가 발생하지 않는 상황에서 외관상 건강상태가 상대적으로 나빠 보이는 개체를 골라서 검사하였으며 각 어장 당 1회 5 마리 검사를 기본으로 하였다. 넙치, 터봇 및 우럭의 진단결과가 있었으나 본 연구에서는 넙치 자료만을 분석에 사용하였다.

이 자료는 워드 문서인 비정형화된 텍스트 데이터로 구성되어 있어 분석 가능한 형태로 정제할 필요가 있었다. 이를 위해 워드 문서를 정형화된 엑셀 형태의 데이터로 구성하는 파서 프로그램을 개발하였고, 개발한 소프트웨어를 이용하여 212개 컬럼 속성을 지닌 7,054개 진단 어류 데이터를 생성하였다. 212개 속성을 크게 분류하면 Table 1과 같이 5개 영역으로 그룹화 할 수 있다. 그룹화된 5개의 항목은 일반항목(양식장명, 일자, 시간, 사육수온, 체장, 체중, 어종, 수조정보, 사전관측기록, 사후관측 기록), 어류 내부 증상, 어류 외부 증상, 현미경 관찰(기생충), 병원체 진단(세균)으로 구성되어 있다. 분석의 효율을 위해 컬럼 위주의 데이터 전처리 과정을 수행하여 컬럼 속성을 212개에서 126개로 줄였다.

전처리 과정은 세균성질병인 병원체 진단 속성에 양(다량, 중량, 소량)으로 구분된 것을 통합하는 과정과 부위별 관측으로 구분된 기생충 정보를 통합하여 12개 기생충 정보로 정리하였다. 특히, 스쿠티카증은 6개 부위(체표, 아가미, 복수, 장, 뇌, 눈)에서 관찰 되었는데 부위에 상관없이 스쿠티카증이 검출 되었

으면 스쿠티카증이 있음으로 처리하였다. 또한 양식장 사육수온, 체장, 체중 등 값이 2개 이상 들어 있는 경우는 평균값으로 대체하였다.

병원체 진단 정보로 세균성 질병은 4종(*Vibrio* spp., *Streptococcus parauberis*, *Edwardsiella piscicida*, *Pseudomonas* spp.) 이고 기생충 정보는 스쿠티카증을 포함한 12종이다. 세균성 질병 진단은 GSPs (gene specific primers)를 사용한 직접 콜로니 PCR (polymerase chain reaction)법이 사용되었다(Sebastiao et al., 2015; Han et al., 2017). 본 연구에서는 기생충성 질병인 스쿠티카증(Scuticociliatosis)에 중점을 두고 연구하였다. 따라서, 스쿠티카증 진단을 위해 필요한 현미경 관찰 정보가 기록되지 않은 2015년 데이터는 분석에 사용하지 않았다.

본 연구의 분석을 위해 총 7,054마리의 어류 진단 데이터로부터 스쿠티카증 진단이 있는 넙치 데이터이면서 결측치가 없는 자료로 제한하니 총 5,747마리의 넙치 데이터가 추출되었다. 분석을 위한 컬럼 속성들은 일반항목 4개(일자, 수온, 체장, 체중), 임상증상 100개, 기생충 속성 18개(스쿠티카 부위별 관측 6개, 기생충 진단 12개), 세균성 질병 속성 4종 등으로 총 125개로 구성하였다. 따라서 본 연구의 분석에 사용된 기본 데이터모형은 5,747개 row, 125개 column 속성을 가졌다.

분석 방법

자료 분석을 위해 데이터 마이닝에 유용한 프로그래밍 언어 R을 이용하였고, 시각화, 연관규칙을 위해 R의 패키지(arules, arulesViz, ggplot2, ggbeeswarm, dplyr, reshape2 등)를 사용하였으며 구체적인 분석 내역은 다음과 같다.

- 1) 스쿠티카증 월별 진단 현황 및 관측 부위를 표로 제시하였다.
- 2) 사육수온, 체장, 체중의 기초 통계 내역과 분포를 제시하였다. 또한 스쿠티카증 유무 두 그룹 간 사육수온, 체장, 체중의 분포 차이 검정을 위해 윌콕슨 순위합 검정(Wilcoxon rank sum test)을 실시하였고 각각의 분포 형태에 사분위수 값 및 평균값을 표시하여 시각화하였다.
- 3) 스쿠티카증과 세균성 질병 4종의 단독 또는 혼합 감염 비율을 그래프로 제시하였고 기본 데이터모형(n=5,747), 단독 스

Table 1. Data Columns

Section	Preprocessing		Items (=Columns)
	Before	After	
General Info.	10	10	Fishfarm name, Date, species, W.temp, Weight, Length,....
Internal signs	48	48	Liver hemorrhages, Kidney hypertrophy, ...
External Signs	52	52	Hernia, Fin hemorrhages, Abdominal distension, Ulcer, .
Parasites	90	12	Scuticociliate, <i>Trichodina</i> ,
Bacteria	12	4	<i>Vibrio</i> spp., <i>Streptococcus parauberis</i> , <i>Edwardsiella piscicida</i> , <i>Pseudomonas</i> spp.
Total Columns	212	126	

쿠티카 데이터모형(n=614)의 상위 30개 임상증상을 표로 제시하였다.

4) 스키투티카증 단독 및 스키투티카증과 세균성 질병 4종과의 혼합 감염에서 임상증상과 질병 사이의 연관성을 분석하기 위해 선협(apriori) 알고리즘 기반의 연관규칙마이닝 기법을 적용하였다. 생성된 연관규칙(association rule) 중 스키투티카증과 세균성질병의 2종 혼합 감염에 대한 연관규칙만 표로 제시하였고, 연관규칙의 빈발항목 집합 즉 자주 발생하는 임상증상 항목들로부터 스키투티카증의 공통 임상증상을 밝히고자 벤다이어그램을 이용하였다.

윌콕슨 순위합 검정은 데이터가 정규성 가정을 만족하지 않거나 혹은 분포 형태를 모르는 경우의 두 그룹의 중심 차이에 대한 비모수 검정(nonparametric test)이다. 본 연구에서는 사육 수온, 체장, 체중 분포가 정규 분포 가정이 충족되지 않아 윌콕슨 순위합 검정을 사용했다.

연관규칙은 데이터로부터 숨겨진 패턴이나 규칙을 탐색하고 찾아내는 대표적인 방법으로 특히 항목 또는 변수(컬럼) 간의 의미 있는 관계를 찾는 것으로 얼마나 자주 함께 발생하는지, 또는 서로 얼마나 연관되어 있는지를 제시해 준다. 즉, 연관규칙은 함께 발생하는 것들의 관계를 알 수 있게 해 준다(Rfriend,

Table 2. Summary of monthly occurrence of Scuticociliatosis in cultured olive flounder *Paralichthys olivaceus* in Jeju

Year	Mon.	P. oliv. (n)	Scut. (n)	Scut. (%)	Microscope Observations (unit : n)					
					Brain	Eye	Intestine	Ascites	Gill	Skin
2016	9	36	2	6	0	0	0	0	0	2
2016	10	57	13	23	0	0	0	0	0	13
2016	11	100	3	3	0	0	0	0	0	3
2016	12	108	14	13	0	0	0	0	0	14
2017	1	195	30	15	0	0	0	0	0	30
2017	2	145	31	21	0	0	0	0	0	31
2017	3	136	25	18	0	0	0	0	0	25
2017	4	137	37	27	0	0	0	0	0	37
2017	5	151	26	17	0	0	0	0	0	26
2017	6	147	51	35	0	0	0	0	0	51
2017	7	177	43	24	0	0	0	0	0	43
2017	8	142	26	18	0	0	0	0	0	26
2017	9	165	31	19	0	0	0	0	13	23
2017	10	142	23	16	0	0	0	0	19	14
2017	11	197	46	23	0	0	0	0	32	37
2017	12	198	41	21	0	0	0	0	31	31
2018	1	201	81	40	0	0	0	0	65	60
2018	2	259	73	28	0	0	2	0	52	60
2018	3	243	68	28	0	0	0	0	50	56
2018	4	255	75	29	0	0	0	0	60	48
2018	5	352	104	30	0	0	0	0	76	67
2018	6	311	81	26	0	0	0	0	62	52
2018	7	295	50	17	0	0	0	0	35	35
2018	8	298	48	16	0	0	0	0	30	32
2018	9	250	59	24	0	0	0	0	42	40
2018	10	397	100	25	0	1	0	0	64	85
2018	11	326	58	18	2	0	1	1	31	51
2018	12	327	60	18	1	0	0	1	31	53
Total		5,747	1,299		3	1	3	2	693	1,045

P.oliv, (n) means the number of *Paralichthys olivaceus* diagnosed per month; Scut, (n) means the number of *Paralichthys olivaceus* infected with Scuticociliatosis per month.

2015; Kang, 2019).

결과 및 고찰

스퀴티카증 월별 진단 현황

기본 데이터 모형(n=5,747)에 근거한 스퀴티카증 발생 현황 정보는 Table 2와 같다. 총 5,747마리의 양식 넙치 중 1,299마리 넙치가 스퀴티카증으로 진단되었다. 월별 진단된 넙치 수, 스퀴티카증이 발생한 넙치수와 비율, 스퀴티카증 관측 부위 건수를 Table 2에 제시하였다. 2017년 9월 이후부터 체표 이외의 부위인 아가미에서도 스퀴티카증의 진단이 실시되어 체표와 아가미에서 중복으로 스퀴티카증이 다수 존재하였다. 2018년에는 체표, 아가미 이외의 4개 부위(장, 복수, 뇌, 눈)에서 9건의 스퀴티카증이 관측되었다. 특히, 복수, 장, 뇌 부위에 스퀴티카증이 있으면 아가미, 체표에도 스퀴티카증이 존재하였다. 2017년과 2018년 기준으로 보면 대부분의 스퀴티카증의 월 발생 비율은 15-30% 사이로 연중 발생하고 있는 것으로 나타났다.

사육수온/체장/체중 분포

기본 데이터모형(n=5,747)에 대한 사육수온, 체장, 체중에 대한 요약 값은 Table 3과 같다. 사육 수온은 20°C 이하 값, 체장은 36 cm 이하 값, 체중은 535 g 이하 값이 각각 전체의 75%를 차지하였다.

Q1, Q2 (median), Q3 값은 전체 데이터를 순위별로 4등분하는 위치의 수이고 중위수(median)와 평균의 차이가 큰 것은 체중으로 나타났고 체중의 표준편차(standard deviation) 또한 매우 컸다. 즉 체중 데이터는 평균(389 g)에서 많이 떨어져 있음을 나타내고 체중의 최대값은 4,000 g, 최소값은 0.85 g으로 범위 또한 컸다.

Fig. 1에서 사육 수온은 16-17°C 구간의 빈도가 높았는데 이는 지하 해수를 사용하는 양식장이 많아 겨울철에도 수온 13°C 이상이 유지되며 여름철에도 25°C 이상 상승하는 날이 많지 않기 때문으로 보였다. 체장은 30 cm 전후 구간에서 빈도가 가장 많았다(Fig. 1a and 1b). 체중은 1000 g이 넘는 경우는 매우 적고 500 g 이하가 가장 큰 빈도를 차지하였다. 체중을 상대도수 히스토그램으로 상세히 살펴보면 기본 넙치 데이터(n=5,747)에서 100 g 이하(18%), 100-200 g (19%), 200-300 g (16%), 300-400 g (11%), 400-500 g (8%), 500-600 g (6%)를 차지하였다(Fig. 1c and 1d). 따라서 200 g 이하 치어가 37%이고 나머

지 63%는 중성어가 차지하였다. 어체의 크기가 작을수록 질병에 의한 증상의 발현과 폐사의 발생이 많아지므로 작은 크기의 넙치를 진단에 사용하는 빈도가 높아 535 g 이하가 75%를 차지하게 된 것으로 보였다. 사육수온, 체장, 체중의 분포가 정규분포 형태가 아니고 한쪽으로 치우쳐진 경향이 있어 그룹 간 분포의 차이를 비교하기 위해서는 평균(mean)보다 중앙값(median)이 더 적절한 것으로 판단되었다.

스퀴티카증 유무 그룹간 사육수온/체장/체중 분포

스퀴티카증 진단을 받은 그룹과 그렇지 않은 그룹 간 환경·생물학적 인자(사육수온, 체장, 체중)의 분포 차이 검정을 위해 윌콕슨 순위합 검정을 수행한 결과 두 그룹간 환경·생물학적 인자들이 모두 유의한 차이를 보였다(P<0.01). 두 집단 간 중위수 값을 비교하면 스퀴티카증 진단을 받은 그룹의 사육수온, 체장, 체중의 값이 스퀴티카증 진단을 받지 않은 그룹에 비해 모두 작았다. P-값이 모두 매우 작아 1% 유의수준에서 두 집단 간 사육수온, 체장, 체중의 분포 차이가 있음을 통계적으로 말할 수 있다.

Fig. 2는 스퀴티카증 진단을 받은 그룹과 받지 않은 두 그룹 간 환경인자(사육수온)의 데이터 분포 및 사분위수를 표시하였다. 오른쪽이 스퀴티카증 진단을 받은 그룹의 분포이고 왼쪽은 스퀴티카증 진단을 받지 않은 그룹의 분포이다. I자 모양 막대는 아래쪽은 Q1, 윗쪽은 Q3를 표시하고 막대 안의 네모는 Q2 중위수(median) 값이며 동그라미는 평균(mean)을 나타냈다. 오른쪽의 사육수온(중위수, 17.40°C)이 왼쪽의 사육수온(중위수, 18.05°C)보다 더 낮았다. 또한 Q1, Q3 값도 오른쪽이 더 낮았고 편차도 오른쪽이 더 작았다.

Fig. 3은 스퀴티카증 유무 그룹 간 체장(length)에 관한 분포를 나타내었다. 오른쪽(스퀴티카증 진단을 받은 그룹)은 30 cm 이하가 75%이지만 왼쪽(스퀴티카증 진단을 받지 않은 그룹)은 37 cm 이하가 75% 차지하고 있었다. 또한 스퀴티카증 진단을 받은 그룹이 그렇지 않은 그룹보다 체장의 중위수 값은 5 cm 작았고, Q1, Q3도 각각 약 4 cm, 7 cm 작았다. 가장 많은 크기의 빈도 값인 최빈값(mode)은 오른쪽은 23 cm, 왼쪽은 30 cm였다. 이를 통해 스퀴티카증은 크기가 작을수록 통계적으로 유의하게 스퀴티카증에 더 잘 걸린다고 보여졌다.

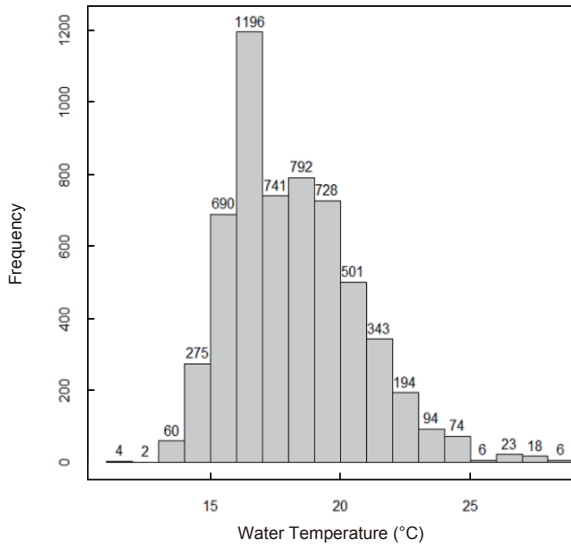
Fig. 4는 스퀴티카증 유무 그룹 간 체중(weight)에 관한 분포를 나타내었다. 오른쪽(스퀴티카증 진단을 받은 그룹)은 335 g 이하가 75% 차지하고 왼쪽(스퀴티카증 진단을 받지 않은 그

Table 3. Distribution summary

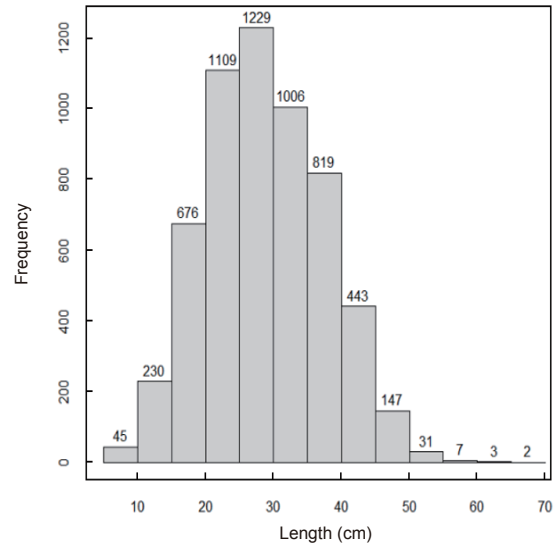
Variables	Min.	Q1	Q2	Mean	Q3	Max.	Mode	SD
W.Temp (°C)	11	16.5	18	18.4	20	28.4	17	2.46
Length (cm)	5	23	29	29.3	36	69	30	8.78
Weight (g)	0.85	140	280	389	535	4000	150	359

Q1, 25th percentile; Q2, 50th percentile; Q3, 75th percentile.

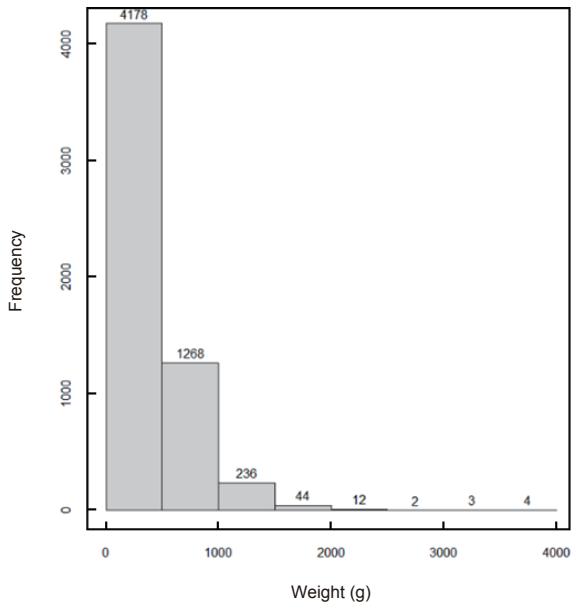
(a) Histogram of W.Temperature



(b) Histogram of Length



(c) Histogram of Weight



(d) Histogram of Weight with a smaller interval (100)

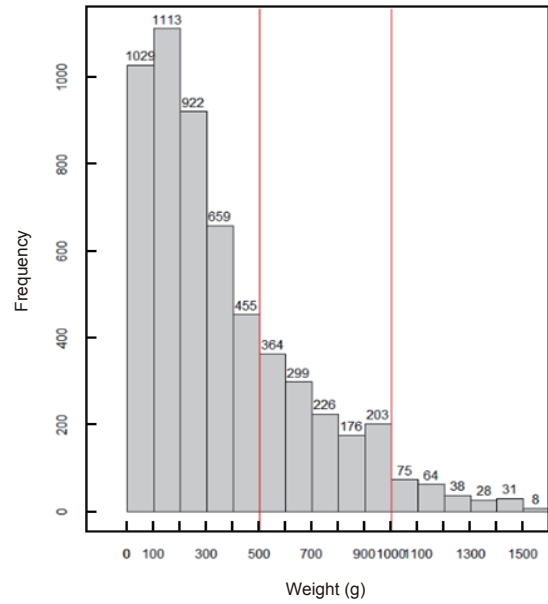


Fig. 1. Histogram of culture water temperature (a), fish length (b), weight (c) and weight with a smaller interval (100) (d) of olive flounder *Paralichthys olivaceus* used for diagnosis.

톱)은 600 g 이하가 75% 차지하고 있다. Q1, Q2 (median) 값도 오른쪽이 왼쪽에 비해 약 2배 정도 작게 나타났다. 따라서 무게가 작을수록 통계적으로 유의하게 스킨카증에 더 잘 걸린다고 할 수 있다.

스킨카증은 15-25°C의 범위에서 16-21시간마다 2분열로 증식하며 수온이 높을수록 이분열의 속도가 빠르므로 고수온기에 큰 피해를 입힐 가능성이 크다(Bae et al., 2009). Jin et al. (2007)의 연구에서도 2000년부터 2004년까지의 스킨카

증 발병 상황은 여름철이 더 심한 것으로 보고하고 있다. 그러나 본 연구에서는 수온이 낮은 경우에 높은 진단율을 보이는 경향이 있는데 그 원인으로 예전에는 봄과 가을에 7-10 cm 크기의 치어를 입식 하였으나 최근에는 20 cm 정도까지 키운 치어를 입식 하여 폐사가 많이 발생하는 여름철에 질병에 저항성을 가지는 500 g 이상의 크기로 성장시키기 때문으로 생각되었다. 즉, 체장과 체중의 분석결과에서 보이는 바와 같이 어체의 크기가 작을수록 스킨카증의 진단율은 높아지는데 크기가 작은 개체

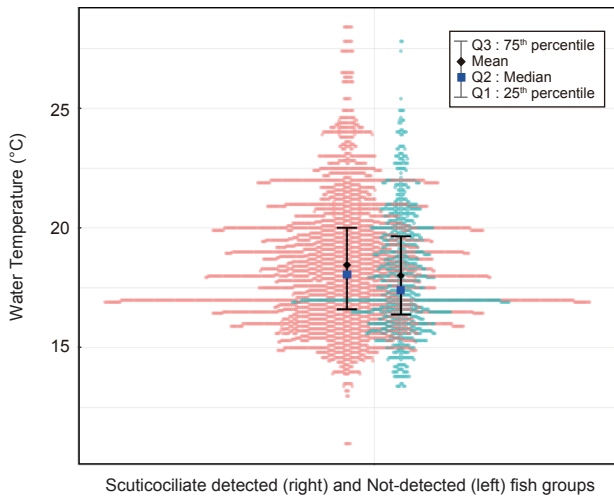


Fig. 2. Culture water temperature distribution of scuticociliate detected (right) and not-detected (left) fish groups. One spot means one fish.

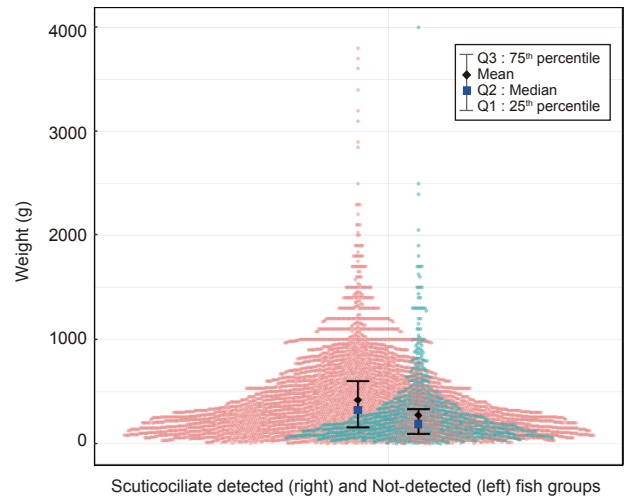


Fig. 4. Fish weight distribution of scuticociliate detected (right) and not-detected (left) fish groups. One spot means one fish.

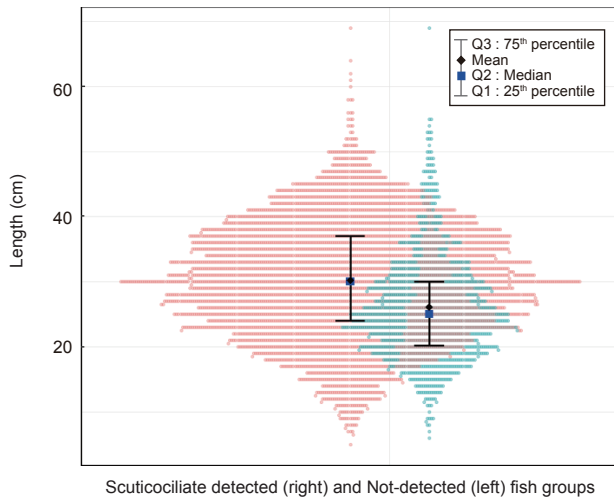


Fig. 3. Fish length distribution of scuticociliate detected (right) and not-detected (left) fish groups. One spot means one fish.

가 봄과 가을철에 많기 때문에 이 시기에 스키테카증 진단율이 높아지는 것으로 판단하였다.

스쿠테카증 단독/혼합 질병 패턴

질병 진단 비율

스쿠테카증, 세균성 질병 4종(*Vibrio* spp., *Streptococcus parauberis*, *Edwardsiella piscicida*와 *Pseudomonas* spp.) 그리고 이들의 혼합 감염의 진단 비율은 Fig. 5에 나타내었다. None의 의미는 스키테카증과 세균성 질병 4종을 포함한 총 5종의 질병이 모두 검출되지 않았다는 의미로 33.18%가 이에 해당되었

다. 단, 진단 데이터에 없는 다른 질병에 감염되어 있을 가능성은 배제할 수 없다. 또한 스키테카증 단독은 총 5종 질병 중 스키테카증만 진단된 경우로 이는 총 넙치 진단 데이터(n=5,747) 중 10.68% (n=614)가 해당되었다. 세균성 질병 중 단독 감염인 경우는 *Vibrio* spp. (Vi, 9.43%), *Streptococcus parauberis* (St, 8.90%), *Edwardsiella piscicida* (Ed, 4.73%), *Pseudomonas* spp. (Ps, 1.18%) 이고 스키테카증과 혼합 감염 발생 비율은 스키테카증과 *Vibrio* spp. (Vi_Scu, 3.86%), 스키테카증과 *Streptococcus parauberis* (St_Scu, 2.61%), 스키테카증과 *Edwardsiella piscicida* (Ed_Scu, 0.61%), 스키테카증과 *Pseudomonas* spp. (Ps_Scu, 0.30%) 순으로 낮았다(Fig. 5).

임상증상 비율

기본 넙치 데이터(n=5,747, 모든 개체의 임상증상)와 단독 스키테카 데이터(n=614, 스키테카증 단독 감염 개체의 임상증상) 증상의 상위 30개 항목과 빈도 비율은 Table 4와 같이 나타났다. 두 데이터 셋에서 발생하는 상위 30개 항목은 한 개 항목을 제외하고는 동일하고 빈도 비율만 달랐다. 단독 스키테카 데이터(n=614)의 증상 중 10번째 체표궤양(25.1%), 13번째 주둥이궤양(20.8%), 15번째 지느러미궤양(20.5%)이 기본 데이터(n=5,747) 증상 비율에 비해 약 2-3배 높은 특징이 있었다. 단독 스키테카 감염으로 진단된 넙치는 스키테카는 검출이 되지만 4종의 세균이 검출되지는 않았음을 나타내었다. 본 분석에서 단독 스키테카증으로 진단된 다수의 개체가 간의 출혈, 신장비대, 비장비대와 같은 스키테카증 특이 증상이 아닌 임상증상을 보였는데 바이러스성출혈성패혈증과 같은 바이러스성 질병의 관련성에 대한 검사가 필요할 것으로 생각된다.

기본 넙치데이터(n=5,747)에서는 총 100개의 임상증상 중 절반이 1% 이하로 나타났고 단독 스키테카(n=614)에서는 27개

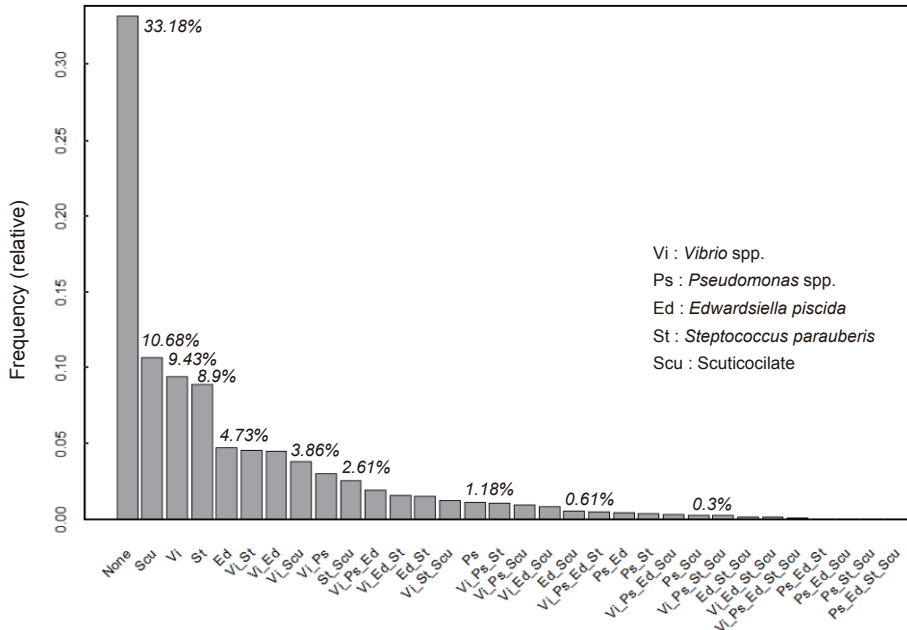


Fig. 5. Disease frequency ratio.

증상은 전혀 관측되지 않았고 26개 증상은 1% 이하로 나타났다. 따라서 연관규칙 생성 시 지지도(support) 값을 매우 낮게 지정하였다.

증상과 질병 연관규칙

스쿠티카충과 세균성 질병의 혼합 감염의 증상을 알고자 임계값(지지도, 0.05%; 신뢰도, 60%) 하에서 연관규칙을 생성하였다. 스쿠티카충과 2종 혼합 감염은 지정된 임계값 하에서 스쿠티카충과 *Vibrio* spp. (Vi+Scu), 스쿠티카충과 *Streptococcus parauberis* (St+Scu), 스쿠티카충과 *Edwardsiella piscicida* (Ed+Scu) 혼합 감염만 연관규칙이 각각 31개, 21개, 14개 생성되었다. 그리고 단독 스쿠티카충(Scu)은 지정 임계값(지지도, 0.05%; 신뢰도, 60%) 하에서 1,194개의 많은 연관규칙이 생성되어 임계값(지지도, 0.09%; 신뢰도, 70%)을 조정하여 재생성하니 40개의 연관규칙이 만들어졌다. 먼저, 지정된 임계값(지지도, 0.05%; 신뢰도, 60%) 하에 생성된 연관규칙 중 스쿠티카충과 세균성 질병과의 2종 혼합 감염 3개에 대해서 신뢰도 70% 이상인 연관규칙들만 Table 5에 제시하였다.

연관규칙 R은 R: X→Y, where X, Y ⊆ I로 표시하며, 'X이면 Y이다'라고 해석한다. 여기서 X, Y는 항목 집합(I={i₁, i₂, ..., i_p})의 부분 집합이며 서로 공통 원소가 없는 항목들의 집합으로 정의한다. X를 선행 항목 집합, Y를 후행 항목 집합이라 한다. 여러 연관 규칙 중 강하고 좋은 규칙을 찾는 기준 지표로 지지도, 신뢰도, 향상도 등이 있는데 지지도(Sup.)는 전체 경우의 수 중 항목 X, Y가 동시에 포함될 확률이고 신뢰도(Conf.)는 1에 가까울수록 관계가 더 의미 있고 유용한 규칙일 가능성이 높

음을 의미한다. 향상도(Lift) 값이 1보다 높을 때 양의 상관으로 항목 X, Y가 연관성이 있음을 나타낸다(Rfriend, 2015; Kang, 2019). Table 5에서 임상증상들은 X, 질병은 Y를 의미한다.

Table 5의 질병(Ed+Scu)에서 1번째 규칙의 의미는 복부팽만, 아가미비대, 탈장, 생식소출혈, 신장비대, 장괴사가 있으면 스쿠티카충과 *Edwardsiella piscicida*의 혼합 감염(Ed+Scu)임을 내포하고 있다. 이는 신뢰도(100%), 향상도(164.2)로 매우 높아 의미 있는 연관 규칙으로 볼 수 있다. Table 5의 질병(St+Scu)에서 1번째 규칙의 의미는 아가미괴사, 아가미창백, 아가미출혈, 주둥이 염증, 간비대가 있으면 스쿠티카충과 *Streptococcus parauberis*의 혼합 질병(St+Scu)일 확률이 신뢰도 75%로 나타났다지만 향상도(28.74)는 높아 해당 임상증상과 질병의 연관성이 있다 할 수 있다. Table 5의 질병(Vi+Scu)에서 1번째 규칙은 주둥이궤양, 지느러미출혈, 간조직변성, 간창백, 생식소내물침이 있으면 스쿠티카충과 *Vibrio* spp.의 혼합 질병(Vi+Scu)임을 신뢰도(100%), 향상도(25.89)로 제시했다. 위의 스쿠티카충과 혼합 감염의 규칙들은 스쿠티카충 단독 감염 증상이라기 보다는 세균성(*Vibrio* spp., *Edwardsiella piscicida*, *Streptococcus parauberis*) 감염증상에 가깝다. 이 의미는 세균성 감염이 일어나서 어제 건강이 약해진 경우 스쿠티카충 감염률이 높아진다고 볼 수 있고 혹은 이와 반대로 스쿠티카충 감염이 일어나면 세균성 감염이 같이 일어난다고 판단할 수도 있다.

마지막으로 스쿠티카충과 세균성 질병과의 2종 혼합 감염으로 생성된 연관규칙들과 단독 스쿠티카충(Scu)일 때 생성된 연관규칙의 빈발항목 집합들을 이용하여 스쿠티카충의 공통 임

상증상들이 무엇인지 밝히고자 하였다. 첫째로, 스키테카증과 *Edwardsiella piscicida*의 혼합 감염(Ed+Scu) 과 단독 스키테카증(Scu)의 공통된 증상을 Fig. 6 벤다이어그램 중앙에 표시했다. 이는 아가미창백, 아가미출혈, 지느러미출혈, 간조직변성, 신장비대, 비장비대, 장출혈로 나타났다. 두번째로, 스키테카증과 *Streptococcus parauberis*의 혼합 감염(St+Scu)과 단독 스키테카증(Scu)의 공통된 증상을 Fig. 7 벤다이어그램 중앙에 표시했다. 이는 간비대, 신장비대, 아가미뚜껑출혈, 아가미뚜껑염증, 간조직변성, 신장염증, 주둥이염증, 비장비대, 생식소내물참, 장출혈, 간출혈, 아가미출혈, 심강 물참, 주둥이궤양, 체표궤양, 아가미창백 등으로 나타났다. 세번째로, 스키테카증

과 *Vibrio spp.*의 혼합 감염(Vi+Scu) 과 단독 스키테카증(Scu)의 공통된 증상을 Fig. 8 벤다이어그램 중앙에 표시했다. 이는 간비대, 신장비대, 비장비대, 간조직변성, 아가미뚜껑염증, 주둥이염증, 장출혈, 장내물참, 생식소내물참, 아가미출혈, 지느러미출혈, 아가미창백, 간창백, 심강 물참, 지느러미궤양, 주둥이궤양, 체표궤양 등으로 나타났다. 최종적으로 스키테카증의 공통 임상증상들을 위 3개의 벤다이어그램 교집합으로부터 추출하니 아가미창백, 아가미출혈, 간조직변성, 신장비대, 비장비대, 장출혈로 나타났다. 위 증상들은 기본적으로는 스키테카증 증상이 아니다. 그러나 위와 관련된 내용이 교집합으로 나타난 것은 위 증상을 보이는 어체가 건강상 문제가 발생할 경우 스쿠

Table 4. Top 30 clinical signs

Rank	Basic data (n=5,747)		Single Scu. (n=614)	
	Clinical Signs	Ratio (%)	Clinical Signs	Ratio (%)
1	Liver hemorrhages	95	Liver hemorrhages	96.4
2	Intestine hemorrhages	78	Gill hemorrhages	78.5
3	Gill hemorrhages	69	Gill pale	73.9
4	Gill pale	62	Intestine hemorrhages	64.8
5	Fin hemorrhages	47	Kidney hypertrophy	45.9
6	Spleen hypertrophy	37	Spleen hypertrophy	45.8
7	Kidney hypertrophy	30	Water in the gonads	37.3
8	Liver hypertrophy	25	Fin hemorrhages	31.1
9	Water in pericardinal cavity	25	Water in pericardinal cavity	28.8
10	Liver tissue degeneration	24	Skin ulcer	25.1
11	Intraperitoneal muscle hemorrhages	23	Gill opeculum inflammation	22.1
12	Gill opeculum inflammation	22	Liver tissue degeneration	22
13	Gill opeculum hemorrhages	19	Snout ulcer	20.8
14	Abdominal distension	18	Gill opeculum hemorrhages	20.8
15	Ascites	17	Fin ulcer	20.5
16	Water in the gonads	16	Liver hypertrophy	19.2
17	Liver pale	15.7	Water in the intestine	18.6
18	Skin ulcer	14	Liver pale	18.4
19	Water in the intestine	13	Abdominal distension	16
20	Gill necrosis	12	Intraperitoneal muscle hemorrhages	15.6
21	Skin hemorrhages	10	Ascites	12.1
22	Gill hypertrophy	9.7	Gill necrosis	9.3
23	Gonad hemorrhages	9.6	Gill hypertrophy	8.6
24	Eye redness	9	Snout inflammation	8.5
25	Eye clouding	8.2	Gonad hemorrhages	8.3
26	Fin ulcer	8.1	Hernia	6.8
27	Snout ulcer	7.3	Eye redness	6.7
28	Hernia	7.2	Gonads water	6.5
29	Snout inflammation	7	Eye clouding	6.4
30	Snout hemorrhages	6.8	Skin hemorrhages	5

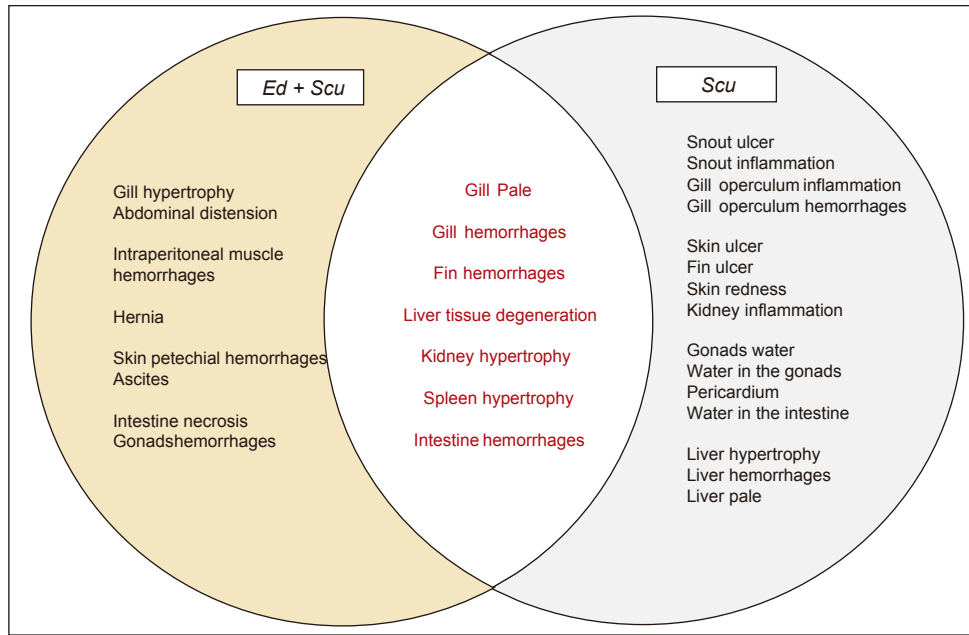
티카충에 의해 더 쉽게 공격을 받고 감염이 잘 이루어진 것으로 보였다. 따라서 면역력이 약하거나 혹은 건강하지 않은 어체에 스쿠티카충은 기회성 감염(opportunistic infection)을 한다고 판단되었다.

본 연구를 통해 임상증상 항목에 대한 표준화 필요성과 질병 중증도에 대한 분류 작업을 통해 경증인 경우와 중증인 경우의 질병 증상들을 구분 분석할 필요성도 대두되었다. 또한, 추후 세균성 감염이 먼저인지 기생충 감염이 먼저 인지에 따라 질병의

Table 5. Association rules of clinical signs mean disease

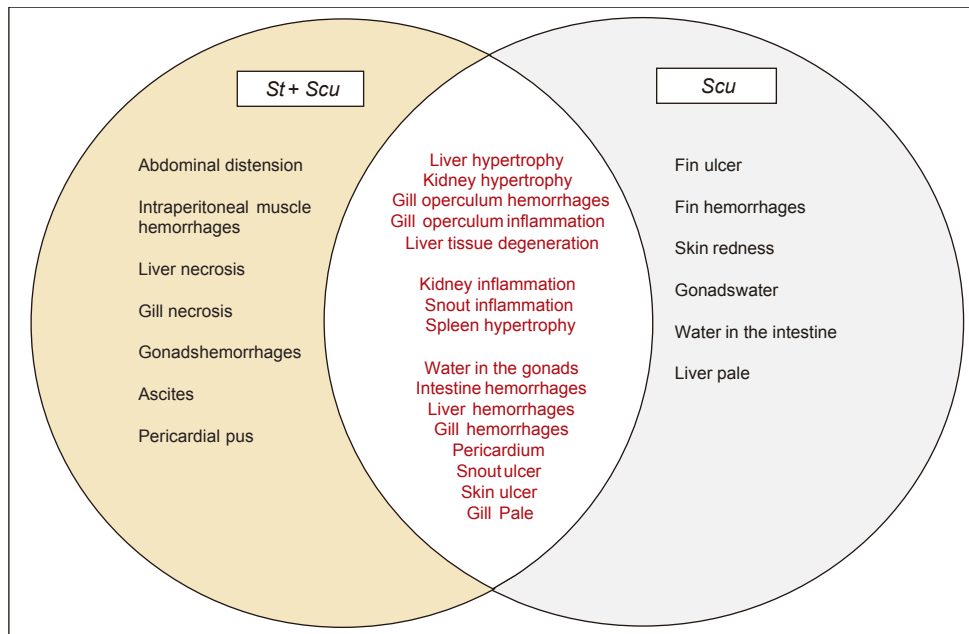
Disease	Clinical signs	Conf.	Lift
Ed+Scu	1 {Abdominal distension, Gill hypertrophy, Hernia, Gonad hemorrhages, Kidney hypertrophy, Intestine necrosis }	1	164.2
	2 {Abdominal distension, Gill hypertrophy, Hernia, Gonad hemorrhages, Intestine necrosis }	0.75	123.15
	3 {Gill hypertrophy, Hernia, Gonad hemorrhages, Kidney hypertrophy, Intestine necrosis }	0.75	123.15
	4 {Gill hypertrophy, Gill pale, Fin hemorrhages, Spleen hypertrophy, Gonad hemorrhages, Kidney hypertrophy, Intestine hemorrhages }	0.75	123.15
	5 {Abdominal distension, Gill hypertrophy, Gill pale, Fin hemorrhages, Liver tissue degeneration, Spleen hypertrophy, Gonad hemorrhages, Intestine hemorrhages }	0.75	123.15
St+Scu	1 {Gill necrosis, Gill pale, Gill hemorrhages, Snout inflammation, Liver hypertrophy }	0.75	28.74
	2 {Intraperitoneal muscle hemorrhages, Gill operculum hemorrhages, Snout ulcer, Liver hypertrophy, Water in the gonads }	0.75	28.74
	3 {Gill operculum hemorrhages, Snout ulcer, Liver tissue degeneration, Spleen hypertrophy, Water in pericardinal cavity }	0.75	28.74
	4 {Abdominal distension, Gill pale, Liver necrosis, Liver hemorrhages, Water in pericardinal cavity, Spleen hypertrophy, Kidney hypertrophy, Intestine hemorrhages }	0.75	28.74
Vi+Scu	1 {Snout ulcer, Fin hemorrhages, Liver tissue degeneration, Liver pale, Water in the gonads }	1	25.89
	2 {Gill necrosis, Skin ulcer, Ascites, Spleen hypertrophy, Intestine hemorrhages }	1	25.89
	3 {Snout ulcer, Fin hemorrhages, Liver tissue degeneration, Liver pale, Water in the gonads }	1	25.89
	4 {Gill necrosis, Gill hemorrhages, Skin ulcer, Spleen hypertrophy, Gonads hemorrhages }	1	25.89
	5 {Gill necrosis, Skin ulcer, Ascites, Spleen hypertrophy, Intestine hemorrhages }	1	25.89
	6 {Gill hemorrhages, Fin ulcer, Fin hemorrhages, Skin ulcer, Liver pale, Spleen hypertrophy, Water in pericardinal cavity }	1	25.89
	7 {Gill pale, Fin ulcer, Fin hemorrhages, Skin ulcer, Liver pale, Spleen hypertrophy, Kidney hypertrophy }	1	25.89
	8 {Gill pale, Fin ulcer, Fin hemorrhages, Skin ulcer, Liver pale, Spleen hypertrophy }	0.8	20.71
	9 {Gill hemorrhages, Fin ulcer, Fin hemorrhages, Skin ulcer, Liver pale, Spleen hypertrophy }	0.8	20.71
	10 {Gill necrosis, Skin ulcer, Ascites, Spleen hypertrophy }	0.75	19.42
	11 {Gill hypertrophy, Gill pale, Gill operculum inflammation, Snout ulcer, Water in the gonads }	0.75	19.42
	12 {Gill necrosis, Gill pale, Skin ulcer, Spleen hypertrophy, Gonads hemorrhages }	0.75	19.42
	13 {Fin ulcer, Fin hemorrhages, Skin ulcer, Liver pale, Spleen hypertrophy, Water in pericardinal cavity }	0.75	19.42
	14 {Fin ulcer, Fin hemorrhages, Skin ulcer, Liver pale, Spleen hypertrophy, Kidney hypertrophy }	0.75	19.42
	15 {Gill pale, Fin ulcer, Fin hemorrhages, Skin ulcer, Liver pale, Kidney hypertrophy }	0.75	19.42
	16 {Gill hemorrhages, Fin ulcer, Skin ulcer, Liver hypertrophy, Liver tissue degeneration, Kidney hypertrophy }	0.75	19.42
	17 {Gill pale, Gill hemorrhages, Fin hemorrhages, Skin ulcer, Liver pale, Spleen hypertrophy, Intestine hemorrhages }	0.75	19.42

Vi, *Vibrio* spp.; Ps, *Pseudomonas* spp.; Ed, *Edwardsiella piscicida*; St, *Streptococcus parauberis*; Scu, Scuticociliate; Min Sup., 0.05%; Min Conf., 70%.



Ed, *Edwardsiella piscicida*; Scu, Scuticociliate.

Fig. 6. Common clinical signs of mixed infection (Scuticociliate and *Edwardsiella piscicida*) and single Scuticociliate are shown in the middle of the Venn diagram. Ed, *Edwardsiella piscicida*; Scu, Scuticociliate.



St, *Streptococcus parauberis*; Scu, Scuticociliate.

Fig. 7. Common clinical signs of mixed infection (Scuticociliate and *Streptococcus parauberis*) and single Scuticociliate are shown in the middle of the Venn diagram. St, *Streptococcus parauberis*; Scu, Scuticociliate.

양상이 어떻게 다른지에 대한 연구도 필요할 것이다.
본 연구는 제주지역 다수의 양식장의 2015-2018년 수천 마리

의 어류 진단 데이터로부터 넙치 어종에 대해 스키테리카 발생과 환경 · 생물학적 인자(수온, 체장, 체중) 관련성을 제시하였고

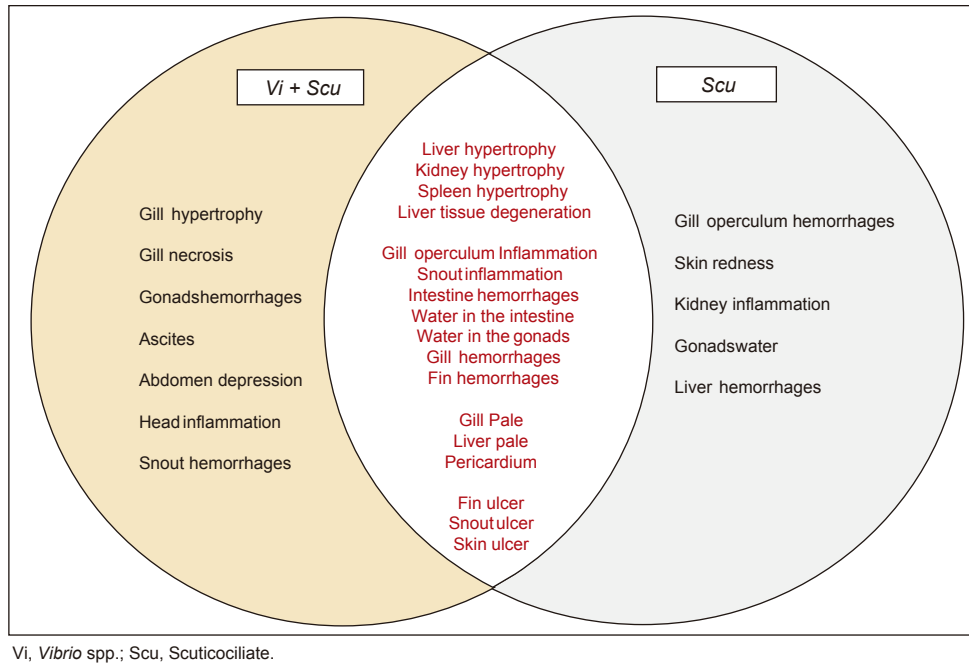


Fig. 8. Common clinical signs of mixed infection (Scuticociliate and *Vibrio* spp.) and single Scuticociliate are shown in the middle of the Venn diagram. Vi, *Vibrio* spp.; Scu, Scuticociliate.

특히, 연관규칙 마이닝을 통해 100가지의 임상증상으로부터 스쿠티카증과 세균성 질병과의 혼합 감염 시 주요 임상증상 및 스쿠티카증의 공통 임상증상을 제시하였다. 그리고 이러한 데이터 기반 분석법이 넓치 질병 분석에 사용될 수 있음을 보였다. 해당 결과는 향후 수산질병관리사나 양식장에서 스쿠티카증 감염을 임상적으로 추정하거나 스쿠티카증의 발생이 많아질 가능성이 높은 시기를 예측하는 것에 있어 도움이 될 수 있을 것이다. 본 연구가 양식생물 질병진단 분야에서 데이터 기반 연구의 시작점이 되어 스쿠티카증 위주의 질병 패턴 분석에서 나아가 다른 질병들의 패턴 추이 분석 및 질병 간 연관성 분석, 질병과 의약품과의 연관성 분석 등으로 확장해 갈 수 있으리라 본다.

사 사

이 논문은 2020년 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 수행된 연구임(스마트 수산양식 연구센터).

References

Bae MJ, Im EY, Kim HY and Jung SJ. 2009. The effect of temperature to scuticociliatida *Miamiensis avidus* proliferation, and to mortality of infected olive flounder *Paralichthys olivaceus*. *J Fish Pathol* 22, 97-105.
 Choi JH, Lee IH, Kim JH, Park SY, Shin AM, Son CS, Park HJ and Kim YN. 2010. Association analysis of comorbidity of

cerebral infarction using data mining. *J Kor Soc Phys Ther* 22, 75-81.
 Cho MY, Kim KI, Min EY and Jung SH. 2019. Global outbreaks and strategies to control the emerging diseases in aquaculture farms in Korea. *Ocean Policy Res* 34, 67-88. <https://doi.org/10.35372/kmiopr.2019.34.1.003>.
 Cho MY, Kim MS, Choi HS, Park GH, Kim JW, Park MS and Park MA. 2008. A statistical study on infectious diseases of cultured olive flounder *Paralichthys olivaceus* in Korea. *J Fish Pathol* 21, 271-278.
 Han SR, Han HS, Evensen Ø and Kim SH. 2017. PCR-based identification of *Pseudomonas fluorescens* in diseased olive flounder *Paralichthys olivaceus*, in Jeju Island, South Korea. *J Fish Pathol* 30, 67-70. <https://doi.org/10.7847/jfp.2017.30.1.067>.
 Jin CN, Kang BJ, Jang YH, Park BH and Jhon BK. 2015. Monitoring of scuticociliatosis of olive flounder *Paralichthys olivaceus* farm in Jeju, Korea from 2007 to 2014. *J Fish Pathol* 28, 165-169. <https://doi.org/10.7847/jfp.2015.28.3.165>.
 Jin CN, Kang HS, Lee CH, Lee YD, Lee JH and Heo MS. 2007. Biological characteristics of Scuticociliate *Philasterides dicentrarchi* isolated from cultured olive flounder, *Paralichthys olivaceus*. *J Aquaculture* 20, 106-113.
 Jin CN, Lee CH, Oh SP, Jung YU, Song CB, Lee J and Heo MS. 2003. Scuticociliatosis in flounder farms of Jeju island. *J Fish Pathol* 16, 135-138.
 Kang BJ. 2019. Machine learning directly in field with Python.

- Acorn press, Seoul, Korea, 377-388.
- Kim JW, Cho MY, Park GH, Won KM, Choi HS, Kim MS and Park MA. 2010. Statistical data on infectious diseases of cultured olive flounder *Paralichthys olivaceus* from 2005 to 2007. *J Fish Pathol* 23, 369-377.
- Kim JW, Jung SH, Park MA, Do JW, Choi DL, Jee BY, Cho MY, Kim MS, Choi HS, Kim YC, Lee JS, Lee CH, Bang JD, Park MS and Seo JS. 2006. Monitoring of Pathogens in Cultured Fish of Korea for the Summer Period from 2000 to 2006. *J Fish Pathol* 19, 207-214.
- Kim SH. 2018. Companion disease analysis of patients with chronic obstructive pulmonary disease using a sequential association rule. In: Report of Health Insurance Review Assessment. HIRA Report 12, 53-59.
- Lee IH, Shin AM, Son CS, Park HJ, Kim JH, Park SY, Choi JH and Kim YN. 2010. Association analysis of comorbidity of cerebral infarction using data mining. *J Kor Soc Phys Ther* 22, 75-81.
- Park HS, Lee MS, Hwang SJ and Oh SY. 2016. TF-IDF based association rule analysis system for medical data. *KIPS Tr. Software Data Eng* 5, 145-154. <https://doi.org/10.3745/KTSDE.2016.5.3.145>.
- Rfriend. 2015. R analysis and programming. Retrieved from <https://rfriend.tistory.com/> on May 7, 2020.
- Sabthami J, Thirumoorthy K and Muneeswaran K. 2016. Mining association rules for early diagnosis of diseases from electronic health records. *Middle East J Sci Res* 24, 248-253. <https://doi.org/10.5829/idosi.mejsr.2016.24.S2.159>.
- Sebastiao FA, Furlan LR, Hashimoto DT and Pilarski F. 2015. Identification of bacterial fish pathogens in Brazil by direct colony PCR and 16s rRNA gene sequencing. *Adv Microbiol* 5, 409-424. <https://doi.org/10.4236/aim.2015.56042>.
- Tai YM and Chiu HW. 2009. Comorbidity study of ADHD: Applying association rule mining (ARM) to National Health Insurance Database of Taiwan. *Int J Med Inform* 78, e75-e83. <https://doi.org/10.1016/j.ijmedinf.2009.09.005>.