리뷰와 전망
eISSN 2586-0860
https://doi.org/10.56786/PHWR.2025.18.34.1

PHWR
Public Health Weekly Report

# 국내 코로나바이러스감염증-19 예측 연구 현황 분석

김현경[1], 류보영[2], 유민규[2], 김재훈[2], 민경덕[3]*

[1]서울대학교 보건대학원 보건학과, [2]질병관리청 질병데이터과학분석관 질병관리연구기획담당관, [3]충북대학교 수의과대학

## 초 록

**목적:** 코로나바이러스감염증-19(코로나19) 발생 초기에, 바이러스 특성에 대한 불확실성이 높은 상황에서 여러 예측연구가 수행되었으며, 해당 결과를 정책 수립에 활용한 바 있다. 본 연구는 국내 코로나19 유행 양상을 예측한 연구 논문에 대한 체계적인 문헌 정보 분석을 수행한 연구이다.

**방법:** 2020년부터 2024년 10월 15일까지 발표된 연구 논문을 138편을 대상으로, 활용한 데이터, 방법론 등을 분석하여 감염병 유행 관리를 위한 방향성과 예측 결과의 활용도를 향상시킬 수 있는 방법을 고찰하였다.

**결과:** 코로나19 유행 양상을 예측하기 위해 수리모형, 통계모형, 머신러닝 및 기계학습 기반 모형 등 여러 분야의 방법론이 활용되었으며, 향후 발생 양상 예측뿐만이 아닌 시행된 정책을 평가하고 시나리오 가정을 통해 효과적인 정책을 제언하는데도 예측모형이 활용되었음을 확인할 수 있었다.

**결론:** 본 연구의 결과를 토대로 향후 감염병 대응·대비를 위해 다학제적 협력을 통한 예측모형 개발을 반영함으로써 국민의 건강 피해를 최소화하는데 기여할 것을 기대한다.

**주요 검색어:** 코로나바이러스감염증-19; 단기 예측; 장기 예측; 모델링

## 서 론

2019년 12월 중국 후베이성 우한시에서 처음 보고된 코로나바이러스감염증-19(코로나19)이 전세계적으로 빠르게 확산함에 따라 2020년 3월 11일 세계보건기구(World Health Organization)는 세계적 감염병 대유행인 팬데믹(pandemic)을 선언하였다. 초기 유행 단계에서 코로나19 바이러스(중증급성호흡기증후군 코로나바이러스 2)의 감염경로, 전파력 등 역학적 특성에 불확실성이 높았다. 이에, 코로나19 전파 양상 및 규모를 예측하고 대응을 위해 개입된 여러 중재의 효과를 평가하기 위해 예측모형을 사용한 연구가 수행되었다. 결과적으로, 이러한 예측모형 연구는 감염병 확산을 효과적으로 통제하고 대응하기 위한 전략을 설계하는 데 중요한 역할을 하였다.

예측은 크게 단기 예측(forecasting)과 장기 예측(projection)으로 나눌 수 있다. 장기 예측은 주로 특정 모델이나 가정

---

**Received** February 17, 2025 **Revised** July 7, 2025 **Accepted** July 7, 2025

***Corresponding author:** 민경덕, Tel: +82-43-261-8393, E-mail: kdmin@chungbuk.ac.kr

KDCA
Korea Disease Control and Prevention Agency

**핵심요약**

**① 이전에 알려진 내용은?**

감염병 예측모형을 활용하여 예측된 유행 패턴을 통해 정책 효과를 파악하고 정책적 근거 마련이 가능하다.

**② 새로이 알게 된 내용은?**

팬데믹 기간 동안 다수의 예측연구가 진행되었으며, 수학, 공학, 사회학 등 여러 분야의 방법론이 활용되었다. 국내 코로나바이러스감염증-19 발생 데이터를 활용하여, 해외에서도 국내 유행 양상을 예측하기 위한 예측모형을 다수 개발하였다.

**③ 시사점은?**

감염병 전파 차단을 위한 예측모형 정교화를 위해 다학제적 협력은 필수이며, 향후 신종감염병 대응·대비 계획에 이를 활용할 수 있다.

을 기반으로 하여 백신 접종률, 행동변화 등 특정 시나리오에서의 장기적인 추이를 보여주는 것이며, 특정 가정이 주어졌을 때 어떤 일이 일어날지 설명하려는 것이다. 반면, 단기 예측은 향후 몇 주 등 짧은 시간 범위 내에서 구체적이고 세밀한 예측치를 제공하여 앞으로의 감염병 발생 추이를 예측하려는 시도를 의미한다[1].

감염병 발생의 단기 예측은 유행 상황 인식 및 즉각적인 대응 마련에 유용하며, 시나리오 기반의 장기 예측은 장기적인 의사 결정 및 다양한 상황에서의 발생 궤적 비교가 가능하다는데 그 의미가 있다. 감염병 예측은 방역 시스템 구축에 과학적 근거를 제공하며, 발생 예측 시뮬레이션을 통해 효과적인 방역정책 수립과 선제적 위기대응 체계 마련에 기여할 수 있다. 1927년 영국의 수학자 커맥과 맥캔드릭이 제안한 S (Susceptible)-I (Infected)-R (Recovered) 구획모형이 개발된 후[2], 감염병 예측 연구가 널리 진행되었으며 최근에는 그 방법이 확장되어 머신러닝 및 딥러닝, 네트워크 분석 등 여러 분야의 방법론이 다양하게 활용되고 있다. 특히, 코로나19 유행 상황에서 미국 질병통제예방센터(Centers for Disease Control and Prevention, CDC), 유럽 질병통제예방센터(European Centre for Disease Prevention and Control) 등 해외 주요 국가 또한 예측모형을 감염병 확산 차단 및 대응 전략 수립에 활용하였다. 특히 미국 CDC의 경우 코로나19 기간 동안 100개 이상의 학계 및 연구 기관과 협력하여 확진자 수, 사망자 수, 환자 수 등의 단기 예측 결과를 웹사이트에 공개하고, 해당 결과를 정책 의사 결정에 활용한 바 있다.

향후 감염병 유행에 대한 선제적 대비·대응을 위해, 감염병 유행 규모 예측을 정교화하는 것은 필수이며[3], 본 연구에서는 이를 위해 필요한 연구 근거를 확보하고, 연구 현황, 수요 및 우선순위를 파악하고자 국내 코로나19 예측 연구 논문에 대한 문헌고찰을 수행하였다. 이를 바탕으로 감염병 유행 관리를 위한 방향성을 고찰하고, 예측 결과의 활용도를 향상시킬 수 있는 방법을 제언하고자 한다.

## 방법

국내 코로나19 예측 연구 현황을 파악하기 위해 보건의료 분야에서 가장 많이 이용되고 있는 국외 학술 논문 데이터베이스인 PubMed를 활용하여 관련된 연구 문헌을 수집하였다.

**표 1.** 국내 코로나바이러스감염증-19 예측 관련 출판 논문 검색식

| 구분 | 검색 단어 |
|---|---|
| 한국 | Korea[Title/Abstract] OR "South Korea"[Title/Abstract] OR "Republic of Korea"[Title/Abstract] |
| 코로나바이러스감염증-19 | covid[Title/Abstract] OR corona[Title/Abstract] OR sars-cov-2[Title/Abstract] OR COVID-19[Title/Abstract] |
| 예측 | predict*[Title/Abstract] OR forecast*[Title/Abstract] OR projecti*[Title/Abstract] OR model*[Title/Abstract] |

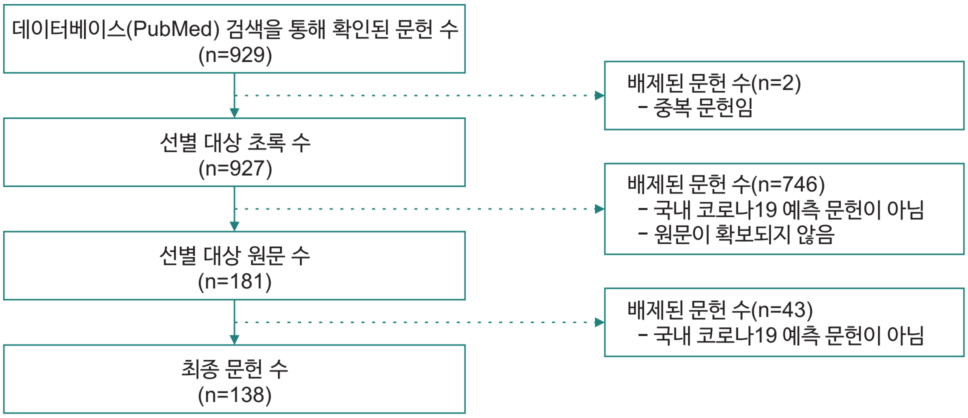

**그림 1.** Preferred Reporting Items for Systematic reviews and Meta-Analyses 문헌 선정 흐름도
코로나19=코로나바이러스감염증-19.

**표 2.** 연도별 연구 논문 발표 현황

| 연도 | 문헌 수(건) |
|---|---|
| 2020년 | 49 |
| 2021년 | 42 |
| 2022년 | 23 |
| 2023년 | 18 |
| 2024년 | 6 |
| 전체 | 138 |

자료검색은 2024년 10월 15일까지 이루어졌으며 검색 시작 기간은 제한 두지 않았다. 문헌을 조사하기 위해 활용한 검색 키워드는 총 3개의 카테고리(한국, 코로나19, 예측)로 구분하여 검색을 수행하였다. 검색 키워드 및 검색식은 연구진 검토를 통해 확정하였다(표 1).

본 연구는 문헌 선택과정의 체계적 기술을 위해 Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 그룹의 흐름도를 활용하여 단계별 문헌 선택과정을 제시하였다. 문헌 선정 기준은 국내 데이터를 사용하여, 국내 코로나19 역학 지표를 예측한 연구이면서 영문으로 출판된 논문으로 그 외 논문은 제외하였다. 검색된 문헌 929개 중 연구 선정 및 제외 기준에 따라 제목과 초록을 검토하여 연구 주제에 부합한 연구 181건을 선정하였으며, 이후 전문을 검토하여 최종적으로 총 138건의 연구 논문을 선정하였다(그림 1, 보충 표 1; available online).

**표 3.** 모형별 연구 논문 발표 현황

| 구분 | 문헌 수(건) |
|---|---|
| 수리모형 | 97 |
| 통계모형 | 20 |
| 머신러닝 및 기계학습 기반 모형 | 11 |
| 혼합모형 | 10 |

## 결 과

국내 코로나19 예측 관련 문헌 검토 결과 최근 5년간(2020년–2024년 10월 15일) 수행된 연구 논문은 총 138건으로 나타났다. 코로나19 발생 직후인 2020년과 2021년이 각각 49건, 42건으로 가장 많았으며 최근 감소하여 2023년에는 18건의 문헌이 발표되었다(표 2). 연도별 코로나19 예측 연구 주제 및 방법론의 동향을 파악하기 위해 선정된 문헌에 대해 방법론을 확인한 결과, 구획모형(compartment model)과 행위자기반모형(agent-based model) 등의 수리모형(70.3%, 97건)이 가장 많이 활용되었다. 그 외에도 통계모형, 머신러닝 및 기계학습 기반의 예측모형, 구획모형과 머신러닝 기반의 모형을 함께 적용하는 혼합모형(hybrid model)도 사용된 것으로 확인되었다(표 3, 그림 2). 국내 코로나19 예측 연구에 활용된 모형별 특징을 보다 구체적으로 살펴본 결과, 연구자들이 각 모형을 선택하는 데 있어 연구의 목적, 데이터의 가용성, 예측 정확도 및 정책 적용 가능성 등을 종합적으로 고려한 것으로 나타났으며 각 모형별 구체적 내용, 장단점 및 주요 특

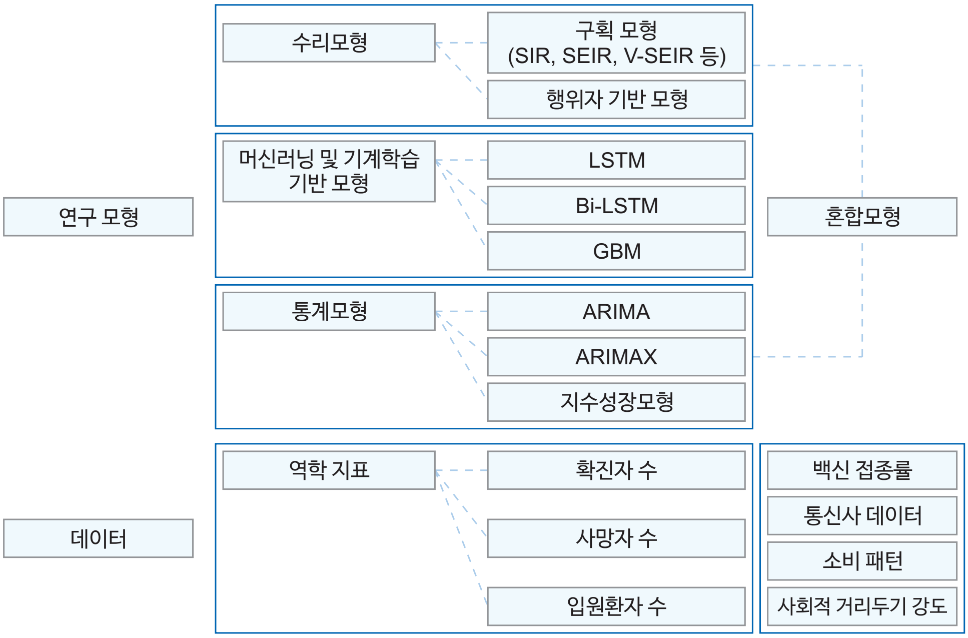

**그림 2.** 연구 모형 및 데이터

SIR=susceptible-infectious-recovered; SEIR=susceptible-exposed-infectious-recovered; V-SEIR=vaccinated-SEIR; LSTM=Long Short-Term Memory; GBM=Gradient Boosting Machine; ARIMA=Autoregressive Integrated Moving Average Model; ARIMAX=Autoregressive Integrated Moving Average with Exogenous Variables.

징을 표 4에 정리하였다.

수리모형은 일반적으로 사용하는 전체 인구를 감염 가능성이 있는 사람(Susceptible), 감염된 사람(Infected), 질병에서 회복 또는 사망하여 감염 위험에서 벗어난 집단(Recovered)으로 나누어 모형을 설계하는 S-I-R 모형 외에, 입원 환자 집단(Hospitalized), 백신 접종 집단(Vaccinated), 격리 집단(Quarantined) 등을 반영한 SEIHR, V-SEIR, SEIHQ, SEIQRDV3P 등 여러 모형으로 발전되어 사용되었다[4-7]. 머신러닝 및 기계학습 기반 모형을 활용한 연구에서는 주로 Long Short-Term Memory (LSTM), Gradient Boosting Machine (GBM) 모형 등을 활용하여 유행 양상을 예측하였다[8-11]. 통계모형으로는 Autoregressive Integrated Moving Average Model (ARIMA), Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) 등의 시계열 모형이 가장 많이 사용되었으며, 로지스틱 성장 모형(logistic growth model), 회귀 모형(regression model) 등도 활용된 것으로 보인다[12-15]. 특히, 최근 감염병 예측을 위해 여러 분야의 방법론을 접목한 혼합모형이 사용되고 있는데, 딥러닝 기반의 수리모형인 m-SIQRD, SIRVD-DL 등을 활용하여 코로나19 유행 양상을 예측하였다[10,16,17]. 코로나19 발생 초기에는 질병관리청에서 공개하는 확진자 수 데이터를 기반으로, 향후 확진자 수를 예측하는 단기 예측을 중심으로 연구가 진행되었으나, 예측 변수가 사망자 수, 위중증 환자 수, 감염재생산지수 등으로 확장되었다. 또한, 한국으로 제한하여 국내의 코로나19 유행 양상을 예측한 연구도 다수 진행되었으나, 이탈리아, 홍콩, 인도 등 타 국가와의 비교 연구 또한 상당수 진행되었다[18-20]. 역학 지표를 활용하여 유행 양상을 예측한 연구 외에도, 백신 접종 및 사회적 거리두기, 등교 중지, 마스크 착용, 개인의 행동 변화 등 중재의 효과를 평가하거나 인구의 이동량 등을 시나리오로 가정하여 유행 패턴의 변화를 확인한 연구도 다수 진행되었다[21-23]. 코로나19 예측을 위해 통신사, 소비패턴 등 여러 종류의 인구 이동 관련 데이터도 활용되었으며 사회적 거리두기 강도 반영을 위해 stringency index, Oxford COVID-19 Government Response

**표 4.** 예측모형의 주요 특징

| 모형 | | 구체적 내용 | 장점 | 단점 |
|---|---|---|---|---|
| 수리모형 | 구획모형 | • SIR, SEIR, SIRD 등 미분방정식 기반<br>• 인구를 S, E, I, R, D로 구분<br>• 구획 간 전이율을 매개변수로 모델링 | • 생물학적 메커니즘 직관적 반영<br>• 매개변수 역학적 의미 명확<br>• 방역 정책 시나리오 분석 가능<br>• 적은 파라미터로 예측 가능 | • 인구 동질성 가정으로 현실성 부족<br>• 단순화된 파라미터 추정으로 인한 불확실성<br>• 공간적 이질성 반영 어려움<br>• 초기 설정 민감도 높음 |
| | 행위자 기반 모형 | • 개별 개체의 행동과 상호작용 모델링<br>• 행위자 속성과 행동 규칙 정의<br>• 미시적 상호작용으로 거시적 패턴 도출 | • 개인 수준 이질성 반영<br>• 공간적 구조와 네트워크 효과<br>• 방역 정책 미시적 효과 분석<br>• 현실적 인구 구조 반영 | • 높은 계산 복잡도<br>• 다수 매개변수로 검증 어려움<br>• 행동 규칙 설정의 주관성<br>• 대규모 시뮬레이션 확장성 문제 |
| 머신러닝 및 기계학습 기반 모형 | Long Short-Term Memory (LSTM) | • 시계열/순차 데이터 분석에 특화된 딥러닝 모형<br>• 순환 신경망의 장기 의존성 학습<br>• 게이트 메커니즘으로 정보 선택적 기억<br>• 과거 감염 데이터 패턴 학습 | • 높은 예측 정확도<br>• 복잡한 비선형 패턴 학습<br>• 장기 의존성 학습 가능<br>• 다변량 시계열 데이터 처리<br>• 통계적 가정 불필요 | • 블랙박스 모델로 해석성 부족<br>• 충분한 학습 데이터 필요<br>• 하이퍼파라미터 튜닝 복잡<br>• 분포 외 상황 일반화 한계 |
| | Gradient Boosting Machine (GBM) | • 약한 학습기(예. 결정트리) 순차적 결합<br>• 오차를 단계적으로 줄이는 방향으로 최종 예측모형 결정<br>• 다양한 특성 변수 활용 예측 | • 높은 예측 정확도<br>• 다양한 데이터 타입 처리에 용이<br>• 변수 중요도 제공 | • 하이퍼파라미터 튜닝 복잡<br>• 순차적 학습으로 병렬화 제약<br>• 시간 의존성 반영 한계 |
| 통계모형 | ARIMA, ARIMAX | • 자기회귀, 차분, 이동평균 성분 모델링<br>• 비정상 시계열 정상화 예측<br>• Box-Jenkins 방법론 활용 | • 통계적 이론 기반 체계적 접근<br>• 매개변수 통계적 해석 가능<br>• 신뢰구간으로 불확실성 정량화<br>• 적은 데이터로도 모델링 가능 | • 선형 관계만 모델링<br>• 구조적 변화 적응성 부족<br>• 모델 차수 선택의 주관성<br>• 장기 예측 시 정확도 저하 |
| | 지수성장모형 | • 감염자 수 지수적 증가 패턴 모델링·성장률 매개변수로 확산 속도 추정<br>• 초기 확산 단계 단순화 모델링 | • 모델 구조 단순성과 해석 용이<br>• 초기 확산 단계 빠른 예측<br>• 적은 매개변수로 효율적 모델링<br>• 성장률의 직관적 이해 | • 복잡한 확산 패턴 반영 한계<br>• 장기 예측 시 과대 추정<br>• 방역 정책 효과 반영 어려움<br>• 포화 효과 미반영 |
| 혼합모형 | | • 서로 다른 모델링 접근법 장점 결합<br>• 수리모형과 머신러닝 모형 결합, 통계모형과 기계학습 모델 앙상블 등<br>• 다중 스케일 모델링 접근 | • 단일 모델의 한계 극복<br>• 예측 정확도와 해석가능성 확보<br>• 다양한 데이터 소스 통합 활용<br>• 안정적인 예측 성능 | • 모델 복잡도 증가로 구현 어려움<br>• 계산 비용 증가<br>• 모델 간 가중치 결정의 복잡성<br>• 검증 및 해석의 어려움 |

SIR=susceptible-infectious-recovered; SEIR=susceptible-exposed-infectious-recovered; SIRD=susceptible-infected-recovered-deceased; ARIMA=Autoregressive Integrated Moving Average Model; ARIMAX=Autoregressive Integrated Moving Average with Exogenous Variables.

Tracker index 등 여러 지표도 함께 반영되었다[24].

## 결론

본 연구는 국내 코로나19 유행 예측 연구의 동향 파악을 통해, 향후 감염병 유행 관리를 보다 효율적으로 수행하기 위한 기반을 마련하고자 하였다. 그 결과 코로나19 유행 양상에 대한 예측 연구는 여러 종류의 데이터와 방법론을 활용한 것으로 보이며, 기존에 잘 알려진 수리모형 외에도 머신러닝, 혼합모형 등 여러 분야의 방법론이 접목되어 활용되었다. 또한, 향후 사망자, 확진자, 입원 환자 등의 추이를 통해 유행을 예측하는 것을 넘어 시행된 정책을 평가하여 향후 의사결정에 활용될 수 있도록 하는 연구도 다수 진행되었다.

본 연구에서 분석한 예측 연구들의 후향적 검토 결과, 전반적으로 중장기 예측(1개월 이상)의 경우 변이 바이러스 출현, 방역 정책 변화, 사회적 행동 변화 등 예측 당시 고려되지 않은 요인들로 인해 실제 유행 양상과 차이를 보이는 경우가 많았다. 특히 바이러스의 변이나 특성 변화가 발생할 때마다 기존 예측모형들의 정확도가 현저히 낮아지는 것을 확인할 수 있었다. 이는 감염병 예측모형의 한계를 보여주는 동시에, 지속적인 모형 보정과 업데이트의 필요성을 시사한다.

본 연구를 통해, 예측모형이 과학적 의사결정에 미치는 영향과 다학제 협력을 통한 예측모형의 확장 가능성에 대해 확인하였으며 감염병 예측 연구의 지속적 발전을 위해 전문가 및 관련 부처 간의 협력은 필수이다. 특히 감염병 예측에 필요한 데이터와 자원을 다양한 분야 간 협력을 통해 확장함으로써 향후 감염병 유행의 불확실성을 감소시키고, 보다 효과적인 대응 체계를 구축할 수 있을 것으로 기대된다. 또한, 본 연구에서 확인된 예측 정확도의 한계를 고려할 때, 단일 모형에 의존하기보다는 다양한 방법론을 활용한 앙상블 예측과 실시간 모형 보정 시스템의 구축이 필요하다. 예측 결과 활용 시에는 예측 불확실성을 명시적으로 고려하고, 다양한 시나리오를 통한 의사결정 지원이 이루어져야 한다. 추가적으로, 연구 결과의 실효성 있는 활용을 위한 제도적 장치 마련이 요구된다. 예측 결과가 실질적으로 감염병 대응 정책에 반영될 수 있도록, 예측모형을 활용해 도출할 수 있는 시사점과 실제 방역 정책의 연계 방안을 마련해야 한다. 이를 통해 보건 당국은, 과학적 근거를 기반으로 한 선제적인 대응을 마련하고, 국민들의 안전을 확보하는 데 기여할 수 있다.

질병관리청은 전국·대규모 감염병 대유행에 대응 가능한 방역·의료 체계를 만들기 위해 국내·외 모든 공중보건위협의 포괄적 위험분석 체계를 구축 중에 있으며, 위기단계 및 상황별 위험평가를 세분화하는 표준운영절차를 마련하였고, 인공지능 모델링 등을 통해 감염병 유행규모 예측을 정교화하는 등 신종감염병 대유행 대비 중장기 계획을 수립하여 단계적 연구 추진을 위한 세부 계획을 수립한 바 있다[3]. 이러한 결과를 바탕으로, 향후 감염병 대응·대비를 위해 수학, 통계학, 인공지능, 보건학, 의학 등 여러 분야의 다학제적 협력을 통해 정교한 예측모형을 지속적으로 개발하여 감염병으로부터 국민들의 건강 피해를 최소화하는데 기여할 것을 기대한다.

## Declarations

**Ethics Statement:** Not applicable.

**Funding Source:** None.

**Acknowledgments:** None.

**Conflict of Interest:** The authors have no conflicts of interest to declare.

**Author Contributions:** Conceptualization: HKK, KDM. Data curation: HKK. Supervision: KDM. Validation: BYR, MGY, JHK. Visualization: HKK. Writing-original draft: HKK. Writing-review & editing: HKK, KDM.

## Supplementary Materials

Supplementary data are available online.

## References

1. Massad E, Burattini MN, Lopez LF, Coutinho FA. Forecasting versus projection models in epidemiology: the case of the SARS epidemics. Med Hypotheses 2005;65:17-22.
2. Kim JK. Achievements and challenges of mathematical model-based AI in COVID-19 response. Seoul: National Assembly Research Service; 2023 Aug. Report No.: 31-9735042-001341-14.
3. Ministry of Health and Welfare, Korea Disease Control and Prevention Agency. Mid- to long-term plan for pandemic preparedness. Government of South Korea; 2023.
4. Choi S, Ki M. Analyzing the effects of social distancing on the COVID-19 pandemic in Korea using mathematical modeling. Epidemiol Health 2020;42:e2020064.
5. Wang F, Cao L, Song X. Mathematical modeling of mutated COVID-19 transmission with quarantine, isolation and vaccination. Math Biosci Eng 2022;19:8035-56.
6. Feng XM, Chen J, Wang K, et al. Phase-adjusted estimation of the COVID-19 outbreak in South Korea under multi-source data and adjustment measures: a modelling study. Math Biosci Eng 2020;17:3637-48.
7. Oh J, Apio C, Park T. Mathematical modeling of the impact of Omicron variant on the COVID-19 situation in South Korea. Genomics Inform 2022;20:e22.
8. Sembiring I, Wahyuni SN, Sediyono E. LSTM algorithm optimization for COVID-19 prediction model. Heliyon 2024;10:e26158.
9. Yang J, Shen Z, Dong X, Shang X, Li W, Xiong G. The prediction of the epidemic trend of COVID-19 using neural networks. IFAC Pap OnLine 2020;53:857-62.
10. Goo T, Apio C, Heo G, et al. Forecasting of the COVID-19 pandemic situation of Korea. Genomics Inform 2021;19:e11.
11. Liao Z, Song Y, Ren S, Song X, Fan X, Liao Z. VOC-DL: deep learning prediction model for COVID-19 based on VOC virus variants. Comput Methods Programs Biomed 2022;224:106981.
12. Choi SB, Ahn I. Forecasting imported COVID-19 cases in South Korea using mobile roaming data. PLoS One 2020;15:e0241466.
13. Duan X, Zhang X. ARIMA modelling and forecasting of irregularly patterned COVID-19 outbreaks using Japanese and South Korean data. Data Brief 2020;31:105779.
14. Chakraborty T, Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: a data-driven analysis. Chaos Solitons Fractals 2020;135:109850.
15. Lee DH, Kim YS, Koh YY, Song KY, Chang IH. Forecasting COVID-19 confirmed cases using empirical data analysis in Korea. Healthcare (Basel) 2021;9:254.
16. Liao Z, Lan P, Fan X, Kelly B, Innes A, Liao Z. SIRVD-DL: a COVID-19 deep learning prediction model based on time-dependent SIRVD. Comput Biol Med 2021;138:104868.
17. Khan JI, Ullah F, Lee S. Attention based parameter estimation and states forecasting of COVID-19 pandemic using modified SIQRD Model. Chaos Solitons Fractals 2022;165:112818.
18. Laha AK, Majumdar S. A multi-type branching process model for epidemics with application to COVID-19. Stoch Environ Res Risk Assess 2023;37:305-25.
19. Fisher A, Xu H, He D, Wang X. Effects of vaccination on mitigating COVID-19 outbreaks: a conceptual modeling approach. Math Biosci Eng 2023;20:4816-37.
20. Tajmirriahi M, Amini Z, Kafieh R, Rabbani H, Mirzazadeh A, Javanmard SH. Statistical inference of COVID-19 outbreak: delay distribution effect in EQIR modeling of epidemic. J Med Signals Sens 2022;12:95-107.
21. Kim S, Kim YJ, Peck KR, Jung E. School opening delay effect on transmission dynamics of coronavirus disease 2019 in Korea: based on mathematical modeling and simulation study. J Korean Med Sci 2020;35:e143.
22. Bai Z, Ma Z, Jing L, et al. Estimation and sensitivity analysis of a COVID-19 model considering the use of face mask and vaccination. Sci Rep 2023;13:6434.
23. Kim K, Kim S, Lee D, Park CY. Impacts of social distancing policy and vaccination during the COVID-19 pandemic in the Republic of Korea. J Econ Dyn Control 2023;150:104642.
24. Chung PC, Chan TC. Impact of physical distancing policy on reducing transmission of SARS-CoV-2 globally: perspective from government's response and residents' compliance. PLoS One 2021;16:e0255873.

PHWR
Public Health Weekly Report

Review & Perspective

# Analysis of Coronavirus Disease 2019 Prediction Studies in the Republic of Korea

Hyun-Kyung Kim[1], Boyeong Ryu[2], Min-Gyu Yoo[2], Jaehoon Kim[2], Kyung-Duk Min[3]*

[1]Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, Seoul, Korea, [2]Division of Disease Control Research Planning, Bureau of Department of Data Science, Korea Disease Control and Prevention Agency, Cheongju, Korea, [3]College of Veterinary Medicine, Chungbuk National University, Cheongju, Korea

## ABSTRACT

**Objectives:** During the initial outbreak of coronavirus disease 2019 (COVID-19), numerous predictive studies were conducted amid high uncertainty regarding the characteristics of the virus, and the study results were considered in the policymaking process.

**Methods:** This study systematically analyzed research papers that predicted the spread of COVID-19 in the Republic of Korea. Focusing on 138 studies published between 2020 and October 15, 2024, it examined the data and methodologies employed and explored ways to enhance the utility of predictive outcomes in managing infectious disease outbreaks.

**Results:** These methodologies included mathematical models, statistical models, and machine learning–based approaches to predict COVID-19 spread patterns. Beyond forecasting future outbreak trends, these predictive models were also instrumental in evaluating existing measures and proposing effective policies through scenario-based assumptions.

**Conclusions:** This study's findings highlight the importance of multidisciplinary collaboration in developing predictive models to effectively prepare for and respond to infectious diseases. By doing so, it aims to minimize the public health impacts of infectious diseases.

**Key words:** Coronavirus disease 2019; Forecasting; Projection; Modelling

***Corresponding author:** Kyung-Duk Min, Tel: +82-43-261-8393, E-mail: kdmin@chungbuk.ac.kr

## Introduction

The coronavirus disease 2019 (COVID-19), which was first detected in Wuhan, Hubei Province, China, in December 2019, rapidly spread across the globe, causing the World Health Organization to officially declare it a global pandemic on March 11, 2020. In the early stages of the outbreak, there was a high level of uncertainty regarding the epidemiological characteristics of the COVID-19 virus (severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2]), such as its routes of transmission and infectiousness. In response, studies using predictive models were conducted to forecast the transmission patterns and scale of COVID-19 and to evaluate the effectiveness of the various interventions implemented. These studies were instrumental in designing strategies to effectively control and respond to the spread of infectious diseases.

**Key messages**

① What is known previously?

Predictive models for infectious diseases can be used to analyze epidemic patterns, assess policy effectiveness, and establish policy-based evidence.

② What new information is presented?

During the pandemic, numerous prediction studies were conducted using methodologies from various fields such as mathematics, engineering, and sociology. Domestic coronavirus disease 2019 data were used to develop predictive models to forecast international epidemic trends.

③ What are implications?

Multidisciplinary collaboration is essential for refining predictive models to control the spread of infectious diseases. These models can be used in future preparedness and response plans for emerging infectious diseases.

Predictions can be broadly divided into short-term forecasting and long-term projection. Long-term projection is primarily based on specific models or assumptions and aims to show long-term trends under certain scenarios, such as vaccination rates or behavioral changes. It is intended to elucidate potential outcomes resulting from specific assumptions. In contrast, short-term forecasting refers to attempts made to predict the future course of an infectious disease over a short time frame, typically spanning a few weeks, by providing specific and detailed estimates [1].

Short-term forecasting of infectious disease outbreaks is useful for understanding the current epidemic situation and preparing prompt responses. In contrast, scenario-based long-term projections are valuable because they support long-term decision-making and facilitate comparisons of outbreak trajectories under various conditions. Infectious disease forecasting provides scientific evidence for establishing public health systems and can contribute to the development of effective prevention policies as well as the preparation of proactive crisis response systems through outbreak prediction simulations. Since the proposal of the S (Susceptible)–I (Infected)–R (Recovered) compartmental model by British mathematicians Kermack and McKendrick in 1927 [2], extensive research has been conducted on infectious disease forecasting. In recent years, methodologies have expanded to include a variety of approaches from fields such as machine learning, deep learning, and network analysis. During the COVID-19 pandemic, major international organizations, such as the United States Centers for Disease Control and Prevention (CDC) and the European Centre for Disease Prevention and Control, also used predictive models to mitigate the spread of the disease and establish effective response strategies. The U.S. CDC, in particular, collaborated with more than 100 academic and research institutions during the pandemic to generate short-term forecasts of confirmed cases, hospitalizations, and deaths. These forecasts were published on CDC's website and served as valuable insights for informing policy decision-making.

It is important to improve the accuracy of outbreak scale predictions to enable proactive preparation for future infectious disease events [3]. To support this effort, the present study conducted a literature review of domestic research papers on COVID-19 forecasting, with the aim of securing the necessary scientific evidence, understanding the current state of research, identifying existing needs, and establishing research priorities. Through this review, the study sought to explore effective strategies for managing infectious disease outbreaks and aimed to propose ways for enhancing the practical utility of forecasting results.

## Methods

To understand the current state of COVID-19 forecasting research in the Republic of Korea (ROK), relevant literature was collected using PubMed, the most widely used international academic database in the field of health and medicine. The literature search was conducted until October 15, 2024, with no restrictions on the start date of the search period. Keywords used for the review were categorized into three main groups, "Korea," "COVID-19," and "Prediction," and combinations of terms from these categories were used during the search. The keywords and query formulas were finalized through a review by the research team (Table 1).

The Preferred Reporting Items for Systematic reviews and Meta-Analyses flow diagram was used to systematically describe the literature selection process, presenting each step of the selection procedure. The literature included studies that used domestic data to predict COVID-19 epidemiological indicators in ROK and were published in English. All other papers were excluded. Out of the 929 retrieved articles, 181 that matched the research topic were selected through title and abstract screening based on the inclusion and exclusion criteria. After a full-text review, a total of 138 research papers were included in the final review (Figure 1, Supplementary Table 1; available online).

## Results

A review of domestic literature on COVID-19 forecasting revealed that a total of 138 research studies were conducted over the past 5 years, from 2020 to October 15, 2024. There were a record number of publications immediately after the

**Table 1.** Search strategy for published papers on COVID-19 predictions in the Republic of Korea

| Topic | Search strategy |
|---|---|
| Republic of Korea | Korea[Title/Abstract] OR "South Korea"[Title/Abstract] OR "Republic of Korea"[Title/Abstract] |
| COVID-19 | covid[Title/Abstract] OR corona[Title/Abstract] OR sars-cov-2[Title/Abstract] OR COVID-19[Title/Abstract] |
| Prediction | predict*[Title/Abstract] OR forecast*[Title/Abstract] OR projecti*[Title/Abstract] OR model*[Title/Abstract] |

COVID-19=coronavirus disease 2019.

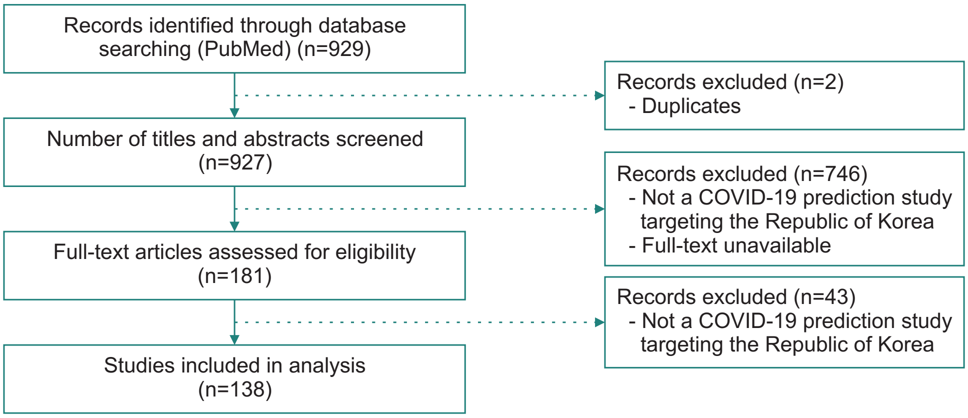

**Figure 1.** Preferred Reporting Items for Systematic reviews and Meta-Analyses flow chart

COVID-19=coronavirus disease 2019.

COVID-19 outbreak, with 49 studies published in 2020 and 42 in 2021. This number has since decreased, with 18 papers published in 2023 (Table 2). Furthermore, an analysis of the selected literature, which aided in identifying trends in research topics and methodologies of COVID-19 forecasting studies by year, revealed that mathematical models, such as compartment and agent-based models, were most frequently used, accounting for 70.3% (97 studies) of the papers. Statistical models, machine learning–based predictive models, and hybrid models that combine compartment models with machine learning approaches were also utilized (Table 3, Figure 2). A closer examination of the characteristics of the models used in domestic COVID-19 forecasting research showed that factors, such as research objectives, data availability, prediction accuracy, and policy applicability, were comprehensively considered when selecting models. Detailed descriptions, advantages and disadvantages, and key features of each model type are summarized in Table 4.

Mathematical models typically categorize the total

**Table 2.** Number of published research papers by year

| Year | Published papers |
|---|---|
| 2020 | 49 |
| 2021 | 42 |
| 2022 | 23 |
| 2023 | 18 |
| 2024 | 6 |
| Total | 138 |

**Table 3.** Number of published research papers by model

| Model | Published papers |
|---|---|
| Mathematical models | 97 |
| Statistical models | 20 |
| Machine learning and artificial intelligence-based models | 11 |
| Mixed models | 10 |

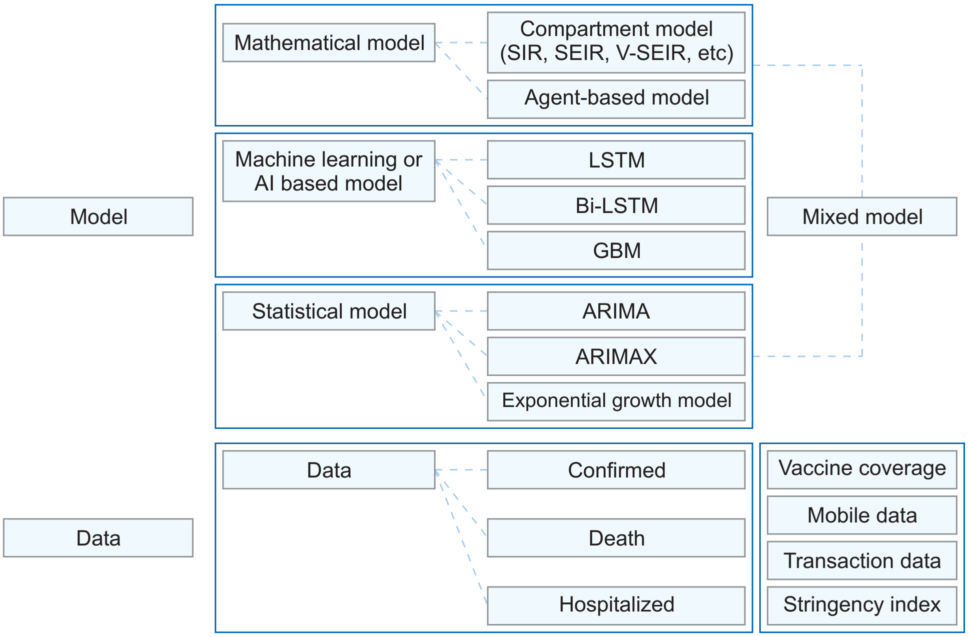

**Figure 2.** Model and data

SIR=susceptible-infectious-recovered; SEIR=susceptible-exposed-infectious-recovered; V-SEIR=vaccinated-SEIR; LSTM=Long Short-Term Memory; GBM=Gradient Boosting Machine; ARIMA=Autoregressive Integrated Moving Average Model; ARIMAX=Autoregressive Integrated Moving Average with Exogenous Variables; AI=artificial intelligence.

Table 4. Summary of prediction models

| Model type | | Specific details | Strengths | Limitations |
|---|---|---|---|---|
| Mathematical models | Compartmental models | • Differential equation-based models (e.g., SIR, SEIR, SIRD)<br>• Population divided into S, E, I, R, D compartments<br>• Transition rates between compartments modeled as parameters | • Intuitive reflection of biological mechanisms<br>• Epidemiologically meaningful parameters<br>• Can simulate intervention scenarios<br>• Predictive with few parameters | • Assumes homogeneous population, reducing realism<br>• Uncertainty due to simplified parameter estimation<br>• Difficult to reflect spatial heterogeneity<br>• Sensitive to initial settings |
| | Agent-based models | • Models individual behaviors and interactions<br>• Defines agent attributes and behavioral rules<br>• Derives macro patterns from micro interactions | • Captures individual-level heterogeneity<br>• Reflects spatial structures and network effects<br>• Analyzes micro-level impacts of interventions<br>• Represents realistic population structures | • High computational complexity<br>• Difficult to validate due to many parameters<br>• Subjectivity in rule setting<br>• Scalability issues in large simulations |
| Machine learning models | Long Short-Term Memory (LSTM) | • Deep learning model specialized for sequential/time-series data<br>• Learns long-term dependencies in recurrent neural networks<br>• Selective memory through gate mechanisms<br>• Learns patterns from past infection data | • High predictive accuracy<br>• Learns complex nonlinear patterns<br>• Capable of learning long-term dependencies<br>• Can handle multivariate time-series data<br>• No need for strict statistical assumptions | • Lack of interpretability (black-box)<br>• Requires sufficient training data<br>• Complex hyperparameter tuning<br>• Limited generalizability to out-of-distribution scenarios |
| | Gradient Boosting Machine (GBM) | • Sequential combination of weak learners (e.g., decision trees)<br>• Builds the final predictive model by gradually reducing errors<br>• Utilizes various feature variables for prediction | • High predictive accuracy<br>• Handles diverse data types<br>• Provides variable importance metrics | • Complex hyperparameter tuning<br>• Limited parallelization due to sequential learning<br>• Limited in capturing temporal dependencies |
| Statistical models | ARIMA, ARIMAX | • Models autoregressive, differencing, and moving average components<br>• Predicts by normalizing non-stationary time series<br>• Uses Box-Jenkins methodology | • Systematic approach based on statistical theory<br>• Parameters are statistically interpretable<br>• Quantifies uncertainty via confidence intervals<br>• Can model with small datasets | • Models only linear relationships<br>• Poor adaptability to structural changes<br>• Subjectivity in order selection<br>• Reduced accuracy in long-term predictions |

**Table 4.** Contiuned

| Model type | | Specific details | Strengths | Limitations |
|---|---|---|---|---|
| Statistical models | Exponential growth models | • Models exponential increase in case counts<br>• Estimates spread speed using growth rate parameters | • Simplicity and interpretability<br>• Rapid prediction in early spread stages<br>• Efficient modeling with few parameters<br>• Intuitive understanding of growth rates | • Limited in reflecting complex spread patterns<br>• May overestimate in long-term predictions<br>• Difficult to incorporate intervention effects<br>• Does not reflect saturation effects |
| Hybrid models | | • Combines advantages of different modeling approaches<br>• Integration of compartmental models with machine learning or ensemble of statistical and machine learning models<br>• Multi-scale modeling approach | • Overcomes limitations of single models<br>• Ensures predictive accuracy and interpretability<br>• Integrates diverse data sources<br>• Provides stable prediction performance | • Increased model complexity and implementation difficulty<br>• Higher computational cost<br>• Complexity in determining weights among models<br>• Challenges in validation and interpretation |

SIR=susceptible-infectious-recovered; SEIR=susceptible-exposed-infectious-recovered; SIRD=susceptible-infected-recovered-deceased; ARIMA=Autoregressive Integrated Moving Average Model; ARIMAX=Autoregressive Integrated Moving Average with Exogenous Variables.

population into groups such as “Susceptible,” “Infected,” and “Recovered,” the classic S-I-R model design. However, these models have evolved to incorporate additional groups, such as the “Hospitalized,” “Vaccinated,” and “Quarantined” populations, resulting in various advanced models such as SEIHR, V-SEIR, SEIHQ, and SEIQRDV3P [4-7]. In studies that used machine learning and algorithmic models, epidemic trends were primarily predicted using techniques such as the Long Short-Term Memory and Gradient Boosting Machine models [8-11]. Among statistical models, time series models such as the Autoregressive Integrated Moving Average Model (ARIMA) and Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) were most commonly used. Logistic growth models and various regression models were additionally employed [12-15]. In recent times, there has been a notable increase in the utilization of hybrid models that integrate methodologies from multiple disciplines for the purpose of infectious disease forecasting. Models, such as m-SIQRD and SIRVD-DL, which combine mathematical modeling with deep learning techniques, have been applied to predict the spread patterns of COVID-19 [10,16,17]. In the early stages of the COVID-19 outbreak, research primarily focused on short-term forecasting of confirmed case numbers by using data released by the Korea Disease Control and Prevention Agency (KDCA). Over time, however, the scope of predictive variables was expanded to include death counts, the number of critically ill patients, and the effective reproduction number ($R_2$). Moreover, although several studies exclusively focused on predicting the spread of COVID-19 within ROK, a considerable number of comparative studies were also conducted to include other countries such as Italy, Hong Kong, and India [18-20]. In addition to studies that predicted epidemic trends using epidemiological indicators, many other studies evaluated the effects of interventions such as vaccination, social distancing, school closures, mask-wearing, and changes in individual behavior. Several also examined the influence of variations in population mobility, modeled through different scenarios, on the patterns of the outbreak [21-23]. Various types of mobility-related data, such as mobile carrier data and consumer behavior patterns, were used for forecasting. In addition, several indices, including the stringency index and the Oxford COVID-19 Government Response Tracker index, were incorporated to account for the intensity of social distancing measures [24].

## Conclusion

The aim of the present study was to establish a foundation for more effective management of future infectious disease outbreaks by identifying trends in COVID-19 forecasting research conducted in ROK. As a result, it was found that forecasting studies on the spread of COVID-19 used a diverse range of data sources and methodologies. Approaches from other fields, such as machine learning, and hybrid models were also integrated and applied in addition to well-established mathematical models. In addition, numerous studies extended beyond simply forecasting trends in deaths, confirmed cases, or hospitalizations to evaluate the effectiveness of implemented policies in order to inform and support future decision-making processes.

A retrospective review of the studies analyzed showed that overall, medium- to long-term predictions (over a 1 month period) often differed from the actual epidemic trends owing to factors not considered at the time of prediction, such as the emergence of new variants, changes in public health policies, and shifts in social behaviors. It was specifically observed that the accuracy of the existing predictive models significantly decreased with each occurrence of virus mutation or alteration in characteristics. This limitation of infectious disease forecasting models highlights the importance of continuous model calibration and updates.

This study confirmed the impact of predictive models on scientific decision-making and the potential for expanding these models through multidisciplinary collaboration. Continuous cooperation among experts and relevant government agencies is essential for the ongoing advancement of infectious disease forecasting research. Specifically, by expanding the data and resources available for infectious disease forecasting through multidisciplinary efforts, it is expected that uncertainties in future outbreaks can be reduced, and more effective response systems can be established. The limitations identified in the study in regard to prediction accuracy also necessitate the development of ensemble forecasting that utilizes multiple methodologies instead of relying on a single model as well as the establishment of real-time model calibration systems. When using forecasting results, it is crucial to explicitly consider prediction uncertainties and to support decision-making with a range of scenarios. There is also a need to establish institutional frameworks to ensure effective utilization of research findings. To ensure that forecasting results are practically reflected in infectious disease response policies, it is crucial to establish strong connections between insights derived from predictive models and the implementation of actual public health policies. This mechanism facilitates health authorities to proactively respond based on scientific evidence and contribute to ensuring public safety.

The KDCA is currently developing an integrated risk analysis system to monitor both domestic and international public health threats, aiming to establish a comprehensive prevention and medical response system for nationwide and large-scale infectious disease pandemics. Standard operating procedures that detail risk assessments according to crisis levels and specific situations have already been established. Moreover, the KDCA has formulated mid- to long-term plans for the prevention of novel infectious disease pandemics by refining outbreak size predictions through methods, such as artificial intelligence modeling, and by formulating comprehensive research plans for implementation in various phases [3]. Future preparedness for infectious diseases is expected to improve by the continuous development of refined predictive models through multidisciplinary collaboration across fields such as mathematics, statistics, artificial intelligence, public health, and medicine, ultimately contributing to reducing the impact on public health.

## Declarations

**Ethics Statement:** Not applicable.

**Funding Source:** None.

**Acknowledgments:** None.

**Conflict of Interest:** The authors have no conflicts of interest to declare.

**Author Contributions:** Conceptualization: HKK, KDM. Data curation: HKK. Supervision: KDM. Validation: BYR, MGY, JHK. Visualization: HKK. Writing–original draft: HKK. Writing–review & editing: HKK, KDM.

## Supplementary Materials

Supplementary data are available online.

## References

1. Massad E, Burattini MN, Lopez LF, Coutinho FA. Forecasting versus projection models in epidemiology: the case of the SARS epidemics. Med Hypotheses 2005;65:17-22.
2. Kim JK. Achievements and challenges of mathematical model-based AI in COVID-19 response. Seoul: National Assembly Research Service; 2023 Aug. Report No.: 31-9735042-001341-14.
3. Ministry of Health and Welfare, Korea Disease Control and Prevention Agency. Mid- to long-term plan for pandemic preparedness. Government of South Korea; 2023.
4. Choi S, Ki M. Analyzing the effects of social distancing on the COVID-19 pandemic in Korea using mathematical modeling. Epidemiol Health 2020;42:e2020064.
5. Wang F, Cao L, Song X. Mathematical modeling of mutated COVID-19 transmission with quarantine, isolation and vaccination. Math Biosci Eng 2022;19:8035-56.
6. Feng XM, Chen J, Wang K, et al. Phase-adjusted estimation of the COVID-19 outbreak in South Korea under multi-source data and adjustment measures: a modelling study. Math Biosci Eng 2020;17:3637-48.
7. Oh J, Apio C, Park T. Mathematical modeling of the impact of Omicron variant on the COVID-19 situation in South Korea. Genomics Inform 2022;20:e22.
8. Sembiring I, Wahyuni SN, Sediyono E. LSTM algorithm optimization for COVID-19 prediction model. Heliyon 2024;10:e26158.
9. Yang J, Shen Z, Dong X, Shang X, Li W, Xiong G. The prediction of the epidemic trend of COVID-19 using neural networks. IFAC Pap OnLine 2020;53:857-62.
10. Goo T, Apio C, Heo G, et al. Forecasting of the COVID-19 pandemic situation of Korea. Genomics Inform 2021;19:e11.
11. Liao Z, Song Y, Ren S, Song X, Fan X, Liao Z. VOC-DL: deep learning prediction model for COVID-19 based on VOC virus variants. Comput Methods Programs Biomed 2022;224:106981.
12. Choi SB, Ahn I. Forecasting imported COVID-19 cases in South Korea using mobile roaming data. PLoS One

2020;15:e0241466.

13. Duan X, Zhang X. ARIMA modelling and forecasting of irregularly patterned COVID-19 outbreaks using Japanese and South Korean data. Data Brief 2020;31:105779.
14. Chakraborty T, Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: a data-driven analysis. Chaos Solitons Fractals 2020;135:109850.
15. Lee DH, Kim YS, Koh YY, Song KY, Chang IH. Forecasting COVID-19 confirmed cases using empirical data analysis in Korea. Healthcare (Basel) 2021;9:254.
16. Liao Z, Lan P, Fan X, Kelly B, Innes A, Liao Z. SIRVD-DL: a COVID-19 deep learning prediction model based on time-dependent SIRVD. Comput Biol Med 2021;138:104868.
17. Khan JI, Ullah F, Lee S. Attention based parameter estimation and states forecasting of COVID-19 pandemic using modified SIQRD Model. Chaos Solitons Fractals 2022;165:112818.
18. Laha AK, Majumdar S. A multi-type branching process model for epidemics with application to COVID-19. Stoch Environ Res Risk Assess 2023;37:305-25.
19. Fisher A, Xu H, He D, Wang X. Effects of vaccination on mitigating COVID-19 outbreaks: a conceptual modeling approach. Math Biosci Eng 2023;20:4816-37.
20. Tajmirriahi M, Amini Z, Kafieh R, Rabbani H, Mirzazadeh A, Javanmard SH. Statistical inference of COVID-19 outbreak: delay distribution effect in EQIR modeling of epidemic. J Med Signals Sens 2022;12:95-107.
21. Kim S, Kim YJ, Peck KR, Jung E. School opening delay effect on transmission dynamics of coronavirus disease 2019 in Korea: based on mathematical modeling and simulation study. J Korean Med Sci 2020;35:e143.
22. Bai Z, Ma Z, Jing L, et al. Estimation and sensitivity analysis of a COVID-19 model considering the use of face mask and vaccination. Sci Rep 2023;13:6434.
23. Kim K, Kim S, Lee D, Park CY. Impacts of social distancing policy and vaccination during the COVID-19 pandemic in the Republic of Korea. J Econ Dyn Control 2023;150:104642.
24. Chung PC, Chan TC. Impact of physical distancing policy on reducing transmission of SARS-CoV-2 globally: perspective from government's response and residents' compliance. PLoS One 2021;16:e0255873.