



A study of artificial neural network for in-situ air temperature mapping using satellite data in urban area

Jeon, Hyunho^a · Jeong, Jaehwan^b · Cho, Seongkeun^c · Choi, Minha^{d,e*}

^aCombined Master / Ph.D, Department of Global Smart City, Sungkyunkwan University, Suwon, Korea

^bResearch Professor, Center for Built Environment, Sungkyunkwan University, Suwon, Korea

^cPh.D Candidate, Department of Water Resources, Sungkyunkwan University, Suwon, Korea

^dProfessor, Department of Water Resources, Sungkyunkwan University, Suwon, Korea

^eProfessor, School of Civil, Architecture Engineering and Landscape Architecture, Sungkyunkwan University, Suwon, Korea

Paper number: 22-053

Received: 23 August 2022; Revised: 5 October 2022; Accepted: 5 October 2022

Abstract

In this study, the Artificial Neural Network (ANN) was used to mapping air temperature in Seoul. MODerate resolution Imaging Spectroradiometer (MODIS) data was used as auxiliary data for mapping. For the ANN network topology optimizing, scatterplots and statistical analysis were conducted, and input-data was classified and combined that highly correlated data which surface temperature, Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), time (satellite observation time, Day of year), location (latitude, hardness), and data quality (cloudness). When machine learning was conducted only with data with a high correlation with air temperature, the average values of correlation coefficient (r) and Root Mean Squared Error (RMSE) were 0.967 and 2.708 $^{\circ}\text{C}$. In addition, the performance improved as other data were added, and when all data were utilized the average values of r and RMSE were 0.9840 and 1.883 $^{\circ}\text{C}$, which showed the best performance. In the Seoul air temperature map by the ANN model, the air temperature was appropriately calculated for each pixels topographic characteristics, and it will be possible to analyze the air temperature distribution in city-level and national-level by expanding research areas and diversifying satellite data.

Keywords: ANN, Air temperature, Seoul, Mapping, AWS, MODIS-Terra

위성 정보를 활용한 도심 지역 기온자료 지도화를 위한 인공지능망 적용 연구

전현호^a · 정재환^b · 조성근^c · 최민하^{d,e*}

^a성균관대학교 글로벌스마트시티융합전공 석박통합과정, ^b성균관대학교 건설환경연구소 연구교수, ^c성균관대학교 수자원학과 박사수료,

^d성균관대학교 수자원학과 교수, ^e성균관대학교 건설환경공학부 교수

요 지

본 연구에서는 서울시 기온 지상관측 자료의 지도화를 위해 Artificial Neural Network (ANN)을 사용하였다. 지도화를 위한 보조자료로는 MODerate resolution Imaging Spectroradiometer (MODIS) 자료를 사용하였다. ANN 모델 설계를 위해 입력자료와 출력자료 간의 산점도 및 통계분석을 수행하였으며, 기온과의 상관성이 비교적 높게 나타나는 입력자료인 지표면온도, Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI)와 시간(위성관측시각, Day of year), 위치(위도, 경도), 데이터 품질(운량)과 관련된 데이터 종류를 분류 및 조합하여 학습을 진행하였다. 기온자료와 상관성이 높은 데이터만으로 학습을 진행하였을 때 상관계수(r)와 Root Mean Squared Error (RMSE)의 평균값이 0.9667, 2.708 $^{\circ}\text{C}$ 로 우수한 성능을 보였다. 학습에 사용된 데이터의 종류가 추가될수록 더 우수한 학습 결과를 보였으며, 모든 데이터가 활용될 때에는 r 과 RMSE의 평균값이 0.9840, 1.883 $^{\circ}\text{C}$ 로 가장 우수한 성능을 보였다. ANN 모델으로 생성한 서울시 기온 지도에서는 픽셀별 지형적 특성에 적절하게 기온이 산정된 것으로 판단되며, 추후 연구지역 확대 및 위성자료의 다양화를 통해 시단위 및 전국단위 기온 분포 분석 연구가 가능할 것이다.

핵심용어: ANN, 기온, 서울시, 지도화, AWS, MODIS-Terra

*Corresponding Author. Tel: +82-31-290-7539

E-mail: mhchoi@skku.edu (Choi, Minha)

1. 서론

과학기술 및 산업의 발전으로 도시 인구 및 도시화 비율이 증가하고 있으며, 대한민국의 경우 지난 65년간 370%의 도시화 비율 상승과 350%의 인구 증가세를 보였다(NGII, 2019). 도시 인구가 밀집되고, 도시면적이 증가함에 따라 다양한 문제들이 발생하고 있으며, 특히 도시 거주민들의 열 스트레스 상승은 최근 사회적으로 큰 이슈가 되고 있다. 지구온난화, 열섬현상, 도시민의 활동 등 무수한 자연적 영향과 인위적 영향으로 도시 내 기온은 꾸준히 상승하는 경향을 보이고 있으며, 스페인과 포르투갈에서는 폭염으로 인한 누적 사망자가 2000명에 달하는 등 유례없는 기온 상승으로 인한 피해가 속출하기도 하였다. 위와 같은 문제의 해결을 위해 세계 각지의 여러 대도시를 대상으로 열섬현상, 한파와 같은 열 스트레스 분석 및 모델링이 수행되고 있으나(Harmay and Choi, 2022; Tewari *et al.*, 2019), 약 1,000만 명의 인구수로 전 세계 도시 중 인구수 21위에 해당하는 서울시의 경우 열 스트레스 모델링 관련 연구가 미비한 실정이다.

도시 지역은 자연적인 초원이나 농지 등에 비해 매우 큰 불균일성을 보이는데, 도시 설계에 따라 지형적 특성의 차이가 크게 나타나 적은 수의 지상관측기기로는 기후 관련 인자들의 분포를 파악하기 어렵다(Mendelsohn *et al.*, 2007; Peterson *et al.*, 2003). 도시 내 기온 또한, 인공구조물의 위치 및 방향과 난방시스템 등에 의한 와류의 변형으로 분석 및 모델링에 어려움이 있으며, 각 인자들의 광범위하며 불규칙한 분포는 모니터링과 변동성 파악을 매우 어렵게 한다.

최근 광범위한 지역의 기후 관련 인자 모니터링을 위해 인공 위성, 드론과 같은 원격탐사기법이 활용되고 있으며, 다수의 플랫폼을 활용하여 광범위한 지역을 연속적으로 관측하는 원격탐사기법의 특성상 기후 관련 인자 분석 및 모니터링에 적합하다(Baik *et al.*, 2018). 원격탐사 기법을 활용하는 방법으로 Rigo *et al.* (2006), Nichol *et al.* (2009), Weng (2009)과 같이 지상관측기기와 원격탐사기법을 같이 활용하여 관련 인자들의 분석 가능하다. 그러나, 인공위성 영상의 경우 상부 대기층의 정보를 1차적으로 수집하여 보정을 수행한 후 지표 인근의 하부 대기층 정보를 산출하는 방법으로 데이터 처리가 수행되기에, 구름과 같은 기상 상황과 지형별 보정 기법의 차이에 영향을 받는다(Richter, 1997; Vanhellemont and Ruddick, 2018).

기후 인자를 분석 및 해석하는 방법의 하나로 경험식과 물리식이 혼합된 알고리즘을 활용하는 방법이 있다(Park *et al.*, 2017; Hao *et al.*, 2019). 경험식이 포함되는 경우 해당 연구가 수행된 지역 혹은 국가에 특화된 경우가 다수이며, 이를 다른

지역에 활용하기 위해서는 해당 지역의 식생, 토질, 경사, 풍속 등 여러 환경적인 요소들을 고려하여 적용성을 검증할 필요가 있다. 예를 들어, 원격탐사기법을 활용한 지표면온도를 산정하는 방법의 하나로 분할창 알고리즘(split-window algorithm)이 있다. 이는 두 개의 다른 파장 범위를 관측하는 열적외선 센서를 활용하여 선형 혹은 비선형으로 지표면온도를 산정하는 기법으로, 대기와 지표면의 상태에 따라 변동하는 여러 계수들을 활용한다(Sobrinho *et al.*, 1993; Yu *et al.*, 2007). 분할창 알고리즘은 현재까지도 다양한 국가 및 지역을 대상으로 그 적용성을 평가하는 연구들이 꾸준히 진행되고 있다(Niclòs *et al.*, 2021; Zheng *et al.*, 2019).

한편 인공지능은 다량의 데이터를 활용하여 블랙박스 형식으로 개별적인 모델을 생성해 원하는 결과를 도출할 수 있어, 최근 데이터 분석 분야에서 주목을 받고 있다. 본 연구에서는, 머신러닝 기법의 하나인 Artificial Neural Network (ANN)와 위성자료를 활용하여 서울시의 대기 온도를 산정하였다. 위성 자료로는 MODerate-resolution Imaging Spectroradiometer (MODIS) 센서를 탑재한 Terra 위성의 MOD11A1과 MOD13A2를 활용하였으며, 효과적인 머신러닝을 위해 2008년부터 2013년까지 총 13년간의 데이터를 사용하였다. 지표면 인근의 지상관측자료를 출력 자료로 상부 대기층의 위성자료를 입력 자료로 활용하여 서울시 내 미관측지역의 기온자료를 위성자료만으로 생성하고자 하였으며, 학습된 ANN 모델을 통해 지점 단위 기온 관측자료를 2D 지도화하였다.

2. 연구 지역 및 데이터

2.1 연구지역

본 연구에서는 대한민국의 수도인 서울시를 대상으로 연구를 수행하였다. 서울시는 약 978만 명이 거주하는 대한민국에서 인구수가 가장 높은 도시이며, 위도 37.41~37.72°, 경도 126.73~127.27°에 위치한다. 연평균 강수량은 1417.9 mm, 연평균기온은 12.8°C이며, 해발고도는 38 m이다. MOIS (2021)에 따르면 2019년 기준 서울시 시가화면적 비율은 61.4%로 서울시 내 개발 가능한 65%의 지역 중 거의 모든 지역이 개발되어있는 상황이다. 서울시에는 29개의 Automatic Weather System (AWS) 지상관측소가 위치하며, 위성 영상자료와 지상관측자료가 장기간 관측되고 있어 연구에 적합하다고 판단하였다(Jee *et al.*, 2014).

지도학습을 통한 ANN 모델의 기온자료 산출을 위해 실제 지상에서 관측된 대기 온도 자료를 활용이 필수적이다. 본 연

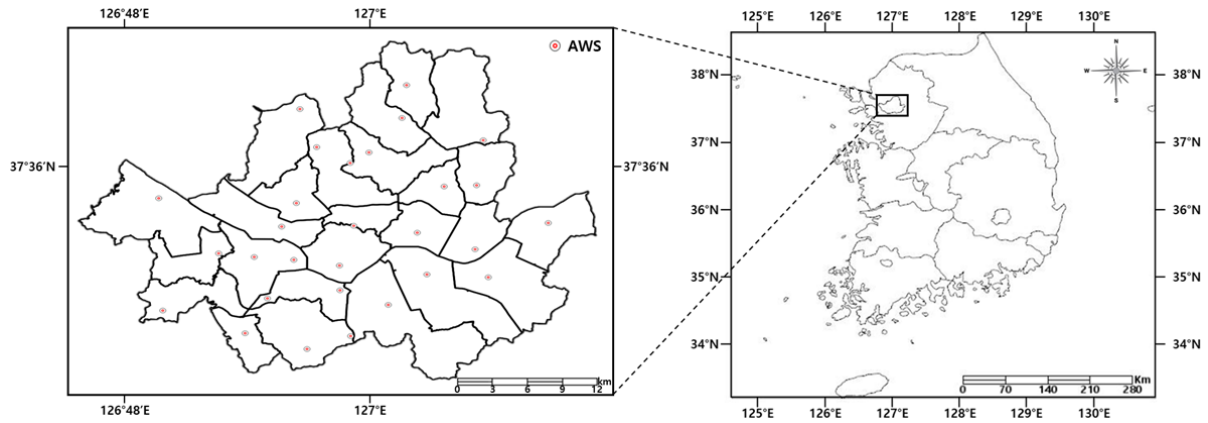


Fig. 1. Location of the study area and AWS sites

구에서는 기상청에서 제공하는 서울시 내 AWS 지상관측소 29개소의 1시간 단위 기온 관측자료 중 2013년부터 2020년 까지 총 8년치의 데이터를 활용하였다(Fig. 1). 또한 관측소별 위·경도 자료를 활용하여, 미 관측지의 위치정보를 반영한 기온자료를 산출하고자 하였다.

2.2 인공위성자료

본 연구에서는 National Aeronautics and Space Administration (NASA)에서 지구관측시스템(Earth Observing System, EOS) 프로젝트로 발사된, MODIS 센서를 탑재한 Terra 위성의 온도 관련 산출물을 활용하였다. MODIS 산출물 중 낮 시간대와 밤 시간대의 지표면온도, 구름량, 천정각, 관측시각, 품질관리정보와 열적외선밴드의 방사율을 제공하는 MOD11A1과 식생활력도와 관계된 지수인 Normalized Difference Vegetation Index (NDVI)와 Enhanced Vegetation Index (EVI), 품질관리정보, 지수산정에 사용되는 밴드별 반사율, 천정각, 방위각, Day of year (DOY), 픽셀안정성정보를 제공하는 MOD13A2를 활용하였다. 본 연구에서 활용된 MOD11A1과 MOD13A2의 공간해상도는 1 km로 동일하나, 시간해상도의 경우 MOD11A1은 낮과 밤 모두 하루 주기, MOD13A2는 16일 주기로 제공되고 있다. 각 산출물의 정보는 Table 1과 같으며, 이들 중 지표면 기준 2 m 상공에서 측정되는 기온자료와 관련된 지표면온도, 태양의 고도 및 방위각과 관련된 위성 관측시각과 DOY, 위성자료 품질 및 태양 복사 에너지량과 관련된 구름량, 픽셀 별 지형 정보와 관련되어 지표 인근 온도변화에 영향을 끼치는 식생지수 자료를 활용하여 머신러닝을 수행하였다. 식생지수로는 NDVI와 EVI를 사용하였다.

Table 1. The scientific data sets of MOD11A1, MOD13A2

| Product | Variable Type |
|---------|---------------------------------|
| MOD11A1 | Daytime LST |
| | Nighttime LST |
| | Day clear sky coverage |
| | Night clear sky coverage |
| | Time of daytime (local solar) |
| | Time of nighttime (local solar) |
| MOD13A2 | NDVI |
| | EVI |

3. 연구방법

3.1 Artificial Neural Network (ANN)

본 연구에서는 머신러닝 기법 중 하나인 ANN을 사용하였다. ANN은 인간의 뉴런과 비슷하게 다수의 노드(node)와 층(layer)으로 구성된 머신러닝 모델이다. ANN의 층은 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)으로 구성되어 있다. 출력층은 연구자가 모의하고자 하는 임의의 데이터들이 해당되며, 입력층은 출력층으로 설정된 데이터의 모의를 위해 활용되는 데이터가 해당된다. 은닉층은 입력층의 데이터로 출력층의 데이터를 모의하는 과정을 위해 필요한 부분으로 은닉층의 개수가 증가할수록 비선형적인 모의가 가능하다. 은닉층의 개수가 2개 이상으로 설정될 경우 최근 활발히 연구되고 있는 딥러닝으로 분류된다.

입력층, 은닉층, 출력층 간의 데이터 전파를 위해 노드가 활용된다. 일반적인 ANN 모델에서는 입력층과 출력층의 노드 수는 모델 설계과정에서 각각 설정된 데이터 개수와 동일하게 설정된다. 은닉층의 경우 데이터 종류, 데이터양, 학습의 복잡

도, 노이즈 양 등에 따라 연구자가 임의로 노드 수를 설정할 수 있다.

데이터 전파 과정에서는 가중치(weight), 편차(bias), 활성화 함수(activation function)가 활용된다. 데이터 전파 과정 전에 위치하는 노드의 값에 가중치와 편차를 적용하고, 적용된 값들의 합에 활성화 함수를 활용하여 전파 과정의 후에 위치하는 노드 값을 산출하는 방식으로 학습이 이루어진다. 본 연구에서 실제 관측값과 추정값의 차이를 계산하는 목적함수로 평균 제곱 오차(Mean Squared Error, MSE)가 사용되었다.

일반적으로 ANN 모델 구조가 복잡하게 설계될수록 더 복잡한 문제를 모의할 수 있으나, 노드의 개수와 은닉층의 수 등이 과도하게 책정될 경우, 과적합(over-fitting)이나 학습비용의 상승을 초래할 수 있다(Lawrence and Giles, 2000; Defornez and Kemsley, 1999). 이를 방지하기 위해 적절한 성능을 나타낼 수 있는 최소의 노드 및 은닉층 수를 설정하는 것이 최선이며, 대부분 학습의 경우 적은 수의 노드만으로도 높은 성능을 보인다(Hunter *et al.*, 2012).

3.2 머신러닝 활용 연구 절차

본 연구에서는 머신러닝 기법을 활용하여 서울시 기온 자료 지도화 알고리즘을 생성하였다. 기온과 관련된 인공위성자료(Table 1)를 입력자료로, AWS 기온 지상관측 자료를 출력 자료로 활용하여 서울시 내에서 기온 지상관측이 수행되지 않는 지역의 대기온도를 인공위성자료만으로 생성하였다. 인공위성자료 중 MOD11A1은 같은 날짜에 낮과 밤을 구분하여 2개의 자료가 제공된다. 이에, 본 연구에서는 2008년부터 2020년까지 총 13년간의 데이터를 낮과 밤을 구분하여 활용하였으며, training set으로 70%, test set으로 30%의 데이터가 활용되었다.

ANN 모델 설계과정에서 최적 학습 모델을 찾기 위해 노드 수, 활성화 함수 종류를 변경하며 반복 학습을 진행하였다.

은닉층의 노드 수는 층마다 10~30개를 변경하며 학습을 진행하였고, 각 층에서의 활성화 함수는 linear, log-sigmoid, tan-sigmoid를 변경하며 학습을 진행하였다. 또한, 데이터의 특성에 따라 기온과 상관성이 높은 자료(지표면온도, NDVI, EVI), 시간 관련 자료(위성관측시각, DOY), 공간 관련 자료(위도, 경도), 자료품질 관련 자료(운량)으로 분류하여 조합별로 학습을 진행하였다. 데이터의 조합은 상관성이 높다고 판단되는 자료들이 학습에 큰 영향을 끼칠 것으로 판단하였기에 이들을 기본적으로 포함하되 나머지 자료들을 추가로 조합하는 형식을 선택하였다. 자세한 조합 방법은 4.2절에 서술하였다. 본 연구에서 설계된 ANN 모델의 개요도는 아래 Fig. 2와 같다.

ANN은 각 노드 간에 임의의 데이터 처리 과정을 거쳐 출력층에 해당하는 데이터를 모의하는 과정으로 학습이 이루어지기에, 데이터 전처리 과정이 매우 중요하다. 기후 인자 관측 데이터 특성상 결측값이 다수 발생하기에, 일부 자료에서 결측값 발생 시 해당 시간대의 다른 자료를 모두 제외하였다. 또한, 각 데이터가 갖는 값의 범위는 인자의 특성과 단위 등에 따라 차이가 크므로, 데이터 정규화 후 머신러닝을 수행하였다(Table 1).

4. 결과 및 고찰

4.1 머신러닝 입출력자료 관계성 분석

머신러닝의 효과적인 학습을 위해서는 사용되는 데이터의 품질이 매우 중요하다. 이를 위해서는 전처리뿐만 아니라, 모델의 학습과 검증을 위해 분리 또한 적절하게 수행되어야 한다. 이에, 학습에 활용된 데이터들이 머신러닝에 끼치는 영향의 다각적인 해석을 위해 산점도(scatter plot) 분석과 통계분석을 수행하였다. 본 연구에서 사용된 데이터량은 약 11만 개로 산점도의 밀도 분석 또한 수행하였다.

Figs. 3 and 4는 각각 낮과 밤의 AWS 기온 지상관측 자료(출

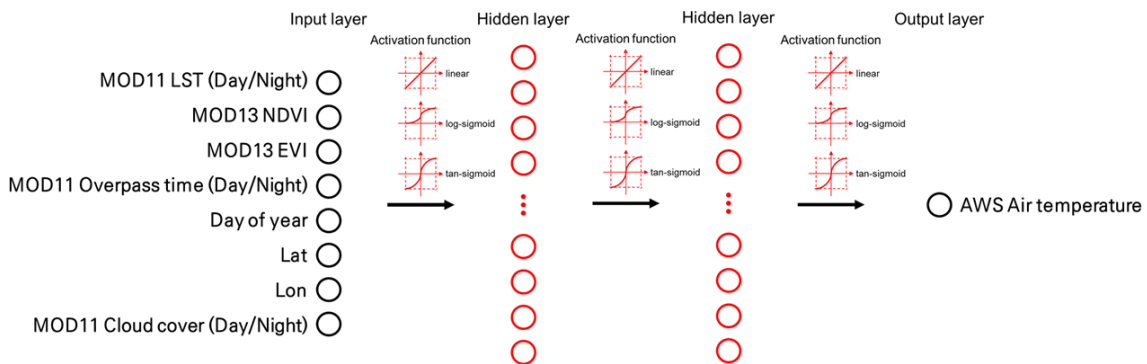


Fig. 2. ANN model structure

력자료)와 인공위성자료(입력자료) 중 상관성이 높게 나타난 자료(지표면온도, NDVI, EVI)의 산점도이다. 각 그림의 1열은 총 연구기간의 데이터(100%)를 활용한 산점도이며, 2열과 3열 순으로 training set (70%)과 test set (30%)에서의 산점도이다. 데이터량의 차이로 인해 밀도가 상대적으로 떨어지는 파란색으로 표시된 부분에서 다소 차이가 있으나, 상대적으로 데이터량이 집중되는 밀도의 분포는 유사한 거동을 보였다. 이는 training set과 test set의 분리가 적절하게 수행되었음을 나타낸다. 또한, 연구기간 동안 계절을 분리하지 않아 밀도에서는 기온이 낮은 부분과 높은 부분으로 양극화가 나타난 것으로 판단된다.

Figs. 3 and 4의 1행에 위치한 AWS 기온자료와 MOD11A1 지표면온도 자료는 높은 상관성을 보였다. 낮 기간 동안은 지표면온도가 기온보다 약 4.2°C가 높게, 밤 기간 동안은 지표면온도가 기온보다 약 1.8°C 정도 낮게 나타났으며 training set에 속하는 약 2008년부터 2016년까지의 기간이 test set에 속하는 약 2017년부터 2020년까지의 기간보다 그 차이가 크게 나타났다.

2행과 3행은 기온자료와 MOD13A2의 식생지수인 NDVI

와 EVI의 산점도이다. 산점도의 전체적인 형태는 기온이 높은 픽셀에서의 식생지수가 넓게 나타나고 기온이 낮은 픽셀에서 식생지수가 좁게 나타나는 사다리꼴 형태를 띠었다. 이는 전체 기간 동안 기온과 식생지수의 분석이 이루어져, 식생활력도가 높게 나타나는 여름의 데이터와 식생활력도가 낮게 나타나는 겨울의 데이터가 혼합되어 이와 같이 나타나는 것으로 판단된다. 또한, 연구지역이 서울시에 한정되어 있기에 상대적으로 식생활력도가 낮은 값을 나타내는 불투수층의 픽셀이 다수를 차지하여 밀도가 높게 나타났으며(노란색 및 초록색으로 표시된 부분), 양의 상관관계를 보였다.

4.2 ANN 학습 분석

본 연구에서는 3.2절에 서술한 바와 같이 머신러닝의 성능을 높이기 위해 ANN 파라미터와 데이터의 조합을 반복하며 학습하였다. 데이터의 조합은 1)지표면온도, NDVI, EVI, 2) 지표면온도, NDVI, EVI, 관측시각, DOY, 3)지표면온도, NDVI, EVI, 위도, 경도, 4)지표면온도, NDVI, EVI, 관측시각, DOY, 위도, 경도, 5)지표면온도, NDVI, EVI, 관측시각, DOY, 위도, 경도, 운량으로 5개의 조합을 구성하였다. Tables 2 and

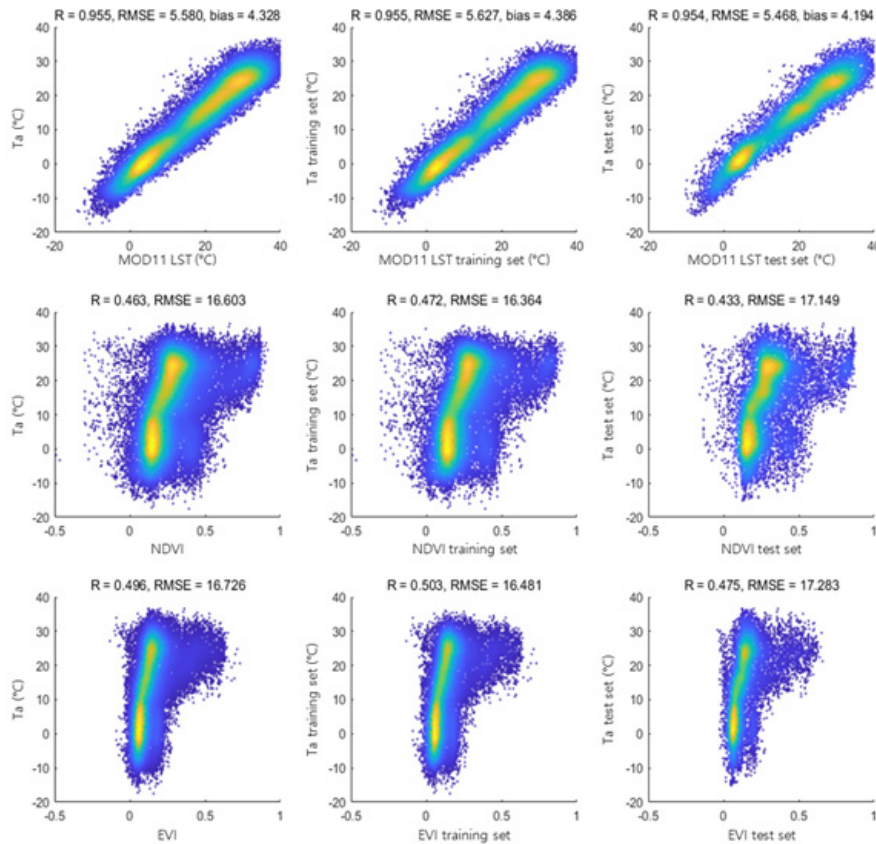


Fig. 3. Scatter plot of AWS air temperature and satellite data (land surface temperature, NDVI, EVI) at day time

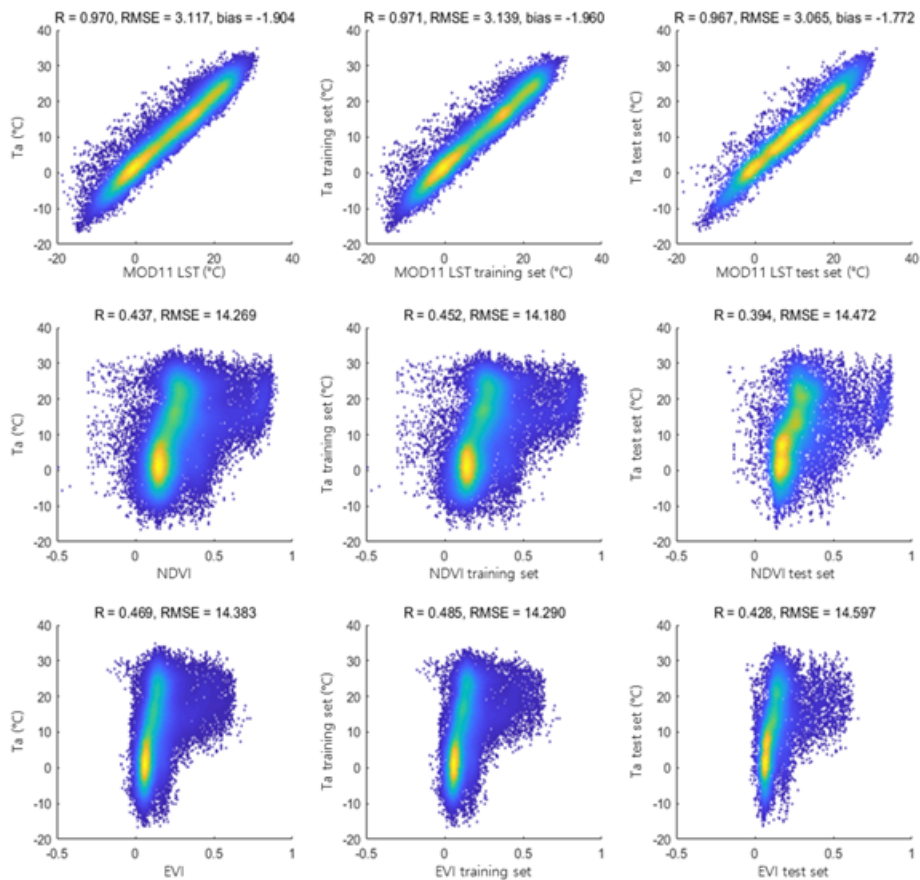


Fig. 4. Scatter plot of AWS air temperature and satellite data (land surface temperature, NDVI, EVI) at night time

Table 2. Statistical summary and parameter of ANN by data combination (day time). I = input layer, H1 = first hidden layer, H2 = second hidden layer, O = output layer

| case | r | | RMSE | | Activation Function | | | Nodes of hidden layer | |
|------|----------|--------|----------|--------|---------------------|---------|--------|-----------------------|----|
| | training | test | training | test | I → H1 | H1 → H2 | H2 → O | H1 | H2 |
| 1 | 0.9642 | 0.9616 | 2.9757 | 3.0005 | tansig | tansig | tansig | 21 | 22 |
| 2 | 0.9806 | 0.9787 | 2.1972 | 2.2405 | logsig | tansig | tansig | 24 | 23 |
| 3 | 0.9702 | 0.9681 | 2.7193 | 2.7339 | logsig | logsig | tansig | 25 | 25 |
| 4 | 0.9840 | 0.9816 | 1.9971 | 2.0833 | logsig | logsig | tansig | 25 | 26 |
| 5 | 0.9849 | 0.9825 | 1.9443 | 2.0318 | tansig | tansig | tansig | 27 | 28 |

Table 3. Statistical summary and parameter of ANN by data combination (night time). I = input layer, H1 = first hidden layer, H2 = second hidden layer, O = output layer

| case | r | | RMSE | | Activation Function | | | Nodes of hidden layer | |
|------|----------|--------|----------|--------|---------------------|---------|--------|-----------------------|----|
| | training | test | training | test | I → H1 | H1 → H2 | H2 → O | H1 | H2 |
| 1 | 0.9732 | 0.9679 | 2.3747 | 2.4812 | logsig | logsig | linear | 21 | 23 |
| 2 | 0.9801 | 0.9759 | 2.0479 | 2.1555 | logsig | linear | linear | 23 | 17 |
| 3 | 0.9786 | 0.9731 | 2.1259 | 2.2810 | logsig | linear | linear | 24 | 25 |
| 4 | 0.9854 | 0.9810 | 1.7596 | 1.9187 | tansig | tansig | linear | 26 | 27 |
| 5 | 0.9863 | 0.9823 | 1.7034 | 1.8528 | tansig | linear | tansig | 27 | 27 |

3은 각각 낮과 밤 시간대의 데이터 조합별 최적 학습 결과의 파라미터와 실제 AWS 기온 관측값과의 통계분석 결과이다.

Tables 2 and 3에서 ANN 모델에서 산출된 기온과 AWS 기온관측 자료의 통계분석 결과는 training set보다 test set의 성능이 다소 낮게 나타난다. 이는 Figs. 3 and 4의 통계분석 결과와 산점도의 분포 양상을 살펴보았을 때, test set에 해당하는 데이터의 품질이 training set에 해당하는 데이터의 품질보다 다소 낮게 나타나는 것과 연관이 있는 것으로 판단된다. 약 11만 개의 많은 데이터양, 2개의 은닉층, 21~28개에 해당하는 노드 수 등으로 인해 training set에서 다소 과적합이 되었다고 볼 수 있으나, training set과 test set에서의 통계분석 결과가 0.96 이상의 높은 상관성과 3°C 이하의 Root Mean Squared Error (RMSE) 값이 나타난 것을 보았을 때 학습이 적절히 수행되었다고 판단하였다.

자료 조합을 기준으로 결과를 분석하였을 때 전체적인 통계분석 결과와 노드 수의 경우 일정한 경향성을 보였다. 먼저, 통계분석의 경우 조합에 포함된 데이터 종류의 수가 증가할수록 자료간의 일치성이 높게 나타났다. 이는 경향성이 높은 지표면온도, NDVI, EVI 만으로도 높은 성능의 학습이 가능하지만, 관련된 유의미한 데이터를 추가할수록 학습의 성능이 더욱 증가하는 것으로 사료된다. 특히, 데이터가 추가될수록 RMSE가 약 0.7~1°C 가까이 감소하는 것으로 유의미한 오차 감소가 확인되었다. 노드 수의 경우 데이터 종류의 수가 증가할수록 노드 수 또한 증가하는 경향이 나타났다. 이는 데이터 종류의 증가에 따른 효과적인 학습을 위해 은닉층의 필요 노드 수가 증가한 것으로 사료된다.

Fig. 5는 ANN 모델로 생성한 2017년 8월 26일의 낮의 서울시 기온을 1 km의 공간해상도로 지도화한 자료이다. 본 연구에서 산출된 ANN 모델을 활용하여 기온 지도를 제작할 시, 픽셀마다 입력자료로 활용된 MOD11A1과 MOD13A2의 값이 하나라도 존재하지 않으면 기온 값이 산출되지 않게 설계하였으며, 2017년 8월 26일은 입력자료의 서울시 내 픽셀의 값이 가장 많은 날이기에 선정되었다. 한강 인근 지역 등의 일부 픽셀을 제외한 나머지 도시 내 불투수층 픽셀에서 비슷한 기온 값이 산출되었다. 두 시간대가 포함된 test set에서 ANN 모델의 성능이 우수하게 나타난 것을 고려하였을 때(Tables 2 and 3), AWS 관측소가 위치한 픽셀뿐만 아니라 관측소 인근 불투수층 픽셀에서의 기온을 우수하게 모의한 것으로 판단된다.

Fig. 6은 연구 기간인 2008년부터 2020년까지 서울시 여름철 NDVI의 평균값을 보여주고 있다. Fig. 6에서 관찰할 수 있는 공간적 패턴에서 NDVI가 높게 나타나는 곳은 서울 내 다른 지역에 비해 식생이 많은 것으로 판단할 수 있다. NDVI값의

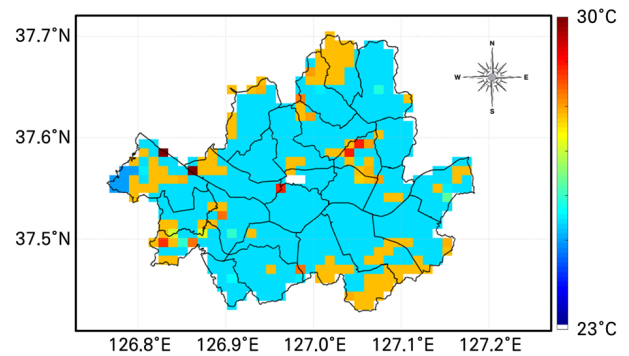


Fig. 5. Spatial distribution of daytime air temperature (Aug. 26, 2017)

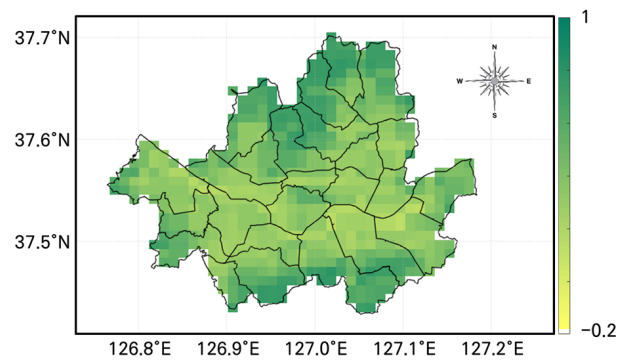


Fig. 6. Mean of summer time NDVI in Seoul

공간분포 패턴과 비슷한 패턴을 Fig. 5에서도 관찰할 수 있다. 이는 불투수층이 지배적인 픽셀의 기온과 식생이 넓게 분포한 픽셀에서의 기온 거동이 다르게 나타나므로, 추후 독립적인 학습을 위한 기준이 필요할 수 있음을 시사한다. 식생이 넓게 분포하는 픽셀에서는 인근 불투수층 픽셀보다 다소 기온이 높게 산정되는 경향을 보였다. 이러한 현상은 Figs. 3 and 4에서 나타난 바와 같이, 1) 학습에 활용된 기온 지상관측자료와 식생 지수가 양의 상관관계가 나타나는 경향을 보였으며, 2) 연구지역인 서울시는 전반적으로 도심지가 지배적이므로 식생 관련 지수들이 매우 낮은 값(NDVI: 0~0.4, EVI: 0~0.2; Figs. 3 and 4의 노란색 및 초록색으로 밀도가 높게 나타난 부분)에 집중되어 있어, 상대적으로 식생지수가 높은 지역에서는 1)의 경향성이 더욱 크게 적용되어 나타난 것으로 판단할 수 있다.

5. 결론

본 연구에서는 ANN 기법의 지도학습을 통해 MOD11A1, MOD13A2 위성자료만을 입력자료로 활용하여 서울시의 기

온자료를 산출하는 모델을 생성하였다. 이를 통해, 하루 단위 기온 자료를 우수한 성능으로 생성할 수 있었으며, 나아가 위성자료의 시공간적 특성을 반영하여 기온 관측소가 설치되지 않은 지역의 기온 자료 또한 산출할 수 있었다. 활용 데이터의 분류 및 조합의 다양화를 통해 기온 산출 모델의 성능을 향상하고자 하였으며, 본 연구의 결론은 다음과 같다.

- 1) ANN 모델에서 산출된 기온자료와 AWS 기온 지상관측 자료의 통계분석 결과 데이터 조합과 무관하게 모든 조합에서 training set의 성능이 test set보다 높게 나타났으며, training set과 test set의 r과 RMSE의 평균값은 각각 0.9787, 0.9753, 2.1845°C, 2.2779°C로 나타났다. 이는 training set과 test set으로 분류된 데이터 중 training set의 품질이 다소 우수한 것이 원인으로 사료되며, 두 데이터셋 모두 양호한 통계 분석 결과를 나타내기에 학습이 적절히 수행되었다고 판단하였다.
- 2) 데이터 조합에 기반한 ANN 성능 분석 결과 데이터 종류가 증가할수록 모델 성능이 우수하게 나타났다. 기온자료와 상관성이 높은 데이터들의 조합만으로도 r과 RMSE의 평균값이 각각 0.9667, 2.708°C로 높은 성능을 보였으며, 관련 데이터를 추가할수록 학습 결과가 더 우수한 경향을 보였으며 최종적으로 모든 데이터 종류를 활용한 조합에서는 r과 RMSE의 평균값이 0.9840, 1.883°C로 가장 높은 성능이 나타났다.
- 3) ANN 모델로 생성된 서울시 기온 지도는 픽셀별 지표면 특성에 기반하여 기온의 변동성이 적절하게 모의된 것으로 판단된다. 또한, 학습에 활용되지 않은 test set에서도 ANN 모델이 우수한 성능을 나타내는 것을 보았을 때, 위성자료만을 활용한 정확도 높은 기온자료의 생성이 가능할 것으로 사료된다.

본 연구에서는 머신러닝 학습의 우수한 성능 확보를 위해 MODIS-Terra 위성의 MOD11A1, MOD13A2 산출물과 기상청에서 제공하는 AWS 기온자료를 사용하였다. 이에 ANN 모델은 높은 성능을 산출할 수 있도록 학습이 되었으며, 1 km 공간해상도와 하루 단위 시간해상도를 갖는 기온 지도를 생성할 수 있었다. 다만, 천리안2A (2 km, 10분), Landsat-8(30 m, 16일), Sentinel-2A/B (10 m, 5일)와 같은 산출물 혹은 Level 1 자료가 제공되는 위성자료들을 활용한다면 시간 단위 기온 분석에 더욱 적절한 지도를 생성할 수 있을 것으로 기대된다. 또한, 본 연구는 위성정보를 활용하여 서울시 기온을 분석한 연구로, 추후 서울시를 비롯한 국내 도시 지역을 종합적으로 관리할 수 있는 기초 기술로 활용가능하다. 추가로, 머신러닝 모

델이 출력값을 도출하는 과정을 알기 어려운 블랙박스 특성으로 실제 입력자료들이 출력자료에 끼치는 영향을 분석하기 어려움이 있었다. 추후 training과 test set의 교차검증 방법과 다중공선성, 주성분 분석, T-distributed stochastic neighbor embedding 등의 데이터 분석 및 해석 방법 등을 활용할 시, 최소한의 비용으로 최적의 효과를 산출할 수 있는 인공지능 모델을 설계할 수 있을 것으로 기대된다.

감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 RS-2022-00155763).

본 연구는 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2022R1A2C2010266)과 교육부 및 한국연구재단의 4단계 두뇌한국21 사업(4단계 BK21 사업)과 2021년도 정부(교육부)의 재원으로 한국연구재단(NRF-2021R1A6A3A01087645)의 지원을 받아 수행된 연구임.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Baik, J., Liaqat, U.W., and Choi, M. (2018). "Assessment of satellite and reanalysis-based evapotranspiration products with two blending approaches over the complex landscapes and climates of Australia." *Agricultural and Forest Meteorology*, Vol. 263, pp. 388-398.
- Defernez, M., and Kemsley, E.K. (1999). "Avoiding overfitting in the analysis of high-dimensional data with artificial neural networks (ANNs)." *Analyst*, Vol. 124, No. 11, pp. 1675-1681.
- Hao, Y., Baik, J., and Choi, M. (2019). "Developing a soil water index-based Priestley - Taylor algorithm for estimating evapotranspiration over East Asia and Australia." *Agricultural and Forest Meteorology*, Vol. 279, 107760.
- Harmay, N.S.M., and Choi, M. (2022). "Effects of heat waves on urban warming across different urban morphologies and climate zones." *Building and Environment*, Vol. 209, 108677.
- Hunter, D., Yu, H., Pukish III, M.S., Kolbusz, J., and Wilamowski, B.M. (2012). "Selection of proper neural network sizes and architectures - A comparative study." *IEEE Transactions on Industrial Informatics*, Vol. 8, No. 2, pp. 228-240.

- Jee, J.B., Lee, K.T., and Choi, Y.J. (2014). "Analysis of land surface temperature from MODIS and Landsat satellites using by AWS temperature in capital area." *Korean Journal of Remote Sensing*, Vol. 30, No. 2, pp. 315-329.
- Lawrence, S., and Giles, C.L. (2000). "Overfitting and neural networks: conjugate gradient and backpropagation." *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, IEEE, Como, Italy, Vol. 1, pp. 114-119.
- Mendelsohn, R., Kurukulasuriya, P., Basist, A., Kogan, F., and Williams, C. (2007). "Climate analysis with satellite versus weather station data." *Climatic Change*, Vol. 81, No. 1, pp. 71-83.
- Ministry of the Interior and Safety (MOIS) (2021). Korea, accessed 28 July 2022, <<https://www.data.go.kr/data/15046541/fileData.do>>.
- National Geographic Information Institute (NGII) (2019). *The national atlas of Korea I*. pp. 68-75.
- Nichol, J.E., Fung, W.Y., Lam, K.S., and Wong, M.S. (2009). "Urban heat island diagnosis using ASTER satellite images and 'in situ' air temperature." *Atmospheric Research*, Vol. 94, No. 2, pp. 276-284.
- Niclòs, R., Puchades, J., Coll, C., Barberà, M. J., Pérez-Planells, L., Valiente, J.A., and Sánchez, J.M. (2021). "Evaluation of Landsat-8 TIRS data recalibrations and land surface temperature split-window algorithms over a homogeneous crop area with different phenological land covers." *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 174, pp. 237-253.
- Park, J., Baik, J., and Choi, M. (2017). "Satellite-based crop coefficient and evapotranspiration using surface soil moisture and vegetation indices in Northeast Asia." *Catena*, Vol. 156, pp. 305-314.
- Peterson, T.C. (2003). "Assessment of urban versus rural in situ surface temperatures in the contiguous United States: no difference found." *Journal of Climate*, Vol. 16, No. 18, pp. 2941-2959.
- Richter, R. (1997). "Correction of atmospheric and topographic effects for high spatial resolution satellite imagery." *International Journal of Remote Sensing*, Vol. 18, No. 5, pp. 1099-1111.
- Rigo, G., Parlow, E., and Oesch, D. (2006). "Validation of satellite observed thermal emission with in-situ measurements over an urban surface." *Remote Sensing of Environment*, Vol. 104, No. 2, pp. 201-210.
- Sobrino, J.A., Caselles, V., and Coll, C. (1993). "Theoretical split-window algorithms for determining the actual surface temperature." *Il Nuovo Cimento C*, Vol. 16, No. 3, pp.219-236.
- Tewari, M., Yang, J., Kusaka, H., Salamanca, F., Watson, C., and Treinish, L. (2019). "Interaction of urban heat islands and heat waves under current and future climate conditions and their mitigation using green and cool roofs in New York City and Phoenix, Arizona." *Environmental Research Letters*, Vol. 14, No. 3, 034002.
- Vanhellemont, Q., and Ruddick, K. (2018). "Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications." *Remote Sensing of Environment*, Vol. 216, pp. 586-597.
- Weng, Q. (2009). "Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends." *ISPRS Journal of Photogrammetry and Remote sensing*, Vol. 64, No. 4), pp. 335-344.
- Yu, Y., Privette, J.L., and Pinheiro, A.C. (2007). "Evaluation of split-window land surface temperature algorithms for generating climate data records." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 46, No. 1, pp. 179-192.
- Zheng, Y., Ren, H., Guo, J., Ghent, D., Tansey, K., Hu, X., Nie, J., and Chen, S. (2019). "Land surface temperature retrieval from sentinel-3A sea and land surface temperature radiometer, using a split-window algorithm." *Remote Sensing*, Vol. 11, No. 6, 650.