Briefings in Bioinformatics, 2026, 27, bbag249
https://doi.org/10.1093/bib/bbag249
Published: 24 May 2026
Review

OXFORD

Comparative review of artificial intelligence for transcriptomic biomarker discovery in coronavirus disease 2019 (COVID-19)

Li Ying Khoo and Sarinder Kaur Dhillon*

Data Science and Bioinformatics Laboratory, Institute of Biological Sciences, Faculty of Science, Universiti Malaya, Lembah Pantai, 50603 Kuala Lumpur, Malaysia

*Corresponding author. E-mail: sarinder@um.edu.my

Abstract

The Coronavirus Disease 2019 (COVID-19) pandemic has highlighted the significance of reliable molecular biomarkers in clinical use. Despite the popularity of traditional statistical approaches, the high dimensionality of transcriptomic data presents challenges for these conventional methods. While artificial intelligence (AI) algorithms have emerged as highly advantageous for handling these complex datasets, there is a lack of evaluation of these approaches in COVID-19 transcriptomic studies. This review aims to provide an evaluation of these studies employed for transcriptomic biomarker discovery in COVID-19 using AI, assessing their study designs, methodologies, and outcomes. Based on a comprehensive search for literature across five databases including Web of Science Core Collection, Scopus, PubMed/MEDLINE, IEEE Xplore Digital Library, and LitCovid from December 2019 to March 2025, this review selected 63 studies for a narrative synthesis of four key sections: (i) The Landscape of AI-Driven COVID-19 Transcriptomics, (ii) Limitations of Studies, (iii) A Proposed AI-Driven Transcriptomics Framework, and (iv) Clinical Translation Challenges, Opportunities, and Future Directions. Our analysis revealed limitations in data quality, sample size, and heterogeneity, as well as methodologies regarding validation and interpretability. Thus, we proposed an evidence-informed workflow that addresses these current limitations in study design, while acknowledging real-world constraints. We further discuss the emerging potential of agentic AI systems as a promising solution to current limitations. By bridging methodological gaps with translation considerations, this review can enhance pandemic response strategies for future emerging infectious diseases.

Keywords COVID-19, artificial intelligence (AI), machine learning (ML), transcriptomics

Key Points

- Applications observed in reviewed studies mainly included applications in diagnosis and severity stratification of COVID-19 patients.
- The limitations of current studies included small sample sizes, the reliance on public datasets lacking detailed metadata, batch effects and data heterogeneity reducing model robustness, the lack of external validation, risks of data leakage and circular validation leading to inflated performance metrics, and challenges in model interpretability.
- An evidence-informed AI-driven framework is proposed, acknowledging real-world constraints including small pandemic cohort sizes, domain shift from viral evolution, and resource-limited settings, with emerging agentic AI systems offering potential solutions.

Introduction

In late 2019, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) outbreak marked the start of a global health emergency, where limitations in our existing diagnostic and prognostic capabilities of infectious diseases were accentuated. The heterogeneous symptoms of Coronavirus Disease 2019 (COVID-19) can manifest as an asymptomatic disease, acute respiratory distress syndrome, or multi-organ failure, making it crucial for molecular biomarkers to stratify patient risk and improve therapeutic interventions. Molecular biomarkers can be obtained using transcriptomic profiling methodologies to understand the underlying interactions and biological pathways in COVID-19 pathogenesis.

Using high-throughput technologies such as bulk RNA sequencing (RNA-Seq), single-cell RNA sequencing (scRNA-Seq), and spatial

Received: January 12, 2026. **Revised:** March 12, 2026. **Accepted:** April 24, 2026

transcriptomics, large amounts of gene expression data from respiratory samples, blood, and infected tissues were generated. These data can provide crucial information on viral tropism, host immune responses, and cellular heterogeneity of SARS-CoV-2 infection. However, the high dimensionality of transcriptomic data, where the number of features (genes) greatly exceeds the number of samples, makes it difficult for traditional statistical methods to comply with the fundamental assumptions. This can lead to overfitting, estimation instability, and suboptimal convergence [1], highlighting the need for specialized computational approaches designed to combat these problems.

Artificial intelligence (AI) is often equipped to handle these computational demands, having already demonstrated robust scalability across diverse omics domains, from genomics, proteomics to metabolomics [2–5]. It can handle a high number of features, solving this problem in transcriptomic data analysis. Under the AI branch, machine learning (ML), neural networks, and deep learning (DL) are some common techniques [6]. A comprehensive review by Cheng *et al.* [7] highlighted the applications of ML in transcriptomics, including to predict disease states, identify gene biomarkers, and deconvolute single-cell data. ML also enables cross-platform integration by learning platform-specific biases and batch effects while facilitating the detection of complex splice variants and transcript isoforms in long-read sequencing [7]. Therefore, complex, nonlinear relationships between gene expression patterns and clinical outcomes can be studied, transforming the challenge of high dimensionality into an opportunity to discover novel biological insights. This has made ML a crucial tool for modern transcriptomic analysis, particularly in complex infectious diseases like COVID-19.

Previous COVID-19 reviews have examined AI applications, including diagnostic, prognostic, therapeutic, epidemiological surveillance, molecular biomarker discovery, and systems biology applications [8–17]. While a systematic review by Sekaran *et al.* [10] has been conducted on multimodal genetic information which includes transcriptomic aspects, critical gaps remain, especially considering the significant challenges of extreme dimensionality, batch effects across studies, platform-specific technical noise, biological variability between patients, and the critical need for interpretable results translatable to clinical applications [1, 18–21].

In addition, the heterogeneity in methodological approaches from the traditional ML to deep neural networks and DL, combined with inconsistent validation strategies and different performance metrics, has raised questions about selecting optimal algorithms for biomarker discovery. This review distinguishes itself from existing reviews by providing a focused, critical, and systematic evaluation of AI-driven transcriptomic biomarker discovery in COVID-19, moving beyond prior broad surveys of AI applications. Without systematic evaluation and comparison of these approaches in COVID-19 transcriptomic studies, researchers risk employing suboptimal approaches and missing clinically translatable biomarkers. Therefore, we aim to address this significant gap by providing a comprehensive review of AI methods specifically applied to COVID-19 transcriptomic biomarker discovery, offering evidence-based recommendations for future emerging infectious diseases or pandemics.

This review starts with the introduction followed by the methodology, describing the search strategy, eligibility criteria, and screening process used to identify the 63 studies. The next section presents the results and discussion, beginning with the landscape of AI-driven COVID-19 transcriptomics, followed by a critical assessment of data and methodological limitations, as well as an evidence-informed AI-driven transcriptomics framework that synthesizes lessons from the reviewed studies. The final sections discuss the challenges in clinical translation, opportunities, and future directions, such as the emerging agentic AI before concluding with a summary of key findings.

Methodology

Search strategy

A comprehensive literature search was conducted across the Web of Science Core Collection, Scopus, PubMed/MEDLINE, IEEE Xplore Digital Library, and LitCovid from 1 December 2019 to 31 March 2025. Our search strategy combined controlled vocabulary terms and keywords in three conceptual domains connected by Boolean operators. The terms searched were listed below:

1) COVID-19 terms: 'COVID-19', 'COVID19', 'coronavirus disease 2019', 'SARS-CoV-2', 'novel coronavirus', or '2019-nCoV'.
2) AI terms: 'artificial intelligence', 'machine learning', 'ML', 'deep learning', 'DL', 'support vector machine', 'random forest', 'XGBoost', 'gradient boost', 'feature selection', 'Boruta', 'SHAP', 'explainable AI', 'computational model', 'predictive model', 'supervised learning', or 'unsupervised learning'.
3) Gene expression terms: 'gene', 'gene expression', 'gene biomarker', 'differential∗ expressed gene∗', 'DEG', 'DEGs', 'RNA-Seq', 'differential gene expression', 'DGE', 'transcriptom∗', 'molecular signature∗' or 'host response'.

Search strings were revised to the requirements for each database's syntax. Filters were used to restrict results to the English language. Conference proceedings were also included to minimize publication bias. The full search strategy with database-specific syntax used in this review was documented in Supplementary Table S1. The records were imported into EndNote software for deduplication and screening.

Eligibility criteria

Studies analyzing any type of transcriptomic data from microarray and next-generation sequencing technologies, including bulk RNA-Seq, scRNA-Seq, microRNA sequencing (miRNA-Seq), and T/B-cell receptor (TCR/BCR) profiling, were included in this review. Studies were included if these inclusion criteria were met:

1) Transcriptomic data were analyzed from human COVID-19 patients.
2) AI or ML algorithms were employed for biomarker discovery, feature selection, or predictive modeling based on gene expression.
3) Quantitative performance metrics such as accuracy, sensitivity, specificity, and the area under the receiver operating characteristic (AUC) were reported.
4) Validation strategies for AI models were employed.

Meanwhile, studies were excluded when:

1) Only animal models or *in vitro* systems without human validation were used.
2) Only traditional statistical methods without AI components were conducted.
3) AI applications were employed on non-transcriptomic data such as genomic variants, radiological images, and clinical parameters.

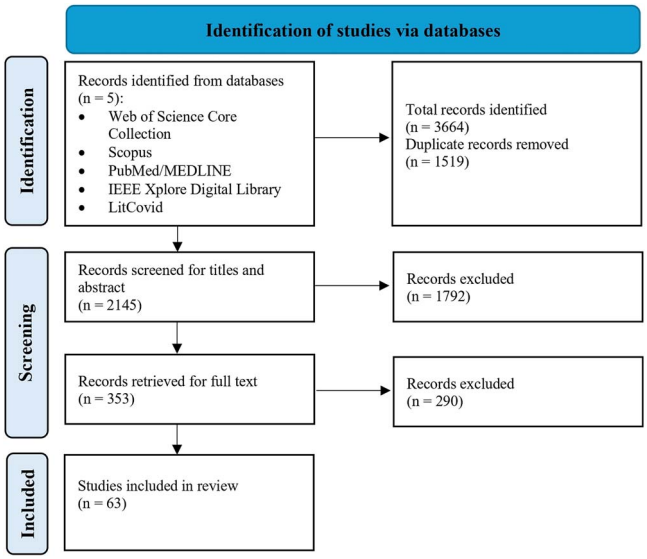

Figure 1 Selection process for literature in this review.

4) Studies included review articles, editorials, or opinion pieces without original data.
5) Studies where full text could not be retrieved.
6) Studies lacked sufficient methodological detail, such as algorithm parameters, training or testing splits, and feature selection methods, for reproducibility assessment.

Screening process

After deduplication, 2145 unique records were subjected to title and abstract screening using the predefined inclusion and exclusion criteria, resulting in 353 studies selected for full-text review. Full-text articles were obtained for all studies meeting the inclusion criteria or where eligibility was unclear from abstract alone, so that it can be assessed again for final inclusion. All full-text articles were assessed for final inclusion. Figure 1 shows the selection process.

Data extraction, synthesis, and analysis

Data were extracted from 63 selected studies. These review findings were synthesized narratively and divided into the following key sections: (i) The Landscape of AI-Driven COVID-19 Transcriptomics, (ii) Limitations of Studies, (iii) A Proposed AI-Driven Transcriptomics Framework, and (iv) Clinical Translation Challenges, Opportunities, and Future Directions.

Results and discussion

The landscape of AI-driven COVID-19 transcriptomics

These studies were published between 2021 and 2025, reflecting the influx of research a year into the pandemic. The predominant transcriptomic datasets, which included RNA-Seq, scRNA-Seq, microarray, and miRNA-Seq datasets, were obtained from databases such as Gene Expression Omnibus (GEO), ArrayExpress, Sequence Read Archive, Genome Sequence Archive (GSA), and CNGB Nucleotide Sequence Archive. Studies analyzed respiratory specimens, whole blood, peripheral blood mononuclear cells, plasma, bronchoalveolar lavage fluid, tissues, and cell lines.

There was a shift in the trends of AI methodologies across the years. In 2021, many studies focused on the prediction of hub genes as potential biomarkers of COVID-19 for diagnosis and patient stratification [22–30]. These studies established the initial, broad candidate biomarkers and pathways of COVID-19, where a majority relied on bulk RNA-Seq datasets. Methodologies were relatively simple, commonly utilizing principal component analysis for dimensionality reduction and paired with foundational classifiers.

In the following years, while there were studies prevalent in identifying broad candidate biomarkers of COVID-19, certain studies had transitioned to refine the discovery of minimal two to three gene signatures [31–33]. Studies became more specialized, highlighting the critical role of different immune cell types in COVID-19 pathogenesis, such as T cells, macrophages or neutrophils, and miRNAs, which utilized scRNA-Seq datasets [34–40]. Instead of obtaining only general candidate biomarkers, studies also investigated specific cell death mechanisms like cuproptosis [41], ferroptosis [42], and immunogenic cell death in COVID-19 [43, 44]. A diverse use of multiple classifiers, including Random Forest (RF), Support Vector Machine (SVM), and gradient boosting variants such as Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost) were the most common, alongside integration of other feature selection, feature extraction, and Explainable AI (XAI) methods.

In addition to the mentioned studies, recent studies reported the involvement in therapeutic applications such as drug target identification and phytomedicine [45, 46]. An emergence of specific tools was identified, where deep learning was dominant [47–51]. The most recent trend showed the maturity of graph learning frameworks with detailed biological knowledge, moving on to provide more translatable tools. Thus, AI models were observed to be applied mainly in domains including the diagnosis of COVID-19, the differentiation of SARS-CoV-2 from other diseases, and patient stratification by severity (Supplementary Tables S2 to S4), while novel tools aimed to address and solve specific problems in data analysis, interpretation, and prediction within the context of COVID-19 (Supplementary Table S5).

These applications of AI for transcriptomic data can be grouped into a few primary functions. From classical ML algorithms to advanced DL algorithms, it can be grouped into methods used for dimensionality reduction, classification, prediction, and XAI. An overview of the methodology used in these studies can be summarized in Fig. 2.

Limitations of studies

The assessment and evaluation of studies revealed major issues that affect the clinical utility of many AI-driven COVID-19 transcriptomics studies. Limitations were related to the data and samples, as well as methodological flaws.

Data and sample-related limitations

In AI studies, a high-quality dataset that considers completeness, accuracy, timeliness, and representativeness is fundamental, especially in clinical research [52]. The model's output reflects the input data used to train the model. Many studies also rely on public datasets, which may lack metadata or any other necessary information [31, 32, 53]. This limited control over research factors such as the sample collection, patient treatment, or other precise metadata. Moreover,

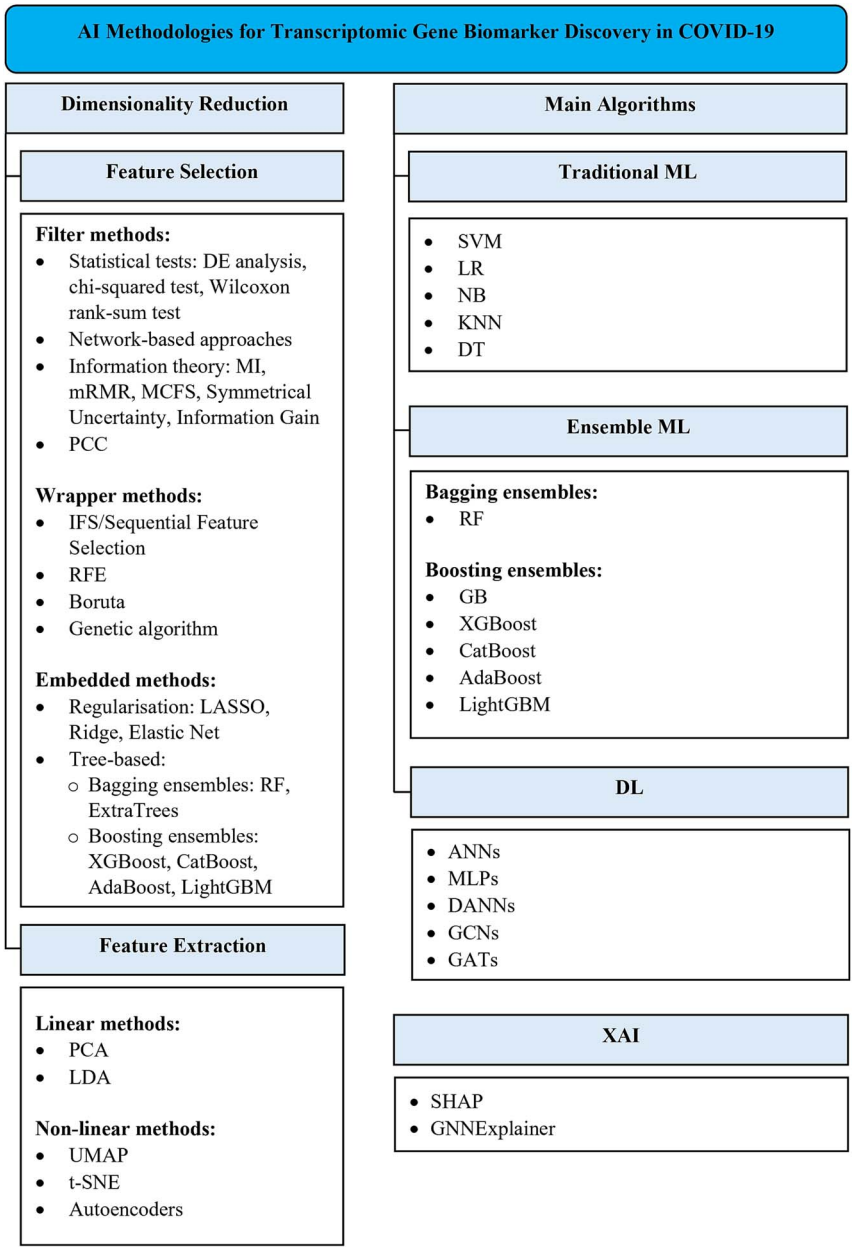

Figure 2 A summary of AI methodologies in this review. Overall methodologies were grouped into dimensionality reduction, main algorithms including ML and DL, as well as XAI methods. *Note:* DE analysis: differential expression analysis; MI: mutual information; mRMR: minimum-redundancy maximum-relevancy; MCFS: Monte Carlo feature selection; PCC: Pearson correlation coefficient; IFS: incremental feature selection; RFE: recursive feature elimination; LASSO: least absolute shrinkage and selection operator; RF: random forest; XGBoost: extreme gradient boosting; CatBoost: categorical boosting; AdaBoost: adaptive boosting; LightGBM: light gradient boosting machine; PCA: principal component analysis; LDA: linear discriminant analysis; UMAP: uniform manifold approximation and projection; t-SNE: t-distributed stochastic neighbor embedding; SVM: support vector machine; LR: logistic regression; NB: naïve bayes; KNN: k-nearest neighbor; DT: decision tree; GB: gradient boosting; ANNs: artificial neural networks; MLPs: multilayer perceptron; DANNs: deep artificial neural networks; GCNs: graph convolution networks; GATs: graph attention network; SHAP: SHapley additive exPlanations.

without accessible information on confounding factors such as age, sex, and comorbidities, the reliability of resulting model outputs may potentially be biased [35].

A major consequence of the reliance on public datasets was the inconsistencies in metadata, particularly for patient stratification [54]. Some studies classified severity based on their qualitative disease presentations, ranging from asymptomatic, mild to severe, or acute and convalescent groups [24, 33], while others utilized outcome-based metrics such as hospital-free days and intensive care unit (ICU) admission [41, 55]. Studies also employed standardized scoring systems such as the World Health Organisation (WHO) Ordinal Scale; Sequential Organ Failure Assessment; and Acute Physiology, Age, and Chronic Health Evaluation (APACHE II) Scores [28, 41, 56, 57]. This inconsistency emphasized the need for standardized definitions to address the complexity and heterogeneity of the disease and allow for cross-study comparison.

Beyond metadata, statistical and technical challenges, such as sample size and data heterogeneity, can affect a model's performance. Among the bulk RNA-Seq studies with reported patient-level sample size, nine studies relied on training their models with small sample numbers below 100 patients [22, 24, 30, 31, 45, 58–61]. With a high feature-to-sample ratio, the risk of overfitting was significantly increased, leading to the model learning the noise rather than true biological signals and having poor performance on new data [1, 62]. This risk is applicable to DL architectures with millions of trainable parameters applied to small sample size [27, 63], as models may memorize the noise structure of a small training set, rather than abstracting generalizable disease biology. With the sparse, heterogeneous nature of single-cell data, it is difficult to achieve a balance between model complexity and the limited data [64]. Critically, dataset size also affects the interpretation of algorithm performance across studies, as model performance can inherently improve when the number of samples increases [65, 66].

Besides, data heterogeneity can introduce batch effects due to the integration of data from multiple sources or different technical platforms [37, 51, 65, 67]. Models trained on one specific tissue, platform, or parameter often fail to generalize across different datasets, indicating poor robustness [65]. There is also a risk of conflating technical batch effects with true biological variance, as models trained on specific cell types rarely generalize to other lineages [51].

To address data quality issues, computational strategies were employed. To fill missing values, data imputation strategies included Multiple Imputation by Chained Equations (MICE) for clinical data [57, 68]. A study by Hausmann *et al.* [51] introduced a novel method, DISCERN, which addressed problems of these traditional imputation methods in scRNA-Seq through novel expression reconstruction. In this study, they compared DISCERN to traditional methods for scRNA-Seq, like Deep Count Autoencoder (DCA), Markov Affinity-based Graph Imputation of Cells (MAGIC), scImpute, DeepImpute, and CarDEC.

Similarly, to handle data imbalance, data augmentation methods, such as Synthetic Minority Over-sampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), and random resampling, were employed [36, 38, 56, 62, 65, 69, 70]. For example, a study by Sethi *et al.* [70] used ADASYN to oversample minority classes for them to be balanced, resulting in a significant increase in accuracy from 40% to 95%. Class weighting was also frequently implemented by assigning higher importance to correctly identify minority samples during training and avoid bias [63, 71]. This is crucial as an imbalance in datasets makes it difficult for models to generalize effectively [26, 36, 38–40, 60, 62, 63, 65, 69, 70, 72–76]. As datasets are skewed to a small minority class compared to a large majority class, this will cause models to be biased towards the majority class.

Multicohort meta-analysis can overcome the sample size problems but risks the introduction of batch effects. Thus, batch effect correction is necessary to ensure that ML models capture true biological signals rather than technical noise, especially due to differences in sequencing platforms, laboratories, and experimental protocols. For instance, studies utilized ComBat-Seq and limma to correct for technical variability before the integration of RNA-Seq count data from different cohorts before performing ML [24, 37, 74]. For scRNA-Seq, Seurat CCA was often used to align samples from different patients [39, 63, 73]. Implementing these corrections is important as prediction accuracy can significantly increase, such as in Xie *et al.* [71], which increased from 0.67 to 1.00, effectively improving cross-dataset prediction accuracy using their scPanel tool.

Although pre-processing techniques such as data imputation, augmentation, and batch effect correction were necessary to enhance data quality, these strategies risk introducing technical artifacts or discarding valuable biological data. For instance, data imputation that addresses missing data can potentially generate artificial correlations that do not exist biologically, while aggressive batch correction can remove true biological signals when batch and biological variables are confounded [19, 77]. Thus, the application of these data refinement strategies should be considered carefully. Ultimately, this highlights the need for high-quality raw data as these computational techniques, while useful, cannot replace it.

Reproducibility remains a concern in AI-driven transcriptomic studies. Most of the reviewed studies did not provide publicly accessible code repositories, and only certain studies provide their codes in GitHub or Zenodo (Supplementary Table S6). Therefore, unavailability of code and data affects the verification of findings and benchmarking, especially cross-institutionally. Simply making code publicly available does not make it reproducible, as model non-determinism, data variations and pre-processing, and computational challenges affect reproducibility as well [78]. While fields such as computer vision have established benchmarks for fair algorithm comparison, COVID-19 transcriptomic studies employ heterogeneous datasets with different sample sizes, pre-processing pipelines, and class definitions [79, 80]. By developing version-controlled benchmark datasets, encompassing diverse tissue types, sequencing platforms, and severity definitions, it would overcome issues in direct performance comparison.

Furthermore, Kapoor and Narayanan [81] also reported the reproducibility crisis caused by code and data availability, data leakage, lack of standard reporting practices, variable data quality, and evaluation metrics choices. To enhance reproducibility, future studies should ensure data transparency and integrity, adopt reporting standards adapted from frameworks such as TRIPOD-AI, and employ version control for code or containerized computational environments like Docker and Singularity to ensure consistent execution across institutions [81–83].

Methodological limitations

A significant challenge in current AI research is the model's generalizability and robustness on other new datasets. Studies that rely on internal validation without testing on an independent validation cohort reported results that may not be reflective of real-world performance [68, 84, 85]. Notably, several studies that relied only on internal cross-validation (CV) had reported near-perfect performance metrics, reflecting overfitting to the training data rather than genuine predictive ability. Furthermore, several researchers underscored the importance of *in vivo* experimental validation, which they did not perform in their own studies [22, 72, 75]. The majority of biomarkers identified across the 43 biomarker-focused studies in Supplementary Table S6 remain exploratory, with only a few studies, such as Albright *et al.* [32] and Carapito *et al.* [61], achieving both external and experimental validation of their gene signatures. This highlights a significant gap between biomarker discovery and clinical translation in the current literature.

Among the reviewed studies, there are methodological patterns that suggest risk of data leakage or circular validation in several studies. Data leakage can happen when the training dataset is not separated from the test dataset during pre-processing, modeling, and evaluation [81]. Performing feature selection using the entire dataset before splitting into training and validation sets may introduce test-set

information into the selected features. For example, there are studies that first conduct differential expression analysis across all samples and then train classifiers on identified genes and evaluate performance using CV on the same data [22, 58, 60, 65]. Such workflows can produce optimistically biased results due to the failure to separate discovery and validation datasets. Consequently, the high performance frequently reported in reviewed studies may be partially attributable to methodological artifacts rather than genuine predictive ability.

While the reviewed literature may share similar computational pipelines, it is critical to distinguish between predictive modeling, biomarker discovery, and clinical translation. Certain studies that were obtained from our search strategy are primarily oriented towards predictive modeling, focusing on optimizing classification accuracy and selecting gene features based on statistical discriminative power, rather than biological function [58, 60]. In contrast, true biomarker discovery aims to identify genes with mechanistic relevance to disease pathogenesis, where AI serves to verify biologically driven hypotheses rather than to maximize classifier performance. It is important to note that high statistical significance does not equate to biological causality or mechanistic plausibility [65].

Furthermore, computationally identifying a biological driver does not guarantee its viability as a clinical biomarker. Clinical translation further requires practical deployability, including robust, real-world assay validation, as demonstrated by the reverse transcription quantitative polymerase chain reaction (RT-qPCR) assay developed by Albright *et al.* [32]. Thus, to provide clarity, Supplementary Table S6 documents experimental validation status, allowing the differentiation of exploratory computational findings from experimentally confirmed and validated biomarkers, as well as help identify studies at the risk of overfitting or data leakage.

In addition to validation gaps, some studies have discussed the black-box nature of models, therefore making interpretability and obtaining biological insights difficult [39, 57, 63]. As years progress, the practical applicability and interpretability of these models have been increasingly utilizing XAI to overcome the black-box nature. The most common XAI in this review, SHapley Additive exPlanations (SHAP), was utilized to explain the output of ML models by quantifying how much each feature contributed to a specific prediction [70, 86]. Additionally, DL interpretability relies on methods like GNNExplainer, designed for Graph Neural Networks (GNNs) as a model-agnostic approach in identifying the most influential subgraph and features that contribute to GNN's prediction [27, 63, 87]. XAI methods, including the Local Interpretable Model-Agnostic Explanations (LIME) that was not found in this review, play a part in interpreting model predictions and identifying the contribution of specific features.

Beyond overfitting and interpretability limitations, DL models, due to their millions of trainable parameters, are particularly sensitive to hyperparameter choices such as learning rate, decay, and dropout rate [88]. Nevertheless, many reviewed studies on DL do not report systematic hyperparameter tuning or sensitivity analyses. Furthermore, model stability remained a challenge. Based on the specific conditions for training of the model, models may not be stable when tested under different circumstances, such as in Ma *et al.* [89]. Additionally, the practical application of certain tools is restricted by requirements. For example, DL models such as moETM require data types to be from the same cells, while DeepDRIM requires a pre-existing set of validated interactions [50, 90]. This highlighted the inflexibility of application of such tools in heterogeneous research settings. Thus, these methodological limitations highlight the need for more rigorous and standardized evaluation frameworks before AI-derived transcriptomic biomarkers can be translated into clinical applications.

A proposed AI-driven transcriptomics framework

In AI workflows, data collection and pre-processing are the foundational steps to ensuring the quality of subsequent analyses. Essential pre-processing steps in AI-driven COVID-19 transcriptomic studies included quality control, feature scaling such as normalization and standardization, and batch effect correction. For instance, quality control involved assessing and removing low-quality elements in raw read data to prevent noise or bias, such as using FastQC, Fastp, Trimmomatic, HTQC, and QTrim [33, 60, 72, 84]. After obtaining gene expression data, low expressed genes and outliers should be filtered to enhance statistical power [61, 63, 84].

Feature scaling is a very crucial step in accounting for technical variations to make it comparable across samples and datasets. Methods of normalization based on data technology and downstream analyses included simple scaling like transcripts per million and fragments per kilobase million, statistical modeling for count data like scTransform, edgeR, and DESeq2, quantile normalization, reference gene normalization, as well as variance and scale transformation like Variance Stabilisation (VST) and StandardScaler [33, 39, 68, 74, 91]. For scale-sensitive ML models, *Z*-score standardization is a common method of pre-processing data [48, 51, 67, 70].

Batch effect correction should be conducted to adjust data for the systematic, non-biological variations introduced. Depending on the data type, there are different methods such as ComBat, ComBat-seq, or limma::removeBatchEffect for bulk RNA-Seq and microarrays, and Seurat CCA for scRNA-Seq [39, 71, 74, 91]. Before feature selection, data imputation to impute missing values and address class imbalances was required. Common methods to impute missing data included using MICE for clinical metadata [57, 68], while scRNA-Seq may use methods like DCA, MAGIC, scImpute, DeepImpute, and CarDEC [51, 92].

In addressing the high dimensionality of data, feature selection is an important step to keep useful features, improving training efficiency and reliability of models [89]. Most studies did not rely on a single method, rather used multiple methods by filtering genes initially using differential expression analysis before refining the genes further using wrapper and embedded methods. To ensure biologically relevant genes, network or functional enrichment analysis tools were used to identify both significant and biologically insightful genes. By conducting these pre-processing methods, clean, comparable, and relevant genes can be obtained to ensure high model performance and interpretability. Before the selection of models, data augmentation (as discussed in the "Data and sample-related limitations" section) should be conducted to address the class imbalance to prevent bias, with strategies such as SMOTE, ADASYN, and random resampling applied within CV folds to prevent data leakage [36, 56, 62, 69, 70].

Next, the selection of these algorithms is primarily driven by identifying the research question, whether to classify, predict, or uncover new hidden patterns. Sample sizes can also affect the models [66]. Hence, choosing models based on sample size is ideal, such as using more complex models including DL for larger sample sizes, while using simpler models like RF and SVM for smaller sample sizes. This can improve performance, generalizability, and reliability. As no specific algorithm is inherently better than another, a few different

models should be used to compare their performance and validate the final selection of the model.

Despite the many algorithms used in this review, we found certain algorithms that were frequently reported as high-performing models in their respective studies such as ensemble methods, XGBoost and RF, as well as the traditional model, SVM (Supplementary Table S6). Due to heterogeneity in datasets, sample sizes, pre-processing pipelines, class definitions, and evaluation metrics, direct quantitative comparison of algorithm performance across studies is constrained. Hence, the observations for model performances reported in this review reflect the frequency and study-level findings, rather than inherent algorithm superiority. It should also be noted that publication and reporting biases may favor the presentation of positive results, so the frequency with which certain models appear as high performing may not fully reflect their comparative advantage across all experimental conditions.

XGBoost was consistently found to be a top performer in different studies. Its success in severity prediction and identifying diagnostic genes was demonstrated in their high performance [29, 45, 70]. Some observations from these studies included the use of XGBoost for feature selection with other methods like Incremental Feature Selection (IFS) before using KNN as final classifier [75], the incorporation of Least Absolute Shrinkage and Selection Operator (LASSO) regularization to prevent overfitting [62, 70], and interpretability with SHAP [29, 65, 70].

Another frequently reported high-performing model in multiple studies was RF, which was applied on a variety of transcriptomic data, including bulk RNA-Seq, scRNA-Seq, TCR, and miRNA [54, 56, 63, 72]. Its performance showed improvement, especially after feature selection. For instance, Bao *et al.* [62] achieved almost perfect scores on internal validation of 10-fold CV after pre-ranking with algorithms like CatBoost, XGBoost, and LASSO; Li *et al.* [40] used RF which yielded higher accuracy than Decision Tree (DT) after IFS; and Maleknia *et al.* [74] achieved higher accuracy after LASSO in their respective studies. Furthermore, traditional models like SVM were often reported as reliable models with high performance, especially with the use of kernels [22, 60]. SVM, when paired with Recursive Feature Elimination, is a known wrapper method, effective in identifying relevant gene signatures with discriminative power [37, 43, 71].

On the other hand, DL models are highly specialized tools designed to solve biological challenges in complex, unstructured data. DL models, particularly GNNs, achieved good performance in specific tasks such as Graph Attention Network in a study by Li *et al.* [63], learning important connections in a network and Graph Convolutional Neural Network in Flores *et al.* [39], incorporating graphs of known gene–gene interactions. Some developments in DL can be observed to include multi-omics analysis and unified frameworks starting from data cleaning and batch correction to prediction in a single model [50].

The choice of CV for model evaluation depends on the data size and the objectives of the study. The most common CV methods, including the train-test splits, also known as holdout validation, are used for large datasets while k-fold CV is more robust with its multiple splits in smaller datasets [93]. Most studies employed 5- and 10-fold CV for model tuning. Another CV method included the leave-one-out CV [30, 58]. To overcome overfitting and data leakage, choosing the correct technique is crucial. Performance metrics, which commonly include accuracy, precision, recall, specificity, F1-score, Matthews Correlation Coefficient, and AUC, can be used to evaluate the performance. Previous mentioned feature selection methods are recommended to be conducted within these CV processes on training datasets to prevent data leakage and biases to model performance.

In the validation of AI models, two commonly used approaches were external validation using independent datasets and experimental validation with wet lab. External validation on independent datasets remains crucial to prevent overfitting as well as ensure the generalizability and reliability of models for translatable clinical use. Meanwhile, experimental validation with wet lab, such as using qPCR, should be conducted to confirm the key hub genes found. Besides, a prediction is not understandable without an explanation. Thus, methods such as XAI and functional enrichment analysis can confirm the biological meaning of the candidate biomarker genes obtained. Both approaches can further strengthen the transition of computationally identified candidate biomarkers using AI models to biologically validated biomarkers, after experimental validation.

Therefore, in addition to the choice of algorithm, the overall methodological workflow, including the pre-processing and quality control of datasets, was crucial in AI COVID-19 transcriptomic studies. The combination of the integration of different types of data, with the selection of robust feature selection, AI models, metrics, and validation strategies, can significantly enhance the identification of gene biomarkers. The overall workflow can be summarized in Fig. 3. However, it is crucial to acknowledge that this framework faces several real-world constraints and potential failure cases. In early stages of emerging pandemics, sample sizes are usually limited, when biomarker discovery is most critical. Under conditions of limited sample size, classical ML and complex DL models are prone to overfitting as they capture the stochastic noise, resulting in biased and overoptimistic performance metrics that fail to generalize to external, real-world datasets [66].

During the pandemic, metadata that is incomplete or inconsistent, common in public datasets, introduced confounders and affected model reliability [35]. Additionally, while batch effect correction is essential, caution must be taken as it can overcorrect to remove genuine biological signals, when applied aggressively or when batch and biological variables are confounded [19, 77]. Over-sampling techniques like SMOTE can also introduce data leakage and inflate performance metrics when applied before CV splitting [81, 94].

Rapid viral evolution, such as in SARS-CoV-2, can induce domain shift, where models trained on one variant may be unreliable and fail to generalize to other variants with distinct transcriptomic profiles [95]. Moreover, in resource-limited settings where computational infrastructure needed for more complex processes like DL models is unavailable, simpler models with pre-selected genes must be prioritized [96]. Nevertheless, emerging systems, such as agentic AI discussed in the “Opportunities and future directions” section, can offer potential solutions.

Clinical translation challenges, opportunities, and future directions

Challenges in clinical translation

The gap between algorithm performance on transcriptomic data in research settings and real-life clinical implementations remains critical. A real-world, multicentre clinical trial needs substantial funding, time, and coordination. Firstly, employing these AI models in clinical settings requires large technical infrastructure demands and computational resources [96]. For instance, complex graph-based or DL algorithms, such as DeepDRIM, can require up to 51GB of memory and 11

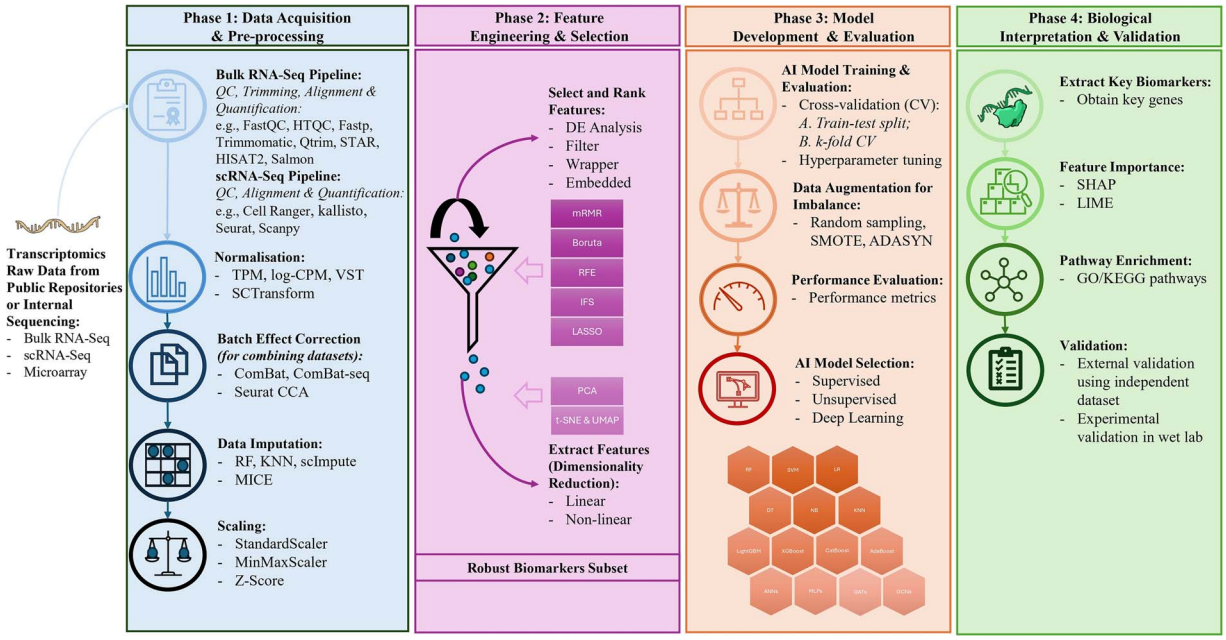

Figure 3 A proposed AI-driven transcriptomics framework in general. *Note:* TPM: transcripts per million; log-CPM: log counts per million; VST: variance stabilizing transformation; CCA: canonical correlation analysis; RF: random forest; KNN: k-nearest neighbor; MICE: multivariate imputation by chained equations; DE: differential expression; SMOTE: synthetic minority over-sampling technique; ADASYN: adaptive synthetic sampling; SHAP: SHapley Additive exPlanations; LIME: local interpretable model-agnostic explanations; GO: Gene Ontology; KEGG: Kyoto Encyclopaedia of Genes and Genomes.

GPU hours to evaluate a single dataset [90]. Predictive models often fail when met with real-world data sparsity, leading to overfitting. Hence, the deployment of these biomarkers in a clinical setting requires reproducible, predictable performance, which cannot be guaranteed when input data are inherently sparse, inconsistently imputed, and compromised by institution-specific batch effects.

High-throughput technologies are costly and time consuming for routine, rapid patient care, causing a need for standardized, cost-effective tests like multiplex RT-qPCR and flow cytometry for clinical utility [32, 51, 71]. However, transitioning to these platforms would often disrupt the stability of targeted gene measurements, hence, obstructing commercial scalability in these diagnostic tools [89]. Among the reviewed studies, only Albright *et al.* [32] developed a clinically deployable RT-qPCR assay from their computationally identified two-gene host-response signature, highlighting the gap between computational discovery and assay development.

Clinical deployment also involves technical and logistical challenges, such as interfacing with various electronic health record systems, ensuring data privacy and security in compliance with Health Insurance Portability and Accountability Act and General Data Protection Regulation, and providing outputs to clinicians in a simple, actionable format [97, 98]. These challenges are applicable to transcriptomic data as well, which contains identifiable genetic information that requires additional security beyond standard clinical data protections. Besides, ethical hurdles and regulations would require extensive processes for approval, as it involves the use of patient data, raising ethical questions on consent and data ownership [97]. Interpretability limitations further add to these issues, as the black-box nature of the model weakens clinician trust and hinders regulatory approval.

For routine clinical use, a rapid and cost-effective assay must be developed. Depending on their intended clinical application, translational feasibility of these assays can differ. Diagnostic assays that detect COVID-19 through gene signatures, such as the two-gene signature in Albright *et al.* [32], will require analytical validation, clinical validation in prospective multicentre trials, and regulatory submission through pathways such as the U.S. Food and Drug Administration 510(k), European Union's In Vitro Diagnostic Regulation, or WHO Prequalification [99]. For prognostic severity stratification tools that predict patient outcomes like ICU admission models by Bello *et al.* [57], a pathway such as clinical decision support software under the International Medical Device Regulators Forum framework for Software as a Medical Device (SaMD) is applicable, as they inform rather than replace clinical judgement [100, 101].

Thus, a concrete roadmap for clinical translation would include (i) the discovery and internal validation of a minimal gene signature with biological significance; (ii) external validation across independent, multicentre cohorts with diverse demographics; (iii) assay development by transitioning from RNA-Seq to clinically deployable platforms such as multiplex RT-qPCR; (iv) prospective clinical validation through clinical trials evaluating clinical utility and patient outcomes; and (v) regulatory submission with accompanying evidence of analytical performance, clinical performance, and software validation for any integrated AI decision support. The reviewed literature remains largely at the discovery phase, with very few studies progressing further, underscoring the significant translational gap.

The final goal is the development of point-of-care testing that integrates assay testing and AI analysis into a single, rapid workflow [102]. The COVID-19 pandemic has accelerated research and development on AI-driven transcriptomic biomarker discovery pipelines that can be studied for future emerging pathogens. Key lessons included the importance of standardized RNA processing protocols, data annotation standardization, and data sharing principles to enable collaboration for transcriptomic research in response to the pandemic.

Opportunities and future directions

Currently, AI trends are rapidly advancing, with an increase in applications to the biological fields, including transcriptomics. There has been a shift from passive analytical pipelines to autonomous agentic workflows. Agentic AI utilizes reinforcement learning, goal-oriented architecture, and adaptive control mechanisms to complete complex goals, with minimal human intervention [103]. A key difference from the conventional AI tools is the autonomous pursuit of research objectives through iterative planning, tool invocation, code generation, and feedback-driven refinement, rather than discrete prompts.

A prominent example is a genomic multi-agent system, known as GenoMAS, which will allow for autonomous processing of gene expression analysis using six distinct agents with orchestration, programming, and advisory roles [104]. Their research showed that GenoMAS was successful in executing end-to-end data pre-processing and statistical analysis, discovering biological relationships on the GenoTEX benchmark. GenoMAS achieved a Composite Similarity Correlation of 89.13% for data pre-processing and an F1 of 60.48% for gene identification, surpassing the best prior methods by 10.61% and 16.85%, respectively.

Similarly, for broader omics data integration, AutoBA has introduced the automation for data analyses, code generation, and execution based on the input data, requiring only data path, description, and analysis goal [105]. Its robustness and adaptability have been validated across diverse omics data, including RNA-Seq, scRNA-Seq, chromatin immunoprecipitation sequencing, and spatial transcriptomics in 40 real-world cases. Additionally, the Agentomics framework has emerged as a specialized solution for the autonomous ML lifecycle, focusing on the rigorous optimization of data splitting and hyperparameter tuning specifically for transcriptomic variables [106]. Meanwhile, instead of large language models (LLMs), BioAgents offers an alternative, and was built on small language models, enhanced with retrieval-augmented generation. Performance on conceptual genomics tasks was found to be comparable to that of human experts, suggesting that resource-efficient, domain-adapted architectures may offer a practical choice, where computational infrastructure or data privacy constrains the use of larger models.

At the level of single-cell and spatial transcriptomics, CellAgent performs end-to-end data analysis through natural language interactions, employing a multi-agent framework that simulates a deep-thinking workflow to ensure each analytical step remains consistent with the overall task objective [107]. Against human experts, CellAgent achieved an improvement in efficiency while maintaining accuracy comparable to existing approaches. For spatial transcriptomics specifically, STAgent is an autonomous multimodal agentic AI integrating multimodal LLMs with specialized computational tools, capable of performing expert spatial analysis in minutes [108].

The use of agentic AI also extends to autonomous data mining and knowledge synthesis. With big data availability, re-analysis of these data could provide insight into molecular signatures beyond the scope of their original study. For instance, in a specific use of mining existing scientific literature, ClockBase Agent uses specialized AI agents to reanalyze human and mouse methylation and RNA-Seq samples with 40 ageing clock predictions, to generate ageing biomarkers and interventions for ageing [109]. Similarly, Paper2Agent addresses challenges of adapting a paper's code, data, and methods by converting the paper into an AI agent acting as a knowledgeable research assistant [110]. Applying Paper2Agent to AlphaGenome, TISSUE, and Scanpy, the agents were successful in transforming research papers and their codebase, constructing a Model Context Protocol server into interactive chat agent via natural language queries. Cell Atria also enables automation of literature-driven metadata extraction and dataset retrieval to standardized scRNA-Seq analysis through containerized pipelines [111].

These systems illustrate the value of the agentic AI in transcriptomics, hence, potentially being the solution for current constraints of AI in transcriptomics. Whereas earlier generations of AI in transcriptomics were largely confined to supervised classification, the transition towards autonomous multi-agent systems enables the orchestration of nonlinear, complex workflows. In the context of evolving SARS-CoV-2, agents such as GenoMAS, AutoBA, and CellAgent provide a framework for real-time molecular surveillance. They are highly analytical, capable of processing data, analyzing data, linking biological relevance, to generating reports autonomously. With minimal user input and performance on standardized benchmarks that meet or exceed manually curated pipelines, these efficiency improvements demonstrate that well-architected multi-agent systems can help streamline workflows and enhance productivity. Furthermore, these systems prevent data leakage and biomarker prediction conflation by enforcing a guided-planning framework, as seen in GenoMAS, where the isolated agent roles ensure that feature selection remains strictly within CV loops.

For instance, CellAtria autonomously ingests literature from PDFs or URLs to extract structured metadata, including sample annotations and accession identifiers from GEO [111]. Agentomics introduced the autonomous splitting and optimizing of train-validation data, as highlighted in the high validation-test score correlation, to ensure models trained on small datasets generalize effectively [106]. CellAgent employs the Evaluator Agent to evaluate and produce the best batch-correction result through iterative optimization [107]. Ten different metrics were employed to quantify batch removal, while preserving biological variation, ensuring that domain shift does not obscure true signal.

Despite these advances, several gaps remain that limit the translational utility of current agentic systems. An end-to-end closed loop system, incorporating processes of interpretability, designing validation experiments, and revising its model accordingly to experimental feedback, would transform the application of agentic AI. The current pipeline remains open ended, without agents being able to revise models accordingly due to a lack of intervention to bridge computational outputs to physical experiments. This is a challenge that requires interdisciplinary collaboration to address.

In addition, there is a gap in inferring causal versus correlational relationships, important for domains like drug development. For discovery of biomarkers such as in COVID-19, a gene whose expression co-varies with disease severity may reflect downstream immunopathology rather than upstream driver, making it a poor target of therapeutic intervention. Hence, combining planning capabilities of agentic AI with causal AI can be a high-potential direction [112], particularly for drug target discovery workflows, where correlation and causation differences has direct translational effects. Diseases like COVID-19 have complex pathophysiology; thus, the development of agentic systems capable of harmonizing multi-omics data, including genomics, transcriptomics, proteomics, and metabolomics, would be a major scientific opportunity. These datasets are often characterized by high dimensionality and heterogeneity, representing a major unresolved integration challenge.

Research on transfer learning approaches must be done for future pandemic preparedness. SARS-CoV-2 continues to evolve and the next pandemic pathogen will require rapid, *de novo* biomarker discovery with limited data, where transfer learning would excel in. Future work should test the transferability of COVID-19 transcriptomic models to other variants or respiratory pathogens. For instance, models like Geneformer pre-trained on transcriptomic data for network biology predictions, which can be fine-tuned for specific tasks [113]. Ultimately, future research should move towards directions outlined above, where the development of AI systems that do not merely assist but actively participate in it should be conducted. Agentic AI offers promise in bridging the gap in transforming transcriptomic biomarker discovery, aligning its technical capabilities with biological and clinical evidence. A shift to prospective validation of these transcriptomic biomarkers, comprising various healthcare systems worldwide, should be carried out to prove the model's utility. Translation of these validated biomarkers to the clinical setting through a simple, rapid, and cost-effective test would be the aim for progressing towards the future of personalized medicine, improving patient outcomes and recovery.

Conclusion

In conclusion, the integration of AI with transcriptomic analysis in studying the complex COVID-19 disease has become more common with the abundant studies reflected in this review. A comprehensive review of these scientific literature reveals the various AI methodologies applied in COVID-19 transcriptomic studies, which mainly focus on diagnosis and disease severity stratification, but are not limited to prognosis, risk stratification, and identifying drug targets. Our understanding of the underlying mechanism of SARS-CoV-2 and the identification of potential candidate biomarkers can be improved through these AI workflows. By using AI on transcriptomic data, large volumes of complex data can be more efficiently handled, accelerating candidate gene biomarker discovery, especially in heterogeneous diseases like COVID-19.

Key challenges to data quality, sample heterogeneity, and methodology can be addressed by well-structured workflows. These workflows can address the problem of extreme dimensionality through robust feature selection and extraction techniques to identify the best set of genes, while correcting for technical variations from multiple sources as needed using batch-correction algorithms. Algorithm selection should be guided by sample size and research objectives, with data augmentation strategies applied strictly within CV folds to prevent data leakage. The performance of models should be evaluated by multi-metric validation frameworks. The identification of specific molecular signatures in consistently high-performing models across diverse datasets, together with XAI techniques, can support the computational findings and improve their interpretability before experimentally validating the findings.

Looking beyond model performance, the translational gap between biomarker discovery and clinical deployment should be addressed. The emergence of agentic AI systems provides significant potential for addressing current methodological constraints in AI-driven transcriptomics, offering autonomous orchestration of complex analytical workflows within predefined, reproducible frameworks that structurally reduce risks of data leakage, inconsistent validation, and analytical bias. By employing a standardized workflow, validating models across various datasets and experimentally, as well as integrating explainability, the development of reliable, validated biomarkers can be accelerated, leading to rapid, cost-effective clinical applications, which is a step forward to precision medicine in infectious diseases. This review offers a critical step towards pandemic preparedness in the future.

Author contributions

Funding acquisition was secured by S.K.D. The conceptualization, data curation, formal analysis, writing of original draft, and visualization were conducted by L.Y.K., with supervision and validation by S.K.D. All the authors were involved in reviewing and in the final editing of the manuscript.

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Conflict of interest

None declared.

Funding

This research was funded by the Ministry of Higher Education, Malaysia, under the Fundamental Research Grant Scheme (FRGS) (FP019-2022).

Data availability

Not applicable.

Ethics approval and consent to participate

Not applicable.

References

1. Clarke R, Ressom HW, Wang A *et al.* The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. *Nat Rev Cancer* 2008;**8**:37–49. https://doi.org/10.1038/nrc2294
2. Avsec Ž, Latysheva N, Cheng J *et al.* Advancing regulatory variant effect prediction with AlphaGenome. *Nature* 2026;**649**:1206–18. https://doi.org/10.1038/s41586-025-10014-0
3. Le NQK, Li W, Cao Y. Sequence-based prediction model of protein crystallization propensity using machine learning and two-level feature selection. *Brief Bioinform* 2023;**24**:24. https://doi.org/10.1093/bib/bbad319
4. Li Q, Hu Z, Wang Y *et al.* Progress and opportunities of foundation models in bioinformatics. *Brief Bioinform* 2024;**25**:25. https://doi.org/10.1093/bib/bbae548
5. Le NQK, Yapp EKY, Nagasundaram N *et al.* Computational identification of vesicular transport proteins from sequences using deep gated recurrent units architecture. *Comput Struct Biotechnol J* 2019;**17**:1245–54. https://doi.org/10.1016/j.csbj.2019.09.005
6. Sarker IH. AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems.

SN Comput Sci 2022;**3**:158. https://doi.org/10.1007/s42979-022-01043-x
7. Cheng Y, Xu S-M, Santucci K *et al.* Machine learning and related approaches in transcriptomics. *Biochem Biophys Res Commun* 2024;**724**:150225. https://doi.org/10.1016/j.bbrc.2024.150225
8. Annan R, Qingge L. Artificial intelligence in COVID-19 research: a comprehensive survey of innovations, challenges, and future directions. *Comput Sci Rev* 2025;**57**:100751. https://doi.org/10.1016/j.cosrev.2025.100751
9. Lv C, Guo W, Yin X *et al.* Innovative applications of artificial intelligence during the COVID-19 pandemic. *Infect Med* 2024;**3**:100095. https://doi.org/10.1016/j.imj.2024.100095
10. Sekaran K, Gnanasambandan R, Thirunavukarasu R *et al.* A systematic review of artificial intelligence-based COVID-19 modeling on multimodal genetic information. *Prog Biophys Mol Biol* 2023;**179**:1–9. https://doi.org/10.1016/j.pbiomolbio.2023.02.003
11. Alballa N, Al-Turaiki I. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: a review. *Inform Med Unlocked* 2021;**24**:100564. https://doi.org/10.1016/j.imu.2021.100564
12. Syeda HB, Syed M, Sexton KW *et al.* Role of machine learning techniques to tackle the COVID-19 crisis: systematic review. *JMIR Med Inform* 2021;**9**:e23811. https://doi.org/10.2196/23811
13. Khan M, Mehran MT, Haq ZU *et al.* Applications of artificial intelligence in COVID-19 pandemic: a comprehensive review. *Expert Syst Appl* 2021;**185**:115695. https://doi.org/10.1016/j.eswa.2021.115695
14. Wang L, Zhang Y, Wang D *et al.* Artificial intelligence for COVID-19: a systematic review. *Front Med* 2021;**8**:8. https://doi.org/10.3389/fmed.2021.704256
15. Rasheed J, Jamil A, Hameed AA *et al.* A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. *Chaos, Solitons Fractals* 2020;**141**:110337. https://doi.org/10.1016/j.chaos.2020.110337
16. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. *Chaos, Solitons Fractals* 2020;**139**:110059. https://doi.org/10.1016/j.chaos.2020.110059
17. Jamshidi M, Lalbakhsh A, Talla J *et al.* Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment. *IEEE Access* 2020;**8**:109581–95. https://doi.org/10.1109/ACCESS.2020.3001973
18. Murdoch WJ, Singh C, Kumbier K *et al.* Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci* 2019;**116**:22071–80. https://doi.org/10.1073/pnas.1900654116
19. Goh WWB, Wang W, Wong L. Why batch effects matter in omics data, and how to avoid them. *Trends Biotechnol* 2017;**35**:498–507. https://doi.org/10.1016/j.tibtech.2017.02.012
20. Conesa A, Madrigal P, Tarazona S *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol* 2016;**17**:13. https://doi.org/10.1186/s13059-016-0881-8
21. Leek JT, Scharpf RB, Bravo HC *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010;**11**:733–9. https://doi.org/10.1038/nrg2825
22. Auwul MR, Rahman MR, Gov E *et al.* Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19. *Brief Bioinform* 2021;**22**:bbab120. https://doi.org/10.1093/bib/bbab120
23. Ng DL, Granados AC, Santos YA *et al.* A diagnostic host response biosignature for COVID-19 from RNA profiling of nasal swabs and blood. *Sci Adv* 2021;**7**:7. https://doi.org/10.1126/sciadv.abe5984
24. Liu Y, Wu YK, Liu B *et al.* Biomarkers and immune repertoire metrics identified by peripheral blood transcriptomic sequencing reveal the pathogenesis of COVID-19. *Front Immunol* 2021;**12**:677025. https://doi.org/10.3389/fimmu.2021.677025
25. Zhang S, Qu RL, Wang PY *et al.* Identification of novel COVID-19 biomarkers by multiple feature selection strategies. *Comput Math Methods Med* 2021;**2021**:1–8. https://doi.org/10.1155/2021/2203636
26. Zhang YH, Li H, Zeng T *et al.* Identifying transcriptomic signatures and rules for SARS-CoV-2 infection. *Front Cell Dev Biol* 2021;**8**:8. https://doi.org/10.3389/fcell.2020.627302
27. Sehanobish A, Ravindra N, Van Dijk D *et al.* Gaining insight into SARS-CoV-2 infection and COVID-19 severity using self-supervised edge features and graph neural networks. *Thirty-Fifth AAAI Conference on Artificial Intelligence, Thirty-Third Conference on Innovative Applications of Artificial Intelligence and the Eleventh Symposium on Educational Advances in Artificial Intelligence*. 2021;**35**:4864–73. https://doi.org/10.1609/aaai.v35i6.16619
28. Overmyer KA, Shishkova E, Miller IJ *et al.* Large-scale multi-omic analysis of COVID-19 severity. *Cell Syst* 2021;**12**:23–40.e7. https://doi.org/10.1016/j.cels.2020.10.003
29. Vázquez-Jiménez A, De León U, Matadamas-Guzman M *et al.* On deep landscape exploration of COVID-19 patients cells and severity markers. *Front Immunol* 2021;**12**:705646. https://doi.org/10.3389/fimmu.2021.705646
30. Zhou YG, Zhang JH, Wang DY *et al.* Profiling of the immune repertoire in COVID-19 patients with mild, severe, convalescent, or retesting-positive status. *J Autoimmun* 2021;**118**:102596. https://doi.org/10.1016/j.jaut.2021.102596
31. Lai GC, Liu H, Deng JL *et al.* A novel 3-gene signature for identifying COVID-19 patients based on bioinformatics and machine learning. *Genes* 2022;**13**:13. https://doi.org/10.3390/genes13091602
32. Albright J, Mick E, Sanchez-Guerrero E *et al.* A 2-gene host signature for improved accuracy of COVID-19 diagnosis agnostic to viral variants. *mSystems* 2022;**8**:e00671-22. https://doi.org/10.1128/msystems.00671-22
33. Li Y, Tao XY, Ye S *et al.* A T-cell-derived 3-gene signature distinguishes SARS-CoV-2 from common respiratory viruses. *Viruses-Basel* 2024;**16**:1029. https://doi.org/10.3390/v16071029
34. Zeng QQ, Qi X, Ma JP *et al.* Distinct miRNAs associated with various clinical presentations of SARS-CoV-2 infection. *ISCIENCE* 2022;**25**:104309. https://doi.org/10.1016/j.isci.2022.104309
35. Schimke LF, Marques AHC, Baiocchi GC *et al.* Severe COVID-19 shares a common neutrophil activation signature with other acute inflammatory states. *Cells* 2022;**11**:11. https://doi.org/10.3390/cells11050847
36. Lu J, Meng M, Zhou X *et al.* Identification of COVID-19 severity biomarkers based on feature selection on single-cell RNA-Seq data of CD8+ T cells. *Front Genet* 2022;**13**:1053772. https://doi.org/10.3389/fgene.2022.1053772
37. Zarei Ghobadi M, Emamzadeh R, Teymoori-Rad M *et al.* Exploration of blood-derived coding and non-coding RNA diagnostic immunological panels for COVID-19 through a co-expressed-based machine learning procedure. *Front Immunol* 2022;**13**:1001070. https://doi.org/10.3389/fimmu.2022.1001070
38. Wu D, Zhang R, Datta S. Unraveling T cell responses for long term protection of SARS-CoV-2 infection. *Front Genet* 2022;**13**:871164. https://doi.org/10.3389/fgene.2022.871164
39. Flores MA, Paniagua K, Huang W *et al.* Characterizing macrophages diversity in COVID-19 patients using deep learning. *Genes (Basel)* 2022;**13**:13. https://doi.org/10.3390/genes13122264

40. Li H, Huang FM, Liao HP *et al.* Identification of COVID-19-specific immune markers using a machine learning method. *Front Mol Biosci* 2022;**9**:9. https://doi.org/10.3389/fmolb.2022.952626
41. Luo H, Yan JS, Zhang DY *et al.* Identification of cuproptosis-related molecular subtypes and a novel predictive model of COVID-19 based on machine learning. *Front Immunol* 2023;**14**:1152223. https://doi.org/10.3389/fimmu.2023.1152223
42. Zhang Z, Pang T, Qi M *et al.* The biological processes of ferroptosis involved in pathogenesis of COVID-19 and core ferroptoic genes related with the occurrence and severity of this disease. *Evol Bioinforma* 2023;**19**:11769343231153293. https://doi.org/10.1177/11769343231153293
43. Zhuo JZ, Wang K, Shi ZJ *et al.* Immunogenic cell death-led discovery of COVID-19 biomarkers and inflammatory infiltrates. *Front Microbiol* 2023;**14**:1191004. https://doi.org/10.3389/fmicb.2023.1191004
44. Li CY, Wu K, Yang R *et al.* Comprehensive analysis of immunogenic cell death-related gene and construction of prediction model based on WGCNA and multiple machine learning in severe COVID-19. *Sci Rep* 2024;**14**:14. https://doi.org/10.1038/s41598-024-59117-0
45. Wu Y, Wu Z, Jin Q *et al.* Identification and analysis of biomarkers associated with lipophagy and therapeutic agents for COVID-19. *Viruses* 2024;**16**:923. https://doi.org/10.3390/v16060923
46. Zhang LH, Li YH, Hu WT *et al.* Computational identification of mitochondrial dysfunction biomarkers in severe SARS-CoV-2 infection: facilitating therapeutic applications of phytomedicine. *Phytomedicine* 2024;**131**:155784. https://doi.org/10.1016/j.phymed.2024.155784
47. Zhao M, Li J, Liu X *et al.* A gene regulatory network-aware graph learning method for cell identity annotation in single-cell RNA-seq data. *Genome Res* 2024;**34**:1036–51. https://doi.org/10.1101/gr.278439.123
48. Ye Q, Zeng YD, Jiang LL *et al.* A knowledge-guided graph learning approach bridging phenotype- and target-based drug discovery. *Adv Sci* 2025;**12**:e2412402. https://doi.org/10.1002/advs.202412402
49. Zheng CY, Wang YX, Cheng YQ *et al.* scNovel: a scalable deep learning-based network for novel rare cell discovery in single-cell transcriptomics. *Brief Bioinform* 2024;**25**:bbae112. https://doi.org/10.1093/bib/bbae112
50. Zhou M, Zhang H, Bai Z *et al.* Single-cell multi-omics topic embedding reveals cell-type-specific and COVID-19 severity-related immune signatures. *Cell Rep Methods* 2023;**3**:100563. https://doi.org/10.1016/j.crmeth.2023.100563
51. Hausmann F, Ergen C, Khatri R *et al.* DISCERN: deep single-cell expression reconstruction for improved cell clustering and cell subtype and state detection. *Genome Biol* 2023;**24**:212. https://doi.org/10.1186/s13059-023-03049-x
52. Koçak B, Cuocolo R, dos Santos DP *et al.* Must-have qualities of clinical research on artificial intelligence and machine learning. *Balkan Med J* 2023;**40**:3–12. https://doi.org/10.4274/balkanmedj.galenos.2022.2022-11-51
53. Özbek M, Toy HI, Takan I *et al.* A counterintuitive neutrophil-mediated pattern in COVID-19 patients revealed through transcriptomics analysis. *Viruses-Basel* 2023;**15**:15. https://doi.org/10.3390/v15010104
54. Potamias G, Gkoublia P, Kanterakis A. The two-stage molecular scenery of SARS-CoV-2 infection with implications to disease severity: an in-silico quest. *Front Immunol* 2023;**14**:1251067. https://doi.org/10.3389/fimmu.2023.1251067
55. Zhu K, Chen Z, Xiao Y *et al.* Multi-omics and immune cells' profiling of COVID-19 patients for ICU admission prediction: in silico analysis and an integrated machine learning-based approach in the framework of predictive, preventive, and personalized medicine. *EPMA J* 2023;**14**:101–17. https://doi.org/10.1007/s13167-023-00317-5
56. Park JJ, Lee KAV, Lam SZ *et al.* Machine learning identifies T cell receptor repertoire signatures associated with COVID-19 severity. *Commun Biol* 2023;**6**:6. https://doi.org/10.1038/s42003-023-04447-4
57. Bello B, Bundey YN, Bhave R *et al.* Integrating AI/ML models for patient stratification leveraging omics dataset and clinical biomarkers from COVID-19 patients: a promising approach to personalized medicine. *Int J Mol Sci* 2023;**24**:6250. https://doi.org/10.3390/ijms24076250
58. Jeyananthan P. SARS-CoV-2 diagnosis using transcriptome data: a machine learning approach. *SN Comput Sci* 2023;**4**:218. https://doi.org/10.1007/s42979-023-01703-6
59. Alarabi AB, Mohsen A, Mizuguchi K *et al.* Co-expression analysis to identify key modules and hub genes associated with COVID-19 in platelets. *BMC Med Genomics* 2022;**15**:83. https://doi.org/10.1186/s12920-022-01222-y
60. Iqbal N, Kumar P. Integrated COVID-19 predictor: differential expression analysis to reveal potential biomarkers and prediction of coronavirus using RNA-Seq profile data. *Comput Biol Med* 2022;**147**:105684. https://doi.org/10.1016/j.compbiomed.2022.105684
61. Carapito R, Li R, Helms J *et al.* Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. *Sci Transl Med* 2022;**14**:eabj7521. https://doi.org/10.1126/scitranslmed.abj7521
62. Bao YS, Ma QL, Chen L *et al.* Recognizing SARS-CoV-2 infection of nasopharyngeal tissue at the single-cell level by machine learning method. *Mol Immunol* 2025;**177**:44–61. https://doi.org/10.1016/j.molimm.2024.12.004
63. Li X, Zhang C, Chen WA *et al.* Identification of single-cell RNA sequencing molecular signatures for COVID-19 infection severity classification. In: Jiang X, Wang H, Alhajj R, Hu X, Engel F, Mahmud M, Pisanti N, Cui X, Song H (eds.), *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2023 Dec 5–8; Istanbul, Türkiye*. IEEE; 2023, pp. 2044–7.
64. Ge S, Sun S, Xu H *et al.* Deep learning in single-cell and spatial transcriptomics data analysis: advances and challenges from a data science perspective. *Brief Bioinform* 2025;**26**:bbaf136. https://doi.org/10.1093/bib/bbaf136
65. Chang YY, Wei AC. Transcriptome and machine learning analysis of the impact of COVID-19 on mitochondria and multiorgan damage. *PLoS One* 2024;**19**:e0297664. https://doi.org/10.1371/journal.pone.0297664
66. Riley RD, Ensor J, Snell KIE *et al.* Importance of sample size on the quality and utility of AI-based prediction models for healthcare. *Lancet Digit Health* 2025;**7**:100857. https://doi.org/10.1016/j.landig.2025.01.013
67. Daamen AR, Bachali P, Grammer AC *et al.* Classification of COVID-19 patients into clinically relevant subsets by a novel machine learning pipeline using transcriptomic features. *Int J Mol Sci* 2023;**24**:4905. https://doi.org/10.3390/ijms24054905
68. Di Pietro P, Abate AC, Izzo C *et al.* Plasma miR-1-3p levels predict severity in hospitalized COVID-19 patients. *Br J Pharmacol* 2024;**182**:451–67. https://doi.org/10.1111/bph.17392
69. Li XH, Zhou XC, Ding SJ *et al.* Identification of transcriptome biomarkers for severe COVID-19 with machine learning methods. *Biomolecules* 2022;**12**:12. https://doi.org/10.3390/biom12121735

70. Sethi S, Shakyawar S, Reddy AS *et al.* A machine learning model for the prediction of COVID-19 severity using RNA-Seq, clinical, and co-morbidity data. *Diagnostics* 2024;**14**:14. https://doi.org/10.3390/diagnostics14121284
71. Xie Y, Yang JF, Ouyang JF *et al.* scPanel: a tool for automatic identification of sparse gene panels for generalizable patient classification using scRNA-seq datasets. *Brief Bioinform* 2024;**25**:bbae482. https://doi.org/10.1093/bib/bbae482
72. Das R, Sinnarasan VSP, Paul D *et al.* Correction: a machine learning approach to identify potential miRNA-gene regulatory network contributing to the pathogenesis of SARS-CoV-2 infection. *Biochem Genet* 2023;**62**:1007. https://doi.org/10.1007/s10528-023-10511-9
73. Goel A, Mudge Z, Bi S *et al.* Identification of COVID-19 severity and associated genetic biomarkers based on scRNA-Seq data. In: *13TH ACM International Conference on Bioinformatics Computational Biology and Health Informatics, BCB*, Vol. **2022**, 2022 Aug 7–10. Northbrook, IL, USA. New York: ACM; 2022.
74. Maleknia S, Tavassolifar MJ, Mottaghitalab F *et al.* Identifying novel host-based diagnostic biomarker panels for COVID-19: a whole-blood/nasopharyngeal transcriptome meta-analysis. *Mol Med* 2022;**28**:86. https://doi.org/10.1186/s10020-022-00513-5
75. Song XB, Zhu JA, Tan XL *et al.* XGBoost-based feature learning method for mining COVID-19 novel diagnostic markers. *Front Public Health* 2022;**10**:926069. https://doi.org/10.3389/fpubh.2022.926069
76. Xu YC, Ma QL, Ren JX *et al.* Using machine learning methods in identifying genes associated with COVID-19 in cardiomyocytes and cardiac vascular endothelial cells. *Life-Basel* 2023;**13**:13. https://doi.org/10.3390/life13041011
77. Somekh J, Shen-Orr SS, Kohane IS. Batch correction evaluation framework using a-priori gene-gene associations: applied to the GTEx dataset. *BMC Bioinformatics* 2019;**20**:268. https://doi.org/10.1186/s12859-019-2855-9
78. Han H. Challenges of reproducible AI in biomedical data science. *BMC Med Genomics* 2025;**18**:8. https://doi.org/10.1186/s12920-024-02072-6
79. Gustafson L, Rolland C, Ravi N *et al.* FACET: Fairness in computer vision evaluation benchmark. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV); 2023 Oct 1–6; Paris, France*. IEEE; 2023, pp. 20313–25.
80. Xiang A, Andrews JTA, Bourke RL *et al.* Fair human-centric image dataset for ethical AI benchmarking. *Nature* 2025;**648**:97–108. https://doi.org/10.1038/s41586-025-09716-2
81. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* 2023;**4**:100804. https://doi.org/10.1016/j.patter.2023.100804
82. de Kanter E, Kaul T, Heus P *et al.* Adherence to TRIPOD+AI guideline: an updated reporting assessment tool. *J Clin Epidemiol* 2026;**191**:112118. https://doi.org/10.1016/j.jclinepi.2025.112118
83. Wilkinson SR, Aloqalaa M, Belhajjame K *et al.* Applying the FAIR principles to computational workflows. *Sci Data* 2025;**12**:328. https://doi.org/10.1038/s41597-025-04451-9
84. Krishnamoorthy P, Raj AS, Kumar H. Machine learning-driven blood transcriptome-based discovery of SARS-CoV-2 specific severity biomarkers. *J Med Virol* 2023;**95**:e28488. https://doi.org/10.1002/jmv.28488
85. Papoutsoglou G, Karaglani M, Lagani V *et al.* Automated machine learning optimizes and accelerates predictive modeling from COVID-19 high throughput datasets. *Sci Rep* 2021;**11**:15107. https://doi.org/10.1038/s41598-021-94501-0
86. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. *Entropy (Basel)* 2020;**23**:23. https://doi.org/10.3390/e23010018
87. Ying R, Bourgeois D, You J *et al.* GNNExplainer: generating explanations for graph neural networks. In: Wallach HM, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox EB (eds.), *Proceedings of the 33rd International Conference on Neural Information Processing Systems; 2019 Dec 8–14, Vancouver, Canada*. Red Hook, NY: Curran Associates Inc, 2019, pp. 9244–55.
88. Xu C, Coen-Pirani P, Jiang X. Empirical study of overfitting in deep learning for predicting breast cancer metastasis. *Cancers (Basel)* 2023;**15**:15. https://doi.org/10.3390/cancers15071969
89. Ma C, Zhang Y, Ding R *et al.* In search of the ratio of miRNA expression as robust biomarkers for constructing stable diagnostic models among multi-center data. *Front Genet* 2024;**15**:15. https://doi.org/10.3389/fgene.2024.1381917
90. Chen JX, Cheong CW, Lan L *et al.* DeepDRIM: a deep neural network to reconstruct cell-type-specific gene regulatory network using single-cell RNA-seq data. *Brief Bioinform* 2021;**22**:bbab325. https://doi.org/10.1093/bib/bbab325
91. Momeni M, Rashidifar M, Balam FH *et al.* A comprehensive analysis of gene expression profiling data in COVID-19 patients for discovery of specific and differential blood biomarker signatures. *Sci Rep* 2023;**13**:5599. https://doi.org/10.1038/s41598-023-32268-2
92. Kumagai Y. BootCellNet, a resampling-based procedure, promotes unsupervised identification of cell populations via robust inference of gene regulatory networks. *PLoS Comput Biol* 2024;**20**:e1012480.
93. Allgaier J, Pryss R. Cross-validation visualized: a narrative guide to advanced methods. *Mach Learn Knowl Extr* 2024;**6**:1378–88. https://doi.org/10.3390/make6020065
94. Demircioğlu A. Applying oversampling before cross-validation will lead to high bias in radiomics. *Sci Rep* 2024;**14**:11563. https://doi.org/10.1038/s41598-024-62585-z
95. Roland T, Böck C, Tschoellitsch T *et al.* Domain shifts in machine learning based Covid-19 diagnosis from blood tests. *J Med Syst* 2022;**46**:23. https://doi.org/10.1007/s10916-022-01807-1
96. Jiang J, Li Y, Cao S *et al.* Artificial intelligence in bioinformatics: a survey. *Brief Bioinform* 2025;**26**:bbaf576. https://doi.org/10.1093/bib/bbaf576
97. Aravazhi PS, Gunasekaran P, Benjamin NZY *et al.* The integration of artificial intelligence into clinical medicine: trends, challenges, and future directions. *Disease-a-Month* 2025;**71**:101882. https://doi.org/10.1016/j.disamonth.2025.101882
98. Kelly CJ, Karthikesalingam A, Suleyman M *et al.* Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;**17**:195. https://doi.org/10.1186/s12916-019-1426-2
99. Kardjadj M. Advances in point-of-care infectious disease diagnostics: integration of technologies, validation, artificial intelligence, and regulatory oversight. *Diagnostics* 2025;**15**:2845. https://doi.org/10.3390/diagnostics15222845
100. Chothani F, Movaliya V, Vaghela K *et al.* Regulatory prospective on software as a medical device. *Int J Drug Regul Aff* 2022;**10**:13–7. https://doi.org/10.22270/ijdra.v10i4.545
101. IMDRF. Software as a Medical Device (SaMD). Key Definitions. https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf (7 March 2026, date last accessed)
102. Han G-R, Goncharov A, Eryilmaz M *et al.* Machine learning in point-of-care testing: innovations, challenges, and opportunities.

Nat Commun 2025;**16**:3165. https://doi.org/10.1038/s41467-025-58527-6
103. Acharya DB, Kuppan K, Divya B. Agentic AI: autonomous intelligence for complex goals—a comprehensive survey. *IEEE Access* 2025;**13**:18912–36. https://doi.org/10.1109/ACCESS.2025.3532853
104. H Liu, Y Li, H Wang. GenoMAS: a multi-agent framework for scientific discovery via code-driven gene expression analysis. arXiv [Preprint]. 2025.
105. Zhou J, Zhang B, Li G *et al.* An AI agent for fully automated multi-omic analyses. *Adv Sci* 2024;**11**:e2407094. https://doi.org/10.1002/advs.202407094
106. Martinek V, Gariboldi A, Tzimotoudis D *et al.* Agentomics: an agentic system that autonomously develops novel state-of-the-art solutions for biomedical machine learning tasks. *bioRxiv* 2001;**2026**:2027, 702049.
107. Xiao Y, Liu J, Zheng Y *et al.* CellAgent: LLM-driven multi-agent framework for natural language-based single-cell analysis. *bioRxiv* 2005;**2025**:2013. 593861.
108. Lin Z, Wang W, Marin-Llobet A *et al.* Spatial transcriptomics AI agent charts hPSC-pancreas maturation in vivo. *bioRxiv* 2025. https://doi.org/10.1101/2025.04.01.646731
109. Ying K, Tyshkovskiy A, Moldakozhayev A *et al.* Autonomous AI agents discover aging interventions from millions of molecular profiles. *bioRxiv* 2002;**2025**:2028. 530532.
110. J Miao, JR Davis, Y Zhang *et al.* Paper2agent: reimagining research papers as interactive and reliable AI agents. arXiv [Preprint] 2025.
111. Nouri N, Artzi R, Savova V. An agentic AI framework for ingestion and standardization of single-cell RNA-seq data analysis. *npj Artif Intell* 2026;**2**:8. https://doi.org/10.1038/s44387-025-00064-0
112. Chakrabarty PK. Causal inference in agentic AI: bridging explainability and dynamic decision making. *Int J Sci Res* 2025;**14**:2112–7. https://doi.org/10.21275/SR25424081718
113. Theodoris CV, Xiao L, Chopra A *et al.* Transfer learning enables predictions in network biology. *Nature* 2023;**618**:616–24. https://doi.org/10.1038/s41586-023-06139-9

Briefings in Bioinformatics, 2026, **27**, bbag249
https://doi.org/10.1093/bib/bbag249
Review