



Research Article

Application of Bayesian network for farmed eel safety inspection in the production stage

양식뱀장어 생산단계 안전성 조사를 위한 베이저안 네트워크 모델의 적용

Seung Yong Cho*

조승용*

Department of Information & Technology Research, National Food Safety Information Service, Seoul 03127, Korea

식품안전정보원 정보기술연구부

Abstract The Bayesian network (BN) model was applied to analyze the characteristic variables that affect compliance with safety inspections of farmed eel during the production stage, using the data from 30,063 cases of eel aquafarm safety inspection in the Integrated Food Safety Information Network (IFSIN) from 2012 to 2021. The dataset for establishing the BN model included 77 non-conforming cases. Relevant HACCP data, geographic information about the aquafarms, and environmental data were collected and mapped to the IFSIN data to derive explanatory variables for nonconformity. Aquafarm HACCP certification, detection history of harmful substances during the last 5 y, history of nonconformity during the last 5 y, and the suitability of the aquatic environment as determined by the levels of total coliform bacteria and total organic carbon were selected as the explanatory variables. The highest achievable eel aquafarm noncompliance rate by manipulating the derived explanatory variables was 24.5%, which was 94 times higher than the overall farmed eel noncompliance rate reported in IFSIN between 2017 and 2021. The established BN model was validated using the IFSIN eel aquafarm inspection results conducted between January and August 2022. The noncompliance rate in the validation set was 0.22% (15 nonconformances out of 6,785 cases). The precision of BN model prediction was 0.1579, which was 71.4 times higher than the non-compliance rate of the validation set.

Keywords farmed eel, aquafarm surveillance, Bayesian network model, conditional probability



OPEN ACCESS

Citation: Cho SY. Application of Bayesian network for farmed eel safety inspection in the production stage. Korean J Food Preserv, 30(3), 459-471 (2023)

Received: May 11, 2023

Revised: June 08, 2023

Accepted: June 14, 2023

*Corresponding author

Seung Yong Cho

Tel: +82-2-744-8112

E-mail: sycho2k@foodinfo.or.kr

Copyright © 2023 The Korean Society of Food Preservation. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

우리나라 1인당 연간 수산물 소비량은 '12년 53.8 kg에서 '20년에 68.1 kg으로 증가 추세에 있지만, 어획을 통한 어업생산량은 수산자원의 남획 등으로 인해 감소하는 경향을 보여 수산물생산에 있어서 양식수산물의 비중이 지속적으로 증가하고 있다(KOSIS, 2022). FAO(2022)에 따르면, 2020년도 전 세계 수산물 소비량 중 양식수산물이 차지하는 비율이 49.2%를 차지하고

있으며 2030년에는 양식수산물의 비중이 60%를 상회할 것으로 전망하고 있어 양식수산물의 중요성이 점점 커지고 있다.

우리나라에서 내수면어업으로 가장 많이 생산되는 어종은 뱀장어로 2021년에는 15,177 M/T이 생산되어 전체 내수면어업으로 생산되는 어종의 36.9%를 차지하고 있으며, 뱀장어 양식장의 형태는 2020년 기준으로 총 232개소 중 약 53%가 순환여과식이며 32%가 지수식 방법을 취하고 있다(Statistics Korea, 2021). 이는 Lee와 Kim(2006)의 보고에서 기술한 2006년 순환여과식 뱀장어양식이 10% 정도에 불과하며 대부분인 90%가 지수식이었던 것에 비해, 보다 친환경적이고 자연환경의 영향을 덜 받으면서도 고밀도 양식으로 생산량을 증대하는 방향으로 변화되고 있다. 한편 고밀도로 뱀장어를 양식하는 방식은 어병 발생 시 이의 확산으로 인한 집단폐사의 위험이 높아 동물용의약품의 사용 등 일련의 위해요소의 투입이 수반되어 수산물 안전에 대한 우려도 증가하게 된다. 수산물 안전을 확보하기 위해 「식품위생법」에 양식 시 사용되어 어체에 잔류하는 경우 인체에 위해성 우려가 있는 동물용의약품 등 잔류물질과 오염된 환경이나 사료로부터 어체로 유입될 수 있는 납, 카드뮴, 수은 등의 중금속과 polychlorinated biphenyls(PCBs), benzopyrene 등 오염물질에 대한 기준을 정하여 관리하고 있다(MFDS, 2021).

우리나라 수산물 및 수산가공식품의 안전관리는 2013년 정부조직법의 개편으로 수산물 안전관리를 식품의약품안전처로 일원화하면서, 생산단계는 해양수산부에서 안전관리를 담당하고, 가공-유통-소비단계는 식약처에서 담당하는 이원화된 체계를 갖고 있다. 국립수산물품질관리원(NFQS, 2022)의 수산물 안전성조사 결과에 따르면, '21년 모니터링조사 및 기타조사를 포함한 안전성조사건수는 총 10,065건이었으며 이 중 25건이 부적합으로 부적합률은 0.25% 수준이다. 수산물 생산단계 안전성조사 정보는 국가 식품안전관리 정보를 망라하는 네트워크인 통합식품안전정보망으로 연계되어 식품의약품안전처에서 식품 등의 안전관리를 총괄하는 데 활용된다(Cho, 2021).

우리나라의 생산단계 수산물 안전관리 계획의 수립은 기존의 현업에 종사하는 전문인력들의 경험적 판단에 의해 이루어지던 것이 빅데이터 분석방법을 적용하여 검사대상의

부적합 발생 확률을 기반으로 대상 선정이 이루어지는 쪽으로 전환되고 있다. 통합식품안전정보망의 데이터를 활용 관련 연구들로는 식품제조가공업소 위생점검(Cho와 Cho, 2020) 및 수입건강기능식품의 통관검사(Lee 등, 2022)의 대상 선정에 기계학습 등 빅데이터 분석을 적용하는 연구가 보고된 바 있다. 이들은 부적합률이 매우 낮은 불균형 데이터의 이상탐지(anomaly detection)에 있어서 소수의 부적합 데이터에서 가상의 데이터를 생성하는 Synthetic minority over-sampling technique(SMOTE) 기법을 적용하여 부적합 사례의 예측력을 높이는 알고리즘개발에 대한 연구로서, 데이터 관점에서 분류 및 예측이 이루어져 부적합 발생을 예방할 수 있는 요인들에 대한 정보는 제공해 주지 못하고 있다.

베이지안 네트워크(Bayesian network, BN)는 베이즈 통계와 결정이론이 그래프이론과 결합된 확률모형의 한 종류로서, 불확실한 상황하에서 변수들 간의 확률적인 관계를 연구자의 전문지식을 활용하여 네트워크의 형태로 표현하고 결론을 예측하고, 인과성을 설명하며 의사결정을 하는데 유용한 도구이다(Jensen과 Nielsen, 2007; Kjaerulff와 Madsen, 2010). 베이지안 네트워크는 금융, 의학분야, 로보틱스, 유전학 및 생태학 등 다양한 분야에서 위험평가 및 원인분석에 활용되어 왔다(Hunte 등, 2022; Meng 등, 2022). 수산분야에서는 Giles(Giles, 2008)가 양식장이 저서환경에 미치는 영향을 베이지안 네트워크를 이용하여 분석하고 예측한 바 있다. 최근 식품분야에서도 원인 설명이 가능하다는 베이지안 네트워크의 장점으로 인해 식품사기 및 부적합의 예측(Bouzembrak 등, 2018; Bouzembrak과 Marvin, 2016; Lupo 등, 2013; Soon, 2020), 식품위해요소에 미치는 영향요인의 예측(Bouzembrak과 Marvin, 2019), 패류독소 등 위해요소의 발생예측(Wang 등, 2022) 등에 베이지안 네트워크 모델의 적용이 보고된 바 있다.

본 연구에서는 통합식품안전정보망의 뱀장어 양식장의 생산단계 안전성 조사 데이터를 통합망 내 안전성조사와 관련한 변수들 및 양식장의 지리정보, 수산물 안전관리인증기준(Hazard Analysis and Critical Control Point, HACCP) 인증정보 및 양식장 소재지 근처의 환경정보를 연계하여 가능한 설명변수를 설정하고, 이에 베이지안 네트워크 모델을 적용하여 생산단계 안전성 조사 부적합과 설명변

수들의 관계를 확률적으로 분석하여 부적합에 관련하는 영향요인을 추론하고자 하였다.

2. 재료 및 방법

2.1. 데이터 수집 및 정리

2012년부터 2021년까지 뱀장어 생산단계 안전성조사 부적합에 관여하는 가능한 설명변수로서 양식장 운영과 관련하여 HACCP 인증, 유해물질 검출이력, 부적합이력 등의 변수를 고려하였으며, 양식장 환경요인으로 양식장위치, 용수의 오염수준과 용수수온을 검토하였고 양식장 질병발생 등의 정보를 검토하였다.

이들 변수에 해당하는 데이터를 각 통합식품안전정보망, 각 주무부처의 웹사이트 및 공공데이터(openAPI)로부터

수집하였다. 통합식품안전 정보망의 수산물안전성 조사 정보를 이용하였으며, 그 외의 정보들은 국립수산물품질관리원(nfqs.go.kr)의 국내수산물 생산위생시설 현황에 포함된 양식장 HACCP 자료 및 환경부 물환경정보시스템의 수질 측정망 자료(water.nier.go.kr) 등 수산물안전과 관련한 데이터를 오픈 API 형태로 이용하였다.

수산물안전성조사의 주요 정보항목은 시료번호, 업체명, 품목명, 조사단계, 조사구분명, 계획년도, 계획월, 차수, 채취일자, 원산지, 채취수량, 채취단위명, 보유수량, 보유수량 단위, 분석시작일시, 분석종료일시, 최종판정, 의뢰기관, 분석기관, 검사대상명, 항목명, 관할기관명 등이다. 정보의 출처로부터 수집한 원천데이터를 수산물안전성조사 분석에 사용하기 위해 각 데이터의 연계와 특성변수의 도출에 대한 과정을 Fig. 1에 도시하였다. 수산물 안전성 조사 정보의

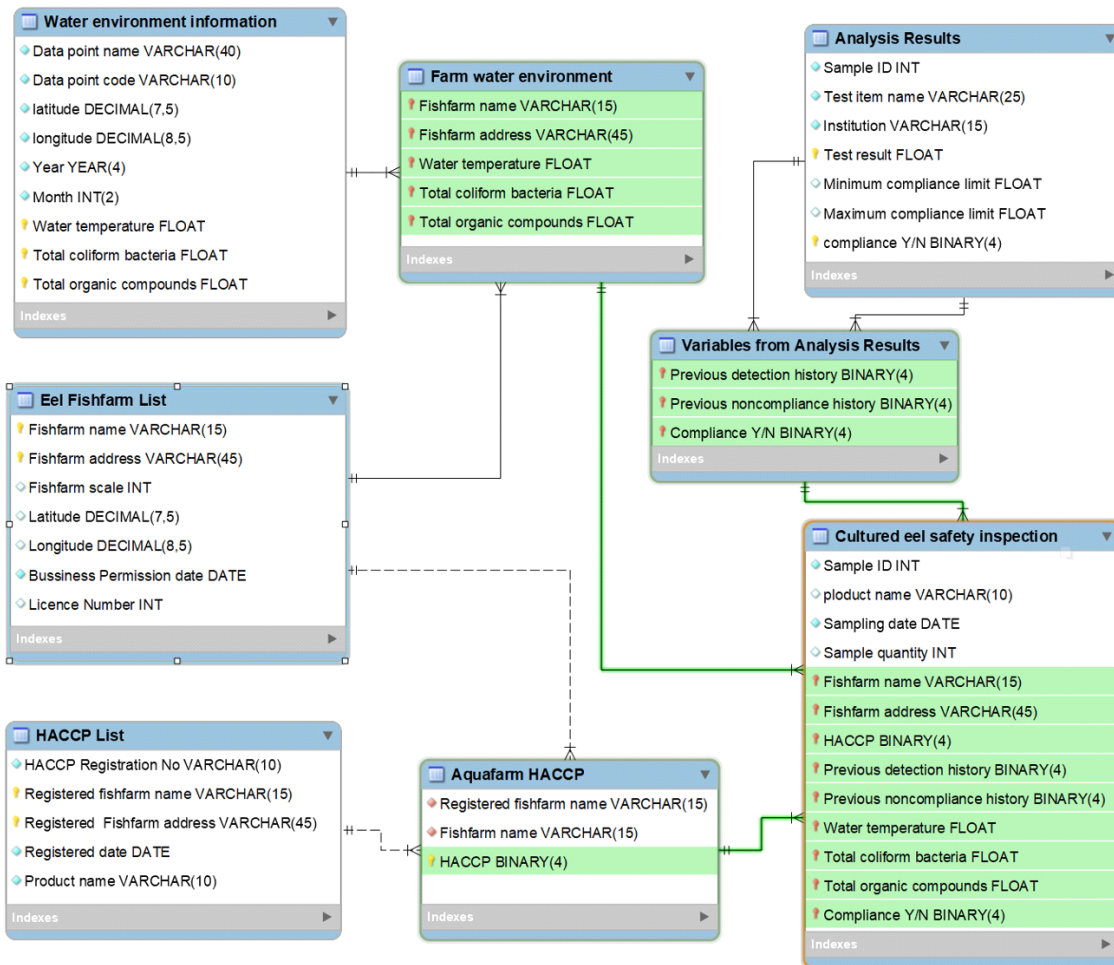


Fig. 1. Derivation of characteristic variables for the Bayesian network model of cultured eel safety inspection from raw data sources.

양식장 소재지 및 양식장업체명을 key값으로 양식장의 HACCP 인증 정보를 연계하였다. 양식장 소재지의 수질환경은 양식장 주소를 위경도 형태의 지리정보로 변환하여 양식장으로부터 지리적으로 유클리드 거리가 가장 가까운 측정망에서 측정한 안전성조사시기의 수질측정망 자료와 연계하여 수질환경 정보를 도출하였다. 수질환경 측정망 정보는 수온, 용존산소, 생물학적산소요구량(BOD), 화학적산소요구량(COD), 부유물질, 총질소, 총인, 총유기탄소, pH를 비롯하여 총대장균수, 카드뮴, 수은 등 중금속에 대한 데이터를 포함하고 있다.

2.2. 데이터 전처리

뱀장어 생산단계 안전성조사 베이지안 네트워크 모델을

구성하는 각 노드와 이의 범주화된 상태를 Table 1에 제시하였다. 뱀장어 생산단계 부적합에 관련되는 변수들로 뱀장어의 양식장 위치(Farm_Location)는 뱀장어 생산량이 많은 전북지역(Region 1), 전남지역(Region 2), 기타지역(Region 3)으로 범주화하였다. HACCP 여부와 직전 5년간 해당검사항목의 검출이력(Detection_History) 및 직전 5년간 부적합이력은 각각 2단계로 범주화하였으며, 용수 수온(Water_Temp)은 고저에 따라 3단계로 구분하였다. 여기서 직전 5년간 해당검사항목의 검출이력이란 비록 부적합 수준으로 유해물질이 검출되지는 않아 적합이긴 하지만 해당검사대상 항목이 잔류하여 불검출이 아닌 경우이다.

한편, 용수의 오염수준을 나타내는 용수환경(Water_Env)은 각각 3단계로 범주화한 총대장균군 수(Ecoli_Lv)와 총유

Table 1. Variables (nodes) and states of nodes of Bayesian network model for farmed eel safety inspection at the production stage

Node name	Description	State level	Number of data counts
Eel_Compliance	Cultured eel safety inspection compliance	Pass	29,986
		Fail	77
Farm_Location	Location of aquafarm	Region 1 (Jeonnam)	14,172
		Region 2 (Jeonbuk)	8,488
		Region 3 (other areas)	7,403
HACCP	HACCP certification	HACCP	15,619
		NonHACCP	14,444
Detection_History	Previous detection history of hazardous materials among the compliant results in 5 years	Detected	1,079
		Nondetected	28,984
Previous_Noncompliance	Previous non-compliant history in 5 years	Yes	2,380
		No	27,683
Water_Temp	Water temperature	High (>27°C)	2,682
		Medium (15~27°C)	15,906
		Low (<15°C)	11,475
Ecoli_Lv	Coli form bacteria level	EColiLv1 (<500/100 mL)	9,573
		EColiLv2 (<5,000/100 mL)	7,968
		EColiLv3 (≥5,000/100 mL)	6,937
TOC	Total organic compound	TOCLV1 (<3 mg/L)	12,351
		TOCLV2 (<6 mg/L)	11,811
		TOCLV3 (≥6 mg/L)	5,539
Water_Env	Water environment	Moderate	18,829
		Poor	10,928

기탄소량(TOC)으로부터 구하였는데, 총대장균군수나 총유기탄소량 중 어느 하나도 오염도가 높은 Level 3에 해당하지 않는 경우의 용수환경을 알맞음(moderate), 어느 하나라도 Level 3에 해당하는 경우를 나쁨(poor)으로 정하여 2단계로 범주화하였다.

2.3. 베이저안 네트워크 모델

베이저안 네트워크(Bayesian network, BN)는 변수들 간의 확률적인 관계를 네트워크 형태로 표현하는 방법으로, 광범위한 데이터를 변수 간의 관계에 따라 네트워크 형태의 그래프로 표시함으로써 단순히 분류하거나 예측할 때에는 반영하지 못하는 데이터의 특성을 이해할 수 있게 해준다(Jensen과 Nielsen, 2007).

베이저안 네트워크는 변수에 해당하는 노드(타원)와 그 노드 간의 인과관계를 나타내는 방향성 간선(arc)들로 구성된 방향성 비순환 그래프(directed acyclic graph, DAG)로 표현된다. 간선의 방향성이 노드 X_i 에서 시작하여 노드 X_j 로 향하는 경우 X_i 는 X_j 의 부모노드가 되며, X_j 는 X_i 의 자식노드가 된다. 베이저정리에 의하면 사건 X_j 가 발생하는 조건하에서 사건 X_i 가 발생하는 확률 $P(X_i | X_j)$ 는 다음과 같이 표현할 수 있다.

$$P(X_i | X_j) = \frac{P(X_i, X_j)}{P(X_j)} = \frac{P(X_j | X_i) \times P(X_i)}{P(X_j)}$$

이때, $P(X_i)$ 는 사건 X_i 의 사전 확률이며, $P(X_j)$ 는 사건 X_j 가 발생할 사후 주변확률, 그리고 $P(X_j | X_i)$ 는 사전 확률을 알고 있는 사건 X_i 가 발생한 조건하에서 사건 X_j 가 발생하는 확률이다. 노드 X_i 의 부모노드를 $pa(X_i)$ 라고 하면, 조건부확률과 노드 간의 연결선들로 이루어진 베이저안 네트워크 구조가 주어지는 경우 각 노드가 동시에 발생할 결합확률분포를 다음 식으로 나타낼 수 있다.

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | pa(X_i))$$

베이저안 네트워크의 노드로 표현되는 변수의 값들은 서로 배반적으로 범주화된 상태들과 부모노드가 사전확률로

주어졌을 때 부모노드의 상태에 따라 자식노드의 각 상태가 발생할 확률로 구성되는데, 이는 각 노드에 대한 조건부확률표(conditional probability table, CPT)로 표현한다. 이러한 베이저안 네트워크는 모든 변수 간의 의존관계를 표현하기 때문에 결측치가 많이 포함된 데이터의 취급이 가능하고 성분들 간의 인과관계를 알 수 있어 특정조건하에서 결과를 예측할 수 있다.

본 연구에서는 오픈 소스툴인 WEKA 3.8.6(The University of Waikato Hamilton, New Zealand)의 Hill-Climbing 알고리즘을 이용하여 베이저안 네트워크 모델을 구성하였으며, Netica 6.0.9(Norsys Software Corp., Canada)를 이용하여 노드가 되는 특성변수들의 발생하는 확률값을 계산하였다.

2.4. 베이저안 네트워크 모델의 적정성 검토

베이저안 네트워크 모델의 적정성은 통합식품안전정보망에 수록된 2022년 1월에서 8월까지의 뱀장어 안전성 조사 데이터 6,785건의 부적합 여부를 예측하는 데 적용하여, 성능지표테이블(confusion matrix)의 정밀도(precision)와 재현율(recall)을 검토하였다.

3. 결과 및 고찰

3.1. 양식뱀장어 안전성조사 부적합 발생추이

2012년부터 2021년까지 양식뱀장어 생산단계 안전성조사에서 나타난 부적합 원인물질은 benzopyrene 부적합 1건을 제외한 76건이 동물용의약품 관련 부적합이었다. 동물용의약품 관련 부적합 사례 중에는 사용이 금지되어 검출되어서는 아니되는 물질인 nitrofurans, ofloxacin, norfloxacin, pefloxacin 등이 총 18회 검출되었으며, 항균제 등 기타 기준치를 초과한 동물용의약품으로 enrofloxacin/ciprofloxacin, oxolinic acid, trimethoprim, erythromycin, sulfonamides (15종), dehydrocholic acid 등이 검출되어 총 58건이 보고되었다(Table 2).

국내에서 양식되는 뱀장어에서는 *Edwardsiella* sp.나 *Aeromonas* sp. 등에 의해 발생하는 세균성 질병과 뱀장어 허피스바이러스(Anguillid herpesvirus-1, AnHV)에 의한 바이러스성 질병이 주로 발생하고 있는데(Kim 등, 2019),

Table 2. Annual number of nonconforming cases detected in farmed eel safety inspection at the production stage from 2012 to 2021

Test items	Number of nonconformity counts/number of inspection counts										
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total
Veterinary drugs with illegal levels											
Enrofloxacin/ciprofloxacin	1/61	0/51	0/84	0/90	1/103	1/131	0/153	2/174	2/238	0/107	7/1,183
Erythromycin	0/15	0/12	0/16	0/21	0/19	0/29	1/72	0/174	0/238	0/107	1/703
Trimethoprim	0/61	0/22	0/79	0/90	0/19	0/29	1/149	0/174	2/238	0/107	3/968
15 Sulfonamides	0/12	0/28	0/80	1/90	0/102	1/126	2/149	0/174	4/238	0/107	8/1,106
Oxolinic acid	1/24	0/42	1/85	3/90	2/103	4/131	5/153	3/174	11/238	3/107	33/1,147
Dehydrocholic acid	0/0	0/0	0/0	0/0	0/0	0/0	1/71	2/74	0/88	3/57	6/290
Praziquantel	0/10	0/18	0/12	0/18	0/33	0/21	0/57	0/58	0/98	0/57	0/382
Banned substances											
Ofloxacin	3/64	0/52	0/85	0/90	1/101	0/127	0/149	0/174	0/228	0/97	4/1,167
Norfloxacin	1/60	0/52	0/85	0/90	0/101	0/126	0/149	0/174	0/228	0/97	1/1,162
Pefloxacin	5/61	1/51	2/85	1/91	1/103	0/126	0/149	0/174	0/228	0/97	10/1,165
Nitrofurans	0/57	0/47	0/85	0/90	0/28	0/29	2/151	1/261	0/227	0/97	3/1,072
Methylene blue	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/38	0/75	0/95	0/208
Benzopyrene	0/15	1/13	0/23	0/84	0/101	0/29	0/148	0/73	0/227	0/98	1/811
Annual nonconformed counts/inspection counts	11/1,112	2/945	3/1,351	5/1,632	5/1,594	6/2,049	12/4,496	8/5,741	19/7,142	6/4,001	77/30,063

이와 관련하여 Table 2에는 동물용의약품 부적합발생 사례 중 에드워드병, 에로모나스병, 적점병 등의 세균성 질병의 치료 목적으로 사용되는 항생제인 oxolinic acid의 잔류 부적합사례가 33건으로 총부적합사례 77건의 43%를 차지하였다. 한편, 사용금지되어 어체에서 검출되면 아니되는 동물용의약품인 ofloxacin, norfloxacin, pefloxacin이 검출되어 부적합되는 사례가 2016년 이전까지는 발견되었으나 이후에는 검출된 경우가 없었으며, 다만 2018년과 2019년 nitrofurans이 검출된 바 있었다. 2021년에는 2016년 기준이 신설된 dehydrocholic acid의 부적합 사례 3건이 새롭게 적발되었다.

3.2. 베이저안 네트워크 모델의 구축

백장어 생산단계 안전성 조사 결과(Eel_Compliance)를 설명하는 베이저안 네트워크 모델은 HACCP 여부(HACCP)와 지난 5년간 해당 검사항목의 검출이력(Detection_History), 지난 5년간 해당 양식장에서의 부적합이력

(Previous_Noncompliance), 그리고 용수환경을 자식노드로 가지며, 다시 용수환경 노드는 총대장균수와 총유기탄소량을 이의 자식노드로 가지는 다층구조를 지닌다(Fig. 2). 한편 각기 이들의 자식노드인 양식장 위치와 용수의 수온을 설명 변수로 검토한 결과, 총대장균수 정보가 주어지면 양식장 위치는 다른 변수들과 조건부독립이며, 총유기탄소량 정보가 주어지면 용수의 수온이 다른 변수들과 조건부독립이다. 따라서 BN 모델을 적용할 때 총대장균수나 총유기탄소량 정보에 결측치가 발생하는 경우, 이들의 자식노드로부터 이들의 상태값들의 발생확률을 계산 가능하다는 장점이 있다. 하지만 본 연구에서는 총대장균수와 총유기탄소량 정보를 알 수 있으므로, BN 모델을 사용하여 양식백장어 생산단계 안전성조사 부적합을 추론하는 데 있어 양식장 위치와 용수의 수온을 제외한 노드만으로 계산할 수 있어 모델의 변수를 줄일 수있었다.

Fig. 2는 각 노드별 상태의 합을 1로 하였을 때 각 상태값들이 차지하는 확률을 BN 모델에 적용한 것이다. 이는 각

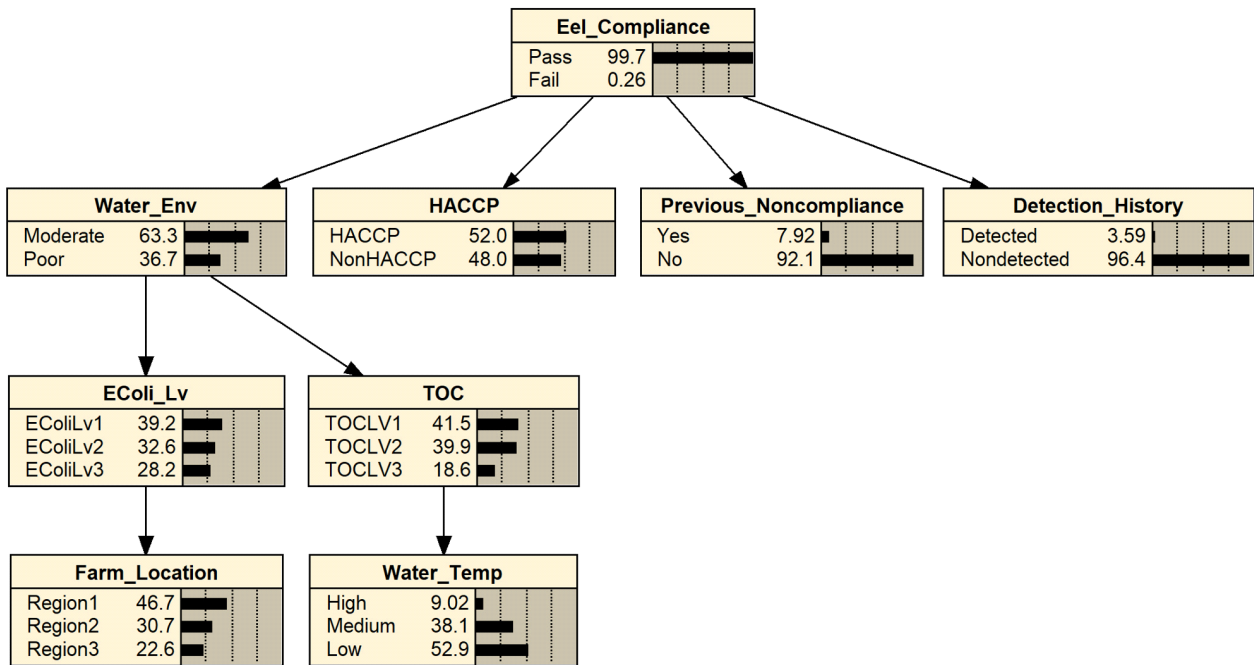


Fig. 2. Bayesian network model for analysis of farmed eel safety inspection results.

변수(노드)들에서 서로 배반적인 각 변수값들이 발생할 경우를 확률로 표현한 것이다. 이를 본 연구의 결과로 설명하면 안전성 조사를 실시할 때 뱀장어의 부적합률은 0.26%라는 것을 알 수 있으며, 뱀장어 양식장 중 HACCP 양식장은 52.0%, HACCP 인증을 받지 않은 일반양식장이 48.0%라는 것을 알 수 있다. 이는 다시 임의의 양식장을 선정하였을 때 그 양식장이 HACCP 양식장일 확률이 52.0%라는 것을 의미하므로 각 노드에서 상태값의 비율은 사전확률에 해당한다.

베이지 정리를 통해 각 설명변수들의 상태변화에 따른 뱀장어 부적합확률의 변화를 탐색하기 위해서는, 우선 부모노드인 뱀장어 안전성조사 결과의 적합 여부를 조건으로 하는 각 자식노드의 서로 배반적인 상태의 조건부확률을 구하여야 한다. 이는 부모노드의 상태가 적합 또는 부적합으로 결정되면 적합 또는 부적합 상태하에서 각 자식노드 상태의 발생비율을 구함으로써 얻을 수 있다. 이를 베이저안 네트워크와 함께 나타내면 Fig. 3과 같다. 각 테이블에서 행은 부모노드의 상태가 조건으로 주어졌을 때, 자식노드의 각 열에 주어진 상태가 발생할 사후확률을 백분율로 나타낸 것으로, 각 행에서 자식노드의 사후확률을 모두 더한 값은

100이 된다.

Fig. 3에서 해당 검사항목 검출이력 노드의 경우 뱀장어 안전성 조사결과가 적합일 경우 검사항목의 검출이력이 있을 확률이 3.139%인 반면, 조사결과가 부적합인 경우는 검출이력이 있을 확률이 81.013%에 달하여, 이 검출이력은 부적합확률과 연관성이 크다는 것을 알 수 있다.

3.3. 양식뱀장어 안전성 조사 부적합확률에 미치는 설명변수의 영향

베이지정리를 이용하여 HACCP 여부, 검출이력, 부적합이력, 용수환경, 총대장균군수, 총유기탄소량 등, 각 설명변수의 값에 따른 뱀장어 안전성조사 부적합 발생 확률을 사후확률로 계산하였다. 여러 조건의 설명변수값에 따라 양식뱀장어 안전성조사 부적합확률을 계산하여 Table 3에 나타내었다. 검출이력은 부적합에 많은 영향을 미쳐 검출이력이 없는 양식장의 부적합확률은 다음 설명변수의 변화에 관계없이 1% 미만이지만, 검출이력이 있는 양식장에서는 다른 설명변수값의 변화에 따라 부적합확률은 4.06%에서 24.28%로 계산되었다. 다음으로 부적합이력이 크게 영향을 미치고 있는데 부적합 이력이 없는 양식장에서 부적합이 발생할 확률



Fig. 3. Conditional probability tables (CPTs) for corresponding nodes of Bayesian network model for farmed eel safety inspection results.

은 최대 6.63%로 계산되는 데 반해, 부적합 이력이 보고된 양식장에서의 부적합 발생확률은 13.14%에서 24.28%로 높았다. 한편 용수환경의 경우 적정인 경우와 부적정인 경우 뱀장어의 부적합확률 차이는 3%에 못 미쳐, 용수환경을 양식장 안전성 조사 결과에 미치는 영향이 크지 않은 것으로 보인다. 이는 내수면 양식장에서 일반적으로 용수를 그대로 쓰지 않고 순환여과시스템 등 시설을 갖추고 있는 데서 기인하는 것으로 사료된다.

양식뱀장어 안전성 조사 시 부적합이 나올 확률이 가장 낮은 경우는 부적합 이력이 없으며 검사항목의 검출이력도 없는 HACCP 양식장에서, 용수의 총대장균수가 적고 총유

기탄소양이 적은 오염이 되지 않은 용수를 사용하는 경우로서 이러한 설명변수조건에서 부적합이 발생할 확률은 0.028%로 뱀장어의 부적합률보다 1/10 수준으로 계산되었다(Fig. 4(B)). 한편 양식뱀장어 부적합이 발생할 확률이 가장 높은 경우는 지난 5년간 검사항목의 검출이력이 있으면서 동시에 부적합 적발이력 있는 HACCP 인증을 받지 않은 양식장으로서 용수환경도 총대장균군 또는 총유기탄소가 높아 오염이 의심되는 용수를 사용하는 경우이다. 이러한 설명 변수의 조건하에서 뱀장어 시료를 채취하여 안전성 조사를 하는 경우 그 부적합 확률이 24.5%로 뱀장어 안전성 조사시 부적합률인 0.26%의 94배 수준으로 계산되었다

Table 3. Changes in conformity rate by explanatory variables

Detection history	Previous noncompliance	Water environment	HACCP	Conformity rate (%)	
				Pass	Fail
Detected	Yes	Moderate	HACCP	86.859	13.141
Detected	Yes	Moderate	NonHACCP	79.768	20.232
Detected	Yes	Poor	HACCP	83.945	16.055
Detected	Yes	Poor	NonHACCP	75.719	24.281
Detected	No	Moderate	HACCP	96.76	3.24
Detected	No	Moderate	NonHACCP	94.685	5.315
Detected	No	Poor	HACCP	95.939	4.061
Detected	No	Poor	NonHACCP	93.373	6.627
Nondetected	Yes	Moderate	HACCP	99.876	0.124
Nondetected	Yes	Moderate	NonHACCP	99.792	0.208
Nondetected	Yes	Poor	HACCP	99.843	0.157
Nondetected	Yes	Poor	NonHACCP	99.737	0.263
Nondetected	No	Moderate	HACCP	99.972	0.0275
Nondetected	No	Moderate	NonHACCP	99.954	0.04662
Nondetected	No	Poor	HACCP	99.965	0.0348
Nondetected	No	Poor	NonHACCP	99.942	0.0584

(Fig. 4(A)).

그러나 부적합 확률 최대조건에 해당하는 경우는 불균형이 심하고 과대적합(overfitting)이 우려되어 BN 모델의 검증을 위한 시험용 데이터의 부적합을 안정적으로 예측하기 위해 부적합확률에 영향을 적게 주는 용수환경을 제외한 검출이력, 부적합이력 그리고 HACCP 인증여부 3개를 사용하여 부적합확률을 계산하도록 하였다.

3.4. 양식뱀장어 안전성 조사 BN 모델의 적정성 확인

시험용 데이터를 대상으로 뱀장어 BN 모델을 이용하여 부적합으로 예측된 데이터에 대한 모델의 분류성능표(confusion matrix)를 Table 4에 제시하였다. 시험용 데이터로 사용된 6,785건의 안전성조사 결과 중에서 실제 부적합은 15건으로, 시험용 데이터세트의 부적합률은 0.22%이었다. BN 모델을 이용하여 예측된 데이터 19개 중에서 3개가 부적합이었으며 16개는 실제로 적합으로 판명되어 이 모델에서의 정밀도(precision)는 $3/19=0.1579$ 로 계산되어, 시험용 데이터세트의 원래 부적합 확률이 0.0022임

을 감안하면 본 BN 모델에 의해 예측된 결과는 무작위적으로 시료를 취하여 부적합을 적발할 확률의 약 71.4배에 해당하는 값이다. 한편, 실제 부적합 데이터 중에서 부적합으로 예측된 비율인 재현율(recall)은 $3/15=0.2$ 로 계산되어, 불균형 데이터인 식품위생점검 결과의 부적합을 예측한 기계학습모델들 중 가장 재현율이 높았던 알고리즘인 decision tree를 사용한 경우와 비슷한 수준으로 나타났다(Cho와 Cho, 2020).

향후 모델의 재현율 개선을 위해 시험용 데이터 세트의 부적합데이터 15건에 대해 각 특성변수의 상태값을 부적합 항목과 함께 Table 5에 제시하였다. 2022년 1월에서 8월까지 양식뱀장어 안전성 조사결과로 구성된 테스트용 데이터 세트 중 부적합으로 적발된 15건의 부적합항목은 oxolinic acid 3건, dehydrocholic acid 8건, methylene blue 1건, 그리고 praziquantel 3건이었다. BN 모델을 적용한 결과 oxolinic acid 부적합은 3건이 성공적으로 예측되었으나, 나머지 항목에 대해서는 부적합을 예측하지 못하였는데 이는 학습데이터 내에는 methylene blue와 praziquantel

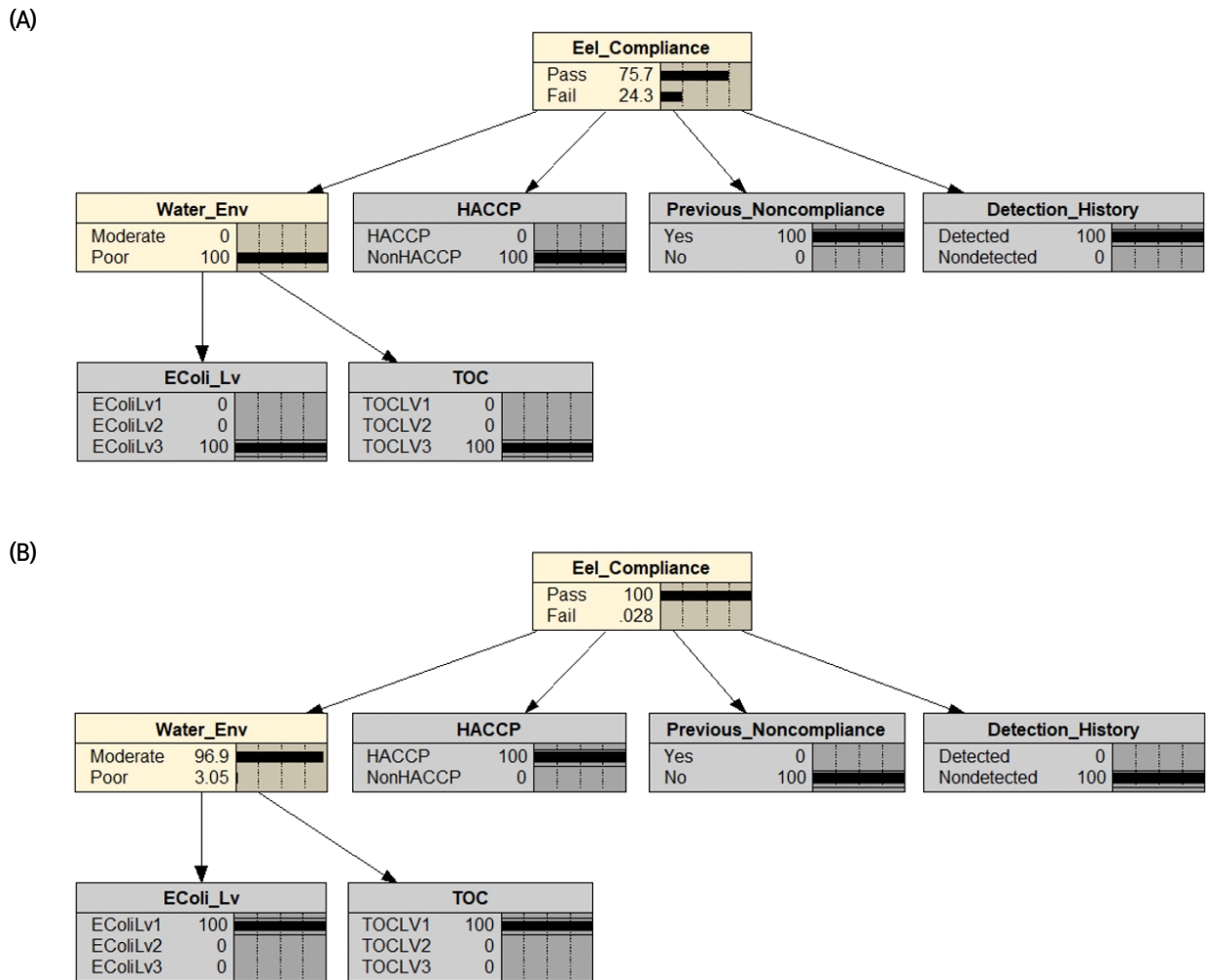


Fig. 4. Predicted probabilities of farmed eel compliance predicted by level of explanatory variables. The highest probability of noncompliance will occur in scenario (A), while the lowest probability of noncompliance will be observed in scenario (B).

Table 4. Confusion matrix of nonconformed eel items in test data set predicted by Bayesian network model

		Actual value	
		Fail (true)	Pass (false)
Predicted value	Fail (true)	3 (true positive)	16 (false positive)
	Pass (false)	12 (false negative)	6,754 (true negative)

의 부적합건이 없었고, dehydrocholic acid의 경우 2018년부터 검사를 시작하여 이전에는 검출이력이 존재하지 않고 '21년도에도 3년 동안의 검출이력만이 있어 충분한 데이터의 축적이 이루어지지 못한 데 기인한다. 이러한 본 BN 모델의 한계는 향후 뱀장어 안전성 조사 결과를 훈련용 데이터세트에 지속적으로 추가함으로써 개선될 것으로 기대

된다.

4. 요약

뱀장어 생산단계 안전성조사 부적합여부에 영향을 미치는 특성변수를 베이지안 네트워크(BN) 모델을 적용하여 분

Table 5. Nonconformity prediction results of Bayesian network model for noncompliance cases in the test dataset

Test item	Explanatory variables			Model prediction
	Detection history	Previous noncompliance	HACCP	
Dehydrocholic acid	Not Detected	Yes	HACCP	Not predicted
Dehydrocholic acid	Not Detected	No	HACCP	Not predicted
Dehydrocholic acid	Not Detected	No	NonHACCP	Not predicted
Dehydrocholic acid	Not Detected	No	NonHACCP	Not predicted
Dehydrocholic acid	Not Detected	No	NonHACCP	Not predicted
Dehydrocholic acid	Detected	No	NonHACCP	Not predicted
Dehydrocholic acid	Not Detected	No	NonHACCP	Not predicted
Dehydrocholic acid	Not Detected	No	NonHACCP	Not predicted
Dehydrocholic acid	Not Detected	No	NonHACCP	Not predicted
Oxolinic acid	Detected	Yes	NonHACCP	Predicted
Oxolinic acid	Detected	Yes	NonHACCP	Predicted
Oxolinic acid	Detected	Yes	NonHACCP	Predicted
Praziquantel	Not Detected	No	HACCP	Not predicted
Praziquantel	Not Detected	No	NonHACCP	Not predicted
Praziquantel	Not Detected	No	NonHACCP	Not predicted

석하였다. 2012년부터 2021년까지의 통합식품안전정보망 (IFSIN)의 뱀장어 생산단계 안전성조사 데이터에 양식장의 HACCP 정보, 지리적 정보 및 용수환경 데이터를 연계하여 BN 모델을 수립하였다. 뱀장어의 부적합여부에 영향을 주는 특성변수로 양식장의 HACCP 인증여부, 양식장의 이전 5년간 검사대상 유해물질의 검출여부, 해당 양식장의 이전 5년간 부적합적발이력, 사용되는 용수환경의 적정성이 제안되었으며, 이때 용수환경의 적정성은 총대장균과 총유기탄소량으로부터 산출되었다. 뱀장어 부적합이 발생할 확률이 가장 높은 경우는 지난 5년간 검사대상 유해물질의 검출이력이 있으면서 동시에 부적합 적발 이력이 있는 HACCP 인증을 받지 않은 양식장으로서, 용수환경도 총대장균 또는 총유기탄소가 높아 오염이 의심되는 용수를 사용하는 경우로 이때 부적합이 발생할 확률은 24.5%로 뱀장어 생산단계 안전성 조사 시 부적합률인 0.26%의 94배 높았다. 2022년 1월부터 8월까지 뱀장어 양식장 안전성조사 결과를 시험용 데이터세트(6,785건 중 부적합 15건)로 하여 BN 모델의 적정성을 검토하였다. 영향강도가 높았던 설명변수인 HACCP, 검출이력, 부적합이력으로 구성된 BN 모델을

시험용 데이터세트에 적용한 결과 부적합일 확률이 15.8%로 시험용데이터의 부적합률인 0.22%의 약 71.4배 개선할 수 있었다. 그러나 이 모델의 재현율은 0.2에 머물렀는데, 이는 특히 부적합항목인 유해물질의 기준·규격이 신설되어 해당 양식장에서 검사기록이 없는 경우와, 매우 드물게 발생하여 10년 동안 검출이력이 없어 학습데이터세트에는 없는 경우이었다. 베이저안 네트워크를 적용하여 부적합확률이 높은 생산단계 안전성 조사대상을 선정하게 되면 설명변수별로 시나리오에 따라 부적합확률을 설명가능하게 되어 다른 머신러닝 알고리즘을 적용하는 경우 지적되어온 설명 불가능이라는 문제점을 해소할 수 있으며, 향후 안전성조사 데이터 축적 시 용이하게 모델 업데이트가 가능하며 이를 통해 모델의 예측성능개선도 기대할 수 있다는 장점이 있다.

Acknowledgements

본 연구는 2022년 식품안전정보원 국가식품안전관리체계 선진화연구사업의 일환으로 수행되었음.

Conflict of interests

The authors declare no potential conflicts of interest.

Author contributions

Conceptualization; Data curation; Formal analysis; Methodology; Validation; Writing: Cho SY.

Ethics approval

This article does not require IRB/IACUC approval because there are no human and animal participants.

ORCID

Seung Yong Cho (First & Corresponding author)
<https://orcid.org/0000-0003-2909-4250>

References

- Bouzembrak Y, Camenzuli L, Janssen E, Van Der Fels-Klerx HJ. Application of Bayesian networks in the development of herbs and spices sampling monitoring system. *Food Control*, 83, 38-44 (2018)
- Bouzembrak Y, Marvin HJP. Prediction of food fraud type using data from rapid alert system for food and feed (RASFF) and Bayesian network modelling. *Food Control*, 61, 180-187 (2016)
- Bouzembrak Y, Marvin HJP. Impact of drivers of change, including climatic factors, on the occurrence of chemical food safety hazards in fruits and vegetables: A Bayesian network approach. *Food Control*, 97, 67-76 (2019)
- Cho S, Cho SY. Improving efficiency of food hygiene surveillance system by using machine learning-based approaches. *JBD*, 5, 53-67 (2020)
- Cho SY. A Study on the Domestic and Foreign Data Utilization Trends for Ensuring Fishery Product Safety. NFSI Research Report. National Food Safety Information Service (2021)
- FAO. The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation. Rome, FAO (2022)
- Giles H. Using Bayesian networks to examine consistent trends in fish farm benthic impact studies. *Aquaculture*, 274, 181-195 (2008)
- Hunte JL, Neil M, Fenton NE. A causal Bayesian network approach for consumer product safety and risk assessment. *J Safety Res*, 80, 198-214 (2022)
- Jensen FV, Nielsen TD. Bayesian Networks and Decision Graphs. Springer, New York, USA, p 32-35 (2007)
- Kim DH, Lee HG, Lim BS, Park SW. Present status of diseases detected from cultured eel, *Anguilla japonica* in Chungcheong and Jeolla provinces during 2004-2017. *JFMSE*, 31, 884-892 (2019)
- Kjaerulff UB, Madsen AL. Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis. Springer, New York, USA, p 8-14 (2010)
- KOSIS. Per capita Korean fish consumption. Available from: https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx_cd=1317. Accessed Mar. 21, 2023.
- Lee JS, Kim DY. The current status and future directions of Korean inland freshwater aquaculture. *J Fish Bus Adm*, 37, 1-24 (2006)
- Lee K, Bak Y, Shin Y, Sohn K, Kwon O. A method of machine learning-based defective health functional food detection system for efficient inspection of imported food. *J Intell Inform Syst*, 28, 139-159 (2022)
- Lupo C, Le Bouquin S, Balaine L, Michel V, Péraste J, Petetin I, Colin P, Jouffe L, Chauvin C. Bayesian network as an aid for food chain information use for meat inspection. *Pre Vet Med*, 109, 25-36 (2013)
- Meng H, An X, Xing J. A data-driven Bayesian network model integrating physical knowledge for prioritization of risk influencing factors. *Process Saf Environ*, 160, 434-449 (2022)
- MFDS. Chapter 2. Common standards and specifications for general. In: Food Code (No.2021-54) (2021)
- NFQS. Fishery product safety inspection results 2021. Available from: <http://www.nfqs.go.kr/hpmsg/board/actionBoardDetail.do?menuId=M0000402>. Accessed Mar. 22, 2023.

Soon JM. Application of Bayesian network modelling to predict food fraud products from China. Food Control, 114, 107232 (2020)

Statistics Korea. Fishery Households by Aquaculture Species/Area of Culture Farm. 2020 Agriculture, Forestry & Fishery Census Report: Inland Water

Fishery, 2 (2021)

Wang X, Bouzembrak Y, Marvin HJP, Clarke D, Butler F. Bayesian networks modeling of diarrhetic shellfish poisoning in *Mytilus edulis* harvested in bantry bay, ireland. Harmful Algae, 112, 102171 (2022)