

Review Article

Artificial Intelligence against COVID-19: A Narrative Review

Chengfan Gu*

RMIT University School of Engineering 264 Plenty Road Mill Park VIC 3082 Australia

ORCID: <https://orcid.org/0000-0002-3510-0162>

Abstract

The COVID-19 pandemic accelerated the development and deployment of artificial intelligence (AI) across clinical medicine, public health, biomedical research, and pharmaceutical development. This narrative review synthesizes representative evidence published primarily from 2024 to 2026, a period chosen to capture the post-emergency phase of COVID-19 in which external validation, multimodal modelling, post-COVID condition, genomic and wastewater surveillance, antiviral discovery, and responsible AI governance have become increasingly prominent. A structured narrative search of PubMed/PMC and World Health Organization (WHO) resources was conducted up to 25 September 2026 using combinations of COVID-19/SARS-CoV-2 and artificial intelligence/machine learning terms together with domain-specific keywords. Evidence was appraised qualitatively according to study population, data source, validation design, reported uncertainty, confounding, and demonstrated clinical or public-health relevance. Recent studies demonstrate that AI can support medical-image interpretation, risk stratification, population surveillance, long-COVID identification, variant characterization, and molecular screening. However, strong internal computational performance does not necessarily translate into external validity or clinical utility. Important 2024 studies showed that apparent diagnostic performance can decline substantially after confounding is controlled or when models are transferred across institutions and imaging devices. Multimodal models integrating electronic health records, patient-reported information, and genomics show modest gains in discrimination, but prospective clinical benefit remains to be established. The evidence indicates that future systems should emphasize external validation, calibration, explainability, privacy, fairness, continuous monitoring, and human clinical oversight. Rather than replacing laboratory testing, clinical judgment, epidemiology, or biomedical experimentation, AI is most appropriately positioned as an integrating and decision-support technology.

Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), produced an unprecedented demand for rapid diagnosis, epidemiological forecasting, treatment discovery, healthcare-resource management, and large-scale disease surveillance. The pandemic also generated enormous volumes of clinical, imaging, laboratory, genomic, behavioral, and population-level data. These characteristics made COVID-19 one of the largest real-world testing environments for artificial intelligence in modern healthcare.

Artificial intelligence includes machine learning (ML), deep learning (DL), natural-language processing, computer vision, generative modelling, causal inference, and related computational approaches capable of identifying patterns

in high-dimensional data. During the early pandemic, AI was investigated for automated interpretation of chest radiographs and computed tomography (CT), prediction of clinical deterioration, epidemic forecasting, contact tracing, and drug repurposing. A 2024 review described AI applications spanning pandemic prediction, diagnosis, public-health decision support, risk analysis, and therapeutic discovery, while also emphasizing continuing limitations involving data quality, infrastructure, ethics, and real-world generalizability [1].

The context has since changed. SARS-CoV-2 continues to circulate globally, but the public-health environment is characterized by higher population immunity, more established clinical management, lower routine community testing, and less comprehensive sequencing than during the emergency phase. WHO surveillance in 2026 continued to

More Information

***Corresponding author:** Chengfan Gu, RMIT University School of Engineering 264 Plenty Road Mill Park VIC 3082 Australia, Email: chengfan.gu@rmit.edu.au

Submitted: September 22, 2026

Accepted: September 25, 2026

Published: September 28, 2026

Citation: Gu C. Artificial Intelligence against COVID-19: A Narrative Review. J Artif Intell Res Innov. 2026; 2(2): 152-159. Available from: <https://dx.doi.org/10.29328/journal.jairi.1001028>

Copyright license: © 2026 Gu C. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords: Artificial intelligence; COVID-19; SARS-CoV-2; Machine learning; Deep learning; Long COVID; Medical imaging; Surveillance; Drug discovery; Pandemic preparedness



document global SARS-CoV-2 activity and changing variant distributions, while cautioning that the available testing and sequencing data are incomplete [2], consistent with WHO's 2026 global risk assessment that continued viral circulation and evolution warrant ongoing monitoring [17].

Accordingly, the key question for contemporary AI research is no longer merely whether an algorithm can distinguish COVID-19 from non-COVID-19 data. The more demanding question is whether a model remains accurate, calibrated, clinically useful, equitable, and safe when prevalence, variants, patient characteristics, clinical practice, and data-acquisition systems change. This distinction is critical because high performance on retrospectively partitioned datasets can overestimate real-world performance when confounding, institutional differences, device characteristics, or population-selection effects are present.

Recent evidence illustrates this challenge. A large 2024 study of respiratory-audio AI reported an ROC-AUC of 0.846 before adjustment, but performance fell to 0.619 after matching on measured confounders [3]. A separate 2024 multi-centre chest-X-ray study demonstrated that equipment manufacturer, image processing, and institutional differences could materially reduce generalization performance [4]. These findings show that high internal accuracy cannot be treated as evidence of deployability without external validation.

At the same time, new opportunities have emerged. Multiscale models integrating electronic health records (EHRs), patient-reported information, and genomics have improved identification of long COVID [5]. AI-assisted genomic surveillance can prioritize emerging variants [6], wastewater data can support community-level prevalence estimation [7], and machine learning can accelerate computational drug screening [8]. WHO guidance published from 2024 to 2026 also places increasing emphasis on safe, ethical, equitable, and evidence-based deployment of AI in health [9-12].

This narrative review examines the current role of AI against COVID-19, with emphasis on evidence published between 2024 and 2026. This time window was selected to focus on the post-emergency phase, when population immunity, variant ecology, testing behavior, clinical management, and AI governance had changed substantially from the early pandemic. Earlier studies are used only where needed for conceptual context. Five application domains are considered: diagnosis and medical imaging; prognosis and post-COVID condition; epidemiological and genomic surveillance; therapeutic discovery; and responsible clinical and public-health implementation. In contrast to reviews that primarily catalogue AI applications, the present review places particular emphasis on the strength of validation, confounding and domain shift, uncertainty reporting, and the distinction between computational performance and demonstrated clinical or public-health impact.

Review methodology

Search scope and strategy

The revised manuscript was developed as a structured narrative review rather than a systematic review. Targeted searches of PubMed/PMC and WHO publication resources were conducted up to 25 September 2026. Search concepts combined (COVID-19 OR SARS-CoV-2) with (artificial intelligence OR machine learning OR deep learning) and domain-specific terms including chest radiograph, computed tomography, respiratory audio, long COVID, electronic health record, genomics, variant surveillance, wastewater, drug discovery, multimodal model, ethics, fairness, and governance. Reference lists of highly relevant recent articles were also used to identify directly related studies. Because the objective was critical synthesis of contemporary evidence rather than exhaustive enumeration, the search was intentionally focused on representative studies with informative validation or translational implications.

Eligibility and study selection

Eligible evidence included peer-reviewed human, population, computational, and translational studies published primarily from January 2024 to September 2026, together with authoritative WHO guidance and surveillance reports relevant to AI use in health. Studies were prioritized when they reported a clearly defined dataset or population, an identifiable AI or computational method, quantitative performance or surveillance outcomes, and sufficient information to judge validation or limitations. Earlier publications, non-COVID applications, opinion pieces without evidentiary content, duplicate reports, and studies lacking enough methodological information for critical interpretation were not emphasized. The 2024-2026 focus was chosen to reflect the contemporary post-emergency context and recent developments in external validation, multimodal modelling, long-COVID research, genomic/wastewater surveillance, and AI governance.

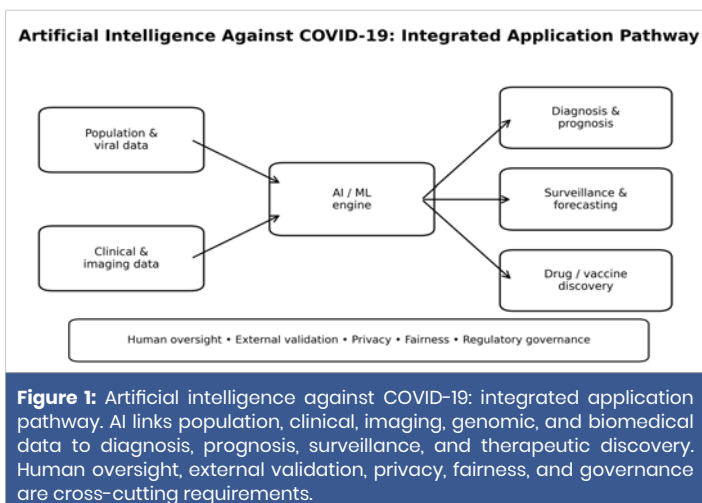
Evidence appraisal and synthesis

Evidence quality was appraised qualitatively rather than through a single pooled score because the included studies span heterogeneous tasks, datasets, and outcome metrics. The appraisal considered sample size and representativeness, internal versus external validation, temporal or cross-institutional testing, control of confounding, reporting of confidence intervals or other uncertainty, subgroup assessment, and whether the study demonstrated only computational performance or also prospective workflow or clinical impact.

Findings were synthesized by application domain, with particular attention to generalizability and translational limitations. Table 1 provides a structured comparison of representative studies.

Table 1: Comparative synthesis of representative recent evidence.

Domain / source	Population or dataset	AI / computational approach	Validation / reported performance	Uncertainty and major limitation
Chest radiography [13]	External validation: 204 radiographs from 107 confirmed COVID-19 patients; second cohort of 50 immunocompromised patients	AI chest-radiograph assessment	174/204 radiographs (85%) identified as COVID-19 pneumonia; 97/107 patients (91%) identified	External validation reported, but no prospective workflow or patient-outcome impact demonstrated.
Imaging generalization [4]	Two centres and multiple X-ray device models	Deep-learning chest-X-ray models	Performance reductions up to 8% with a different manufacturer, 18.9% with image-processing differences, and 9.8% inter-institutionally	Demonstrates device/institution domain shift; reported percentages are performance losses, not directly comparable with diagnostic identification rates.
Respiratory audio [3]	67,842 individuals; 23,514 PCR-positive	Audio classifiers using cough/breathing/speech	ROC-AUC 0.846 unadjusted; 0.619 after matching measured confounders	Strong evidence of confounding; symptom-based prediction outperformed audio-only screening under realistic evaluation.
Long COVID [5]	NIH All of Us; >17,200 SARS-CoV-2-infected participants	XGBoost using EHR alone versus EHR + survey + genomic data	AUROC 0.736 (95% CI 0.730-0.741) versus 0.748 (95% CI 0.741-0.755); repeated train/test splits with cross-validation	Gain is statistically supported but modest; precision remained low and prospective clinical benefit was not demonstrated.
Wastewater surveillance [7,14]	Community and sub-sewershed SARS-CoV-2 RNA time series	Mathematical prevalence estimation and online trend/deviation detection	Population-level prevalence/trend estimation rather than patient-level classification	Uncertainty from shedding variability, rainfall/sewer flow, sampling, assays and changing viral targets.
Genomic surveillance [6]	SARS-CoV-2 sequence data	CoVerage in-silico framework for variant prioritization and characterization	Computational prioritization of variants of interest	Predicted biological relevance still requires laboratory and epidemiological confirmation; sequence sampling is uneven.
Therapeutic discovery [8,15]	Approved compounds and antiviral discovery literature	Molecular docking, ML regression and generative/AI discovery approaches	Computational ranking of candidate compounds and design hypotheses	Docking/predicted affinity is not therapeutic evidence; biochemical, toxicological, pharmacokinetic, animal and clinical validation remain necessary.



Conceptual and theoretical framework

Learning and inference framework

AI against COVID-19 can be formulated as a data-driven inference problem in which heterogeneous observations x are transformed into an output y representing a diagnosis, clinical outcome, transmission state, viral characteristic, or therapeutic property. In supervised learning, the predictor may be written as $\hat{y} = f_{\theta}(x)$, where f_{θ} is a parameterized machine-learning or deep-learning model. Model parameters are estimated by minimizing an empirical objective, $\theta^* = \arg \min_{\theta} [(1/N) \sum_i L(y_i, f_{\theta}(x_i)) + \lambda \Omega(\theta)]$, in which L is the task-specific loss and $\Omega(\theta)$ regularizes model complexity. Depending on the application, x may contain clinical and laboratory measurements, chest images, respiratory

audio, viral genomes, wastewater signals, or molecular descriptors. The same theoretical mapping therefore supports classification, regression, forecasting, anomaly detection, and candidate ranking across the COVID-19 applications summarized in this article [1,16].

Multimodal fusion and temporal prediction

A central theoretical advantage of AI is its ability to combine information that exists at different biological and population scales. Let $x = \{x_C, x_I, x_G, x_E, x_M\}$ denote clinical, imaging/audio, genomic, epidemiological, and molecular inputs. Modality-specific encoders $g_m(\cdot)$ first convert each input into a latent representation z_m ; a fusion operator F then forms $z = F(z_C, z_I, z_G, z_E, z_M)$, from which an output model $h(z)$ estimates $P(y|x)$ or a continuous response. This formulation explains why multimodal systems can outperform single-source models when complementary information is available, as illustrated by recent long-COVID modelling that combines EHR, survey, and genomic data [5]. For surveillance and longitudinal disease, the input becomes time-dependent, x_t , and the model estimates a future state y_{t+h} , allowing sequential measurements from wastewater, variant frequencies, hospital activity, or patient trajectories to support forecasting and early-warning functions [6,7].

Generalization, uncertainty, and human decision support

The theoretical objective is not only accurate fitting of the training data but reliable inference under changing conditions. In practice, the target distribution may differ from the training distribution, $P_{\text{train}}(x,y) \neq P_{\text{target}}(x,y)$, because SARS-CoV-2 variants, disease prevalence, vaccination,



patient populations, hospitals, imaging equipment, and data-collection procedures change over time. Recent COVID-19 studies demonstrate that confounding and domain shift can substantially reduce performance after external evaluation [3,4]. Robust AI therefore requires external validation, calibration, subgroup assessment, uncertainty estimation, and monitoring for model drift. If the model produces a predictive distribution $P(y|x)$ together with uncertainty $u(x)$, its output can be incorporated into a decision rule that selects an action a to maximize expected clinical or public-health utility rather than replacing expert judgement. Human oversight, privacy, fairness, explainability, and governance thus form part of the theoretical system boundary, consistent with WHO guidance on responsible AI for health [9-12].

Figure 2. Theoretical framework of artificial intelligence against COVID-19. Heterogeneous clinical, imaging, genomic, epidemiological, and molecular data are transformed through representation learning and AI inference into probabilistic predictions. External validation, calibration, confounding control, privacy, fairness, explainability, and human oversight determine whether these predictions can be translated into reliable clinical, public-health, or research decisions.

Evidence synthesis

The evidence below is interpreted across four distinct levels: computational performance on the development dataset; external or cross-domain validation; prospective evaluation in real workflows; and demonstrated clinical or public-health impact. These levels should not be treated as equivalent. A higher AUROC or classification accuracy can indicate better discrimination without establishing calibration, transportability, usefulness in practice, or improved health outcomes.

AI-assisted diagnosis and medical imaging

Medical imaging remains one of the most mature areas of AI application to COVID-19. Convolutional neural networks and related architectures can detect pulmonary patterns associated with COVID-19 pneumonia on chest radiographs and CT images. Recent studies, however, provide a more

nuanced understanding of their practical value: a model can show clinically useful discrimination while still being sensitive to institutional and technical domain shifts.

A 2024 external-validation study of AI assessment of chest radiographs reported clinically useful identification of COVID-19-associated pneumonia in an external cohort [13]. Such systems can support radiographic triage when rapid review is required. However, Fernandez-Miranda, et al. evaluated deep-learning generalization across institutions and X-ray equipment and found that adding images from a device manufacturer different from the training devices reduced internal performance by up to 8%. Generalization losses associated with image-processing differences and inter-institutional variation reached 18.9% and 9.8%, respectively, in the reported comparisons [4].

These results expose a central weakness of early COVID-19 imaging research: algorithms may learn acquisition- or institution-specific features in addition to disease biology. External validation should therefore include diverse hospitals, device manufacturers, image-processing pipelines, disease prevalences, and patient groups. Calibration and subgroup performance are also important because a model with a high overall AUC can still produce clinically problematic probability estimates or unequal error rates.

Figure 3. Representative 2024 imaging-AI evidence with explicit denominators and metric types. In an external validation study [13], 174 of 204 chest radiographs (85%) were identified as COVID-19 pneumonia, corresponding to 97 of 107 patients (91%). In a separate two-centre/device-generalization study [4], reported performance losses reached up to 8% for a different equipment manufacturer, 18.9% for image-processing differences, and 9.8% for inter-institutional variation. The identification proportions and performance-loss percentages have different denominators and are displayed together only to contrast external validation with domain-shift sensitivity; they should not be interpreted as a common accuracy scale.

Respiratory audio and the problem of confounding

Cough, breathing, and speech recordings attracted interest because they could theoretically enable inexpensive non-contact screening through smartphones or wearable devices.

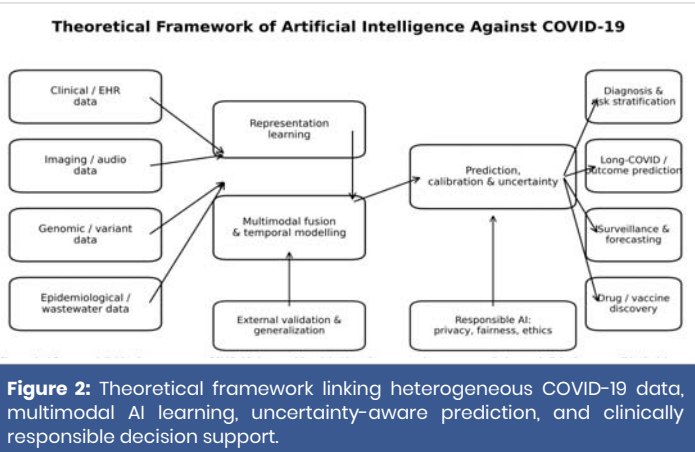


Figure 2: Theoretical framework linking heterogeneous COVID-19 data, multimodal AI learning, uncertainty-aware prediction, and clinically responsible decision support.

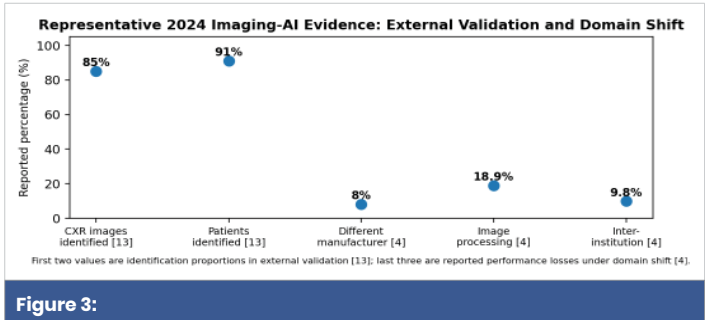


Figure 3:



However, recent evidence demonstrates why representative evaluation is essential. Coppock, et al. analyzed 67,842 individuals, including 23,514 PCR-positive participants. In unadjusted analysis, respiratory-audio classifiers achieved an ROC-AUC of 0.846. After matching on measured confounders such as symptoms, age, and gender, performance fell to 0.619; under realistic evaluations, symptom-based prediction was more useful than audio-only classification [3].

The implication extends beyond acoustic diagnosis. Machine-learning systems can exploit correlations that are statistically predictive within a dataset but not causally related to the target disease. Recruitment procedures, symptoms, recording devices, geography, and population characteristics can all become shortcuts. Carefully designed external and temporally separated validation, together with explicit confounder analysis, is therefore essential before deploying AI screening systems.

Figure 4. COVID-19 respiratory-audio AI and confounding. Coppock, et al. [3] analysed 67,842 individuals, including 23,514 PCR-positive participants. ROC-AUC declined from 0.846 in the unadjusted analysis to 0.619 after matching on measured confounders including symptoms, age, and gender. The figure therefore illustrates the effect of confounder control on apparent discrimination rather than a prospective clinical-impact comparison.

Prognosis, multimodal prediction, and long COVID

AI approaches have been used to predict hospitalization, intensive-care requirements, mortality, and longer-term sequelae using demographic, laboratory, physiological, and EHR variables. More recent work extends these methods to post-COVID condition, where prediction is more difficult because symptoms are heterogeneous, longitudinal, and influenced by pre-existing health, reinfection, vaccination, and social factors.

A 2026 study using more than 17,200 SARS-CoV-2-infected participants from the NIH All of Us Research Program investigated whether combining EHR information with survey and genomic data improved long-COVID identification. Using XGBoost, the EHR-only model achieved an AUROC of 0.736 (95% CI 0.730-0.741), whereas the multiscale model reached 0.748 (95% CI 0.741-0.755) [5]. The improvement was statistically

supported but modest in absolute magnitude. Importantly, the study evaluated predictive discrimination under repeated model-development and validation procedures; it did not demonstrate prospective clinical benefit, and the reported precision remained low. These findings therefore support incremental predictive value from additional data modalities rather than evidence that multimodal AI is ready to replace clinical assessment.

This work illustrates a potential direction for healthcare AI, in which baseline medical history, acute-infection characteristics, laboratory data, vaccination and reinfection status, symptoms, lifestyle, social determinants, wearable measurements, and genomics may be integrated when such data are available and clinically justified. Whether the additional complexity produces meaningful benefit will depend on external validation, calibration, missing-data handling, subgroup performance, privacy, implementation cost, and prospective evaluation. Multimodal integration should therefore be viewed as a promising but context-dependent strategy rather than an established superior approach for all COVID-19 tasks.

Figure 5. Multiscale AI for long-COVID identification in the NIH All of Us cohort (n > 17,200) [5]. The XGBoost EHR-only model achieved AUROC 0.736 (95% CI 0.730-0.741), while the EHR + survey + genomic model achieved AUROC 0.748 (95% CI 0.741-0.755). Error bars show the reported 95% confidence intervals. The absolute gain is modest and does not by itself establish prospective clinical benefit.

Epidemiological surveillance and wastewater analytics

As routine population-wide diagnostic testing declines, confirmed-case counts become less reliable indicators of community transmission. Wastewater surveillance provides an alternative because infected individuals can shed viral material regardless of whether they seek clinical testing. Mohring et al. showed in 2024 that SARS-CoV-2 wastewater measurements could be integrated with mathematical modelling to estimate community prevalence [7]. Ensor et al. separately demonstrated online trend estimation and detection of trend deviations in sub-sewershed SARS-CoV-2 RNA time series [14].

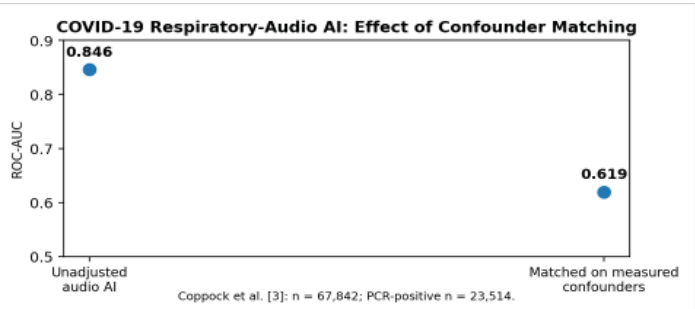


Figure 4:

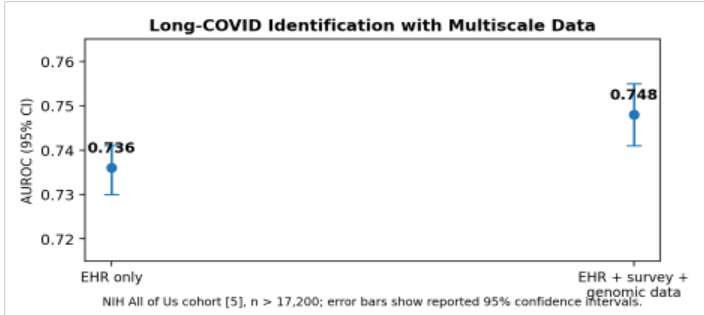


Figure 5:

These systems are naturally compatible with AI. Machine-learning models can combine wastewater viral concentrations with weather, mobility, clinical admissions, variant composition, and historical transmission patterns. Such integration may be especially valuable when individual testing is inconsistent. Nevertheless, wastewater modelling has important uncertainties: viral shedding varies across individuals and disease stages; rainfall and sewer flow can dilute measurements; laboratory assays and sampling protocols vary; and mutation can affect detection. AI therefore needs to model measurement uncertainty rather than obscure it.

AI-assisted genomic surveillance

SARS-CoV-2 continues to evolve, creating a continuing need for rapid assessment of emerging lineages. Conventional genomic surveillance identifies mutations and lineage frequencies, while AI can prioritize sequence changes that may alter viral fitness, immune escape, or epidemiological behavior. Norwood et al. introduced CoVerage, an in silico genomic surveillance framework designed to predict and characterize SARS-CoV-2 variants of interest [6]. This represents a shift from passive sequence description toward computational prioritization of potentially important variants.

WHO's 2026 epidemiological reporting also illustrates the dynamic surveillance environment. In the 28 days ending 7 June 2026, NB.1.8.1 represented 50% of the sequences included in the summarized global dataset, XFG 17%, and BA.3.2 9%. WHO emphasized that sequence counts were limited and geographically uneven and that the monitored variants had not shown evidence of increased overall public-health risk in that reporting period [2]. AI can help integrate mutation patterns, structural modelling, epidemiological growth rates, neutralization data, and geographic spread, but predictions of biological behavior still require laboratory and epidemiological confirmation.

Figure 6. WHO-reported SARS-CoV-2 variant distribution for the 28 days ending 7 June 2026 [2]. The denominator is the set of SARS-CoV-2 sequences included in the summarized WHO global dataset for that reporting period: NB.1.8.1 accounted for approximately 50%, XFG 17%, and BA.3.2 9%. WHO cautioned that sequencing volume was limited and geographically uneven; these percentages therefore describe submitted sequence data and are not direct estimates of population prevalence.

AI in antiviral and therapeutic discovery

Drug discovery is another important AI application. AI can accelerate target identification, virtual screening, drug repurposing, molecular-property prediction, de novo molecular generation, and resistance analysis. Aqeel, et al. combined molecular docking and machine-learning regression to screen approved compounds against the SARS-CoV-2 3CL protease, illustrating how computational approaches can prioritize candidate compounds before laboratory validation [8].

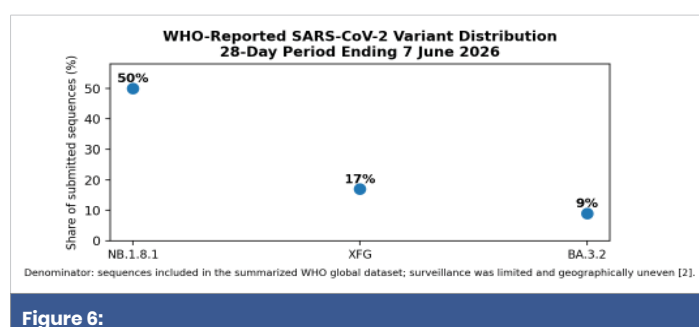


Figure 6:

The field is moving beyond ranking existing compounds. Modern generative models can propose molecular structures conditioned on desired binding, toxicity, pharmacokinetic, and synthetic properties. A 2026 review of antiviral drug discovery identified four major AI application areas: target identification using host-virus interactions and genome-wide screening; drug repurposing; de novo molecule design; and prediction of resistance-associated mutations [15]. WHO has also emphasized both the potential and the ethical and governance risks of AI throughout pharmaceutical development [10].

Computational acceleration should not be confused with therapeutic evidence. Docking scores, predicted affinities, and generated molecules remain hypotheses until they are verified by biochemical assays, toxicity testing, pharmacokinetic evaluation, appropriate animal studies, and ultimately clinical trials.

Discussion

From proof-of-concept accuracy to real-world robustness

The central trend in the 2024-2026 literature is a maturation of the scientific question. Early COVID-19 AI research frequently emphasized whether sophisticated neural networks could achieve high classification accuracy. More recent work asks whether predictions remain useful when transferred to different populations, devices, hospitals, and epidemiological periods. This transition is necessary because COVID-19 presents substantial dataset shift: dominant variants, vaccination coverage, prior immunity, treatment protocols, testing behavior, disease prevalence, and healthcare utilization all change over time.

External validation should therefore be treated as a core part of model development rather than an optional final experiment. The respiratory-audio and imaging studies reviewed here show that recruitment bias, symptoms, institution, and device characteristics can materially alter apparent performance [3,4]. Prospective evaluation is equally important because deployment changes the context in which predictions are used and may expose calibration, workflow, or subgroup problems that are not visible in retrospective studies.

Multimodal AI: potential value and evidentiary limits

COVID-19 is not represented by any single measurement. Imaging captures pulmonary structure; laboratory tests reflect physiological responses; EHRs describe clinical history; viral and host genomics characterize biological variation; wastewater reflects community transmission; wearables provide longitudinal physiology; and patient-reported information captures symptoms that may not appear in clinical records. Multimodal AI can in principle integrate these heterogeneous sources, but added modalities do not automatically improve clinically relevant performance and may increase missingness, implementation burden, privacy exposure, and opportunities for bias.

The 2026 long-COVID study provides a practical example in which adding survey and genomic data to EHR data increased AUROC from 0.736 (95% CI 0.730-0.741) to 0.748 (95% CI 0.741-0.755) [5]. This is a measurable but modest improvement and should not be interpreted as evidence of clinical impact without prospective evaluation. Large multimodal models may eventually integrate imaging, text, laboratory measurements, and molecular information, but their value must be demonstrated against simpler baselines and assessed for calibration, subgroup reliability, cost, privacy, hallucination, hidden bias, and automation error. WHO guidance on large multimodal models accordingly emphasizes governance throughout design, development, evaluation, and deployment [11].

Integrated surveillance as an increasingly relevant use case

The relative value of AI applications changes with the epidemiological environment. When diagnostic capacity was scarce, automated diagnosis attracted substantial attention. In the current environment, laboratory testing is well established, while routine community testing and sequencing are less comprehensive. AI may consequently offer increasing public-health value through integrated surveillance rather than stand-alone classification.

Wastewater measurements, hospital admissions, syndromic surveillance, sequence databases, and other respiratory-pathogen data can be fused to estimate disease activity and identify unusual patterns. The same infrastructure can extend beyond SARS-CoV-2 to influenza, respiratory syncytial virus, novel coronaviruses, and future zoonotic infections. COVID-19 can therefore be viewed as a development and stress test for AI-enabled pandemic intelligence.

AI should augment rather than replace human expertise

AI is most valuable when it reduces information-processing

burden while preserving human accountability. In radiology, AI may flag suspicious images or quantify abnormalities, but interpretation still depends on clinical context and differential diagnosis. In epidemiology, AI can identify trends, but public-health decisions also require behavioral, social, and economic context. In drug discovery, AI can prioritize candidates, but laboratory and clinical validation remains indispensable. Human-AI collaboration is therefore preferable to autonomous decision-making in high-stakes applications.

Systems should communicate uncertainty rather than simply return categorical outputs. Explainability is especially important when predictions influence clinical intervention, resource allocation, or public-health policy. Users need to understand not only what the model predicts, but whether the prediction is well supported for the relevant patient, population, and current epidemiological context.

Ethical, regulatory, and equity considerations

The rapid adoption of AI during the pandemic revealed significant ethical challenges. Training data frequently overrepresent locations with strong digital infrastructure, while underrepresenting communities with fewer resources. Models developed in one setting may therefore perform poorly in populations that differ in healthcare access, language, equipment, socioeconomic conditions, or disease prevalence. Privacy is also critical because COVID-19 AI may combine health records, genomics, mobility information, voice recordings, and wearable-sensor data.

WHO guidance published from 2024 to 2026 emphasizes safe, ethical, equitable, and science-based AI adoption, appropriate regulation, and strengthened ethics oversight [9-12]. The 2026 WHO report on AI-related health research highlights issues including fairness, benefit sharing, power imbalances, capacity building, and risks affecting low- and middle-income countries [12]. Responsible COVID-19 AI should therefore include representative data, independent multicentre validation, subgroup evaluation, explicit uncertainty reporting, privacy protection, transparent documentation, monitoring for model drift, and human oversight of consequential decisions.

In practical terms, responsible deployment should include prespecified subgroup analyses (for example by age, sex, relevant comorbidity, site, and equipment), subgroup-specific calibration where appropriate, explicit reporting of missing data and uncertainty, data minimization and access controls for sensitive health/genomic information, documentation of model provenance and intended use, independent external validation, drift monitoring after deployment, and predefined escalation to human review when confidence is low or outputs conflict with clinical evidence. These measures link fairness, privacy, explainability, and oversight to auditable operational requirements rather than treating them only as abstract principles.

Limitations of the current evidence

The evidence base has several limitations. First, studies are highly heterogeneous in patient populations, pandemic periods, variants, vaccination status, outcome definitions, input variables, and performance metrics. Second, publication bias can favor models reporting strong accuracy. Third, prospective clinical-impact studies remain less common than retrospective technical evaluations. Fourth, rapidly evolving SARS-CoV-2 epidemiology creates temporal dataset shift, meaning that a model trained on one variant era may not maintain the same calibration in another. Fifth, many algorithms predict correlation rather than causation; the respiratory-audio literature demonstrates how confounding can produce apparently strong performance without robust disease-specific information. Finally, technical accuracy does not automatically improve patient outcomes or public-health decisions. Future evaluations should therefore measure workflow impact, clinical utility, cost-effectiveness, and health outcomes in addition to discrimination metrics.

Conclusion

Artificial intelligence has become an important component of the scientific and technological response to COVID-19. Research from 2024 to 2026 has progressed beyond early proof-of-concept classification toward more demanding applications involving external validation, multimodal prediction, long-COVID assessment, wastewater and genomic surveillance, antiviral discovery, and responsible AI governance.

Medical-imaging systems continue to demonstrate useful computational and external-validation performance in selected datasets, but recent evidence confirms that performance can deteriorate across hospitals and devices. Respiratory-audio studies provide an important warning that confounding can produce misleadingly high apparent accuracy. Multimodal AI integrating clinical, behavioral, and genomic information can yield statistically supported, but sometimes modest improvements in discrimination, and its prospective clinical value remains to be established. AI-assisted wastewater and genomic surveillance can strengthen situational awareness when interpreted with sampling and measurement limitations, while machine-learning and generative approaches can accelerate identification of therapeutic candidates before laboratory and clinical validation.

The principal lesson is not that increasingly complex AI models will automatically defeat COVID-19. AI is most effective when it integrates reliable data, operates within a well-defined clinical or epidemiological task, undergoes rigorous external validation, quantifies uncertainty, and supports rather than replaces expert judgment. The infrastructure and methodological lessons developed during COVID-19 extend beyond SARS-CoV-2 and can contribute to faster, more adaptive, and more resilient responses to future infectious-disease emergencies.

References

1. Innovative applications of artificial intelligence during the COVID-19 pandemic. *Infect Med.* 2024;3(1):100095. Available from: <https://dx.doi.org/10.1016/j.imj.2024.100095>.
2. World Health Organization. Weekly epidemiological record: SARS-CoV-2 (COVID-19) global epidemiological update. 2026;101(28). Data reported for the period ending 21 June 2026.
3. Coppock H, Nicholson G, Kiskin I, et al. Audio-based AI classifiers show no evidence of improved COVID-19 screening over simple symptoms checkers. *Nat Mach Intell.* 2024;6:229–42. Available from: <https://dx.doi.org/10.1038/s42256-023-00773-8>.
4. Fernandez-Miranda PM, Marques Fraguola E, de Linera-Alperi MA, et al. A retrospective study of deep learning generalization across two centers and multiple models of X-ray devices using COVID-19 chest X-rays. *Sci Rep.* 2024;14:14657. Available from: <https://dx.doi.org/10.1038/s41598-024-64941-5>.
5. Guardo C, Xinmeng Z, Gangireddy S, et al. Multi-scale data improves performance of machine learning model for long COVID identification. *Commun Med.* 2026;6:389. Available from: <https://dx.doi.org/10.1038/s43856-026-01621-7>.
6. Norwood K, Deng ZL, Reimering S, et al. In silico genomic surveillance by CoVerage predicts and characterizes SARS-CoV-2 variants of interest. *Nat Commun.* 2025;16:6281. Available from: <https://dx.doi.org/10.1038/s41467-025-60231-4>.
7. Mohring J, Leithauser N, Wlazlo J, et al. Estimating the COVID-19 prevalence from wastewater. *Sci Rep.* 2024;14:14384. Available from: <https://dx.doi.org/10.1038/s41598-024-64864-1>.
8. Aqeel I, Majid A, Albanyan A, et al. Drug repurposing targeting COVID-19 3CL protease using molecular docking and machine learning regression approaches. *Sci Rep.* 2025;15:18722. Available from: <https://dx.doi.org/10.1038/s41598-025-02773-7>.
9. World Health Organization. Artificial intelligence for health: supporting countries to deploy responsible AI technologies to accelerate equitable health for all. Geneva: WHO; 2024.
10. World Health Organization. Benefits and risks of using artificial intelligence for pharmaceutical development and delivery. Geneva: WHO; 2024. ISBN 978-92-4-008810-8.
11. World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Geneva: WHO; 2025. ISBN 978-92-4-008475-9.
12. World Health Organization. Artificial intelligence-related health research: ethics review and oversight. Geneva: WHO; 2026. ISBN 978-92-4-012407-3.
13. Artificial intelligence assessment of chest radiographs for COVID-19. 2024. PubMed PMID: 39710565.
14. Ensor KB, Schedler JC, Sun T, et al. Online trend estimation and detection of trend deviations in sub-sewershed time series of SARS-CoV-2 RNA measured in wastewater. *Sci Rep.* 2024;14:5575. Available from: <https://dx.doi.org/10.1038/s41598-024-56175-2>.
15. Tirosoyan I, Gabrielyan Y, Petrosyan V, Vignuzzi M, Zakaryan H. Can artificial intelligence transform antiviral drug discovery? *Drug Discov Today.* 2026;31(3):104648. Available from: <https://dx.doi.org/10.1016/j.drudis.2026.104648>.
16. Ghorbian M, Ghorbian S, Ghoabaei-Arani M. AI-driven techniques for detection and mitigation of SARS-CoV-2 spread: a review, taxonomy, and trends. *Clin Exp Med.* 2025;25:204. Available from: <https://dx.doi.org/10.1007/s10238-025-01753-5>.
17. World Health Organization. COVID-19 global risk assessment – version 10. Geneva: WHO; 6 August 2026. Global public-health risk assessed using data through 30 July 2026.