

# LDA 토픽 모델링을 활용한 글로벌 기상 분야의 인공지능 연구동향 분석 (2018~2024년 논문 주제 중심)

표성훈<sup>1),2)</sup> · 김태종<sup>1),2)\*</sup>

<sup>1)</sup>국립기상과학원 인공지능기상연구과, <sup>2)</sup>기상청 인공지능기상기술연구회

(접수일: 2024년 10월 14일, 수정일: 2024년 11월 8일, 게재확정일: 2024년 12월 10일)

## Analysis of Global Research Trends in Artificial Intelligence for Meteorology Using LDA Topic Modeling (Focus on Papers from 2018 to 2024)

Seong-Hun Pyo<sup>1),2)</sup> and Tae-Jong Kim<sup>1),2)\*</sup>

<sup>1)</sup>AI Meteorological Research Division, National Institute of Meteorological Sciences, Jeju, Korea

<sup>2)</sup>Artificial Intelligence Meteorological Technology Research Society,  
Korea Meteorological Administration, Deajeon, Korea

(Manuscript received 14 October 2024; revised 8 November 2024; accepted 10 December 2024)

**Abstract** This study aims to analyze research trends related to Artificial Intelligence (AI) in the global meteorological field from 2018 to 2024 and identify the major research topics and keywords. By utilizing the Web of Science database, a total of 5,846 papers related to AI in meteorology were identified and analyzed. The study employed latent Dirichlet allocation (LDA) topic modeling to extract the main research topics. The optimization of topic modeling parameters was performed by adjusting document-topic density (alpha) and word-topic density (beta) distributions, which control the concentration of topics in documents and words in topics, respectively. Through comprehensive parameter optimization, the model achieved the coherence score of 0.639 with alpha value of 0.08, beta value of 0.01, and 6 topics, indicating clear and well-separated research themes in the field. These optimal parameter values were used for the topic modeling analysis. The analysis revealed that (1) research on 'AI-based prediction of hydrological variables' encompasses studies applying AI techniques to predict hydrological variables such as rainfall and evaporation, aiming for more precise meteorological forecasting. (2) Studies on 'AI-based analysis of the impacts of climate change' utilize AI models to analyze the effects of climate change on various regions and ecosystems, assessing potential impacts under different climate change scenarios and predicting future environmental changes. (3) Research on 'AI-based prediction of oceanic and surface temperatures' focuses on improving the accuracy of meteorological and environmental observations by predicting ocean and land surface temperatures using satellite data. (4) Studies on 'Machine learning-based risk assessment and prediction of natural disasters' evaluate and predict the likelihood of natural disasters such as floods and landslides, providing crucial information for disaster management and prevention. (5) Research on 'AI and meteorological data utilization for real-time rainfall prediction' aims to enhance the

\*Corresponding Author: Tae-Jong Kim, National Institute of Meteorological Sciences, 33 seohobuk-ro, Seogwipo-si, Jeju-do 63568, Korea.  
Phone: +82-64-780-6763, Fax: +82-64-738-4914  
E-mail: k2boy3@naver.com

accuracy of real-time rainfall forecasting by combining meteorological radar data with AI techniques, playing a critical role in rapidly changing weather conditions. (6) Studies on 'AI utilization in wind power forecasting and meteorological condition analysis' aim to optimize wind energy production by predicting wind speed and weather conditions, contributing to efficient energy management. This study systematically analyzes research trends related to the application of AI in meteorology, contributing to the academic development of the field and suggesting future research directions. Specifically, by identifying research trends through topic modeling, this study provides a structured understanding of the convergence of meteorology and AI, offering valuable foundational data to researchers in the field.

**Key words:** Artificial intelligence, Weather forecasting, Climate change, Topic modeling, Machine learning

## 1. 서 론

기후 변화는 기존의 기상현상이 더욱 강렬하게 변화하거나 새로운 양상의 극한 기상 현상을 통해 나타나면서 지속적으로 증가하는 추세를 보인다(IPCC, 2021; Clarke et al., 2022). 이에 따라 기상 예측과 자연재해 대응은 전 세계적으로 중대한 과제로 떠오르고 있다. 기후 변화는 인간 사회와 생태계에 광범위한 영향을 미치며(Rosenzweig et al., 2007), 그로 인해 기상학적 연구의 중요성은 더욱 부각되고 있다. 특히, 급격한 기상변화로 인한 재난을 사전에 예측하고 대비하기 위한 기술적 진보는 환경 관리와 재해 대응의 필수적 요소로 자리 잡고 있다(Majumdar et al., 2021; IPCC, 2022). 이러한 배경에서 기상 데이터의 양적 증가와 복잡성은 더욱 큰 도전 과제로 작용하고 있으며(Fathi et al., 2022), 이를 효과적으로 처리하고 분석할 수 있는 새로운 방법론에 대한 요구가 증가하고 있다.

최근 인공지능(artificial intelligence, AI)은 기상학의 이러한 문제를 해결하기 위한 유망한 도구로 주목받고 있다(Hewage et al., 2021). 인공지능, 특히 기계학습과 딥러닝 기법은 방대한 양의 데이터를 효율적으로 분석하고 패턴을 추출하는 특징으로, 기존의 물리 기반 모델을 보완하거나 대체할 수 있어 기상 예측, 자연재해 예측, 기후 변화 분석 등 다양한 분야에서 인공지능의 도입은 새로운 가능성을 열어주었으며, 이를 통해 예측 정확도가 개선되고, 실시간 분석 능력이 강화되고 있다(Hinton, 2006; Jozefowicz et al., 2015; Kim and Reiter, 2017; Lea et al., 2017).

따라서 그 활용 범위와 성과를 체계적으로 평가하고 분석하는 연구는 필수적이다. 특히, 기상학에서 인공지능 기술이 구체적으로 어떤 연구 주제와 방법론에 적용되고 있는지에 대한 명확한 동향 분석이 제대로 이루어지면 기상과 인공지능을 결합하는 연구를 시작함에 있어서 발전 방향과 성과를 명확히 알고 더 진보된 기술 개발이 가능할 것이다. 본 연구는 2018

년부터 2024년까지의 글로벌 기상 분야에서 인공지능 관련 연구 동향을 분석하고, 주요 연구 주제와 시사점을 도출하는 것을 목적으로 한다.

본 연구에서는 Web of Science (WoS) 데이터베이스에서 기상 분야의 인공지능 관련 논문을 수집하고, 잠재 디리클레 할당(latent Dirichlet allocation, LDA) 기법을 적용하여 주요 연구 주제를 도출하였다. 이를 통해 기상 예측, 기후 변화 분석, 자연재해 위험 평가 등에서 인공지능이 어떤 방식으로 기여하고 있는지를 체계적으로 분석하고, 향후 기상학과 인공지능 간의 융합 연구가 나아갈 방향을 제시하고자 한다. 따라서 본 연구는 인공지능이 기상학에서 어떻게 활용되고 있으며, 그로 인해 기상 예측 및 분석의 정확성과 효율성이 어떻게 향상되고 있는지를 명확히 이해하고, 향후 기상학과 인공지능 연구의 발전 방향을 제시하는 데 중요한 학술적 기초를 제공할 것이다.

## 2. 연구 데이터 및 방법

### 2.1 데이터 수집

본 연구는 2018년부터 2024년까지 글로벌 기상 분야에서 인공지능 관련 연구 동향을 분석하기 위해, WoS 데이터베이스에서 관련 논문을 수집하여 분석하였다. WoS는 다양한 학술 분야에서 고품질의 논문을 제공하는 신뢰할 수 있는 데이터베이스로, 기상학 및 인공지능 관련 연구를 폭넓게 포함하고 있어 본 연구에 적합하다. 데이터 수집 기간은 인공지능이 기상학 분야에서 활발히 연구되기 시작한 2018년부터 현재 연구시점인 2024년 6월까지로 설정하여 검색을 수행하였다.

초기 검색을 통해 총 9,332건의 논문이 수집되었으며, 이 중 기상학과 인공지능 간의 직접적인 연관성이 낮은 3,486건의 논문을 제외하였다. 논문 제목과 초록을 검토하여, 기상학적 응용과 인공지능 기법의 활용 여부를 기준으로 논문을 선별한 결과, 최종적으로

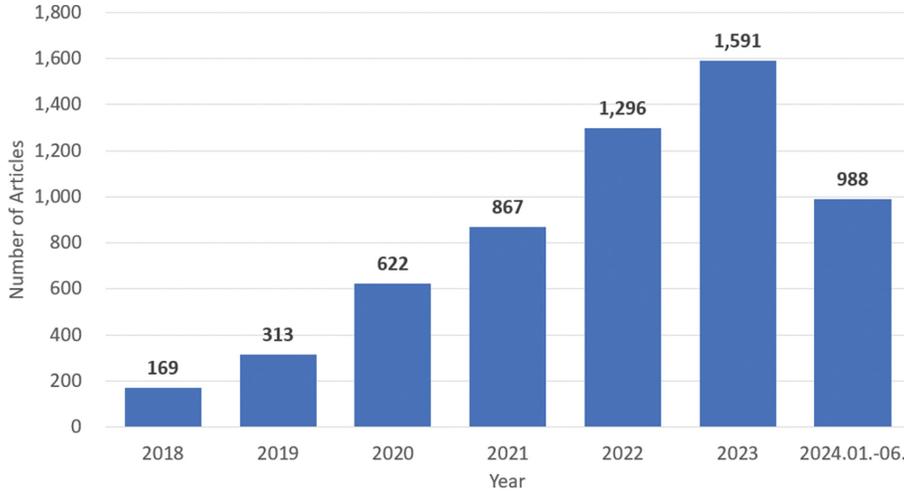


Fig. 1. The number of research papers utilizing artificial intelligence in the field of meteorology from 2018 to June 2024, categorized by year.

로 Fig. 1과 같이 5,846건의 논문이 분석 대상으로 선정되었다. 이러한 방법론은 대규모 학술 데이터에서 주요 연구 경향을 효과적으로 파악하는 접근법으로 활용되었다(Blei et al., 2003).

## 2.2 키워드 정제 및 분석

본 연구에서 수집된 논문 초록에 대한 데이터 전처리는 주요 키워드의 정제 과정을 중심으로 이루어졌다. 우선, 각 논문의 핵심 키워드를 도출하기 위해 NTF (normalized frequency analysis)과 TF-IDF (term frequency-inverse document frequency) 분석을 적용하였다. 빈도 분석은 논문 초록에서 자주 등장하는 키워드를 추출하는 데 사용되었으며, TF-IDF 분석은 특정 단어가 전체 문서 집합 내에서 얼마나 중요한지를 정량적으로 평가하는 기법으로, 연구 키워드의 상대적인 중요도를 파악하는 데 유용하였다(Jones, 1972). 이러한 분석 기법은 대규모 텍스트 데이터를 처리할 때 연구의 핵심 용어와 주제를 효과적으로 추출하는 데 중요한 역할을 한다.

$$NTF(t, d) = \frac{f_{t,d}}{N_d}$$

$f_{t,d}$ : 단어  $t$ 가 문서  $d$ 에 등장한 횟수

$N_d$ : 문서  $d$ 의 총 단어 수

$$IDF(t, D) = \log\left(\frac{|D|}{|\{d \in D : t \in d\}|}\right)$$

$|D|$ : 전체 문서의 수

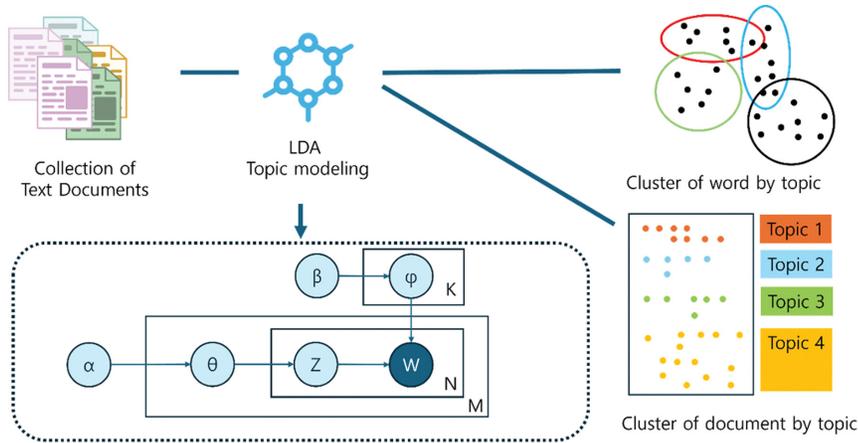
$|\{d \in D : t \in d\}|$ : 단어  $t$ 가 등장하는 문서의 수

키워드 분석의 정확성과 일관성을 높이기 위해 유사어 사전, 정의어 사전, 제외어 사전을 구축하였다. 유사어 사전은 동일한 개념을 나타내는 다양한 표현을 통합하여 키워드 간의 중복을 제거하고, 각 용어의 의미를 통일시켜 분석의 일관성을 유지하였다. 정의어 사전은 연구 주제와 밀접하게 연관된 기상학 및 인공지능 관련 용어를 체계적으로 정의하여, 논문 분석 결과가 연구 목적에 부합하도록 한다. 제외어 사전은 기상학 및 인공지능 연구와 관련성이 낮거나 일반적으로 자주 사용되는 의미 없는 단어들을 배제하는 데 사용되었으며, 이를 통해 노이즈를 제거하고 분석의 정확성을 향상시켰다. 이러한 전처리 과정을 통해 논문 초록에 포함된 주요 키워드가 연구 주제를 정확히 반영하고, 분석 결과의 신뢰도를 높일 수 있었다.

이와 같은 데이터 전처리 방법론은 대규모 텍스트 분석에서 필수적인 단계로, 기상학적 용어와 인공지능 기법의 활용에 대한 연구 주제의 일관성을 확보하는 데 기여하였다. 이를 통해 본 연구는 글로벌 기상 분야에서 인공지능 관련 연구 동향을 보다 체계적이고 정확하게 분석할 수 있었다.

## 2.3 토픽 모델링 분석

본 연구에서는 LDA 기법을 활용한 토픽 모델링을 핵심 분석 방법으로 적용하였다. Figure 2의 LDA는 대규모 텍스트 데이터에서 잠재적인 주제를 자동으로 추출하는 데 널리 사용되는 확률적 기법으로, 문서 내에서 다양한 주제가 혼합되어 있을 때 각 문서가 어느 정도의 확률로 특정 주제에 속하는지를 분석하는 방법이다(Blei et al., 2003). 이 기법은 특히 대규모 학



- M: 문헌의 수
- K: 주제의 수
- N: 문헌에 속한 단어의 수
- $\Theta$ : 문헌의 주제 분포
- $\Phi$ : 주제의 단어 분포
- Z: 해당 단어가 속한 주제의 번호

**Fig. 2.** A diagram illustrating the structure of the Latent Dirichlet Allocation (LDA) algorithm and the method for determining topics.

술 데이터에서 주제 분포를 분석하고 연구 경향을 파악하는 데 유용하며, 기상학과 인공지능 연구의 주요 흐름을 체계적으로 분류할 수 있는 강력한 도구로 활용된다. 아래 식은 LDA의 식이며 각 변수별 의미를 설명하며, 디리클레 분포의 특성은 LDA 토픽 모델링에서 문서 집합의 잠재적 주제 구조를 발견하는 데 핵심적인 역할을 한다. 특히 alpha 값의 조정을 통해 토픽의 분포 특성을 제어할 수 있어, 연구 목적에 맞는 최적의 토픽 구조를 도출하는 데 활용된다.

$$Dir(x; \alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^K \chi_i^{\alpha_i - 1}$$

$\chi$ : K-차원 확률 벡터, 즉 K개의 범주에 대한 확률을 나타내는 값들로  $0 \leq x \leq 1$ 이고  $\sum_{i=1}^K x_i = 1$  을 만족하며 예를 들어, 특정 문서에서 각 토픽이 나타날 확률을 의미함

$\alpha$ : 디리클레 분포의 매개변수로 각 범주에 대한 집중도를 결정하는 값으로 문서나 단어가 특정 토픽에 얼마나 집중될지를 제어함.

$B(\alpha)$ : 베타함수로 분포의 정규화 상수

$$B(\alpha) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K \alpha_i)}$$

$\alpha_i$ : 각 토픽이 문서 내에서 차지하는 비율에 대한 변수  
 $\Gamma(\alpha_i)$ :  $\alpha_i$ 의 값에 따라 확률값을 변형시키는 함수  
 $\Gamma(\sum_{i=1}^K \alpha_i)$ : 여러 개의 토픽이나 단어 분포가 하나의 문서나 하나의 토픽에서 정규화 되어 더해져 1이 되도록 보정하는 값

본 연구에서는 5,846건의 논문을 대상으로 LDA 토픽 모델링을 사용하여 주요 연구 주제 및 키워드를 분석하였다. LDA는 확률적 생성 모델로, 문서 내 단어의 출현 패턴을 바탕으로 잠재적 주제를 도출한다. 이를 위해, 본 연구는 다음과 같은 과정을 통해 키워드와 논문의 중요도를 평가하였다.

첫째, 키워드별 확률 분석 과정으로, LDA 모델은 각 주제에 속하는 단어들의 분포를 확률적으로 추정한다. 각 키워드는 특정 주제 내에서 나타날 확률을 기반으로 중요도가 결정된다. 확률이 높을수록 해당 키워드가 주제를 대표할 가능성이 높다는 것을 의미한다. 예를 들어, ‘deep learning’ 키워드가 특정 주제에서 상대적으로 높은 확률로 나타난다면, 이는 해당 주제의 주요 특징을 잘 나타내는 주요 키워드로 간주된다. 이러한 확률 기반 접근법은 주제의 의미적 일관성을 확보하며, 키워드 간의 상관성을 분석하는 데 효과적이다. 본 논문에서는 LDA 모델에 의해 결정된 토픽별 주요 키워드를 출현확률과 함께 바 차트로 제

시켰다.

둘째, 논문별 확률 분석 과정으로, LDA 모델은 각 논문이 특정 주제에 속할 확률도 계산한다. 논문별 확률이 높을수록 해당 논문이 주제와 강하게 연관되어 있음을 나타낸다. 예를 들어, GANs 기법을 활용한 논문이 실시간 강수 예측 관련 주제에 상대적으로 높은 확률로 속한다고 분석되었다면, 이 논문은 해당 주제의 주요 연구로 평가된다. 이를 통해 논문들이 각 주제와 어떻게 연관되는지를 체계적으로 분석할 수 있다. 본 논문에서는 LDA 모델에 의해 결정된 토픽별 주요 논문을 출현확률과 함께 표로 제시하면서, 주요 논문들에서 사용된 세부 인공지능 기법들을 제시했다.

셋째, 확률 기반 중요성 평가 과정으로, LDA 모델의 특성상, 높은 확률을 가진 키워드 및 논문은 주제의 핵심적인 구성 요소로 간주된다. 키워드 확률이 높다는 것은 해당 용어가 주제를 정의하는 데 중요한 역할을 한다는 것을 의미하며, 논문 확률이 높다는 것은 해당 논문이 주제의 중심 연구라는 것을 나타낸다. 이러한 확률 기반 평가 방법은 연구자의 주관을 최소화하고, 데이터를 기반으로 연구 주제의 주요 흐름을 파악하는 데 기여한다. 이를 통해 본 연구에서는 인공지능과 기상학 분야의 상호작용을 보다 명확하게 이해하고, 두 분야 간의 연구 동향을 분석할 수 있다.

LDA 모델링 과정에서는 토픽 수(number of topics)를 사전에 설정하는 것이 중요한 단계이다. 본 연구에서는 다양한 토픽 수를 시도하여 최적의 주제 개수를 결정하였으며, 최종적으로 가장 높은 토픽 일관성(topic coherence)을 보이는 토픽 수를 선택하였다. 토픽 일관성은 주제의 해석 가능성과 관련된 지표로, 각 주제 내에서 연관성이 높은 단어들이 함께 나타나는 정도를 측정한다. 이를 통해 본 연구는 LDA 모델이 도출한 주제가 실제로 논문 내의 주요 연구 흐름을 잘 반영하고 있음을 검증하였다.

LDA는 기상학 및 과학적 문헌에서 연구 주제를 추출하는 데 자주 사용되는 기법이며, 본 연구에서도 이를 통해 기상학과 인공지능의 교차 연구 분야에서 중요한 주제를 도출할 수 있었다. LDA 모델의 결과는 향후 기상 예측, 기후 변화 연구 등에서 인공지능이 어떤 역할을 할 수 있는지에 대한 통찰을 제공하는 중요한 기초 자료로 활용될 수 있다.

## 2.4 모델 최적화 및 검증

LDA 토픽 모델링의 최적화 과정을 Coherence Score를 사용하여 수행하였다. Coherence Score는 도출된 토픽이 의미적으로 일관성 있고 해석 가능성이 높은지를 평가하는 지표로, 주제 모델링의 품질을 판단하는 데 중요한 역할을 한다(Röder et al., 2015). 높은 Coherence Score는 도출된 주제들이 서로 강하게 연

관되어 있으며, 논리적인 해석이 가능함을 나타낸다. 이를 통해 본 연구는 LDA 모델의 성능을 최적화하고, 보다 신뢰할 수 있는 분석 결과를 도출하고자  $C_v$  계산법을 사용하였다.  $C_v$  계산법을 이용하여 Coherence Score를 계산하는 방법은 상위 단어 쌍의 빈도 기반으로 단어 간의 의미적 유사성을 측정하는 방법이다.

$$PMI(\omega_i, \omega_j) = \log \frac{P(\omega_i, \omega_j)}{P(\omega_i) \cdot P(\omega_j)}$$

$$NPMI(\omega_i, \omega_j) = \frac{PMI(\omega_i, \omega_j)}{-\log P(\omega_i, \omega_j)}$$

$$Coherence(T, D) = \sum_{i=2}^{|T|} \sum_{j=1}^{i-1} NPMI(\omega_i, \omega_j)$$

$T$ : 토픽 내 단어들의 집합

$PMI(\omega_i, \omega_j)$ : 점별 상호정보량(pointwise mutual information), 두 단어가 독립적일 때에 비해 실제로 함께 등장하는 빈도가 얼마나 높은지를 나타내는 지표

$P(\omega_i, \omega_j)$ : 두 단어가 동시에 문서 내에 등장할 확률

$NPMI(\omega_i, \omega_j)$ : 두 단어  $\omega_i, \omega_j$ 가 정규화된 점진적 상호 정보량으로, 두 단어가 함께 등장할 확률과 각 단어가 단독으로 등장할 확률 간의 비율

$\log P(\omega_i, \omega_j)$ : 두 단어가 함께 등장할 확률의 로그 값

LDA 모델 최적화의 주요 단계는 알파( $\alpha$ ) 값, 베타( $\beta$ ) 값, 그리고 토픽 수를 조정하는 것이었다. 알파 값은 문서별로 할당되는 주제의 분포를 제어하며, 값이 낮을수록 문서가 소수의 주제에 더 집중되는 경향을 보인다. 반면, 베타 값은 주제 내에서 단어들이 얼마나 고르게 분포되는지를 결정하며, 낮은 베타 값은 주제 내에서 소수의 핵심 단어가 주도하는 경향을 나타낸다. 본 연구에서는 알파 값을 0.01에서 0.1까지, 베타 값을 0.01에서 0.02까지 설정하였으며, 토픽 수는 2개에서 11개까지 다양하게 조정하였다. 이 과정을 통해 총 200개의 경우의 수에 대해 Coherence Score를 측정하였다.

## 3. 연구 결과

LDA 토픽 모델링의 각 조합에서의 Coherence Score를 비교한 결과, Fig. 3과 같이 알파 값 0.08, 베타 값 0.01, 토픽 수 6개일 때 가장 높은 Coherence Score인 0.639를 기록하였다. 이러한 최적의 파라미터 설정은 주제 간의 일관성을 최대화하고, 논문 초록에서 도출된 주제가 연구의 본질을 가장 잘 반영할 수 있도록 돕는다. 특히, 알파 값 0.08은 문서 내에서 다수의 주제 간 균형을 유지하면서도 소수의 핵심 주제에 집중될 수 있는 환경을 제공하며, 베타 값 0.01은 각 주제

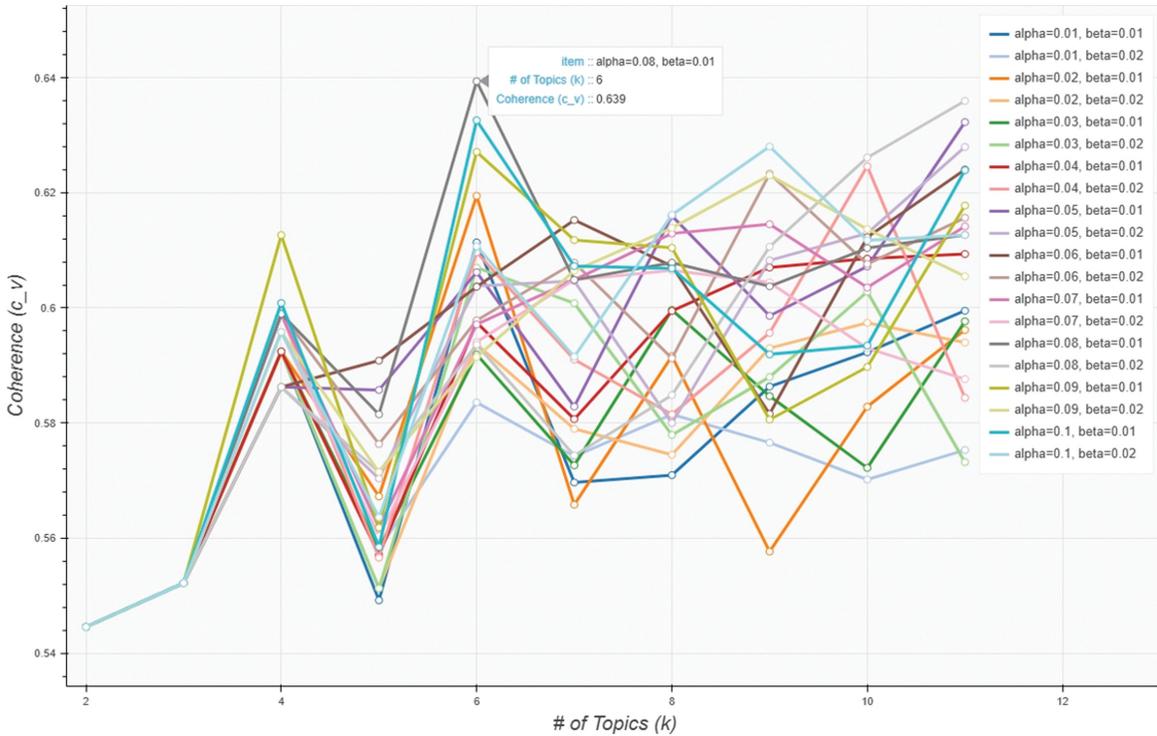
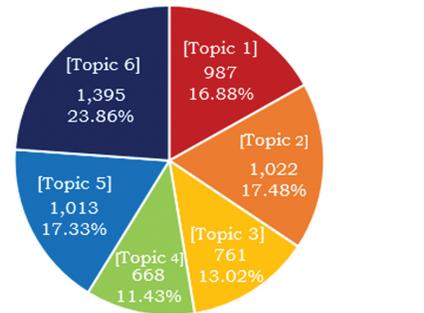


Fig. 3. A diagram showing the variation in c\_v coherence values (alpha, beta) based on the number of topics.

내에서 소수의 관련성이 높은 키워드들이 주도적으로 작용하도록 설정되었다. 최종적으로 도출된 6개의 토픽은 인공지능이 기상학 분야에서 어떻게 응용되고 있는지를 설명하는 데 중요한 역할을 하였다. 이 최적화 과정은 LDA 모델의 신뢰성을 높이고, 본 연구의 결과가 기상 분야와 인공지능 연구 간의 주요 교차점을 효과적으로 반영할 수 있도록 기여하였다. Coherence Score를 통한 성능 검증은 주제 모델링의 일관성을 평가하는 중요한 방법으로, 이를 통해 본 연구는 더 명확하고 해석 가능한 연구 주제를 도출할 수 있었다.

총 6개의 각 주제의 비중은 전체 논문에서 해당 주제가 차지하는 비율을 의미하며, 이를 통해 인공지능이 기상학 분야에서 어느 정도로 다루어지고 있는지를 알 수 있다. 또한, 각 주제의 시계열적 변화 추이를 분석함으로써, 특정 주제가 시간이 지남에 따라 증가하는 경향이 있는지, 감소하는 경향이 있는지를 파악할 수 있다.

Figure 4에서 각 토픽의 비중을 살펴보면, 토픽 6(풍력 예측 및 기상 조건 분석에서의 인공지능 활용)이 전체 논문 중에서 23.86%로 가장 높은 비중을 차지하고 있다. 이는 풍력 예측과 관련된 연구가 기상학에서 매우 중요한 역할을 하고 있음을 시사한다. 그



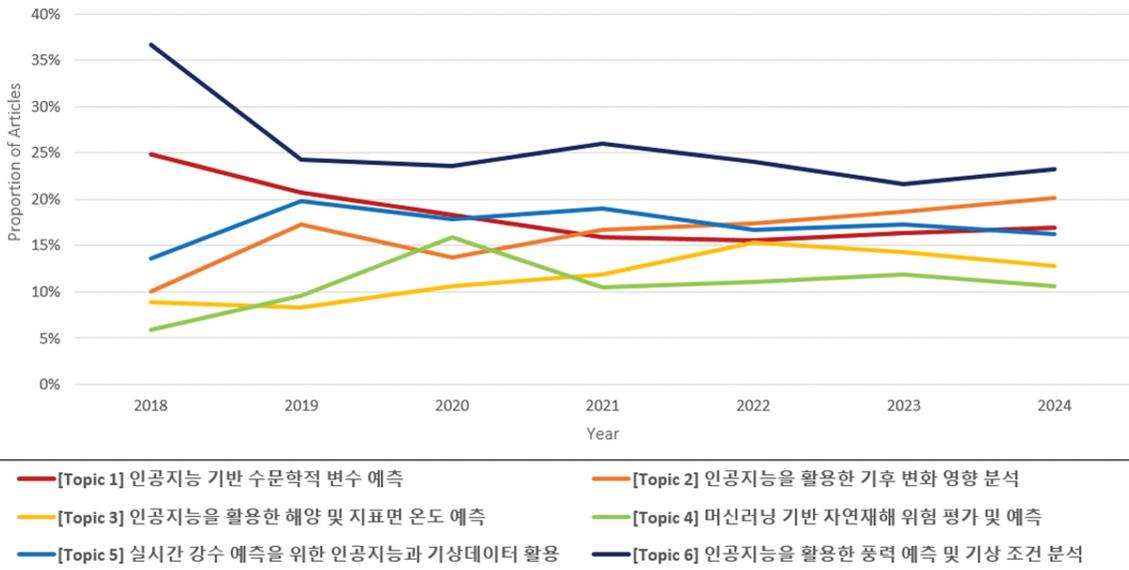
- [Topic 1] 인공지능 기반 수문학적 변수 예측
- [Topic 2] 인공지능을 활용한 기후변화 영향 분석
- [Topic 3] 인공지능을 활용한 해양 및 지표면 온도 예측
- [Topic 4] 머신러닝 기반 자연재해 위험 평가 및 예측
- [Topic 5] 실시간 강수 예측을 위한 인공지능과 기상 데이터의 활용
- [Topic 6] 풍력 예측 및 기상 조건 분석에서의 인공지능 활용

Fig. 4. A diagram illustrating the proportion of each of the six topics.

뒤를 이어 토픽 2(기후 변화 영향 분석)가 17.48%, 토픽 5(실시간 강수 예측을 위한 인공지능 활용)가 17.33%의 비중을 차지하고 있다. 이러한 결과는 기후 변화

**Table 1.** A table showing the proportion, trend (Increase/Decrease), and trajectory (Emerging/Declining) of each topic.

No.	Topic	Proportion (%)	Increase/Decrease Trend (Predicted Equation from Linear Regression)	Emerging/Declining Topic
1	[Topic 6] 풍력 예측 및 기상 조건 분석에서의 인공지능 활용	23.86	Decrease ( $y = -0.0161x + 0.3209$ )	Declining Topic
2	[Topic 2] 인공지능을 활용한 기후 변화 영향 분석	17.48	Increase ( $y = 0.0131x + 0.1102$ )	Emerging Topic
3	[Topic 5] 실시간 강수 예측을 위한 인공지능과 기상데이터의 활용	17.33	Stable ( $y = 0.0007x + 0.1696$ )	Sustained Topic
4	[Topic 1] 인공지능 기반 수문학적 변수 예측	16.88	Decrease ( $y = -0.0127x + 0.2345$ )	Declining Topic
5	[Topic 3] 인공지능을 활용한 해양 및 지표면 온도 예측	13.02	Increase ( $y = 0.0101x + 0.0768$ )	Emerging Topic
6	[Topic 4] 머신러닝 기반 자연재해 위험 평가 및 예측	11.43	Increase ( $y = 0.005x + 0.088$ )	Emerging Topic



**Fig. 5.** A time-series graph illustrating the change in the proportion of papers corresponding to each topic over the years.

와 강수 예측이 기상학에서 중요한 연구 주제를 보여준다. 그 외에 토픽 1(수문학적 변수 예측)은 16.88%, 토픽 3(해양 및 지표면 온도 예측)은 13.02%, 토픽 4(자연재해 위험 평가)는 11.43%의 비중을 보였다.

Table 1과 Fig. 5에서 시간에 따른 각 토픽의 비중 변화를 분석한 결과, 비율의 증가 추세를 보이는 주제와 비율의 감소 추세를 보이는 주제로 나눌 수 있다. 토픽 2(기후 변화 영향 분석), 토픽 3(해양 및 지표면 온도 예측), 토픽 4(자연재해 위험 평가)는 시간이 지남에 따라 비중이 점차 증가하고 있는 것으로

나타났으며, 이는 기상학에서 인공지능을 활용한 기후 변화 분석과 자연재해 예측 연구의 비율이 지속적으로 증가하고 있음을 의미한다.

반면, 토픽 6(풍력 예측 및 기상 조건 분석에서의 인공지능 활용)과 토픽 1(수문학적 변수 예측)은 시간이 지남에 따라 비율이 상대적으로 감소하는 경향을 보였다. 하지만 이는 해당 분야에서 인공지능을 활용한 연구가 상대적으로 덜 중요해 졌다고 해석할 순 없다. 절대적인 논문의 수가 감소한 것이 아니라 오히려 증가하였으나 다른 주제의 연구활동이 증가하여

상대적으로 비율이 감소한 결과로 해석할 수 있다.

### 3.1 토픽 1(인공지능 기반 수문학적 변수 예측) 분석결과

토픽 1은 인공지능 기반 수문학적 변수 예측을 주제로 한다. 이 주제에서는 인공지능 기법을 적용하여 강수량 예측, 증발량 추정 등 다양한 수문학적 변수를 예측하는 연구들이 포함된다. 이러한 연구들은 기상학적 수문학 데이터의 복잡성을 다루기 위해 머신러닝과 딥러닝 기법을 사용하고 있으며, 이를 통해 기존의 물리적 모델보다 더 정밀한 예측을 가능하게 하고 있다. Figure 6의 주요 키워드는 rainfall prediction,

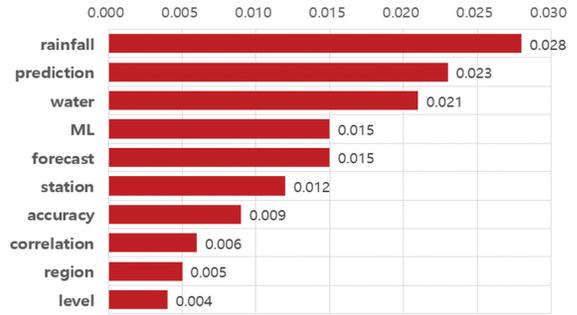


Fig. 6. Key keywords and their occurrence probabilities for Topic 1.

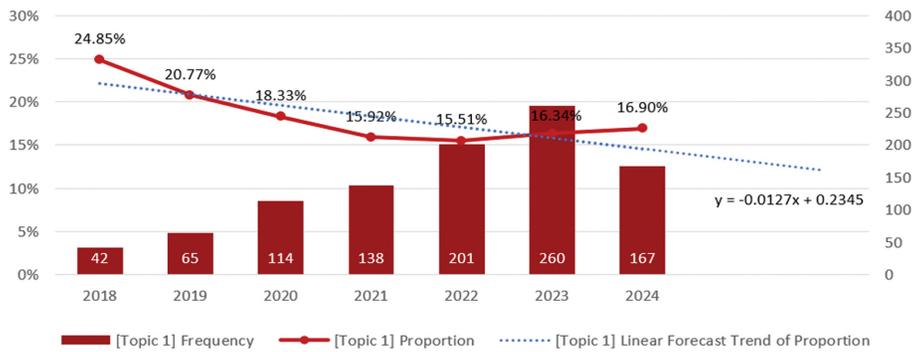


Fig. 7. A graph showing the change in the frequency and proportion of papers corresponding to Topic 1 over the years, along with a linear regression analysis.

Table 2. A table of key papers corresponding to Topic 1.

Topic	Prob.	Article Title	Year
Topic 1	0.976	Monthly and seasonal hydrological drought forecasting using multiple extreme learning machine models	2022
	0.966	Physical and artificial intelligence-based hybrid models for rainfall-runoff-sediment process modelling	2023
	0.963	Short-Term Hydrological Drought Forecasting Based on Different Nature-Inspired Optimization Algorithms Hybridized With Artificial Neural Networks	2020
	0.952	Integration of extreme gradient boosting feature selection approach with machine learning models: application of weather relative humidity prediction	2022
	0.949	Development of machine learning models for estimation of daily evaporation and mean temperature: a case study in New Delhi, India	2024
	0.942	A Comparative Evaluation of the Use of Artificial Neural Networks for Modeling the Rainfall-Runoff Relationship in Water Resources Management	2021
	0.939	Deep learning model for daily rainfall prediction: case study of Jimma, Ethiopia	2022
	0.937	Reservoir Inflow Prediction: A Comparison between Semi Distributed Numerical and Artificial Neural Network Modelling	2023
	0.933	Potential of Hybrid Data-Intelligence Algorithms for Multi-Station Modelling of Rainfall	2019
	0.932	Short term rainfall-runoff modelling using several machine learning methods and a conceptual event-based model	2021

evaporation estimation, artificial neural networks (ANN), hybrid models, short-term forecasting, machine learning 등으로 도출되었다. 인공지능 기법으로는 주로 machine learning, deep learning, ensemble learning, extreme learning machines (ELM), ANN, hybrid models 등이 활용되었으며, 특히 hybrid models와 ANN 기반의 모델들이 수문학적 변수 예측에서 중요한 역할을 하고 있다. 이러한 기법들은 수문학적 변수 간의 복잡한 관계를 모델링하고 비정형적인 데이터를 처리하는 데 매우 유용하게 사용된다.

토픽 1은 전체 연구에서 16.88%의 비중을 차지하며, 이는 수문학적 변수 예측에 관한 인공지능 연구가 여전히 중요한 분야로 남아 있음을 나타낸다. 그러나 Fig. 7의 시계열 분석 결과, 비중 감소 추세를 보이고 있으며, 선형 회귀 분석 결과( $y = -0.0127x + 0.2345$ )로 나타났다. 수문학적 변수 예측 연구의 비율은 2018년부터 2024년 사이에 감소 추세( $y = -0.0127x + 0.2345$ )를 보인다. 그러나 실제 논문 수는 2018년 42편에서 2023년 260편으로 증가하였다. 이는 수문학적 변수 예측 연구가 지속적으로 성장하는 가운데, 기후변화 영향 분석이나 자연재해 위험 평가와 같은 새로운 연구 분야도 함께 급격히 성장하고 있음을 보여준다.

토픽 1의 연구 동향은 인공지능 기법을 통해 수문학적 변수의 예측 정확도를 높이려는 노력이 주된 흐름이다. ANN, hybrid models와 같은 기법들이 주로 활용되었으며, 특히 물리 기반 모델과 인공지능 모델을 결합한 하이브리드 방식이 효과적으로 사용되고 있다. 이는 기존 물리 모델의 한계를 보완하면서 인공지능 기법이 기상학적 문제 해결에 기여할 수 있음을 잘 보여준다. 시사점으로는, 수문학적 변수 예측 연구는 기상학과 인공지능의 융합 연구에서 중요한 역할을 해왔으나, 최근 다른 기상 연구 분야로 연구

의 초점이 이동하고 있음을 확인할 수 있다. 이러한 연구 동향은 기상 예측의 다양한 영역에서 인공지능의 가능성이 계속해서 확장되고 있음을 시사하며, 향후 연구자들이 수문학적 예측 외의 분야로도 관심을 확대할 필요성을 보여준다.

### 3.2 토픽 2(인공지능을 활용한 기후변화 영향 분석) 분석결과

토픽 2는 인공지능을 활용한 기후변화 영향 분석을 주제로 한다. 이 토픽에서는 인공지능 기법을 통해 기후변화가 생태계, 지역사회, 그리고 경제에 미치는 영향을 분석하고 예측하는 연구들이 포함된다. 기후변화로 인해 발생하는 복잡한 기상 패턴과 상호작용을 분석하기 위해 인공지능이 활용되고 있으며, 이를 통해 기후변화에 따른 장기적 영향 평가 및 적응 전략 수립이 이루어지고 있다. Figure 8에서 주요 키워드는 climate change, impact analysis, machine learning, maximum entropy models, adaptation strategies, ecosystem

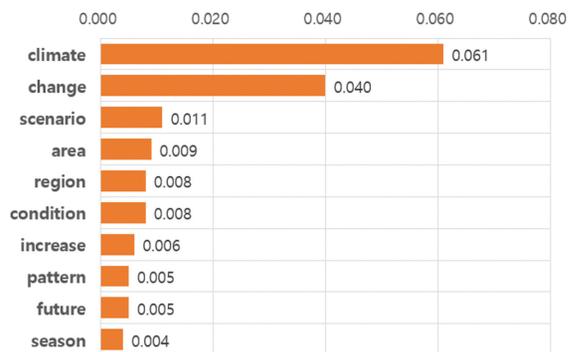


Fig. 8. Key keywords and their occurrence probabilities for Topic 2.

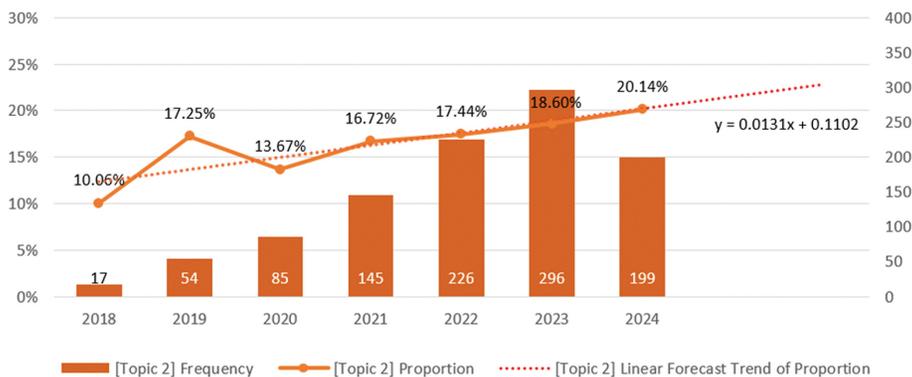


Fig. 9. A graph showing the change in the frequency and proportion of papers corresponding to Topic 2 over the years, along with a linear regression analysis.

**Table 3.** A table of key papers corresponding to Topic 2.

Topic	Prob.	Article Title	Year
Topic 2	0.987	Predicting the Potential Distribution of Haloxylon ammodendron under Climate Change Scenarios Using Machine Learning of a Maximum Entropy Model	2024
	0.976	Modeling of the Potential Geographical Distribution of Three Fritillaria Species Under Climate Change	2022
	0.976	Identifying links between monsoon variability and rice production in India through machine learning	2023
	0.972	Identifying climate thresholds for dominant natural vegetation types at the global scale using machine learning: Average climate versus extremes	2022
	0.972	Climate change, asset redeployability, and sustainability: evidence from earnings conference calls	2024
	0.958	AI-orientation and company climate action: The moderating role of dependency structure and innovation capability	2024
	0.956	Unleashing the power of artificial intelligence for climate action in industrial markets	2024
	0.956	Revisiting the importance of temperature, weather and air pollution variables in heat-mortality relationships with machine learning	2024
	0.953	Deploying artificial intelligence for climate change adaptation	2022
	0.953	Prediction of global temperature anomaly by machine learning based techniques	2023

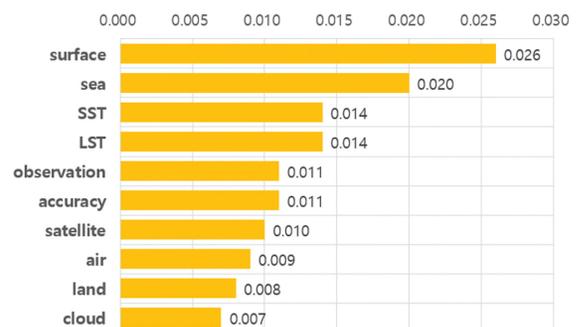
responses, predictive modeling 등으로 도출되었다. 주로 다양한 machine learning 기법(maximum entropy models, random forests, ANN)과 그 하위 기법인 Deep learning과 함께 활용되었다. 이러한 기법들은 대규모 기후 데이터와 다양한 환경 변수를 분석하여 기후변화 시나리오에 따른 변화를 예측하고, 기후 변화로 인한 위험성을 평가하는 데 중요한 역할을 한다. 토픽 2의 비중은 전체 연구에서 17.48%를 차지하며, 이는 기후변화에 대한 관심이 인공지능 연구에서도 중요한 위치를 차지하고 있음을 나타낸다. Figure 9에서 시계열 분석 결과, 토픽 2는 증가 추세를 보이며, 선형 회귀 분석 결과( $y = 0.0131x + 0.1102$ )로 나타났다. 이는 2018년부터 2024년까지 기후변화 영향 분석에서 인공지능 연구가 점점 더 활발히 진행되고 있음을 시사한다.

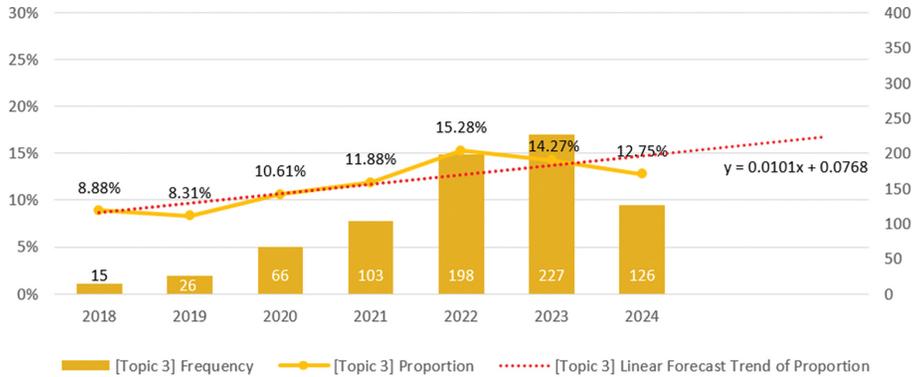
토픽 2의 연구 동향은 인공지능 기법을 통해 기후변화가 특정 지역이나 생태계에 미치는 영향을 정교하게 분석하려는 노력이 두드러진다. 다양한 기후변화 시나리오를 평가하기 위해 machine learning 기법 중 maximum entropy models을 포함한 구체적인 기법들이 적용되었다. 이러한 연구들은 기후변화로 인한 재해 위험을 사전에 평가하고, 기후 적응 및 완화 전략을 수립하는 데 기여하고 있다. 시사점으로는, 기후변화가 글로벌 사회에 미치는 영향을 분석하고 이를 예측하기 위한 인공지능 기법의 도입이 앞으로 더욱 확대될 것으로 보인다. 기후변화에 따른 생태계 및 인

간 사회의 변화를 예측하고 대응 전략을 마련하는 데 있어 인공지능의 역할이 매우 중요해지고 있으며, 이 분야에서의 연구는 향후 지속적인 성장이 예상된다. 특히, 기후변화 관련 정책 수립과 대응 방안 마련에 인공지능 기법이 중요한 도구로 자리매김할 가능성이 높다.

### 3.3 토픽 3(인공지능을 활용한 해양 및 지표면 온도 예측) 분석결과

토픽 3은 인공지능을 활용한 해양 및 지표면 온도 예측을 주제로 한다. 이 주제에서는 인공지능 기법을 활용하여 해양 및 육지 표면의 온도를 예측하고, 이

**Fig. 10.** Key keywords and their occurrence probabilities for Topic 3.



**Fig. 11.** A graph showing the change in the frequency and proportion of papers corresponding to Topic 3 over the years, along with a linear regression analysis.

**Table 4.** A table of key papers corresponding to Topic 3.

Topic	Prob.	Article Title	Year
Topic 3	0.986	Comparison of Three Convolution Neural Network Schemes to Retrieve Temperature and Humidity Profiles from the FY4A GIIRS Observations	2022
	0.986	Generation of Spatial-Seamless AMSR2 Land Surface Temperature in China During 20122020 Using a Deep Neural Network	2023
	0.978	Fusion Retrieval of Sea Surface Barometric Pressure from the Microwave Humidity and Temperature Sounder and Microwave Temperature Sounder-II Onboard the Fengyun-3 Satellite	2022
	0.974	A two-step deep learning framework for mapping gapless all-weather land surface temperature using thermal infrared and passive microwave data	2022
	0.969	Downscaling Aster Land Surface Temperature over Urban Areas with Machine Learning-Based Area-To-Point Regression Kriging	2020
	0.969	Applicability Evaluation of Automated Machine Learning and Deep Neural Networks for Arctic Sea Ice Surface Temperature Estimation	2023
	0.965	A New Approach for Surface Urban Heat Island Monitoring Based on Machine Learning Algorithm and Spatiotemporal Fusion Model	2020
	0.957	Artificial Neural Networks to Retrieve Land and Sea Skin Temperature from IASI	2020
	0.957	Temperature and Relative Humidity Profile Retrieval from Fengyun-3D/VASS in the Arctic Region Using Neural Networks	2023
	0.955	Artificial Neural Network Model for Estimating Ocean Heat Content in the Sea Ice-Covered Arctic Regions Using Satellite Data	2022

를 통해 기후 변동성 및 기상 현상의 영향을 분석하는 연구들이 포함된다. 해양과 지표면 온도는 기상 예측 및 기후 분석에서 중요한 요소로 작용하며, 위성 데이터 및 지상 관측 데이터를 결합하여 더 정확한 예측을 가능하게 한다. Figure 10에서 주요 키워드는 sea surface temperature (SST) prediction, land surface temperature (LST) prediction, remote sensing, deep learning, convolutional neural networks (CNNs), data fusion, spatial-temporal modeling 등이 도출되었다. 사

용된 기법으로는 크게 deep learning과 data fusion technique으로 구분할 수 있다. Deep learning 분야를 분석한 결과, spatial-temporal 데이터 처리를 위한 3D-CNN, 시계열 예측을 위한 RNN (recurrent neural network)과 LSTM (long short-term memory) 등이 주로 사용되었다. Data fusion technique 분야에서는 multi-modal learning, feature-level fusion, decision-level fusion 등의 기법이 활용되었다. 이러한 기법들은 위성 및 센서 데이터를 결합하여 해양과 육지의

표면 온도를 예측하는 데 중요한 역할을 하며, 특히 고해상도 데이터를 바탕으로 보다 정밀한 예측이 가능하게 한다. 토픽 3의 비중은 전체 연구에서 13.02%를 차지하며, 이는 해양 및 지표면 온도 예측이 기상 예측 및 기후 분석에서 중요한 연구 주제임을 나타낸다. 시계열 분석 결과, 토픽 3은 증가 추세를 보이며, Fig. 11에서 선형 회귀 분석 결과( $y = 0.0101x + 0.0768$ )로 나타났다. 이는 2018년부터 2024년까지 해양 및 지표면 온도 예측 연구에서 인공지능의 활용이 꾸준히 증가하고 있음을 시사한다.

토픽 3의 연구 동향은 주로 deep learning과 spatial-temporal models를 활용하여 해양 및 지표면 온도의 시공간적 변화를 예측하는 연구들이 활발히 이루어지고 있다. 특히 CNNs, RNNs와 같은 딥러닝 기법들이 위성 데이터의 해석 및 예측 정확도를 높이는 데 중요한 역할을 하고 있으며, data fusion techniques를 통해 여러 센서 데이터를 결합하여 예측 성능을 향상시키는 연구들이 주목받고 있다. 이러한 연구는 기상 관측 및 기후 변화 분석에서 해양과 지표면 온도의 변동성을 이해하는 데 필수적이다. 시사점으로는, 해양 및 지표면 온도 예측 연구는 기후변화와 해양 기상 현상의 영향을 보다 정밀하게 분석할 수 있는 인공지능 기법의 도입이 필수적임을 보여준다. 향후 기상 예측 및 기후 분석에서 위성 및 센서 데이터를 바탕으로 한 고정밀 예측 모델의 개발이 더욱 중요해질 것으로 예상된다. 이러한 연구는 해양 및 육지의 기후 변동성을 보다 정확하게 예측하는 데 기여할 것이며, 특히 기후변화 대응 및 자연재해 예방에 중요한 역할을 할 것이다.

### 3.4 토픽 4(머신러닝 기반 자연재해 위험 평가 및 예측) 분석결과

토픽 4는 머신러닝 기반 자연재해 위험 평가 및 예

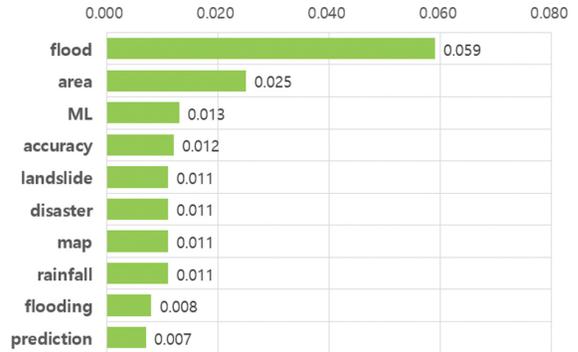


Fig. 12. Key keywords and their occurrence probabilities for Topic 4.

측을 주제로 한다. 이 토픽에서는 머신러닝과 인공지능 기법을 활용하여 홍수, 산사태, 폭풍과 같은 자연재해의 발생 가능성을 평가하고, 이를 통해 재해 예측 및 대응 전략을 수립하는 연구들이 포함된다. 자연재해는 기후 변화로 인해 발생 빈도와 강도가 증가하고 있으며, 이에 따라 정확한 재해 예측 및 위험 평가가 중요한 과제로 떠오르고 있다. Figure 12에서 주요 키워드는 flood risk assessment, landslide prediction, disaster risk management, machine learning, ensemble models, spatial analysis, hazard susceptibility mapping 등이 도출되었다. LDA 알고리즘이 통계적으로 도출한 주요 논문에 사용된 인공지능 기법으로는 주로 machine learning과 deep learning 기법이 사용되었으며, 각각의 기법은 random forest, SVM, XGBoost와 같은 머신러닝 기법과 CNN, RNN, LSTM, Transformer 등의 딥러닝 기법으로 구체화된다. 이러한 기법들은 다양한 기상 데이터 및 지형 데이터를 분석하여 자연재해 위험 지역을 파악하고, 재해 발생 가능성을 예측

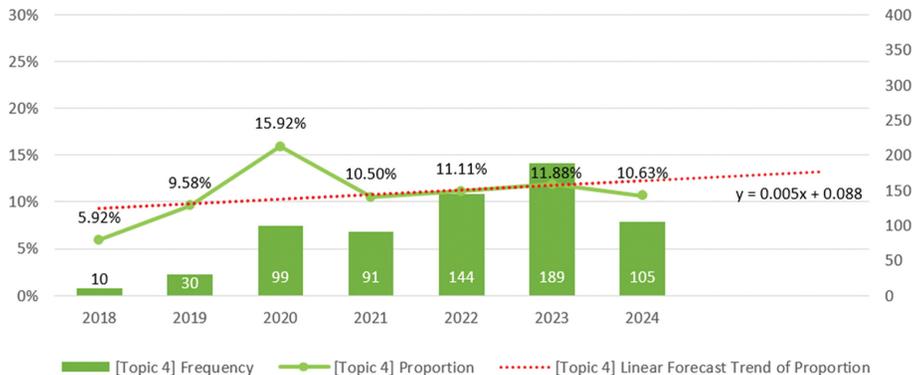


Fig. 13. A graph showing the change in the frequency and proportion of papers corresponding to Topic 4 over the years, along with a linear regression analysis.

**Table 5.** A table of key papers corresponding to Topic 4.

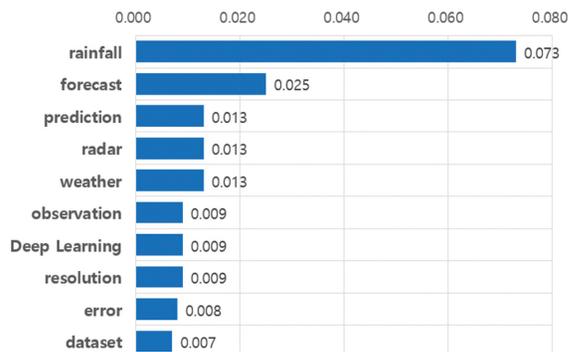
Topic	Prob.	Article Title	Year
Topic 4	0.983	Evaluating urban flood risk using hybrid method of TOPSIS and machine learning	2021
	0.982	GIS Based Hybrid Computational Approaches for Flash Flood Susceptibility Assessment	2020
	0.978	Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran	2019
	0.978	Multi-Hazard Exposure Mapping Using Machine Learning for the State of Salzburg, Austria	2020
	0.972	Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania	2019
	0.971	Flood susceptibility prediction using tree-based machine learning models in the GBA	2023
	0.969	Detection of areas prone to flood-induced landslides risk using certainty factor and its hybridization with FAHP, XGBoost and deep learning neural network	2022
	0.965	Flash Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms	2020
	0.962	Interpretable machine learning for predicting urban flash flood hotspots using intertwined land and built-environment features	2024
	0.956	Assessing Coastal Flood Susceptibility in East Java, Indonesia: Comparison of Statistical Bivariate and Machine Learning Techniques	2022

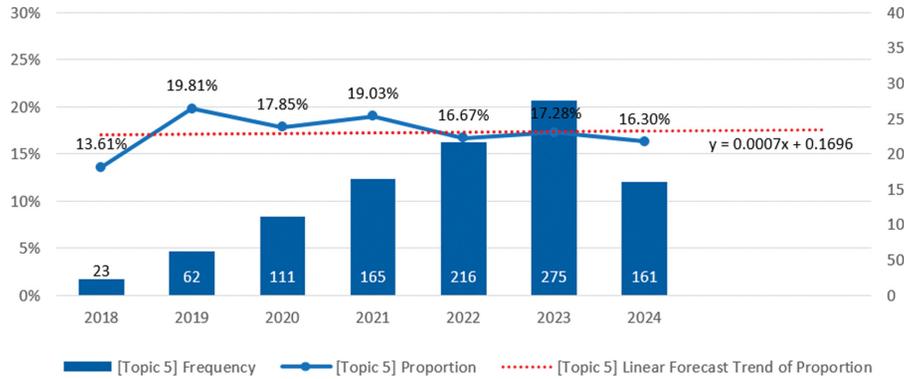
하는 데 효과적으로 사용된다. 토픽 4의 비중은 전체 연구에서 11.43%를 차지하며, 이는 자연재해 위험 평가와 예측이 기상 분야에서 중요한 연구 주제임을 나타낸다. Figure 13에서 시계열 분석 결과, 토픽 4는 증가 추세를 보이고 있으며, 선형 회귀 분석 결과( $y = 0.005x + 0.088$ )로 나타났다. 이는 2018년부터 2024년까지 자연재해 위험 평가 및 예측 연구가 점차 활발해지고 있음을 시사한다.

토픽 4의 연구 동향은 machine learning과 ensemble learning을 사용하여 자연재해 위험을 보다 정확하게 평가하고 예측하는 방향으로 발전하고 있다. 특히 hybrid models와 ensemble models이 많이 사용되고 있으며, 여러 알고리즘을 결합하여 재해 위험 지역을 보다 정밀하게 파악하는 연구들이 진행되고 있다. 또한 spatial analysis와 hazard susceptibility mapping을 통해 자연재해가 발생할 수 있는 지역을 미리 예측하고, 이에 따라 재해 대비 및 관리 시스템을 강화하는 연구가 주목받고 있다. 시사점으로는, 자연재해의 빈도와 강도가 증가함에 따라 인공지능을 활용한 재해 예측 및 위험 평가의 중요성이 더욱 커지고 있음을 확인할 수 있다. 특히 기후 변화로 인해 예측 불가능한 기상 현상이 증가하면서, 이러한 연구는 재해 예방 및 관리의 필수적인 요소로 자리잡고 있다. 향후 재해 관리와 예방을 위한 정책 수립에도 인공지능 기반 위험 평가 기법이 중요한 역할을 할 것이며, 다양한 자연재해 유형에 적용 가능한 통합적인 재해 예측 모델의 개발이 필요할 것으로 보인다.

### 3.5 토픽 5(실시간 강수 예측을 위한 인공지능과 기상 데이터의 활용) 분석결과

토픽 5는 실시간 강수 예측을 위한 인공지능과 기상 데이터의 활용을 주제로 한다. 이 주제에서는 인공지능 기법과 기상 데이터를 결합하여 실시간으로 강수량을 예측하는 연구들이 포함되며, 특히 급변하는 기상 상황에서 강수 예측의 정확도를 높이는 것을 목표로 하고 있다. 실시간 강수 예측은 기상 재난 대응, 홍수 예방, 항공 및 교통 관리 등 다양한 분야에서 매우 중요한 역할을 한다. Figure 14에서 주요 키워드는 precipitation nowcasting, quantitative precipitation estimation (QPE), radar data, deep learning,

**Fig. 14.** Key keywords and their occurrence probabilities for Topic 5.



**Fig. 15.** A graph showing the change in the frequency and proportion of papers corresponding to Topic 5 over the years, along with a linear regression analysis.

**Table 6.** A table of key papers corresponding to Topic 5.

Topic	Prob.	Article Title	Year
Topic 5	0.978	Improving Nowcasting of Convective Development by Incorporating Polarimetric Radar Variables Into a Deep-Learning Model	2021
	0.975	Multi-Task Learning for Simultaneous Retrievals of Passive Microwave Precipitation Estimates and Rain/No-Rain Classification	2023
	0.969	RN-Net: A Deep Learning Approach to 0-2 Hour Rainfall Nowcasting Based on Radar and Automatic Weather Station Data	2021
	0.968	Key factors for quantitative precipitation nowcasting using ground weather radar data based on deep learning	2023
	0.967	Effective training strategies for deep-learning-based precipitation nowcasting and estimation	2022
	0.966	Improving the heavy rainfall forecasting using a weighted deep learning model	2023
	0.961	Deep-Learning-Based Gridded Downscaling of Surface Meteorological Variables in Complex Terrain. Part II: Daily Precipitation	2020
	0.961	Deep Learning for Polarimetric Radar Quantitative Precipitation Estimation during Landfalling Typhoons in South China	2021
	0.960	Artificial Intelligence-Based Techniques for Rainfall Estimation Integrating Multisource Precipitation Datasets	2021
	0.958	spateGAN: Spatio-Temporal Downscaling of Rainfall Fields Using a cGAN Approach	2023

CNNs, polarimetric radar, weather radar integration 등이 도출되었다. LDA 알고리즘이 통계적으로 도출한 주요 논문에 사용된 인공지능 기법으로는 주로 deep learning, CNNs, recurrent neural networks (RNNs), generative adversarial networks (GANs), ensemble learning, random forests (RFs) 등이 활용되고 있다. 이러한 기법들은 기상 레이더 데이터 및 다양한 기상 관측 데이터를 통합하여 실시간으로 강수량을 예측하는 데 효과적으로 사용된다. 토픽 5의 비중은 전체 연구에서 17.33%를 차지하며, 이는 실시간 강수 예측이 기상학 분야에서 중요한 연구 주제 중 하나임을 나타

낸다. Figure 15의 시계열 분석 결과, 토픽 5는 안정적인 추세를 보이고 있으며, 선형 회귀 분석 결과  $y = 0.0007x + 0.1696$ 로 나타났다. 이는 2018년부터 2024년까지 실시간 강수 예측 연구가 지속적으로 높은 비중을 유지하고 있음을 의미한다.

토픽 5의 연구 동향은 deep learning과 CNNs를 활용하여 실시간 강수 예측의 정확도를 높이는 연구들이 주로 진행되고 있다. 특히 weather radar data와 polarimetric radar data를 통합하여 실시간으로 강수량을 추정하는 연구가 활발하며, QPE 모델을 개선하는 방향으로 연구가 집중되고 있다. 또한, GANs와 같은

신경망 기법들이 도입되면서 기상 데이터의 시공간적 변화를 더 정밀하게 예측할 수 있는 모델이 개발되고 있다. 시사점으로는, 실시간 강수 예측에서 인공지능과 기상 데이터의 결합이 기상 재난 대응 및 교통 관리 등 다양한 실생활 영역에서 중요한 역할을 하고 있음을 알 수 있다. 특히 기상 변화가 급격하게 이루어지는 상황에서 실시간으로 강수량을 정확하게 예측하는 능력은 재해 예방과 경제적 손실을 줄이는 데 큰 기여를 할 수 있다. 향후 연구에서는 더욱 정밀한 레이다 데이터와 인공지능 기법의 융합을 통해 실시간 예측 모델의 정확도가 향상될 것으로 기대되며, 이 분야의 연구는 지속적으로 중요한 역할을 할 것으로 보인다.

### 3.6 토픽 6(풍력 예측 및 기상 조건 분석에서의 인공지능 활용) 분석결과

토픽 6은 풍력 예측 및 기상 조건 분석에서의 인공지능 활용을 주제로 한다. 이 주제에서는 인공지능 기법을 통해 풍력 예측 및 기상 조건을 분석하여 풍력 에너지 발전의 효율성을 극대화하는 연구들이 포함된다. 풍력 에너지는 재생 가능 에너지로서, 정확한 풍속 예측이 풍력 발전의 성과를 높이는 데 매우 중요한 요소로 작용하며, 기상 조건의 변화를 정밀하게 분석하는 것은 에너지 관리의 핵심 과제가 되고 있다. Figure 16에서 주요 키워드는 wind speed prediction, wind energy forecasting, deep learning, LSTM, hybrid models, transformer models, spatial-temporal analysis 등이 도출되었다. LDA 알고리즘이 통계적으로 도출한 주요 논문에 사용된 인공지능 기법으로는 주로 deep learning, LSTM, transformer models, hybrid models, ensemble learning 등이 활용되고 있으며, 이러한 기법들은 풍속 및 기상 조건의 시공간적 변화를 예측하고, 이를 통해 풍력 발전의 효율성을 높이는 데 사용된다.

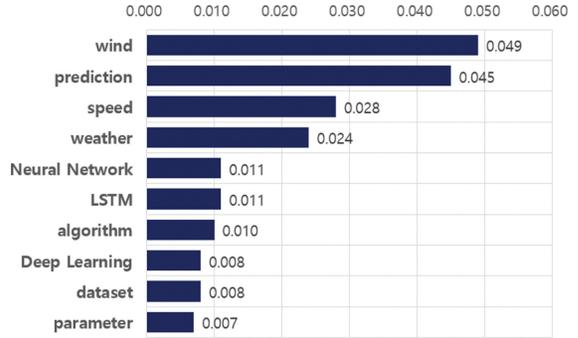


Fig. 16. Key keywords and their occurrence probabilities for Topic 6.

토픽 6의 비중은 전체 연구에서 23.86%로, 가장 높은 비중을 차지하고 있다. 이는 풍력 예측과 기상 조건 분석에서 인공지능의 활용이 매우 활발하게 이루어지고 있음을 나타낸다. 시계열 분석 결과, 토픽 6은 감소 추세를 보이고 있으며, Fig. 17의 선형 회귀 분석 결과( $y = -0.0161x + 0.3209$ )로 나타났다. 토픽 1과 마찬가지로 해당 주제와 관련된 연구의 중요성이 감소한 것이 아니라 다른 주제로의 연구가 더욱 활발히 진행되며 상대적 비율이 감소한 것으로 보이므로 중요성의 감소를 의미하진 않음을 고려해야 한다.

토픽 6의 연구 동향은 deep learning과 LSTM, transformer models을 활용하여 풍속 및 기상 조건을 정밀하게 예측하는 연구들이 주를 이루고 있다. 특히 LSTM과 같은 시계열 예측 모델들이 풍력 발전의 성과를 예측하는 데 중요한 역할을 하고 있으며, hybrid models과 ensemble learning을 결합하여 예측 정확도를 높이는 연구들이 활발히 진행되고 있다. 또한, transformer models와 같은 최신 기법을 도입하여 복잡한 시공간 데이터를 처리하는 연구들이 주목받고

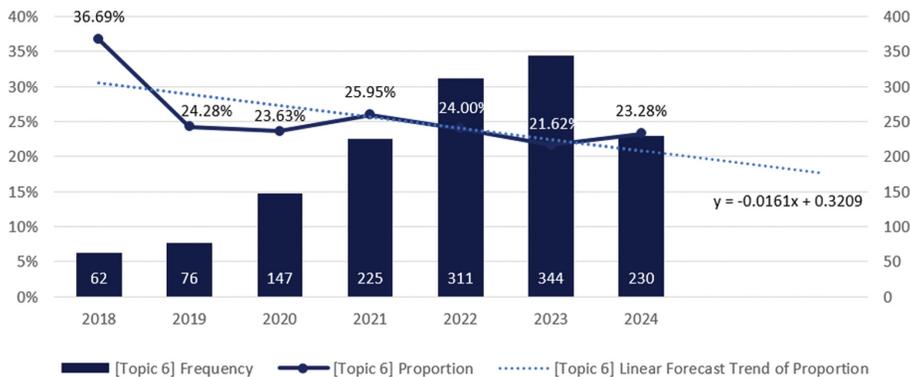


Fig. 17. A graph showing the change in the frequency and proportion of papers corresponding to Topic 6 over the years, along with a linear regression analysis.

**Table 7.** A table of key papers corresponding to Topic 6.

Topic	Prob.	Article Title	Year
Topic 6	0.991	Hybrid wind speed forecasting using ICEEMDAN and transformer model with novel loss function	2023
	0.990	An efficient wind speed prediction method based on a deep neural network without future information leakage	2023
	0.989	A new hybrid ensemble deep reinforcement learning model for wind speed short term forecasting	2020
	0.989	Ultra-short-term wind-speed bi-forecasting system via artificial intelligence and a double-forecasting scheme	2021
	0.989	Promoting wind energy for sustainable development by precise wind speed prediction based on graph neural networks	2022
	0.989	Short-term wind speed forecasting based on spatial-temporal graph transformer networks	2022
	0.988	A novel combined model for wind speed prediction - Combination of linear model, shallow neural networks, and deep learning approaches	2021
	0.987	A Hybrid Model Based on A Modified Optimization Algorithm and An Artificial Intelligence Algorithm for Short-Term Wind Speed Multi-Step Ahead Forecasting	2018
	0.987	Hybrid wind speed forecasting model based on multivariate data secondary decomposition approach and deep learning algorithm with attention mechanism	2021
	0.987	A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers	2021

있다. 시사점으로는, 풍력 예측 연구에서 인공지능의 역할이 매우 중요한 위치를 차지하고 있음을 확인할 수 있으며, 이를 통해 풍력 발전의 효율성을 극대화하고자 하는 연구가 활발히 진행되고 있음을 알 수 있다. 다만, 최근 연구에서 풍력 예측에 대한 관심이 상대적으로 감소하고 있다는 점은 향후 연구 방향에 있어 에너지 관리 및 풍력 발전 분야에서 새로운 기술적 도전이 필요함을 시사한다. 기상 조건의 정확한 예측이 에너지 산업에서 중요한 역할을 하기 때문에, 지속적인 연구가 요구되며, 풍력 에너지와 기상학의 융합 연구는 앞으로도 중요한 연구 주제로 남을 것이다.

## 4. 결론 및 시사점

### 4.1 요약

본 연구는 2018년부터 2024년까지 글로벌 기상 분야에서 인공지능 기술이 어떻게 활용되고 있는지에 대한 연구 동향을 분석하였다. LDA 토픽 모델링 기법을 적용하여 5,846건의 논문 초록을 분석한 결과, 주요 연구 주제는 인공지능 기반 수문학적 변수 예측, 기후변화 영향 분석, 해양 및 지표면 온도 예측, 자연재해 위험 평가 및 예측, 실시간 강수 예측, 풍력 예측 및 기상 조건 분석으로 도출되었다. 이러한 연구 주제들은 기상학에서 인공지능이 중요한 도구로 자리

잡고 있음을 나타내며, 각 주제별로 인공지능 기법이 적용되어 기상 현상 예측과 분석의 정확성을 크게 향상시킨 것을 확인할 수 있었다. 시간에 따른 비중 변화 분석 결과, 기후변화 영향 분석, 해양 및 지표면 온도 예측, 자연재해 위험 평가 등의 주제는 증가 추세를 보였으며, 이는 기후 변화 대응이 기상학에서 더욱 중요한 연구 분야로 자리 잡고 있음을 시사한다. 반면, 수문학적 변수 예측과 풍력 예측 관련 연구는 감소 추세를 보였으나, 여전히 중요한 연구 분야로 남아 있다.

### 4.2 시사점

기상학에서 인공지능 기술이 어떤 방식으로 활용되고 있으며, 각 분야에서 어떠한 발전이 이루어졌는지를 체계적으로 분석하였다. 연구 결과는 몇 가지 중요한 시사점을 제공한다. 첫째, 기후변화와 재해 관리와 같은 글로벌 환경 문제에 대한 대응이 인공지능 기술을 통해 더 정밀하고 신속하게 이루어지고 있으며, 이 분야에서의 연구가 지속적으로 증가하고 있음을 보여준다. 둘째, CNN, LSTM, Transformer 등의 딥러닝 기법과 random forest, support vector machine, XGBoost 등의 머신러닝 기법, 그리고 이들을 결합한 하이브리드 모델 등 다양한 인공지능 알고리즘이 기존의 물리 기반 예측 모델을 보완하거나 대체하며, 기

상 예측의 정확도와 효율성을 크게 향상시키고 있다. 특히 시계열 데이터 처리에는 LSTM과 Transformer가, 이미지 기반 기상 데이터 분석에는 CNN이, 복잡한 기상 변수 간 관계 분석에는 random forest와 XGBoost가 주로 활용되고 있다. 셋째, 실시간 기상 예측과 같은 분야에서는 빠르고 정확한 예측이 중요한 만큼, 인공지능 기반 기법의 도입이 기상학의 실질적 적용에 기여하고 있음을 확인할 수 있다. 이러한 시사점들은 기상학 연구자 뿐만 아니라, 정책 결정자와 산업 관계자들에게도 유용한 정보를 제공하며 LDA 토픽 모델링 기법을 통해 대규모 기상학 관련 논문 데이터를 분석함으로써, 인공지능이 기상학에서 어떻게 활용되고 있는지를 구체적으로 제시하여 향후 기상학 연구에서 인공지능 기술의 적용 가능성을 더욱 확대하는데 중요한 기초 자료로 활용될 수 있다. 이를 통해 기상학 내에서 주요 연구 주제를 명확하게 구분하여 기상학과 인공지능 융합 연구의 향후 방향성을 제시한다는 연구의의를 지닌다.

본 연구는 인공지능과 기상학 연구의 융합이 지속적으로 발전할 가능성을 제시하였으며, 인공지능이 기상 예측, 자연재해 대응, 에너지 관리 등의 영역에서 실제 기상 예측 및 대응 시스템에 적용하는 연구의 추세와 앞으로 어떤 연구가 필요한지, 정책적으로 어떤 연구에 지원이 필요한지에 대한 중요한 자료가 될 것이다.

## 감사의 글

이 연구는 기상청 국립기상과학원 「AI 기상예측기술개발」(KMA2021-00121)의 지원으로 수행되었습니다.

## REFERENCES

- Blei, D. M., A. Y. Ng, and M. I. Jordan, 2003: Latent Dirichlet Allocation. *J. Machine Learning Res.*, **3**, 993-1022, doi:10.5555/944919.944937.
- Clarke, B., F. Otto, R. Stuart-Smith, and L. Harrington, 2022: Extreme weather impacts of climate change: an attribution perspective. *Environ. Res. Climate*, **1**, 012001, doi:10.1088/2752-5295/ac6e7d.
- Fathi, M., M. Haghi Kashani, S. M. Jameii, and E. Mahdipour, 2022: Big data analytics in weather forecasting: A systematic review. *Arch. Computat. Methods Eng.*, **29**, 1247-1275, doi:10.1007/s11831-021-09616-4.
- Hewage, P., M. Trovati, E. Pereira, and A. Behera, 2021: Deep learning-based effective fine-grained weather forecasting model. *Pattern Anal. Applic.*, **24**, 343-366, doi:10.1007/s10044-020-00898-1.
- Hinton, G. E., and R. R. Salakhutdinov, 2006: Reducing the dimensionality of data with neural networks. *Science*, **313**, 504-507, doi:10.1126/science.1127647.
- IPCC, 2021: *Climate Change 2021: The Physical Science Basis*. In Masson-Delmotte, V., et al. Eds., Cambridge University Press, 3949 pp.
- \_\_\_\_\_, 2022: *Summary for policymakers*. In Pörtner, H.-O., et al. Eds., *Climate Change 2022: Impacts, Adaptation, and Vulnerability*, Cambridge University Press, 3-33.
- Jones, K. S., 1972: A statistical interpretation of term specificity and its application in retrieval. *J. Documentation*, **28**, 11-21, doi:10.1108/eb026526.
- Jozefowicz, R., W. Zaremba, and I. Sutskever, 2015: An empirical exploration of recurrent network architectures. *Proceedings of the International Conference on Machine Learning*, 2342-2350.
- Kim, T. S., and A. Reiter, 2017: Interpretable 3D human action analysis with temporal convolutional networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 10-14 July 2017*, 1623-1631.
- Lea, C., M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, 2017: Temporal convolutional networks for action segmentation and detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, July*, 1003-1012, doi:10.1109/cvpr.2017.113.
- Majumdar, S. J., and Coauthors, 2021: Multiscale forecasting of high-impact weather: Current status and future challenges. *Bull. Amer. Meteor. Soc.*, **102**, E635-E659, doi:10.1175/BAMS-D-20-0111.1.
- Röder, M., A. Both, and A. Hinneburg, 2015: Exploring the space of topic coherence measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*, 399-408, doi:10.1145/2684822.2685324.
- Rosenzweig, C., and Coauthors, 2007: *Assessment of observed changes and responses in natural and managed systems*. Cambridge University Press, Cambridge, UK, 79-131, doi:10.5167/uzh-33180.