

기후 변화 적응을 위한 벡터매개질병의 생태 모델 및 심층 인공 신경망 기반 공간-시간적 발병 모델링 및 예측

김상윤[‡] · 남기전[‡] · 허성구[‡] · 이선정 · 최지훈 · 박준규 · 유창규[†]

경희대학교 공과대학 환경학 및 환경공학과
 17104 경기도 용인시 기흥구 덕영대로 1732
 (2020년 1월 24일 접수, 2020년 3월 3일 수정본 접수, 2020년 3월 6일 채택)

Spatio-Temporal Incidence Modeling and Prediction of the Vector-Borne Disease Using an Ecological Model and Deep Neural Network for Climate Change Adaption

SangYoun Kim[‡], KiJeon Nam[‡], SungKu Heo[‡], SunJung Lee, JiHun Choi, JunKyu Park and ChangKyoo Yoo[†]

Department of Environmental Science and Engineering, College of Engineering, Kyung Hee University,
 1732, Deogyong-daero, Giheung-gu, Yongin-si, Gyeonggi-do, 17104, Korea
 (Received 24 January 2020; Received in revised form 3 March 2020; accepted 6 March 2020)

요 약

본 연구에서는 발병 횟수가 빠르게 증가하고 있는 벡터매개질병(vector-borne disease) 중 하나인 쯔쯔가무시증의 발병 특성을 공간적 그리고 시간적으로 분석하고 기후변화 시나리오에 따른 미래 발병 특성을 예측하였다. 쯔쯔가무시증의 공간적 분포와 발병률을 예측하기 위하여 환경 그리고 사회 변수의 공간적 특성을 이용하여 maximum entropy (MaxEnt) 생태 모델을 구성하고, 주요 변수의 쯔쯔가무시증 발병에 관한 상관관계를 분석하였다. 공간 특성 중 환경 변수인 고도 및 기온이 주요한 변수로 분석되었으며, 이는 쯔쯔가무시증의 매개체인 털진드기의 생육 환경과 주요 관련이 있는 것으로 나타났다. 쯔쯔가무시증의 시간적 발병 횟수는 심층 인공 신경망 모델기반 예측을 하였으며, 특히 쯔쯔가무시증의 주요 특성인 지연 효과를 고려하여 모델을 구성하였다. 심층 인공 신경망을 이용한 예측 결과 여름철의 기온, 강수량, 그리고 습도가 털진드기의 활동에 주된 관련이 있으며 가을철의 쯔쯔가무시증 발병 횟수에 영향을 끼치는 것으로 확인 되었다. 또한, 기존 통계적 예측 모델과 비교하였을 때, 심층 인공 신경망 기반 예측 모델의 예측 정확성이 우수함을 확인하였다. 공간적 그리고 시간적 모델에 기후 변화 시나리오를 이용하여 2040년의 쯔쯔가무시증 발병 특성을 예측한 결과, 최대 발병률이 8% 증가, 발병률이 높은 지역이 9% 확대, 그리고 주된 발병 기간이 2개월 증가하였다. 본 연구 결과를 통해 쯔쯔가무시증의 공간적 및 시간적 발병 특성 분석을 통하여, 공중보건 측면에서 벡터매개질병 발병 요인 규명을 통해 주민 건강을 위한 질병 관리 및 예측에 기여할 수 있을 것으로 기대한다.

Abstract – This study was carried out to analyze spatial and temporal incidence characteristics of scrub typhus and predict the future incidence of scrub typhus since the incidences of scrub typhus have been rapidly increased among vector-borne diseases. A maximum entropy (MaxEnt) ecological model was implemented to predict spatial distribution and incidence rate of scrub typhus using spatial data sets on environmental and social variables. Additionally, relationships between the incidence of scrub typhus and critical spatial data were analyzed. Elevation and temperature were analyzed as dominant spatial factors which influenced the growth environment of *Leptotrombidium scutellare* (*L. scutellare*) which is the primary vector of scrub typhus. A temporal number of diseases by scrub typhus was predicted by a deep neural network (DNN). The model considered the time-lagged effect of scrub typhus. The DNN-based prediction model showed that temperature, precipitation, and humidity in summer had significant influence factors on the activity of *L. scutellare* and the number of diseases at fall. Moreover, the DNN-based prediction model had superior performance

[†]To whom correspondence should be addressed.

E-mail: ckyoo@khu.ac.kr

[‡]이 논문은 POSTECH 이인범 교수님의 정년을 기념하여 투고되었습니다.

[‡]The first, second, and third authors have identical collaboration in this research paper.

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.

compared to a conventional statistical prediction model. Finally, the spatial and temporal models were used under climate change scenario. The future characteristics of scrub typhus showed that the maximum incidence rate would increase by 8%, areas of the high potential of incidence rate would increase by 9%, and disease occurrence duration would expand by 2 months. The results would contribute to the disease management and prediction for the health of residents in terms of public health.

Key words: Scrub typhus, Maximum entropy model, Deep neural network, Climate change, Public health

1. 서 론

벡터매개질병은 곤충 및 설치류 등의 감염 매개체 생물로부터 인간에게 전염되는 질병으로, 계속하여 발생률과 분포가 크게 증가하고 있다. 전 세계의 전염병 중 벡터매개질병은 약 17%를 차지하고 있으며, 이 질병에 대해 치료를 받지 못할 경우 환자의 외형 손상 및 장애를 일으킬 가능성이 높다. 이에 따라서 세계보건기구는 벡터매개질병을 국제사회가 대처해야 할 중요한 공중 보건 문제로 지정하였으며, 벡터매개질병을 예측하고 대응하기 위한 방안의 중요성을 강조하고 있다[1,2]. 벡터매개질병은 질병을 옮기는 매개체 생물이 기후 환경에 직간접적으로 영향을 받기 때문에, 변화하는 기후변화의 규모와 강도에 매우 민감한 영향을 받는다. 기후변화로 인한 온도 상승, 날씨 변화, 그리고 매개체 생물의 서식지 영향 등의 환경적 변화가 발생하고, 이에 따른 결과로 벡터매개질병의 발병률은 크게 증대될 것으로 예상된다[3].

쯔쯔가무시증은 벡터매개질병 중 하나로 쯔쯔가무시 병원균에 감염된 털진드기 유충을 매개체로 하여 사람에게 감염되는 질병이다. 쯔쯔가무시증은 세계적으로 연간 약 백만 건의 발병이 발생하며, 적절한 치료가 병행되지 않을 시 약 10%의 사망률을 나타낸다. 이는 대표적 벡터매개질병인 뎡기열 보다 더 많은 사망을 초래할 가능성이 높다[4]. 쯔쯔가무시증은 대부분 시베리아 연해주, 히말라야, 북부 호주, 동아시아 지역 등에서 국지적으로 발생하는 것으로 나타난다. 특히, 쯔쯔가무시증은 한국에서 발병하는 대표적인 벡터매개질병으로 환자의 수가 꾸준히 증가하고 있는 추세이다. 쯔쯔가무시증의 환자 수는 2001년부터 2003년까지 연간 2,700명, 2010년에는 연간 약 5,700명, 2016년에는 연간 약 11,000명으로 발병률이 증가한 것으로 관측된다. 이는 기후변화의 영향으로 인하여 온도 변화 및 서식지의 변화가 나타난 것으로 판단되며, 이로 인하여 최근 10년 동안 환자 수가 약 300% 증가한 실정이다[5]. 따라서 기후 변화에 대응하여 변화하는 온도 및 서식지 환경을 예측하고 그에 따른 쯔쯔가무시증의 발병을 예측하는 것은 공중보건 측면에서 질병 예방 및 주민의 건강 유지를 위하여 중요하다.

기후변화에 따라 증가하고 있는 쯔쯔가무시증의 발병에 대응하기 위하여, 환경 및 사회 변수와 질병의 시간적-공간적 분포 분석을 통하여 쯔쯔가무시증의 분포 분석에 관한 연구들이 진행되어 왔다. 아이젠벡터 공간필터링 모델을 통한 쯔쯔가무시증의 환경 및 사회 요소 간의 상관관계를 분석하여 쯔쯔가무시증의 공간적 분포가 분석이 실시되었다. 쯔쯔가무시증의 발병에서 가을 환경 및 논 경작 인구비율이 주요한 요인으로 나타났다[5]. 하지만, 환경 요소 중 여름과 가을의 정적 상태만을 고려한 환경 요소는 기후변화에 대응하여 동적으로 증가하고 있는 쯔쯔가무시증의 발병을 예측 및 분석하기에 한계점이 있다. 쯔쯔가무시증의 환경 요소 및 사회 요소의 관측된 데이터를 기반으로 상관관계 분석을 통하여 발병 요인이 기후와 토지 이용 변화에 큰 영향을 받는 것으로 연구 되었다. 또한, 토

지 이용 변화에 따라 농촌지역에서의 쯔쯔가무시증 발병률이 증가하는 것으로 나타났다[6]. 하지만, 이는 쯔쯔가무시증의 발병을 시군구 규모에서 분석을 실시하여, 그 이하 규모의 지역에 대한 공간 분포 분석에 대한 한계점이 존재한다.

쯔쯔가무시증과 같은 벡터매개질병은 보건지리학의 측면에서 인간 및 자연환경은 공간적 요인인 토지이용변화 그리고 인구분포변화 등의 영향으로 발생과 확산 가능성이 직간접적으로 변동되는 특징이 있다. 또한, 기후 환경의 변화는 벡터매개질병의 매개체에 직접적인 영향을 미치므로, 쯔쯔가무시증의 발병은 기후변화의 규모와 강도에 매우 민감한 영향을 받는다[2,7]. 게다가, 쯔쯔가무시증은 벡터매개질병의 특성으로 기후 요소와 인간의 활동에 대해 즉각적으로 발병 반응을 보이지 않고, 기후 요소와 인간의 활동 - 질병 매개체 - 질병 발병의 일련의 메커니즘으로 쯔쯔가무시증 발병에 지연 효과(time-lagged effect)의 특성을 가지고 있다[7,8]. 따라서 쯔쯔가무시증의 발병에 관한 연구는 공간 분포 및 시간적 분석이 필수적이며, 또한, 급격한 기후 변화 및 지연 효과를 반영한 쯔쯔가무시증의 시공간적 분포를 고려해야 한다.

본 연구에서는 환경 변수와 사회 변수 중 쯔쯔가무시증 발병에 영향을 미치는 주요한 인자를 분석하고, 이를 통해 쯔쯔가무시증의 발병 확률이 높은 지역 및 시기를 시간적-공간적으로 평가한다. 이를 위해, 생태 모델링 방법인 최대 엔트로피 모델을 이용하여 쯔쯔가무시증의 공간 분포를 구성하고, 심층신경망을 활용하여 쯔쯔가무시증의 시간적 분포를 분석하였다. 추가적으로, 본 연구에서 활용한 최대 엔트로피-심층신경망의 공간적-시간적 모델을 이용하여, 미래 기후 변화 시나리오를 통해 기후변화가 한국에서의 쯔쯔가무시증 발병에 끼치는 영향을 예측 및 평가하였다.

2. 연구 이론

2-1. 쯔쯔가무시증 공간-시간적 특성 분석

쯔쯔가무시증은 쯔쯔가무시 병원균에 감염된 털진드기 유충(Fig. 1(a))이 매개체로 작용되어 사람에게 감염되는 대표적인 벡터매개질병이다[6]. 쯔쯔가무시증의 증상으로는 Fig. 1(b)와 같이 털진드기 유충에게 물린 부위에 가피가 형성되고, 1~3주의 잠복기를 거친 후 감기와 유사한 심한 두통, 발열, 그리고 오한 등의 급성 열성 질환과 림프절이 붓고 간, 비장의 비대 등의 증상이 나타난다[9]. 쯔쯔가무시증의 매개체인 털진드기는 알-유충-자충-성충의 4단계 생활사를 거치고, 돌 밑, 흙덩이, 벽 틈, 땅과 목재의 빈틈 등 은폐된 곳에 산란하며 초가을에 알에서 부화하여 성충이 되기까지 약 3~4달이 소요된다[6]. 털진드기 유충은 성충으로 변태하기 위해 동물의 체액을 반드시 필요하기 때문에 유충 시기에 사람의 몸에 붙어 체액을 섭취하며, 이 과정에서 쯔쯔가무시균이 체내로 유입되어 사람이 쯔쯔가무시증에 감염된다[10]. 따라서 털진드기의 개체가 증가하는 시기와 털진드기의 성장 과정 중 유충 시기에 사람에게 쯔쯔

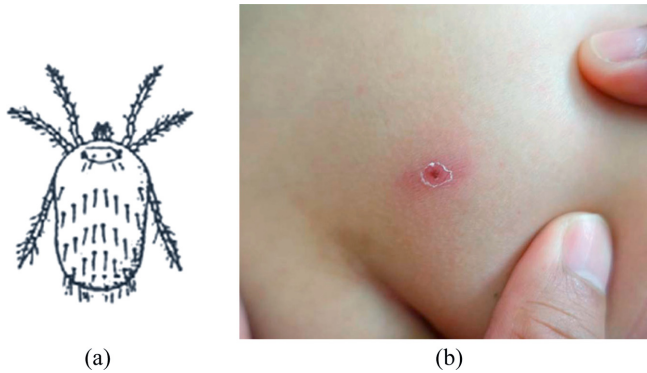


Fig. 1. (a) Picture of *L. scutellare* the main vector of scrub typhus and (b) eschar by the Scrub typhus [6,9].

가무시증을 감염시키는, 기후자료와 찌르가무시증 발병의 지연 효과 존재한다[8].

찌르가무시의 공간적 특성으로, 한국에서는 찌르가무시 매개체 중 털진드기가 전국적으로 우점하며, 특히 인구 10만 명 당 찌르가무시증 발병 횟수가 높은 지역과 털진드기의 비율이 높은 지역이 일치하는 것으로 나타난다. 따라서 찌르가무시증이 발병하는 위치와 털진드기의 출현 지역 간의 높은 지역적 상관관계가 있다[6,10]. 찌르가무시증의 시간적 특성 측면에서는, 앞서 언급과 같이 사람이 털진드기의 증가 시기와 찌르가무시증의 감염으로 인한 급성 열성 질환들의 발병 시기에 지연 효과 특성을 가진다[7]. 또한, 찌르가무시증의 발병은 기후적 측면에서 환자의 90% 이상이 10~11월인 가을에 집중되고, 주된 감염 발생 지역은 털진드기의 서식지 특성으로 인하여 농촌에서 주로 발생하는 경향이 있다[10,11].

2-2. 생태 정보 기반 최대 엔트로피 모델(MaxENT)

생태 모델은 기후 변화에 따른 생태계의 예측 및 동식물 종의 효율적인 생태 관리를 위한 유용한 모델이다(Fig. 2). 공간적 정보를 바탕으로 대상으로 하는 동식물 종의 분포를 생태 모델을 통해 추정할 수 있는 장점이 있다[12,13]. 본 연구에서는 생태 모델 중 하나인 종 분포 모델을 응용하여 털진드기 정보 기반 찌르가무시증의

발병 확률 분포를 예측하였다. 종 분포 생태 모델은 크게 데이터 유형 및 모델의 학습 원리에 따라 분류된다[14]. 먼저, 사용하는 데이터 유형에 따라 대상 종의 관측된 출현 정보 데이터를 주 도메인으로 이용하는 artificial neural network (ANN), generalized linear model (GLM), 그리고 generalized additive model (GAM) 등의 생태 모델이 있다. 그리고 종의 출현 정보와 환경 정보를 동시에 사용하는 maximum entropy model (MaxENT)와 ecological niche factor analysis (ENFA) 모델로 분류된다. 모델의 학습 원리에 따라서는 통계적 방법과 기계적 학습 방법으로 나뉜다. 통계적 방법 기반 생태 모델은 관측 출현 정보와 환경 변수를 연관시켜, 대상 생물 종에게 적합한 환경조건을 추정하고 그를 통해 출현 정보를 예측하는 GLM과 GAM 모델이 있으며, 기계적 학습 방법 기반 생태 모델은 환경 조건의 변화에 따른 생물 종의 물리적 변화를 고려하는 ANN과 MaxENT 모델이 있다[14].

본 연구에서 분석하는 찌르가무시증의 경우 털진드기의 출현 정보 및 기후 환경 정보에 큰 영향을 받으며, 이에 따라 찌르가무시증의 분석을 위해서는 털진드기의 생물학적 특성을 먼저 고려해야한다[5]. 따라서 본 연구에서는 찌르가무시증과 털진드기의 공간적 변수 내의 사회적 그리고 기후 변수의 상관관계를 고려하고 관련 환경 정보를 활용하기 위해서, 털진드기 출현 정보와 환경 정보를 활용할 수 있는 MaxENT 모델을 주 생태 모델로 선택하여 사용하였다. MaxENT 모델은 출현 정보와 환경정보를 활용 할 경우 ENFA 보다 더 높은 생태 모델링 능력을 나타낸 바 있다[15]. 본 연구에서는 생태 모델인 MaxENT를 이용하여 털진드기의 출현 정보 및 기후 환경 변수의 상관관계를 고려하고 그에 따른 털진드기 출현 확률을 정량적으로 추정한다.

MaxENT는 환경 정보에 종속 되어있는 연구 대상 변수의 출현 정보를 이용하여 연구 대상 변수의 출현 확률 분포 모델을 구성할 수 있다[16,17]. 따라서 털진드기의 공간적 출현 정보로 찌르가무시증의 발병 위치를 유추하고 찌르가무시증의 발병에 영향을 미치는 환경 변수와 대상 지역의 사회 변수를 활용하면 찌르가무시증의 발병 자료가 없는 공간에 대한 찌르가무시증의 발병 확률을 예측할 수 있다. 또한 MaxENT는 단위 크기의 정사각형(cells)에서 연구 대상 종의 출현 확률을 계산하기 때문에 행정구역에 상관없이 찌르가무시증의 발병 자료가 없는 지점의 찌르가무시증 발병 확률을 계

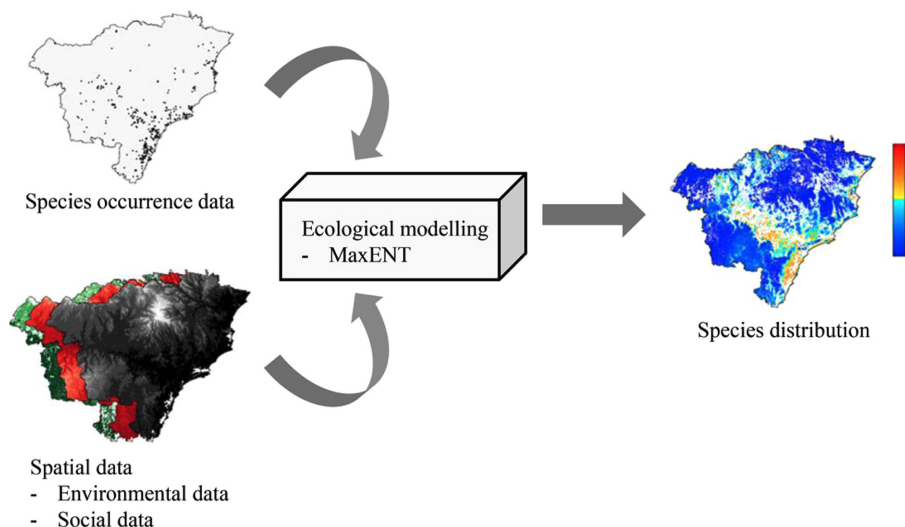


Fig. 2. Graphical description of spatio-temporal ecological modeling modified from [12].

산할 수 있는 장점이 있다[14]. MaxENT의 모델링은 식 (1)-(4)와 같이 나타낼 수 있다[17,18]:

$$H(\pi) = -\sum_{x \in X} \pi(x) \ln \pi(x) \quad (1)$$

$$q_{\lambda}(x) = \frac{\exp(\lambda \cdot f(x))}{Z_{\lambda}} \quad (2)$$

$$|-\ln(q_{\lambda})| = \sum_{x \in X} \pi(x) \cdot (-\ln(q_{\lambda})) \quad (3)$$

$$\pi = |-\ln(q_{\lambda})| + \sum_j B_j |\lambda_j| \quad (4)$$

여기서 H 는 대상 종 출현에 따른 엔트로피, π 는 대상 종의 출현 확률, x 는 대상 종의 출현 위치, $q_{\lambda}(x)$ 는 Gibbs 분포를 가지는 π 의 확률 분포, $f(x)$ 는 출현 위치에 따른 특성 벡터, λ 는 각 변수들에 대한 가중치 벡터, 그리고 Z_{λ} 는 총 $q_{\lambda}(x)$ 의 합이 1이 되게 하는 정규화 상수다. 이때, q_{λ} 는 식 (3)에서 x 에 따른 로그우도의 음의 값이 최소가 되도록 결정되며, 최종적으로 과적합을 방지하기 위한 상수 B_j 가 추가되면 MaxENT 확률 분포를 식 (4)와 같이 나타낼 수 있다.

2-3. 심층 인공 신경망(Deep neural network, DNN)

인공 신경망 기반의 심층 신경망 모델은 비선형적이고 비정상성의 특성을 가지는 데이터를 예측하기 위해 많은 분야에서 적용되고 있다. 심층 인공 신경망은 기존 시계열 데이터 예측모델에 비하여 데이터의 숨은 경향을 파악하고, 데이터 간의 복잡한 연관성을 해석하여 미래 데이터 예측 성능이 높은 특성을 가진다. 본 연구에서는 심층 인공 신경망을 적용하여 월별 찌꺼기무시증의 발병 횟수를 예측하였다. 심층 인공 신경망은 자료를 입력하는 입력층, 결과를 나타내는 출력층, 입력층과 출력층 사이에 여러 개의 은닉층, 그리고 각각의 층을 구성하는 노드로 구성되어 있다. 심층 인공 신경망 모델은 각 노드의 관계를 나타내는 가중치가 있고, 이 가중치의 값을 반복적으로 학습하여 데이터 예측 성능을 증가시킨다. 가중치의 값을 학습하는 방법으로 오류 역전파 알고리즘을 사용하며, 이 알고리즘은 계산 방향이 가중치가 연결된 방향과 반대로 오류를 줄이도록 값이 전파되고, 이는 기존의 방식에 비해 계산 시간이 크게 단축되며 오버 피팅의 문제를 극복할 수 있는 장점이 있다[19]. 심층 인공 신경망에서 l 번째 층을 계산하는 방법은 다음과 같다:

$$z^l = \tau_l(U^l z^{l-1}), \quad 1 \leq l \leq L \quad (5)$$

이 때, z^l 는 l 번째 층의 노드 값을 나타내는 벡터, τ_l 은 l 번째 층의 활성화 함수, 그리고 U^l 은 l 번째 층의 가중치 행렬을 나타낸다.

2-4. 미래 기후 변화 시나리오(Representative concentration pathway, RCP)

Representative concentration pathway (RCP)는 미래 기후 변화의

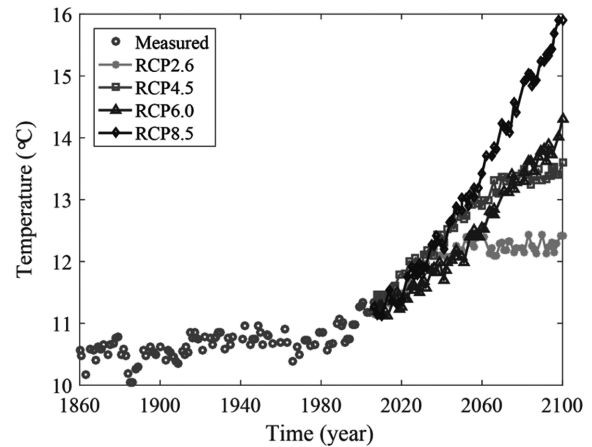


Fig. 3. Annual predicted temperatures according to representative concentration pathway 2.6, 4.5, 6.0, and 8.5 scenarios [20,21].

강도에 따른 시나리오로, 사회 및 경제적 인간 활동이 대기 중의 온실가스 농도에 미치는 영향에 따라 구성되어 있다. RCP는 사회 및 경제적 측면과 시간에 따른 변화에 따라 여러 가지 시나리오로 나뉘게 된다. Table 1과 같이 RCP 시나리오는 미래 온실가스 농도에 의한 온실효과에 따라 크게 네 가지로 분류된다. 첫째로, RCP 2.6은 인간 활동에 의한 대기 중의 온실 가스 발생을 지구의 자체 자정 능력으로 회복하는 시나리오를 나타내며, 둘째로, RCP 4.5는 온실가스 저감 정책이 상당히 실현될 경우의 기후 변화 시나리오, 셋째로, RCP 6.0은 온실가스 저감 정책이 어느 정도 실현될 경우를 나타내고, 넷째로, RCP 8.5는 온실가스 저감 정책 없이 현재의 온실 가스 배출 추세가 지속될 경우의 시나리오를 나타낸다. 여기서 RCP 시나리오의 각 숫자는 복사 강제력을 나타내는 것으로, 온실가스 등으로 에너지의 평형을 변화시키는 영향력의 정도를 W/m^2 로 표시한 것이다[20,21]. 온실가스 농도에 따른 평균기온은 지속적으로 상승할 것으로 예측되며 시간에 따른 연평균기온의 편차는 Fig. 3과 같다. RCP 2.6 그리고 4.5 시나리오의 경우 21세기 중 후반까지 지속적으로 상승하다 일정하게 유지되며, RCP 6.0과 8.5 시나리오 경우 2100년 까지 지속적으로 온도가 증가하는 것으로 예측된다. 본 연구에서는 국내 기후 변화 정책의 정도를 반영하여, 기후 변화 대응 정책이 어느 정도 수행되어 2100년에 복사 강제력이 $6.0 W/m^2$ (이산화탄소 농도 670 ppm)이 될 경우를 나타내는 RCP 6.0 시나리오를 사용하였다.

3. 연구 방법

공간적 생태 모델과 시간적 심층 신경망 모델을 이용하여 찌꺼기무시증의 공간적-시간적 발생 확률을 파악하고, 기후변화 시나리오에 따른 찌꺼기무시증의 발생률을 예측하기까지의 제시된 연구 순

Table 1. Descriptions of representative concentration pathway-climate change scenarios according to the intensities of greenhouse effects

Scenarios	Greenhouse effects (CO ₂ concentration) in 2100	Expected environment conditions in 2040			
		Temperature (°C)	Precipitation (mm)	Humidity (%)	Wind speed (m/s)
RCP2.6	2.6W/m ² (420 ppm)	12.5	1232.3	70.1	1.9
RCP4.5	6.0W/m ² (540 ppm)	12.3	1096.9	69.1	1.9
RCP6.0	6.0W/m ² (670 ppm)	11.8	1164.3	69.9	1.9
RCP8.5	8.5W/m ² (940 ppm)	12.3	1153.5	69.4	1.9

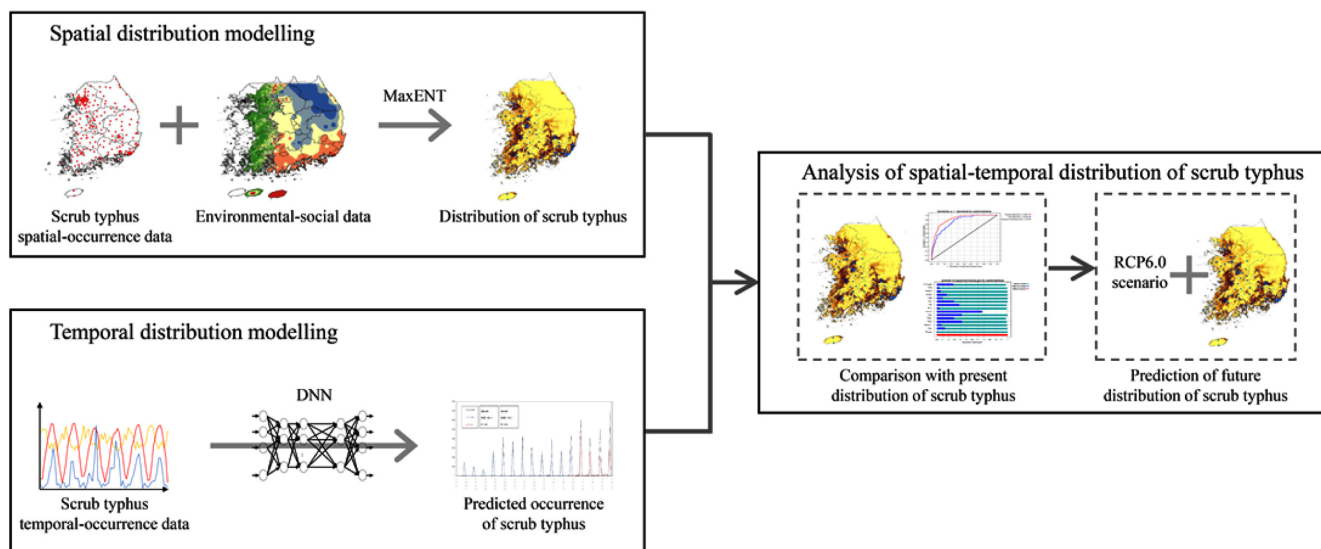


Fig. 4. Graphical scheme of the spatial and temporal distribution analysis for scrub typhus.

서는 Fig. 4와 같다. 먼저, 쯔쯔가무시증의 공간적 모델을 생성하기 위하여 행정 구역에 따른 쯔쯔가무시증의 발병 횟수 그리고 털진드기의 서식 환경과 연관되는 해당 행정 구역의 환경 변수와 사회 변수를 수집하였다. 다음으로, 수집된 공간 자료에서 사회-환경 변수를 이용하여 생태 모델인 MaxENT 모델을 구성하여 쯔쯔가무시증의 공간적 발병 확률 분포 지도를 나타내고, 측정된 쯔쯔가무시증의 공간적 발병 데이터와 비교하여 모델의 성능을 area under the curve (AUC)를 이용하여 평가하였다. 추가적으로, 사용된 사회 변수 및 환경 변수에서 주요한 변수를 분석하기 위하여, Jackknife test의 응답 곡선을 통해 변수에 따른 쯔쯔가무시증의 발병 영향 정도를 비교 분석하였다.

두 번째로, 쯔쯔가무시증의 시간에 따른 발병 예측을 위하여 월별 쯔쯔가무시증 발병 횟수와 기후 환경 변수를 이용하여 심층 신경망 기반 예측 모델을 구성하였다. 통계적 시계열 모델과 단층 신경망은 모델의 복잡성, 구조 선택 문제성, 데이터 내 복잡한 경향 추출의 한계점을 가지는 반면, 심층 인공 신경망은 고도의 데이터 추출 성능을 통해 데이터 내의 복잡한 특징을 반영하여 예측 성능을 증가시키는 장점이 있다[22]. 예측 모델에서 쯔쯔가무시증의 주요한 특징 중 하나인 지연 효과를 반영하기 위하여 기존 연구결과[17]를 참고하여 0~6개월의 지연 효과를 고려하였다. 발병 지연 효과 중 적절한 지연 효과 기간을 선정하기 위하여 피어슨 상관관계 (Pearson's correlation) 분석을 이용하였다. 환경 변수에 0~6개월의 시간 지연을 각각 적용시키고, 월별 쯔쯔가무시증 발병 증감에 대한 영향을 알기 위한 상관관계 분석을 실시하였다. 이를 통해 환경 변수와 쯔쯔가무시증의 발병에서 상관관계 값이 가장 큰 기간을 환경 변수에 따른 털진드기의 생육환경 및 쯔쯔가무시증 발병 변화에 주요한 지연 변수로 선택하였다. 따라서 본 연구에서는 심층 인공 신경망 모델에서 지연 효과를 고려한 기후 환경 변수를 입력 데이터로 하여 추후 월별 쯔쯔가무시증의 발병 횟수를 예측하였다. 심층 인공 신경망 기반 쯔쯔가무시증의 월별 발병 횟수 예측 모델의 평가는 측정값과 모델 기반 예측 값을 이용하여 결정 계수(coefficient of determination, R^2)와 평균 제곱근 오차(root mean squared error, RMSE)를 계산하여 실시하였다.

마지막 단계로, 미래의 쯔쯔가무시증 질병 예방 및 주민 건강 유지를 위한 공중보건적 목적을 위하여 미래의 쯔쯔가무시증 발병 정보를 예측하였다. 미래의 환경 조건 시나리오는 RCP6.0을 이용하였다. 온실가스 저감 정책이 적용되지 않을 경우의 2040년 대기 중 CO_2 농도와 비교 해볼 때, 온실 가스 저감 정책이 적용된 RCP6.0 시나리오는 CO_2 농도를 약 270 ppm 저감 시키게 된다. 이는 현재 한국의 온실가스 저감 정책에 상응하는 미래 기후변화 시나리오다. 따라서 RCP6.0 시나리오를 기반으로 2040년의 한국의 연 평균 기후 조건과 월 평균 기후 조건을 공간적 생태 모델인 MaxENT와 시간적 모델인 심층 인공 신경망 모델에 각각 사용하였다. 이에 따른, 2040년의 쯔쯔가무시증의 발병 장소 및 발병 확률의 변화를 현재 기후 조건에서의 쯔쯔가무시증 발병 특성과 비교 분석하였다.

3-1. 공간-시간적 데이터 수집

본 연구에서 사용한 환경적 그리고 사회적 변수들을 Table 2에 나타내었다. 쯔쯔가무시증의 공간 분포를 나타내는 생태 모델 MaxENT와 시간에 따른 발병 횟수를 나타내는 심층 인공 신경망 예측모델을 개발하기 위해서 환경 변수와 공간 변수를 독립 변수로 수집하였으며, 2001~2018년 사이의 행정 구역별 쯔쯔가무시증 발병 횟수의 확정 자료를 종속 변수로 지정하였다. 쯔쯔가무시증의 공간 분포를 나타내는 MaxENT 모델은 Table 2와 같이 환경 변수 및 사회 변수를 입력 값으로 사용하며, 행정 구역에 따른 데이터의 총 관측수는 3,600개를 사용하였다. 쯔쯔가무시증의 발병 횟수를 예측하는 심층 인공 신경망 모델의 경우, 18년 동안 수집된 1,296개의 관측 데이터를 사용하였다. 수집 데이터 중 환경 변수의 경우 쯔쯔가무시증의 주 매개체인 털진드기의 활동에 영향을 끼치는 것으로 알려진 기온, 강수량, 습도, 풍속, 그리고 고도의 환경 정보를 수집하였다[5,6]. 이때, 환경 정보는 기상청에서 제공하는 각 지역별 2001~2018년의 월별 평균 자료로 수집하였으며, 환경 변수 중 지형 정보를 나타내는 고도는 shuttle radar topography mission (SRTM)을 통해 수집하였다. 본 연구에서 각 행정구역별로 측정된 데이터의 세부 공간적 단점을 극복하고자, 세부 공간별 환경 정보는 거리 반비례 가중치(inverse distance weighted, IDW)법을 이용하여 계산

Table 2. Environmental and social variables used to model the scrub typhus spatial distribution model and the scrub typhus prediction model

	Variables	Unit	Abbreviation	Mean	Standard deviation	Observations for model 1	Observations for model 2	Data source	Model
Environmental variables	Maximum temperature	°C	Tmax	18.08	1.13	225	216	Korea Meteorological Administration	1,2
	Average temperature	°C	Tavg	12.39	1.31	225	216		1,2
	Minimum temperature	°C	Tmin	7.49	1.66	225	216		1,2
	Precipitation	mm	Prec	109.5	12.03	225	216		1,2
	Humidity	%	Hum	67.78	3.13	225	216		1,2
	Average wind speed	m/s	WS	1.89	0.50	225	216		1,2
	Elevation	m	El	262.5	248.80	225	-	Consortium for Spatial Information	1
Social variables	Population density	person/km ²	Pop	528	1860	225	-	Korea Statistical Information Service	1
	Ratio of under 15 years of age	%	U15	11.9	2.75	225	-		1
	Ratio of over 65 years of age	%	O65	21.7	8.07	225	-		1
	Ratio of farm population	%	Farm_pop	21.2	12.1	225	-		1
	Ratio of female population	%	Woman	49.7	1.60	225	-		1
	Ratio of park area	%	Park	0.20	0.71	225	-		1
	Ratio of orchard area	%	Orchard	0.59	1.34	225	-		1
	Ratio of paddy area	%	Paddy	7.63	3.31	225	-		1
	Ratio of field area	%	Field	11.4	8.20	225	-		1

Note that the model 1 and 2 indicate the spatial scrub typhus distribution model and the temporal scrub typhus prediction model, respectively.

하였다. IDW는 데이터의 측정 위치와 가까이 있는 위치에 더 큰 가중 값을 부여하여 데이터를 보간하는 방법으로 공간적 위치가 데이터 측정 위치에서 멀어질수록 가중 값을 감소시켜 영향력을 줄인다[23]. 본 연구에서 사용된 사회변수들은 사람이 탈진드기와 접촉하여 쯔쯔가무시증에 감염될 위험성에 영향을 미친다고 알려진 요소인 인구밀도, 15세 이하 인구 비율, 60세 이상 비율, 농촌 인구 비율, 여성 인구 비율, 공원 부지 비율, 과수원 부지 비율, 논 부지 비율, 그리고 밭 부지 비율을 수집하였다[5,6]. 종속 변수인 쯔쯔가무시증 발병 횟수는 질병관리본부로부터 시, 군, 구에 따른 월별 발병 횟수 평균값을 수집하였다. 그리고 생태 모델 MaxENT로부터 공간 분포를 나타내기 위하여, 행정 구역별 월별 쯔쯔가무시증 발병 횟수를 위치 좌표에 따라 지정하였다. 미래 환경 조건은 RCP 6.0 시나리오에 따른 2040년의 기온, 강수량, 습도 그리고 풍속을 기상청을 통해 수집하였으며, 수집된 미래 환경 정보는 각 공간 정보 자료와 월별 자료로 처리하여 생태 모델과 예측 모델에 사용하였다.

3-2. 발병 확률 분포와 MaxENT 생태 모델

쯔쯔가무시증 발병 확률 분포는 공간적 MaxENT 생태 모델을 이용하여 다음과 같이 모델링을 실시하였다. MaxENT 모델 구성시 입력 데이터 3600개 중 75%인 2,700개의 환경 및 사회 변수는 모델을 구성하기 위한 학습(train) 데이터로 사용하고, 나머지 25%인 900개의 변수는 구성된 모델의 성능을 평가하기 위한 테스트(test) 데이터로 사용하였다. 모델의 성능 평가는 통계적 예측 모델 평가인자인 AUC를 이용하여 평가하였다. AUC는 x축에 거짓인 값을 모델이 참이라 잘못 예측한 비율을, y축에 참인 값을 모델이 참이라 옳게 예측한 비율로 나타낸다. 따라서 모델이 실제 참인 값과 거짓인 값을 얼마나 잘 구분하여 예측하는지 보여주는 그래프이다. AUC 그래프의 곡선이 좌상단으로 향할수록 모델이 실제 값을 정확하게 구분하고 예측하고 있음을 의미하고 AUC값이 커지게 되어 예측 모델의 성능이 좋다고 할 수 있다. 특히, AUC의 값이 1.0-0.9이면 매우 우수, 0.9-0.8이면 우수, 0.8-0.7이면 보통, 0.7이하이면 나

쁨으로 평가한다[24]. MaxENT 모델에 입력 값으로 사용된 환경 변수 및 공간-사회 변수의 상대적 중요도 평가는 Jackknife test를 이용하여 평가하였다. Jackknife test는 n개의 변수로 이루어진 집합에서 i번째 변수를 제거하여 새로운 n-1개의 변수로 이루어진 집합을 재배열하여, 원래 집합과 변수 하나가 제거된 집합의 분산의 차이로부터 제거된 변수의 상대적 중요성을 확인하는 방법이다[25]. 선정된 상대적으로 중요한 변수는 추가적으로 쯔쯔가무시증 발병 확률과의 관계를 응답 곡선을 통하여 분석하였다. 응답 곡선은 다른 변수들은 고려하지 않고 해당 변수가 변할 때 쯔쯔가무시증의 발병 확률이 어떻게 변화하는지 분석할 수 있다.

3-3. 쯔쯔가무시증 발병 횟수 예측을 위한 심층 인공 신경망 모델링

심층 인공 신경망을 이용하여 월별 쯔쯔가무시증의 발병 횟수를 파악하기 위한 시계열 모델을 개발하기 위하여 다음과 같은 과정을 실시하였다. 쯔쯔가무시증은 탈진드기로부터 사람에게 균이 옮겨졌을 때와 쯔쯔가무시증의 발병 시기에 지연 효과 특성이 나타난다. 따라서 본 연구에서는, 쯔쯔가무시증의 발병 지연 효과를 고려하기 위해 환경 변수에 0개월부터 6개월까지의 시간 차이를 고려하였다. 0~6개월까지의 발병 지연 효과 및 탈진드기의 활동과 밀접한 관계를 가지는 환경 변수와 쯔쯔가무시증의 월별 발병 횟수의 데이터간의 상관관계를 피어슨 상관관계 검정을 통해 분석하였다. 각 지연 효과 및 환경 변수 그리고 쯔쯔가무시증의 발병 횟수 상관관계 값에서 가장 큰 상관관계 값을 나타내는 지연 효과의 시간을 도출하여, 이를 심층 인공 신경망 기반 예측 모델에서의 발병 지연 효과 입력 값으로 활용하였다. 심층 인공 신경망 기반 예측 모델의 학습을 위해서 2001~2018년 데이터 중 2001~2012년의 864개의 데이터를 모델 학습 자료로 사용하였으며, 특히 모델의 성능을 증가시키기 위하여 2010~2012년의 자료의 내부 검증(internal validation) 자료로 사용하였다. 나머지인 2013~2018년의 432개의 데이터를 모델 테스트 자료로 사용하였으며, 그 중 2017년 이후의 데이터를 외부

검증(external validation) 자료로 사용하였다. 침충 인공 신경망의 내부 은닉층의 개수는 5개의 층으로 선정하였으며, 모델의 학습은 오류 역전파 알고리즘과 미니 배치 경사 하강법을 사용하였다. 추가적으로, 본 연구에서 제안된 침충 인공 신경망 모델의 예측 우월성을 확인하기 위하여 기존 통계적 모델인 부분 최소 제곱 회귀법(partial least squares, PLS) 기반 모델과 비교 분석하였다. 모델의 예측 성능 평가 지표로는 R^2 와 RMSE를 이용하였다. R^2 는 시계열 모델의 적합성을 나타내는 지표이다. R^2 값이 높을수록 예측 모델이 측정된 데이터 표본을 잘 설명할 수 있다는 것을 의미한다. RMSE는 모델을 통한 예측 값과 측정된 데이터의 차이를 계산하게 되며 그 값이 낮을수록 예측 모델의 성능이 우수하다고 할 수 있다. R^2 와 RMSE는 식 (6) 그리고 (7)과 같이 표현할 수 있다[25]:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (7)$$

여기서 y_i 는 i 번째 측정값, \bar{y} 는 측정값의 평균, 그리고 \hat{y}_i 는 예측 모델을 통한 예측 값을 나타낸다.

3.4. 기후변화 시나리오에 따른 쯔쯔가무시증 공간-시간적 분포 예측

본 연구에서는 공간적 생태 모델 MaxENT와 시간적 예측 모델인 침충 인공 신경망을 이용하여 2040년의 쯔쯔가무시증 발병 예측을 다음과 같은 과정을 실시하였다. 기후 변화에 따라 변화된 환경 변수는 RCP 6.0 시나리오를 이용하여 2040년의 기온, 강수량, 습도, 그리고 풍속의 예측 값을 사용하였다. RCP 6.0 시나리오에 따른 2040년의 기후 조건은 평균 기온은 13.2 °C, 강수량은 1346.0 mm, 상대습도는 68.8%, 그리고 풍속은 2.0 m/s로 예측 되었다. 이는, 현재 기후 대비 평균 기온, 강수량, 그리고 상대습도가 각 5.6%, 2.9%, 0.3% 증가하는 것이다. 이에 따라서 폭염 일수, 열대야 일수, 그리고 여름 일수와 같은 고온 관련 지수가 증가할 것으로 전망된다[21]. 기후 변화에 따른 각 지역별 환경 정보를 토대로 앞서 학습된 MaxENT 생태 모델의 구성 파라미터를 이용하여 2040년의 한국에서의 각 지역별 쯔쯔가무시증 발병 확률 분포를 예측하였다. 또한, 학습된 침충 인공 신경망 모델의 구조 및 신경망 가중치를 이용하여 환경 조건 변화에 따른 쯔쯔가무시증의 월별 발병 확률을 예측하였다.

4. 결 과

4.1. MaxENT 생태 모델 기반 쯔쯔가무시증 발병 공간분포 분석

쯔쯔가무시증의 발병 확률 분포 분석을 위한 공간 정보 기반 MaxENT 모델의 결과는 Fig. 5(a)에 도식하였다. MaxENT를 이용한 쯔쯔가무시증 모델은 학습 데이터(AUC=0.90)와 테스트 데이터(AUC=0.86)에서 모두 발병 확률 분포를 잘 예측하는 것으로 나타났다. 따라서 MaxENT 모델로 예측한 쯔쯔가무시증의 발병 확률이 높은 지역과 실제 쯔쯔가무시증의 발병이 일어난 위치를 비교하

였을 때 일치하였다(Fig. 5(a)). 분석 지역 중 쯔쯔가무시증 발병률의 최댓값은 0.74로 나타났으며 발병률 0.7 이상인 지역의 총 면적은 2053.12 km²를 차지한다. 발병률 0.7 이상의 행정 구역은 김해시, 창원시, 합천군, 구례군, 진도군, 전주시, 포항시, 성주군, 부산광역시, 광주광역시, 그리고 울산광역시로 나타났다. 따라서 경남, 전남, 그리고 전북 지역에서 높은 쯔쯔가무시증 발병률이 나타나는데, 이는 쯔쯔가무시증의 주 매개체인 털진드기가 주로 10 °C 이상에서 성장 및 활동하기 때문에 중부 지역에 비하여 주로 남부지역에서 높은 발병률이 예측 된다[5]. 또한, 쯔쯔가무시증의 발병률이 높은 지역은 상대적으로 고도가 낮은 곳에 분포하는데, 이는 인구 분포와 밀접한 연관성을 가지며 털진드기의 생장 중 유충 시기의 체액 섭취가 필수적이므로 인구 밀도가 높은 지역에서 주로 활동하는 것으로 사료된다. 따라서 털진드기의 생육 환경과 사람과의 접촉을 위한 환경에서 주요한 공간적 인자가 무엇인지 평가하는 것은 쯔쯔가무시증의 공간-지역적 발병 특성을 이해하는데 중요하다.

본 연구에서는 공간적 변수 중 환경 변수와 사회 변수의 쯔쯔가무시증 발병 확률 분포에 대한 영향 정도를 Jackknife test를 이용하여 평가하였다. Fig. 5(b)와 같이 Jackknife test 결과 총 16개의 공간 변수 중 사회적 변수인 농가 인구 비율, 공원 면적 비율, 그리고 인구 밀도와 환경적 변수인 고도, 평균 기온, 최고 기온, 그리고 최저 기온의 7개의 변수가 상대적으로 중요한 변수로 나타났다. Jackknife test 결과의 중요한 변수는 쯔쯔가무시증 발병 확률 증가에 상대적으로 높은 기여도를 가지게 된다[16]. 이중 3개의 중요 변수인 농가 인구 비율, 공원 면적 비율, 그리고 인구 밀도는 사회적 변수의 범주에 포함되며 이는 사람이 거주하는 농촌과 도시에서 쯔쯔가무시증에 감염되는 경로를 의미한다. 농촌의 풀밭 및 공원의 잔디는 털진드기 및 숙주가 될 수 있는 설치류가 활동하는데 지표가 포함되어 있는 도시에 비해 적합하기 때문이다. 또한, 농업 작업자가 앉아서 농촌의 풀밭에서 있거나, 시민이 공원의 잔디에 앉아 있기 때문에 털진드기에 보다 쉽게 노출되어 쯔쯔가무시증 발병의 위험성이 증가한다[5]. 인구 밀도의 경우 사람이 털진드기의 직접적인 숙주가 될 수 있기 때문에, 중요한 변수로 선정된 것으로 사료된다. 전체 변수 중 환경적 변수인 고도가 가장 중요한 변수로 나타났는데, 이는 고도가 낮은 지역과 쯔쯔가무시증의 주 매개체인 털진드기의 분포 지역이 일치하기 때문이다. 또한, 환경 변수 중 평균 기온, 최저 기온, 그리고 최고 기온은 털진드기의 생육에 직접적인 영향을 미친다. 털진드기는 땅 속에서 산란하는 특징으로 온도에 민감하며, 20~30 °C의 온도에서 산란하고 15 °C 이하의 온도에서는 산란을 하지 않는다[27]. Jackknife test의 전체적인 분석 결과, 사회 변수보다 환경 변수가 상대적인 중요성이 더욱 높게 평가되었으며, 이를 통해 쯔쯔가무시증의 발병에 인간의 활동보다 털진드기의 생육 환경이 더 중요한 영향을 끼친다고 판단된다.

MaxENT 생태모델을 통한 쯔쯔가무시증의 공간 분포 예측에서 중요한 변수로 선정된 7가지 변수들의 각 영향도를 살펴보기 위해 응답 곡선을 통한 분석을 실시하였다(Fig. 5(c)). 먼저, 사회 변수인 농가 인구 비율의 응답곡선에서는 쯔쯔가무시 발병 확률이 농가 인구 비율이 0.05%에서 발병 확률이 최대가 된 후, 20%에서 최저점을 보인 후, 다시 증가하여 35%에서 높은 발병 확률을 나타냈다. 농가 인구 비율이 낮은 지역에서 오히려 높은 발병 확률을 나타낸 것은 해당하는 지역인 부산 그리고 광주 등 남부 대도시 지역의 영향이 큰 것으로 나타났다. 즉, 농가 인구 비율이 낮더라도 남부에 위치하

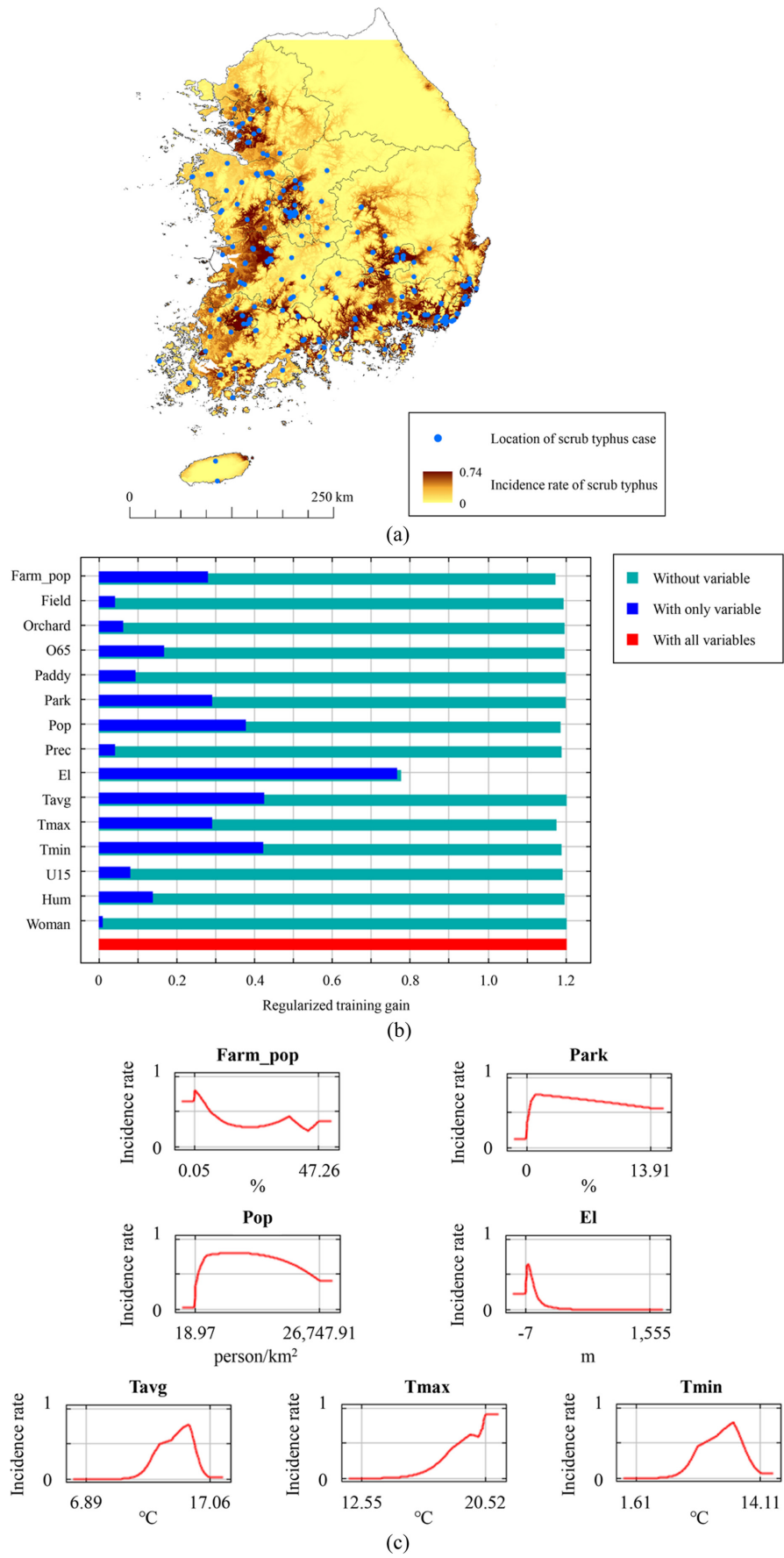


Fig. 5. MaxENT-based spatial distribution of scrub typhus; (a) potential distribution map of scrub typhus, (b) Jackknife test for spatial values, and (c) response curves of influential spatial values on to incidence rate of scrub typhus.

여 털진드기의 생육 환경에 적합한 기후 환경으로 인해 높은 발병 확률을 나타낸 것으로 사료된다. 또한, 농가 인구 비율이 20%인 정선, 철원, 논산, 그리고 연천 등은 낮은 쯔쯔가무시증의 발병 확률로 털진드기가 거의 분포하지 않는 것으로, 농가 인구 비율이 35%인 합천, 청도, 그리고 곡성 등의 남부 농촌 지역에서는 쯔쯔가무시증 발병 확률이 다소 높은 것으로 나타났다. 또한, 농가 인구 비율이 40% 이상인 지역인 영양, 그리고 의성 등의 지역은 오히려 발병 확률이 낮은 것으로 나타났는데, 해당하는 지역은 인구 밀도가 50명/km² 이하인 규모가 작은 농촌 지역이다. 따라서 농가 인구 비율이 높더라도 쯔쯔가무시증의 숙주가 되는 사람의 인구수가 적으면 오히려 쯔쯔가무시증의 발병 확률이 낮게 나타나는 것을 확인할 수 있다. 공원 면적 비율의 응답 곡선에서는 쯔쯔가무시증 발병 확률이 증가했다가 2% 이후부터 감소하는 것을 확인할 수 있다. 이는 도시 지역의 사람이 공원을 이용하다 털진드기에 의해 쯔쯔가무시증에 감염 될 수 있지만, 공원 면적이 증가할수록 도시화된 지역을 나타내므로 포장된 지표의 면적 또한 늘어나 털진드기의 생육 환경이 줄어든다. 이러한 이유로 공원 면적이 증가하여도 쯔쯔가무시증의 발병률이 감소하는 것으로 사료된다. 이와 비슷한 현상으로 인구 밀도의 응답곡선에서 쯔쯔가무시증의 발병 확률이 인구 밀도 증가에 따라 증가하다가 감소하는 것을 확인할 수 있다. 실제, 인구 밀도가 18,000 명/km² 이상인 지역은 서울특별시 강동구 그리고 송파구 등의 도시 지역으로, 이 지역들은 농촌 지역과 비교하여 부족한 털진드기의 생육 환경으로 인하여 상대적으로 낮은 쯔쯔가무시증 발병 확률을 나타낸다.

환경 변수인 고도의 경우 고도가 높아질수록 쯔쯔가무시증 발병 확률이 급격히 감소하는 것을 볼 수 있는데, 이는 고도가 높아지면서 낮아지는 기온과 변화하는 식생이 털진드기가 선호하지 않는 환경이기 때문이다. 평균 기온과 최저 기온의 응답 곡선 경우 온도 증가에 따라 쯔쯔가무시증 발병 확률이 증가하다가 최고점을 보인 후 감소하는 특징을 가지고 있다. 이는 털진드기의 설치류 및 사람에 대한 기생물에 주요한 요인 중 하나인 온도가 10 °C 이상임을 나타낸다[28]. 최고 기온의 응답 곡선 경우 온도 상승에 따라 쯔쯔가무시증 발병 확률이 지속적으로 증가한다. 이는 기온이 상승함에 따라 털진드기의 활동이 점차 활발해지며 쯔쯔가무시증 발병 확률이 증가하는 것을 나타내며, 최고 기온 20.52 °C에서도 증가하는 경향은 털진드기 유충의 가장 활발한 활동 기온인 24 °C를 반영하고 있다[5,6,28]. 추가적으로, Fig. 5(a)에 나타난 쯔쯔가무시증의 확률 분포와 비교해 볼 때, 일부 남부 해안 지역은 상대적으로 높은 평균 기온, 최고 기온, 그리고 최저 기온을 가지고 있지만 발병 확률 분포가 낮게 나타난다. 이는 해안 지역이 내륙 지역과 비교하였을 때 털진드기가 선호하는 식생 조건인 관목 및 수풀이 부족하기 때문에 높은 온도 조건임에도 불구하고 낮은 쯔쯔가무시증 발병 확률 분포를 가지는 것으로 사료된다. 반면 밀양 그리고 합천 등 남부 내륙 지역은 털진드기에 적합한 생육 환경을 가지고 있으며, 이에 따라서 비교적 고온의 기상 조건과 상승하게 높은 쯔쯔가무시증의 발병 확률을 가지는 것으로 나타난다.

4.2. 심층 인공 신경망 모델 기반 쯔쯔가무시증 발병 횟수 예측

환경 변수와 쯔쯔가무시증 발병 사이의 지연 효과를 고려하기 위하여, 0부터 6개월까지의 지연 효과를 더한 환경 변수와 쯔쯔가무시증 발병 사이의 피어슨 상관관계 분석을 실시하였다. Fig. 6은 각

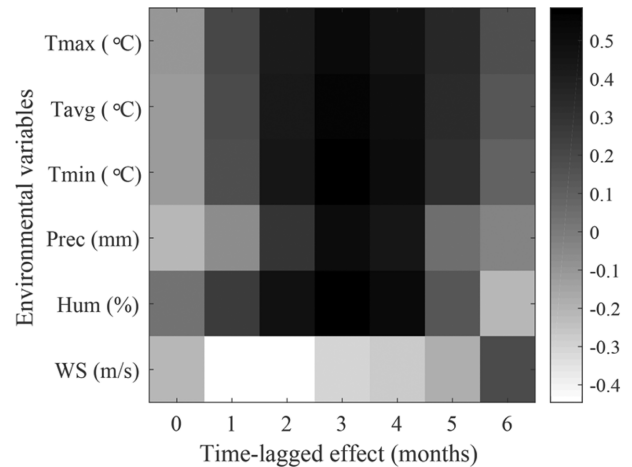


Fig. 6. Pearson's correlation coefficient between the environmental variables and time-lagged effect on the occurrence of scrub typhus.

환경 변수의 시간 지연 효과에 따른 쯔쯔가무시증 발병 횟수 사이의 피어슨 상관 계수 결과를 보여준다. 이때, 열은 색일수록 강한 음의 상관관계를, 짙은 색일수록 강한 양의 상관관계를 나타낸다. 기온의 최댓값, 평균값, 그리고 최솟값의 경우 2~4개월의 지연 효과를 고려하였을 때, 쯔쯔가무시증 발병과 강한 양의 상관관계를 가졌다. 또한, 강수량은 3~4개월의 시차 그리고 습도는 2~4개월의 시차를 두고 양의 상관관계를 보였으며, 평균 풍속은 1~2개월의 시차를 가지고 음의 상관관계를 나타냈다.

이는 쯔쯔가무시증의 주 매개체인 털진드기는 변온 동물이기 때문에 번식 및 성장 등에 기후변화의 영향을 많이 받기 때문이다[7]. 한국에서 털진드기의 유충은 주로 가을인 10월과 11월에 주로 나타나는데, 이를 통해 털진드기 성충이 주로 산란하는 시기는 여름철로 알려져 있다[28]. 여름철의 기온, 강수량, 그리고 습도는 털진드기의 산란에 큰 영향을 끼치는데, 기온이 상승하면 털진드기의 활동이 증가하게 되어 가을철의 유충 개체수 증가에 영향을 줄 수 있다. 반면 강우량이 줄어들어 습도가 낮아지게 되면 털진드기가 땅속 깊이 들어가 산란 활동을 중지하게 되어 결국 가을철 털진드기 유충 개체수를 감소시키게 된다. 또한 여름철의 높은 기온, 강수량, 그리고 습도는 털진드기의 숙주인 설치류가 생활하기 적합한 환경을 조성하여 쯔쯔가무시증 발병을 증가시킬 수 있다[29]. 이와 같은 이유로 여름철(2~4개월의 시차)의 기온, 강수량, 그리고 습도는 털진드기 유충 및 쯔쯔가무시증 발병에 강한 양의 상관관계를 가진다. 풍속의 경우 다른 환경 변수와 비교했을 때, 1~2개월의 짧은 시차를 두고 음의 상관관계를 가지고 있다. 하지만, 풍속의 쯔쯔가무시증 발병 및 털진드기에 관한 생물학적 메커니즘은 기존 연구에서 밝혀진 바가 거의 없다. 따라서 풍속에 따른 털진드기의 생애 주기 변동에 관한 연구가 추가로 이루어져야 할 것으로 보인다. 본 연구에서는 피어슨 상관관계 분석을 통한 결과를 활용하여 최대, 최소 및 평균 기온, 강수량, 그리고 습도는 3개월의 시차를 이용하고 풍속은 1개월의 시차를 부여하여 시간에 따른 쯔쯔가무시 발병 횟수 예측 모델 구성에 사용하였다.

시간에 따른 환경 변수와 지연 효과를 고려한 심층 인공 신경망 및 부분 최소 회귀법을 이용한 쯔쯔가무시증 발병 횟수 예측 모델의 결과를 Fig. 7(a)-(b)에 도시하였다. 예측 값과 측정값의 차이를 나

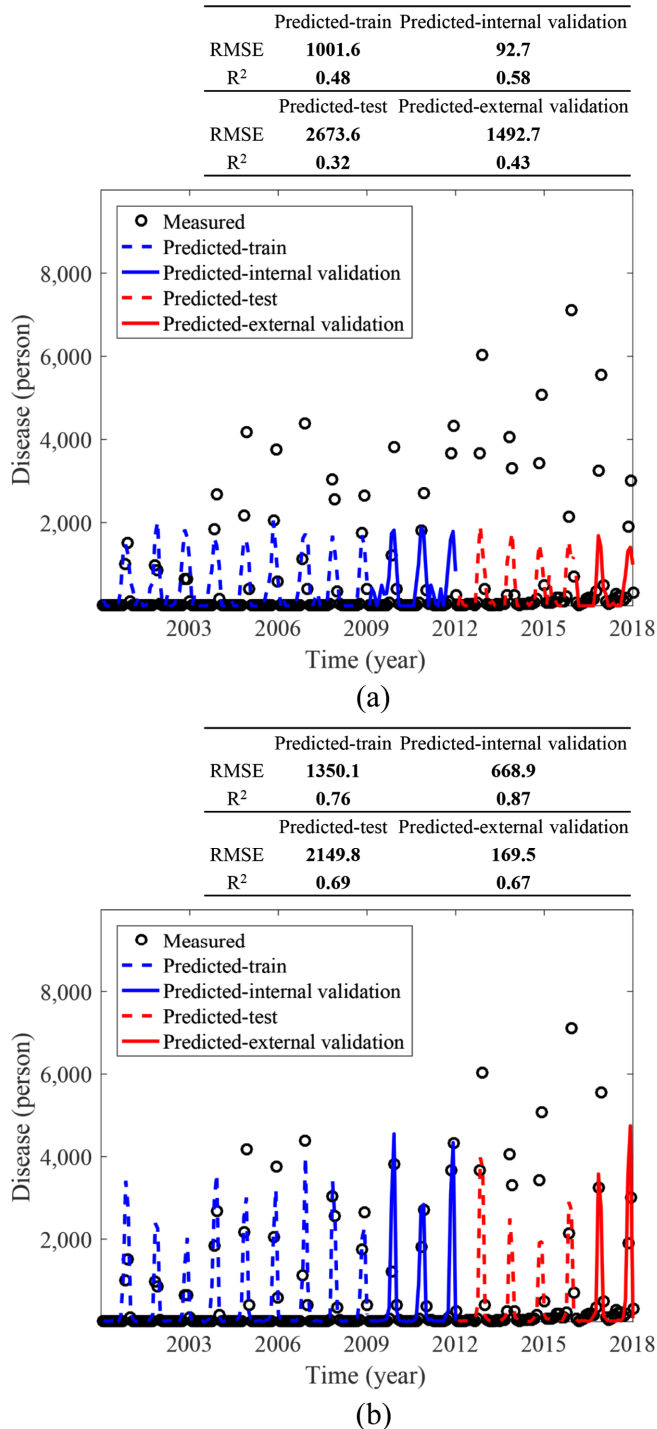


Fig. 7. The prediction number of diseases by scrub typhus by using (a) partial least squares and (b) deep neural network.

타내는 RMSE를 각 비교해보면, 심층 인공 신경망 기반 모델의 학습 데이터에 관한 RMSE 값이 부분 최소 회귀법 기반 모델보다 높게 나타났다. 이는 심층 인공 신경망이 쯔쯔가무시증 발병 횟수 측정 데이터에서 10~11월에 최고점들을 가지는 것을 반영하기 위해 변동성을 가지도록 예측이 되어 RMSE 값이 높게 나타났고, 부분 최소 회귀법의 경우 변동성을 미약하게 반영하여 데이터의 최고점 변화를 잘 예측하지 못하였지만 결과적으로 낮은 RMSE 값을 가진 것으로 사료된다. 이와 같은 특징은 R²값을 비교해보았을 때, 비슷

한 경향이 나타난다. R²값은 모델을 통한 예측 값이 실제 측정 데이터의 분산 정도를 얼마나 잘 설명하는지를 나타내는 척도이다 [30]. 따라서 심층 인공 신경망을 사용했을 때 0.76 R²로, 이는 부분 최소 회귀법의 0.48 R²값보다 높은 값을 나타내며 측정 데이터의 경향 및 분산 정도를 비교적 잘 표현한다고 할 수 있다.

테스트 데이터와 외부 검증 데이터를 비교해 보았을 때, 심층 인공 신경망이 낮은 RMSE 및 높은 R²을 나타내어 결과적으로 부분 최소 회귀법 기반 모델 보다 높은 예측 성능을 나타냄을 알 수 있다. 또한, 부분 최소 회귀법 기반 모델의 경우 10~11월에 쯔쯔가무시증 발생의 최고점들을 반영하지 못하며, 이는 쯔쯔가무시증이 10~11월에 감염 환자수의 90%가 발생한다는 주 발병 특징을 예측하는데 한계점이 존재한다. 따라서 본 연구에서 데이터간의 복잡하고 숨겨진 특징들을 추출 및 반영할 수 있는 심층 인공 신경망의 구조적 특징으로 쯔쯔가무시증 발병 횟수 예측 모델을 생성시 예측 성능이 비교적 우수함을 확인할 수 있었다. 하지만 쯔쯔가무시증 발병 횟수의 측정 데이터에서 최고점이 4,000명 이상으로 급증하는 경우 심층 인공 신경망 기반 모델을 이용한 예측에 한계점이 존재하였다. 이는 환경 변수뿐만 아니라 사회 변수 및 쯔쯔가무시증에 대한 도메인 전문 지식을 활용 및 주요 데이터를 추출하여 추후 심층 인공 신경망의 구조 개선을 통해 극복이 가능할 것으로 사료된다.

4-3. 기후변화에 따른 국내 쯔쯔가무시증 발병의 spatio-temporal 변화 미래 예측 및 분석

앞서 구성된 쯔쯔가무시증의 공간적 발병 확률 분포 모델과 시간적 발병 횟수 예측 모델에 2040년의 기후변화에 따른 환경 조건을 이용하여 미래의 쯔쯔가무시증 발병 변화에 따른 분석을 실시하였다. 2040년의 변화되는 환경 조건은 RCP 6.0 시나리오에 따라 제시된 기온, 강수량, 습도, 그리고 풍속의 예측 값을 사용하였다. 2040년의 예상되는 쯔쯔가무시증의 공간적 발병 확률 분포 지도는 Fig. 8(a)에 나타내었으며, 현재의 발병 확률과 비교하여 발병률이 증가할 것으로 예상되는 지역은 점선 원으로 표시하였다. 먼저, 기후변화 시나리오 RCP 6.0에 따른 2040년의 쯔쯔가무시증 최대 발병 확률이 Fig. 4(a)에 나타난 현재 기후조건에서의 최대 발병 확률 0.74에 비해 0.08 증가한 0.82로 나타났다. 따라서 2040년의 기후 환경 조건에서 평균 기온, 강수량, 그리고 습도가 각 5.6%, 2.9%, 0.3% 증가하는 것이, 앞서 언급한 바와 같이 털진드기 생육 환경 및 쯔쯔가무시증 발병에 양의 상관관계를 가져 최대 쯔쯔가무시증 발병 확률이 증가한 것으로 분석 된다. 또한, 발병 확률이 70% 이상인 지역은 현재 환경 조건에서는 2053.12 km²이지만, 미래 기후 조건에서는 9.1% 증가한 2240.06 km²로 예상된다. 특히, Fig. 5(a)의 현재 기후에 따른 발병 확률 분포와 Fig. 8(a)에 나타난 2040년의 발병 확률 분포를 비교해볼 때, 경기 남부, 충청도 북부, 전라남도, 그리고 경상남도에서 발병 확률이 크게 증가하는 것을 확인할 수 있다. 비교적 내륙에 위치하고 어느 정도의 인구 밀도를 보유하고 있을 때, 기온, 강수량, 그리고 습도 등의 털진드기 생육에 대한 환경 조건이 변화하면 쯔쯔가무시증 발병 확률이 증가하는 것을 나타낸다. 따라서 기후 변화에 따른 환경 조건의 변동으로 한국에서의 쯔쯔가무시증 발병 지역 및 확률이 확대 될 것으로 예측된다.

기후변화 시나리오 RCP 6.0에 따라 2040년의 예상되는 쯔쯔가무시증의 시간에 따른 발병 횟수는 Fig. 8(b)에 나타내었다. 심층 인공 신경망 기반 시간적 모델에 의해 예측된 쯔쯔가무시증의 발병

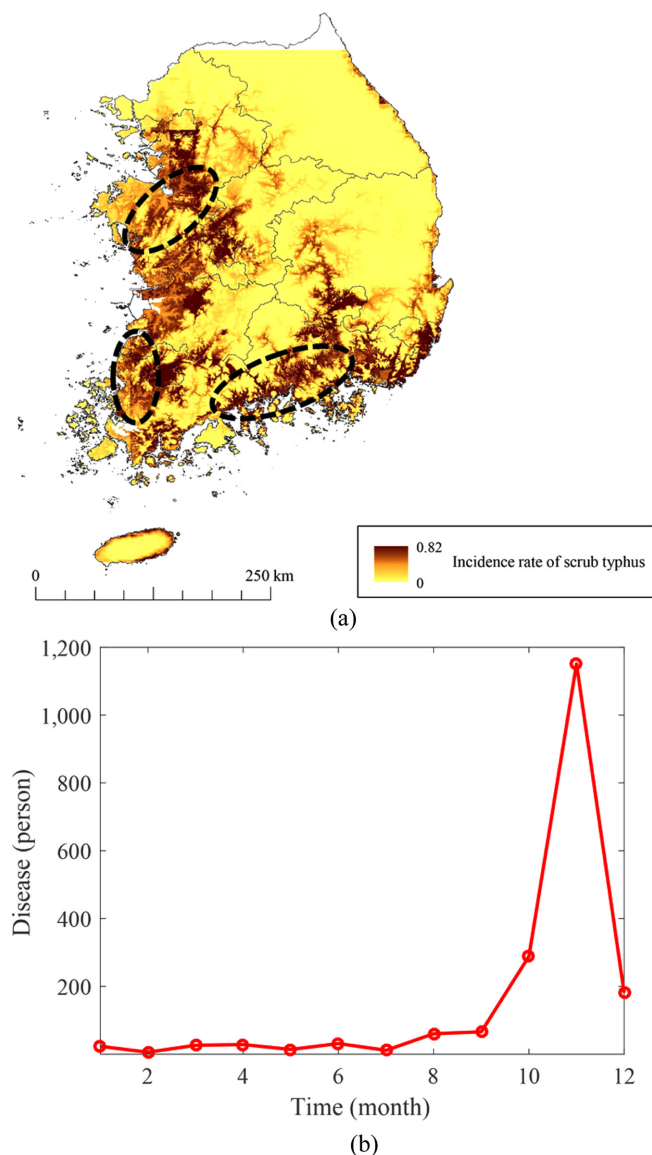


Fig. 8. Predicted incidence of scrub typhus in 2040: (a) potential distribution map and increasing incidence rate area indicated by the circles and (b) the number of diseases.

횡수는 가을인 10~11월에 최고점을 보이는 현재 기후 조건에 대한 경향과 비슷한 경향이 관찰되었다. 하지만, 현재 기후 조건에서 10~11월에 쯔쯔가무시증의 발병 횡수가 90%로 보고된 반면[10,11], 예측된 2040년의 쯔쯔가무시증 발병의 경우 9~12월의 발병 횡수가 전체 발병 횡수의 89%를 차지하고 12월에는 약 200명의 쯔쯔가무시증 환자가 발생할 것으로 예측 된다. 그 이유는 기온에 관한 3개월의 지연 효과를 고려해볼 때, 기후변화 시나리오에 의해 증가한 기온 때문인 것으로 사료된다. 2040년 9월에 예상되는 최저 기온, 최고 기온, 그리고 평균 기온은 각 17.01 °C, 24.18 °C, 그리고 20.54 °C 이다. 이는 털진드기의 주된 활동 온도인 24 °C에 근접하는 온도로 여름철뿐만 아니라 9월에도 털진드기의 산란 활동이 활발해지고 털진드기의 유충이 12월에도 존재할 수 있음을 의미한다. 즉, 기후 변화로 인한 환경 변화는 동절기인 12월에도 털진드기의 개체수를 증대시켜 쯔쯔가무시증의 발병 분포 및 확률을 증대시킬 것으로 사료된다. 이는 곧 증대하는 쯔쯔가무시증 발병의 위험성을 감지할

수 있다. 특히 평균 연령대가 높은 농촌은 쯔쯔가무시증의 노출 확률이 높을뿐만 아니라 노인들의 질병에 대한 면역력이 낮기 때문에 [5], 쯔쯔가무시증의 위험성은 크게 증가할 수 있다. 쯔쯔가무시증 발병 분포와 확률의 증대 그리고 질병 취약계층을 고려할 때, 감염 매개체 생물인 털진드기의 퇴치 및 쯔쯔가무시증 방제를 위한 관리가 필요할 것으로 판단된다.

공중보건 측면에서 기후변화 대비 쯔쯔가무시증 방제를 효율적으로 실시하기 위해서는 공간-시간적 쯔쯔가무시증 예측 모델의 성능 향상이 더욱 필요하다. 쯔쯔가무시증의 발병 요인으로 환경적 변수뿐만 아니라 사회적 변수의 영향이 있다는 점을 고려할 때 기후 변화로 인한 환경 변화 이외의 인구 밀도, 노년층 증가, 그리고 유아층 감소 등의 사회 변수의 변화를 고려하는 등의 후속 연구가 필요할 것으로 사료된다. 기후 변화로 인한 다양한 변수들의 고려가 필요하며, 추가적으로 털진드기가 주로 기생하고 생육 환경에 밀접한 관련이 있는 설치류에 관한 데이터 및 정보 추가를 통해서 보다 향상된 쯔쯔가무시증의 발병 예측력을 기대할 수 있을 것으로 보인다.

5. 결 론

본 연구는 털진드기의 감염 매개체로부터 인간에게 전염되는 쯔쯔가무시증을 공간적 생태 모델인 MaxENT와 시간적 예측 모델인 심층 인공 신경망을 통해 그 발병 특징을 분석하고자 하였다. MaxENT 모델을 적용하였을 때, 한국에서의 쯔쯔가무시증 발병 확률 분포는 남부 지역에서 높게 나타나는 것을 알 수 있었고, 중부 지역은 비교적 낮은 발병률을 보이는 것을 확인할 수 있었다. 또한, Jackknife test를 통해 16개의 공간 및 환경 변수 중 농가 인구 비율, 공원 면적 비율, 인구 밀도, 고도, 평균 기온, 최고 기온, 그리고 최저 기온의 7가지 변수가 상대적으로 주요한 것으로 나타났다. 추가적으로 응답 곡선 분석을 통하여 환경 변수가 공간 변수 보다 쯔쯔가무시증 발병에 상대적으로 큰 중요도를 나타냈다. 시간에 따른 쯔쯔가무시증의 발병 횡수 예측 모델을 위한 피어슨 상관 계수 분석 결과 기온, 강우량, 그리고 습도에서 3개월의 지연 효과가 나타남을 확인할 수 있었다. 쯔쯔가무시증 발병 횡수 예측 모델은 심층 인공 신경망의 성능을 분석하기 위해 기존 통계 모델인 부분 최소 회귀법을 비교하였으며, 이를 통하여 심층 인공 신경망의 예측 성능의 강건성을 확인하였다. MaxENT와 심층 인공 신경망 모델을 이용하여 2040년의 기후변화에 따른 쯔쯔가무시증의 발병 특징을 예측한 결과, 최대 발병률은 0.82로 증가하였으며 발병률 0.7 이상인 지역의 면적은 2053.12 km²으로 증가하였다. 또한, 쯔쯔가무시증 발병의 집중적인 시기가 현재 기후 조건인 경우 10~11월이나, 2040년의 경우 9~12월로 증가하는 것으로 예측되었다. 미래의 쯔쯔가무시증 발병 예측에 환경 요인을 주로 사용 하였으나 변화하는 공간 변수 및 설치류에 관한 고려를 통해 보다 향상된 예측 정확도를 얻을 수 있을 것으로 기대된다. 본 연구 결과는 향후 보다 정확한 쯔쯔가무시증의 발병 모니터링 및 예측, 그에 따른 공중보건 측면에서 발병 요인 관리, 질병 예방 및 주민 건강 유지를 위한 연구의 바탕이 될 것으로 기대된다.

감 사

본 연구는 환경부 「기후변화특성화대학원사업」의 지원으로 수행되었습니다.

References

- World Health Organization, "Vector-borne Diseases," WHO Regional Office for South-East Asia (2014).
- Park, S. and Han, D., "Reviews in Medical Geography: Spatial Epidemiology of Vector-Borne Diseases," *J. Korean Geogr. Soc.*, **47**(5), 677-699(2012).
- Medlock, J. M. and Leach, S. A., "Effect of Climate Change on Vector-borne Disease Risk in the UK," *Lancet Infect. Dis.*, **15**(6), 721-730(2015).
- Paris, D. H., Shelite, T. R., Day, N. P. and Walker, D. H., "Unresolved Problems Related to Scrub Typhus: a Seriously Neglected Life-threatening Disease," *Am. J. Trop. Med. Hyg.*, **89**(2), 301-307(2013).
- Kim, S. W. and Kim, Y. H., "Spatial Analysis Modeling on Scrub Typhus Disease Occurrence in Korea," *J. Korean Cart. Assoc.*, **14**, 41-54(2014).
- Kong, W. S., Shin, E. H., Lee, H. I., Hwang, T. S., Kim, H. H., Lee, N. Y., Sung, J. H., Lee, S. G. and Yoon, K. H., "Time-spatial Distribution of Scrub Typhus and its Environmental Ecology," *J. Korean. Geogr. Soc.*, **42**(6), 863-878(2007).
- Yang, L. P., Liu, J., Wang, X. J., Ma, W., Jia, C. X. and Jiang, B. F., "Effects of Meteorological Factors on Scrub Typhus in a Temperate Region of China," *Epidemiol. Infect.*, **142**(10), 2217-2226(2014).
- Lee, Y. G., Choi, K. H. and Kwak, J. W., "A Study on the Public Health Disasters Using Meteorological Factor: Scrub Typhus in South Korea," *J. Korean Soc. Hazard Mitig.*, **18**, 343-350(2018).
- Weng, S. C., Lee, H. C., Chen, J. J., Cheng, Y. J., Chi, H. and Lin, C. Y., "Eschar: a Stepping Stone to Scrub Typhus," *J. Pediatr.*, **181**, 320(2017).
- Lee, I. Y., Ree, H. I. and Hong, H. K., "Seasonal Prevalence and Geographical Distribution of Trombiculid mites (Acarina: Trombiculidae) in Korea," *Korean J. Zool.*, **36**(3), 408-415(1993).
- Xu, G., Walker, D. H., Jupiter, D., Melby, P. C. and Arcari, C. M., "A Review of the Global Epidemiology of Scrub Typhus," *PLoS Negl. Trop. Dis.*, **11**(11), e0006062(2017).
- Kujala, H., Whitehead, A. L. and Wintle, B. A., "Identifying Conservation Priorities and Assessing Impacts and Trade-offs of Potential Future Development in the Lower Hunter Valley in New South Wales," *A Rep. by NERP Environ. Decis. Hub*, (2015).
- Seo, C. W., Park, Y. R. and Choi, Y. S., "Comparison of Species Distribution Models According to Location Data," *J. Korean Soc. Geospatial Inf. Syst.*, **16**(4), 59-64(2008).
- Pearson, R. G., "Species' Distribution Modeling for Conservation Educators and Practitioners," *Synth. Am. Museum Nat. Hist.*, **50**, 54-89(2007).
- Braunisch, V. and Suchant, R., "Predicting Species Distributions Based on Incomplete Survey Data: the Trade off Between Precision and Scale," *Ecography*, **33**(5), 823-840(2010).
- Jácome, G., Vilela, P. and Yoo, C., "Social-ecological Modelling of the Spatial Distribution of Dengue Fever and its Temporal Dynamics in Guayaquil," *Ecol. Inform.*, **49**, 1-12(2019).
- Phillips, S. J., Anderson, R. P. and Schapire, R. E., "Maximum Entropy Modeling of Species Geographic Distributions," *Ecol. Modell.*, **190**(3-4), 231-259(2006).
- Elith, J. and Leathwick, J. R., "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time," *Annu. Rev. Ecol. Evol. Syst.*, **40**, 677-697(2009).
- Kihoro, J., Otieno, R. O. and Wafula, C., "Seasonal Time Series Forecasting: A Comparative Study of ARIMA and ANN Models," *Afr. J. Sci. Technol.*, **5**(2), 41-49(2004).
- Enhofer, O., Pichs-Madruga, R., Sokona, Y., Seyboth, K., Matschoss, P., Kadner, S., Zwickel, T., Eickemeier, P., Hansen, G., Schlömer, S. and Von Stechow, C., "IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation," Cambridge University Press, Cambridge, NY(2011).
- Korea meteorological administration, "Climate Change Forecasting Report of Korea," (2018).
- Nam, K., Hwangbo, S. and Yoo, C. "A Deep Learning-based Forecasting Model for Renewable Energy Scenarios to Guide Sustainable Energy Policy: A Case Study of Korea," *Renew. Sust. Energy. Rev.*, **122**, 109725(2020).
- Shahbeik, S., Afzal, P., Moarefvand, P. and Qumarsy, M., "Comparison Between Ordinary Kriging (OK) and Inverse Distance Weighted (IDW) Based on Estimation Error. Case study: Dardevey Iron ore Deposit, NE Iran," *Arab. J. Geosci.*, **7**(9), 3693-3704(2014).
- Brzezinski, D. and Stefanowski, J., "Prequential AUC: Properties of the Area Under the ROC Curve for Data Streams with Concept Drift," *Knowl. Inf. Syst.*, **52**(2), 531-562(2017).
- McIntosh, A., "The Jackknife Estimation Method," *ArXiv Prepr ArXiv*, 1606.00497(2016).
- Nam, K. J., Kim, M. J., Lee, S., Hwangbo, S. and Yoo, C. K. "Interpretation and Diagnosis of Fouling Progress in Membrane Bioreactor Plants Using a Periodic Pattern Recognition Method," *Korean J. Chem. Eng.*, **34**(11), 2966-2977(2017).
- Lee, H. J. and Park, C., "Distribution of Chigger Mites as Tsutsugamushi Vectors Sampled in Seogwipo," *Korean J. Clin. Lab. Sci.*, **51**(3), 344-350(2019).
- Song, H. J., "Environmental Survey on the Vectors and Hosts of Tsutsugamushi Disease in Jeonnam province, Korea," *Korean J. Vet. Serv.*, **35**(3), 183-189(2012).
- Kim, S. H. and Jang, J. Y., "Correlations Between Climate Change-Related Infectious Diseases and Meteorological Factors in Korea," *J. Prev. Med. Public Heal.*, **43**(5), 436-444(2010).
- Hung, P. V. X., Cho, S. H. and Moon, S. H., "Prediction of Boron Transport Through Seawater Reverse Osmosis Membranes Using Solution-diffusion Model," *Desalination*, **247**(1-3), 33-44(2009).