



Review

Software Applications in Biomedicine: A Narrative Review of Translational Pathways from Data to Decision

Gabriela Georgieva Panayotova

Division of Physiology, Department of Physiology and Pathophysiology, Medical University of Varna, 9002 Varna, Bulgaria; gabrielapanayotova95@gmail.com

Abstract

Background/Objectives: Software is now core infrastructure in biomedical science, yet fragmented workflows across subfields hinder reproducibility and delay the translation of data into actionable decisions. There is a critical need for a cross-disciplinary synthesis to bridge these silos and establish a unified framework for software maturity. This narrative review addresses this gap by synthesizing representative software ecosystems across three major pillars: bioinformatics, molecular modeling/simulations, and epidemiology/public health. **Methods:** A narrative review of articles indexed in PubMed/NCBI, Web of Science, and Scopus between 2000 and 2025 was conducted. Domain-specific terms related to bioinformatics, molecular modeling, docking, molecular dynamics, epidemiology, public health, and workflow management were combined with software- and algorithm-focused keywords. Studies describing, validating, or applying documented tools with biomedical relevance were included. **Results:** Across domains, mature data standards and reference resources (e.g., FASTQ, BAM/CRAM, VCF, mzML), widely adopted platforms (e.g., BLAST+ (v2.16.0, NCBI, Bethesda, USA), Bioconductor (v3.20, Bioconductor Foundation, Seattle, USA), AutoDock Vina (v1.2.5, Scripps Research, La Jolla, USA), GROMACS (v2024.3, GROMACS Team, Stockholm, Sweden), Epi Info (v7.2.6, CDC, Atlanta, USA), QGIS (v3.40, QGIS.org, Gossau, Switzerland), and increasing use of workflow engines were identified. Software pipelines routinely transform molecular and surveillance data into interpretable features supporting hypothesis generation. **Conclusions:** Integrated, standards-based, and validated software pipelines can shorten the path from measurement to decision in biomedicine and public health. Future progress depends on reproducibility practices, benchmarking, user-centered design, portable implementations, and responsible deployment of machine learning.

Keywords: bioinformatics; molecular modeling; epidemiology; workflows; reproducibility; machine learning

Academic Editor: Hans Binder

Received: 5 January 2026

Revised: 21 January 2026

Accepted: 29 January 2026

Published: 4 February 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Over the past three decades, the biomedical sciences have undergone a profound turning point: highly specialized instruments—from sequencers and mass spectrometers to microscopes and portable biosensors—now generate data of such volume and complexity that manual analysis is no longer feasible [1–3]. As a result, software has moved from a historically supporting role to a central driver of discovery, translation, and decision-making [4–6]. Today, fundamental questions in biology and medicine—how genes

and proteins function, how molecules interact, how diseases spread, and how interventions should be implemented—are routinely addressed using diverse programs that combine statistics, machine learning, and data storage and analysis [7–9].

Three classes of software have been particularly important historically. First, bioinformatics tools now convert raw biological signals into interpretable features: short reads become assembled genomes; expression matrices become differential expression profiles; clinical records become systematic cohorts [10–12]. These workflows depend on standardized formats (e.g., FASTQ, BAM/CRAM, VCF for genomics; mzML for proteomics), reference databases, and algorithmic building blocks such as sequence alignment, multiple sequence analysis, clustering, dimensionality reduction, and predictive modeling [13–15]. The result is a relatively defined path from asynchronous measurements to hypotheses about function, mechanism, and phenotype [16].

Second, molecular modeling and simulation software connect structure with function. Visualization tools enable researchers to inspect macromolecular architecture and binding interfaces; docking programs propose plausible protein–ligand poses and rank them using scoring functions; molecular dynamics simulations probe conformational landscapes and temporal stability [17,18]. In combination with experimental restraints (e.g., cryo-EM density maps and NMR-derived constraints), these software systems support structure-based design, shrink the search space, and accelerate the optimization of lead compounds in drug discovery [9,19].

Third, epidemiological and public health software provides situational awareness and guides population-level interventions [20]. Statistical packages and specialized tools allow analysts to ingest clinical case reports, laboratory confirmations, mobility data, and environmental signals, then fit compartmental or agent-based models, estimate reproduction numbers, forecast incidence, and analyze counterfactual scenarios [9,21]. Geographic information systems and interactive dashboards translate these data into spatially explicit, time-sensitive visualizations for decision-makers—capabilities that proved essential during recent years of pandemics and uncertainty in disease spread [22,23].

The convergence of these domains is transforming translational pipelines. For example, a pathogen genome assembled and annotated using bioinformatics tools can immediately inform molecular modeling software about potential inhibitor candidates, while epidemiological platforms assess where, how, and in whom these interventions would have the greatest impact [24]. Likewise, integrating multi-omics (genomics, transcriptomics, proteomics, metabolomics) with electronic health records and imaging data requires software that can link heterogeneous sources, ensure traceability, and provide explainable outputs suitable for clinical review [25,26].

Despite this progress, several challenges demand a critical view of the current landscape. Data heterogeneity and missingness complicate integration [27]. Many tools are powerful but difficult for practitioners to learn, creating barriers for smaller laboratories and health services with limited staff [28]. Data analysis itself can be inconsistent—varying datasets, metrics, and preprocessing steps can create an impression of progress without a clear real-world benefit [29]. In addition, the sustainability of open-source projects—on which the community heavily relies—depends on funding, governance, and data stewardship, raising concerns about long-term maintenance and the responsible use and sharing of sensitive information [30,31].

Accordingly, this review maps representative software ecosystems across bioinformatics, molecular modeling, and epidemiology/public health, and critically evaluates their strengths, limitations, and translational readiness from measurement to decision.

Priority is given to widely adopted, well-documented tools with proven value, while also noting emerging approaches—such as machine learning for structure prediction and

graph-based methods for multimodal integration—that substantially extend current capabilities. With an emphasis on reproducibility and usability, the article summarizes domains, tasks, licensing, computational requirements, and validation in order to align tools with specific scientific or public health objectives.

While numerous existing reviews provide deep technical evaluations of specific domains—such as dedicated surveys of genomic variant callers, protein docking algorithms, or epidemiological forecasting models—they often remain focused within their respective silos. This review differs by shifting the focus from individual tool performance to the cross-domain integration required for translational medicine. Unlike domain-specific reviews, this work applies a uniform evaluative framework to three historically distinct pillars, highlighting how software components must ‘lock into’ one another to bridge the gap between raw data generation and actionable public health or clinical decisions. This holistic approach allows for the identification of systemic bottlenecks, such as metadata handoff failures, that are often overlooked in reviews restricted to a single discipline.

This review makes three primary contributions to the literature. First, it establishes a unified “translational readiness” framework that allows for the comparative evaluation of software ecosystems across distinct biomedical pillars—bioinformatics, molecular modeling, and epidemiology—which are typically treated in isolation. Second, it identifies specific technical and structural bottlenecks, such as the “usability-performance” trade-off and metadata standardization gaps, that currently hinder the flow of data from raw measurement to clinical or public health action. Finally, it provides a curated roadmap of mature and emerging tools, serving as a practical guide for researchers and practitioners to select software that aligns with both scientific rigor and real-world deployment requirements. By bridging these domains, the article offers a holistic perspective on the computational infrastructure necessary for modern translational medicine.

2. Materials and Methods

2.1. Study Design

This narrative review synthesizes information on software applications used in three foundational pillars of the biomedical sciences: bioinformatics, molecular modeling and simulations, and epidemiological/public health analytics. This approach was chosen to address the significant methodological heterogeneity in the current literature, which includes diverse formats such as algorithm development papers, performance benchmarks, and clinical case studies. A narrative design is particularly suited to this scope, as the included software spans conflicting methodological paradigms—ranging from stochastic simulation frameworks to deterministic statistical modeling—that do not allow for the uniform outcome measures required for a systematic meta-analysis.

Central to this study is the evaluation of “translational readiness,” which is defined as the capacity of a software system to reliably transform primary biological or clinical measurements into interpretable outputs that directly inform downstream research, clinical, or public health decisions. To operationalize this concept, the review utilizes a shared cross-domain framework that assesses tools based on their maturity along the path from data generation to action. This framework evaluates translational readiness across five core dimensions:

- **Reproducibility:** The extent to which a tool supports auditable and consistent results through the use of workflow engines and environment encapsulation via containerization.
- **Validation:** The availability of evidence from independent benchmarks, blinded evaluations, or replication studies that confirm the tool’s performance using representative datasets.

- **Interoperability:** The degree to which software adheres to established data standards and provides application programming interfaces (APIs) to facilitate integration into multimodal pipelines.
- **Usability and Human Capacity:** The presence of high-quality documentation, user-friendly interfaces (GUIs), and training resources that lower the barrier for adoption in non-computational settings, such as clinical labs or local health departments.
- **Decision Relevance:** The proven ability of the software's output to support specific decision-making tasks, such as lead optimization in drug discovery, cohort stratification in medicine, or intervention targeting in public health.

By applying these criteria uniformly across the three pillars, the objective is not to quantify effect sizes, but to identify where software ecosystems successfully facilitate translation and where friction—such as poor documentation or opaque validation—remains a bottleneck to real-world impact.

2.2. Information Sources and Search Strategy

The search covered the period 2000–2025. The primary databases were PubMed/NCBI, Web of Science, and Scopus. To ensure a systematic and inclusive retrieval of relevant studies, a complex Boolean search string was developed. This string combined domain-specific terminology with keywords focused on software architecture, implementation, and evaluative frameworks. The primary search string employed was as follows:

(bioinformatics OR “molecular modeling” OR “molecular dynamics” OR “docking” OR “epidemiology” OR “public health”) AND (software OR algorithm OR workflow) AND (reproducibility OR validation OR benchmarking OR “translational readiness”).

In addition to the database queries, the reference lists of included articles were screened manually using a “snowballing” technique to identify significant tools or foundational methodological papers that might have been omitted by the electronic search. This process ensured that both established gold-standard software and emerging platforms were captured within the analysis.

2.3. Inclusion and Exclusion Criteria

Sources were included if they:

- described, validated, or applied software central to one of the three pillars;
- provided sufficient methodological detail on core algorithms/approaches;
- demonstrated scientific or translational relevance.

Conversely, specific exclusion criteria were applied to maintain the technical quality and focus of the review. Studies were excluded if they:

- presented purely conceptual frameworks without an accompanying software implementation or accessible codebase;
- focused on tools that lacked minimal documentation or public repositories, rendering them non-reproducible for the community;
- described applications without clear significance or measurable utility for the biomedical and public health sciences.

2.4. Study Selection

The initial database search yielded 412 records (PubMed: 156; Web of Science: 134; Scopus: 122). After removing 88 duplicates, 324 unique titles and abstracts were screened for domain relevance. During this primary screen, 196 records were excluded because they did not involve software implementation or were outside the biomedical scope. Of

the remaining 128 records assessed for full-text eligibility, 56 were excluded based on the established criteria: 22 lacked sufficient methodological detail or documentation, 18 focused on obsolete software versions, and 16 were conceptual frameworks without an accessible codebase. This rigorous selection process resulted in a final corpus of 78 primary sources and software systems synthesized in this review. Selection prioritized tools with sustained adoption and documented validation to ensure the ‘translational readiness’ rubric was applied to the most impactful platforms in the field.

2.5. Data Extraction

For each tool, the following were extracted: domain and main tasks; typical inputs/outputs and supported data standards; core algorithms/approaches; validation evidence; usability characteristics; interoperability; computational profile; licensing and sustainability; and known limitations. Validation evidence was assessed with attention to benchmark design, dataset representativeness, evaluation metrics, and independent replication, recognizing that reported performance is context-dependent.

The quality of the included studies was assessed using a modified version of the SANRA (Scale for the Assessment of Narrative Review Articles) framework to ensure scientific rigor. Each study was evaluated based on six criteria:

- (1) justification of the article’s importance;
- (2) statement of concrete aims;
- (3) description of the search strategy;
- (4) referencing of appropriate evidence;
- (5) scientific reasoning;
- (6) clear presentation of data/software performance.

For technical validation studies, additional weight was given to the transparency of benchmarking protocols and the availability of open-source code. Only studies that met at least 80% of these quality indicators were included in the final synthesis to ensure that the ‘translational readiness’ rubric was applied to a robust and reliable corpus of evidence.

2.6. Synthesis Methods

To ensure a structured and transparent synthesis, extracted data were cross-referenced using a formal three-stage thematic synthesis approach. This methodology was chosen to move beyond a simple descriptive catalog and instead provide a critical evaluation of software maturity:

1. **Technical Coding and Performance Extraction:** The initial stage involved line-by-line coding of the technical capabilities and validation results for each of the 78 included sources. Quantitative performance metrics (e.g., docking scores, alignment accuracy, or forecast error rates) were extracted and contextualized alongside the specific datasets and evaluation protocols used in the original studies.
2. **Development of Descriptive Themes:** Coded data were then grouped into descriptive themes organized around the five core dimensions of the “translational readiness” framework: reproducibility, validation, interoperability, usability, and decision relevance. This allowed for a cross-domain comparison, identifying whether specific strengths—such as mature data standards—were consistent across bioinformatics, modeling, and epidemiology.
3. **Generation of Analytical Themes:** In the final stage, these descriptive themes were synthesized into overarching analytical findings that link software technicalities to real-world impact. This involved mapping how specific software outputs (e.g., variant panels or risk maps) directly enable downstream clinical and public health decisions.

To minimize selection and reporting bias, the synthesis purposefully incorporated a “critical appraisal” component. This involved documenting not only successful applications but also “failure modes,” such as scoring-function biases in molecular modeling or data-latency issues in epidemiological forecasting. By balancing positive validation results with these known limitations, the review provides a nuanced roadmap for practitioners to select tools that are fit for purpose.

3. Results

The final synthesis included a corpus of 78 primary sources and software systems identified through the search strategy. A quantitative breakdown of the selected literature reveals that 38% of the sources pertain to bioinformatics workflows, 33% to molecular modeling and simulation platforms, and 29% to epidemiological and public health analytics. Temporally, the corpus shows a significant concentration of recent literature, with approximately 85% of the referenced tools and validation studies published between 2010 and 2025, reflecting the rapid acceleration of software-driven research in the last decade. The remaining 15% consists of foundational software and established data standards developed prior to 2010 that remain central to modern translational pipelines.

3.1. Bioinformatics

In the reviewed sources, bioinformatics software supports sequence search/alignment, genome/transcriptome analysis, variant detection and interpretation, and multi-omics statistics. Representative examples of widely used bioinformatics tools, their typical inputs/outputs, and core characteristics are summarized in Table 1.

Table 1. Major bioinformatics software systems, pipelines, and use cases.

Tool	Version	Main Role	Typical Inputs → Outputs	Key Features
BLAST/BLAST+	v2.16.0, NCBI, Bethesda, USA	Local similarity search, support for annotation	Sequence(s) → ranked alignments, statistics	Web interface and command line; supports local databases; integrates into automated pipelines.
Clustal Omega	v1.2.4, University College Dublin, Dublin, Ireland	Multiple sequence alignment at scale	Unaligned sequences → multiple sequence alignment	Designed for large datasets; fast and accurate alignment.
Bioconductor	v3.20, Bioconductor Foundation, Seattle, USA	Omics analyses (RNA-sequencing, methylation, proteomics), statistics	Count/intensity matrices → differential expression, enrichment, reports	Open source; strict packaging, training materials, and reproducible workflows.

Common inputs include FASTQ (raw reads), BAM/CRAM (alignments), VCF (variants), and count/intensity matrices; outputs include annotated genomes, prioritized variants, and differential expression or pathway signatures [32–37]. Tools such as BLAST/BLAST+ remain fundamental for similarity search and functional inference, while local pipelines enable large-scale, automated analyses [38].

Large-scale multiple sequence alignment is routinely handled by Clustal Omega (version 1.2.4, Conway Institute, Dublin, Ireland), which emphasizes speed and accuracy for large protein sets, whereas Bioconductor provides a broad, open ecosystem of R packages for omics workflows, statistics, and reproducibility [39,40].

Key strengths include mature conventions for file formats and reference resources; rich ecosystems of packages; and support for containerization/workflows for deployment on HPC or cloud environments—critical for reproducibility and research scalability [41,42].

Limitations include sensitivity to preprocessing choices and reference biases; batch effects when integrating results; and variable usability, including limited language/geographic support. This creates a need for skills in command-line interfaces and workflow management [43].

Collectively, these outputs represent the first translational step, converting high-dimensional molecular measurements into structured features suitable for downstream mechanistic modeling and population-level interpretation.

3.2. Molecular Modeling and Simulations

This pillar encompasses molecular visualization, docking/virtual screening, and molecular dynamics simulations. Inputs include 3D structures, force fields, and ligand libraries; outputs include predicted binding poses and/or scores, as well as trajectories and derived observables [44,45].

Key examples include AutoDock Vina, a widely used tool for receptor–ligand docking that improves speed and pose accuracy over earlier AutoDock versions; PyMOL (version 3.1., Schrödinger, Inc., New York, USA) a user-sponsored and globally adopted visualization system; and GROMACS, a high-performance, actively maintained package for molecular dynamics with extensive user documentation [46].

Recent years have seen a paradigm shift in the availability of high-quality structural models due to the emergence of deep-learning-based protein structure prediction tools [8,39]. Most notably, AlphaFold (and its successor AlphaFold2) and RoseTTAFold have demonstrated the ability to predict protein architectures with near-experimental accuracy [47]. These tools are highly cited and widely used to generate ‘starting structures’ for proteins that lack experimental coordinates in the Protein Data Bank (PDB) [48,49]. In translational workflows, these predicted models serve as foundational substrates for subsequent docking campaigns and molecular dynamics simulations, significantly expanding the range of druggable targets available for virtual screening [50]. However, while these platforms provide remarkable backbone accuracy, users must still account for the local uncertainty of side-chain orientations and the conformational flexibility required for ligand binding, which remains a key area for integrated simulation-based refinement [48,49,51].

Strengths include diverse structure-based design workflows that integrate visualization, docking, and molecular dynamics; and acceleration via GPU support and parallelization, which makes demanding simulations more accessible to a broader community, especially when paired with abundant training materials [52].

Limitations include biases in scoring functions, sensitivity to protonation states and other preparation steps, and the computational cost of exhaustive sampling, which remains only partially mitigated by advanced sampling schemes [53]. Table 2 outlines key molecular modeling platforms, detailing for each tool its primary function, the types of data it processes and produces, and the main features that support its widespread use.

Table 2. Representative tools for molecular modeling and their roles.

Tool	Version	Main Role	Typical Inputs → Outputs	Key Features
AlphaFold	v2.3.2, DeepMind/Google, London, UK	Protein structure prediction	Amino acid sequence → 3D coordinates	Deep-learning-based; provides near-experimental accuracy; foundational for targets lacking experimental PDB structures.
RoseTTAFold	v1.1.0, University of Washington, Seattle, USA			
AutoDock Vina	v1.2.5, The Scripps Research Institute, La Jolla, USA	Docking/virtual screening	Protein and ligand structures → binding poses and scores	Multithreaded; substantial speed-up with competitive pose accuracy.
PyMOL	v3.1, Schrödinger, Inc., New York City, USA	Molecular visualization	Structures/trajectories → annotated figures/videos	Widely used in academia and industry; actively maintained documentation and regular releases.
GROMACS	v2024.3, GROMACS Development Team, Stockholm, Sweden	Molecular dynamics	Topologies, coordinates, force fields → trajectories, observables	High performance; up-to-date releases and extensive user manuals.

These tools translate molecular-scale hypotheses into ranked and testable candidates, providing a mechanistic bridge between omics-derived targets and experimental or clinical validation.

3.3. Epidemiology and Public Health Analytics

Software in this pillar supports surveillance data management, situational analysis, statistical/cluster modeling, mapping, and monitoring dashboards. Processed data include case counts, laboratory confirmations, mobility/environmental signals, and metadata. Outputs comprise estimates, forecasts, maps, and scenario analyses to support decisions [4,54,55].

Epi Info is a prominent example: it provides public-domain tools for questionnaire design, data entry and analysis, visualization, and mapping, with extensions such as mobile apps for vector surveillance. QGIS is a community-driven geographic platform for spatial analysis and map production, widely used in public health [56,57]. Projects such as Nextstrain demonstrate real-time phylogenetics and visualization for pathogens [58].

Strengths include rapid situational awareness through integrated data capture, analytics, and geographic visualization, as well as the availability of public-domain and open-source platforms that lower barriers for health professionals [59].

Weaknesses include limitations related to data quality/latency; interoperability issues with laboratory information systems and electronic health records; and wide variation in statistical capacity across institutions [60].

Overall, the reviewed software supports reproducible pipelines from raw biological signals to actionable outputs: bioinformatics pipelines that convert sequences/omics into features and hypotheses; modeling environments that translate structural hypotheses into ranked ligand poses and evidence of dynamic stability; and epidemiological platforms that transform case and contextual data into forecasts and spatially explicit guidance [10,61,62]. Table 3 summarizes representative software platforms used in epidemiology

and public health, detailing their main roles, typical data flows from input to output, and key distinguishing features.

Table 3. Representative tools for epidemiology/public health and their roles.

Tool	Main Role	Typical Inputs → Outputs	Key Features
Epi Info	Capture and analysis of surveillance data	Case/questionnaire data → statistics, graphs, maps	Public domain; widely used for investigating current epidemiological situations.
Epi Info (Vector App)	Mobile vector surveillance	Field observations → datasets ready for analysis	Android application for vector surveillance and analysis.
QGIS	Spatial analysis and cartography	Spatial layers → analytical maps, geoprocessing	Free and open source; extensive plugin ecosystem and comprehensive documentation.

In practice, results from one pillar “lock into” the others—for example, an annotated pathogen genome produced by bioinformatics tools can initiate docking campaigns against key proteins, while public health software quantifies the expected population-level impact and targeting of interventions [63,64]. To make this linkage more explicit, Table 4 summarizes the typical software outputs in each domain and illustrates the kinds of downstream research and public health decisions they inform.

Here, software completes the translational pathway by contextualizing biological and clinical signals into population-level risk estimates and decision-support outputs.

Table 4. Software outputs across domains and examples of downstream research and public health decisions enabled.

Domain	Main Outputs/Products	Examples of Downstream Decisions Enabled
Bioinformatics	Annotated genomes and gene sets; differential expression and other multi-omics signatures; variant/biomarker panels with pathway and network annotations.	Prioritization of candidate targets and biomarkers for functional studies or clinical trials; cohort stratification and patient subtyping; design of diagnostic panels and validation experiments; selection of follow-up omics or imaging assays.
Molecular modeling	Ranked docking poses and scores; molecular-dynamics–derived stability and flexibility profiles; interaction energy decompositions; high-quality 3D structure visualizations.	Triage and optimization of lead compounds or protein variants; formulation of mechanistic hypotheses about binding and allostery; selection of candidates for synthesis, biophysical testing, and in vivo studies; generation of publication- and submission-ready structural figures.
Epidemiology and public health	Time-series estimates of incidence and reproduction numbers; short- and medium-term forecasts under alternative scenarios; spatial risk maps at different administrative levels; quantitative evaluations of past or ongoing interventions.	Allocation of clinical, laboratory, and public health resources; timing and geographic targeting of interventions (e.g., vaccination, screening, NPIs); setting trigger thresholds for escalation or de-escalation of measures; design and optimization of surveillance and monitoring systems.

The broad adoption of domain standards, common file formats, and evidence-based workflow solutions enables integration of software components across the full chain from

research to translation. To systematically apply the “translational readiness” rubric described in the methodology, Table 5 provides a cross-domain comparison of representative software systems. This mapping illustrates how core dimensions—reproducibility, validation, interoperability, usability, and decision relevance—manifest across the three pillars, offering a structural overview of software maturity in the current landscape.

Table 5. Systematic mapping of representative tools across the translational readiness framework.

Domain	Example Tool	Reproducibility	Validation	Interoperability	Usability	Decision Relevance
Bioinformatics	Bioconductor	High (R-based packaging)	Extensive (Peer benchmarks)	High (Standard omics formats)	Moderate (Requires R knowledge)	High (Biomarker discovery)
Molecular Modeling	GROMACS	High (Standard topologies)	High (Force field verification)	Moderate (PDB/GRO formats)	Moderate (CLI-based)	High (Lead optimization)
Epidemiology	QGIS	Moderate (Project files)	High (Geospatial standards)	High (GIS/Spatial APIs)	High (GUI-based)	High (Risk mapping)

4. Discussion

This review highlights how software has become an indispensable substrate of modern biomedical science, enabling large-scale studies, structure-informed design, and timely situational awareness in epidemiology. Across all pillars, standardized file formats, widely adopted toolchains, and growing workflow support enhance interoperability and reproducibility. At the same time, persistent challenges—data heterogeneity, uneven usability, variable validation practices, and sensitive data governance—continue to shape what is practically achievable in research and translation.

4.1. Reproducibility and Alignment with Good Data/Software Management Principles

Workflow engines (e.g., CWL (v1.2, CWL Consortium, International, Global) /WDL (v1.1, OpenWDL/Broad Institute, Cambridge, USA) /Nextflow (v24.10, Seqera Labs, Barcelona, Spain) / Snakemake(v8.25, Johannes Köster, Essen, Germany) and containerization (Docker/Singularity) now underpin many production pipelines. These practices reduce the classic “works on my machine” failure mode and render analyses auditable [42]. However, reproducibility still depends on complete reporting: exact versions, reference databases, parameters, and traceability of intermediate files [65]. Common gaps include incomplete documentation of preprocessing steps (e.g., read trimming, normalization, batch effect correction) and of environment details that substantially affect results. Community templates and checklists can raise the baseline without stifling innovation [66,67].

4.2. Validation and Benchmarking

Performance claims are most useful when linked to transparent datasets, well-justified metrics, and baselines reflecting current practice [68]. In bioinformatics, reference biases, class imbalance, and data leakage can inflate reported accuracy [64]. In docking and molecular dynamics, scoring functions may favor certain chemotypes, and insufficient conformational sampling limits generalizability [52]. In epidemiology, forecast quality must be interpreted alongside data latency and revision effects [69]. Validation across multiple datasets, blinded or hold-out evaluation, and replication by independent groups are practical safeguards [70].

Across domains, insufficiently transparent benchmarking remains a key translational bottleneck. Performance metrics reported without access to datasets, preprocessing pipelines, or baseline comparisons limit reproducibility and impede independent verification. From a translational perspective, validation quality is often more decisive than marginal algorithmic improvements, particularly when outputs inform experimental prioritization or public health action.

4.3. Artificial Intelligence and Machine Learning

Machine learning is increasingly central—from variant effect prediction and multi-omics integration, to structure prediction and epidemic “nowcasting” [7]. The advantages are clear: enhanced sensitivity/specificity and automation of labor-intensive steps [71]. The risks are equally concrete: opaque decision boundaries, domain shifts, and propagation of biases from historical data. Priority directions include interpretable features or post hoc explanations suitable for scientific and clinical scrutiny; quantitative uncertainty estimation (calibrated probabilities, interval forecasts); and rigorous external validation across laboratories, instruments, and populations [3,72].

A recurring theme across all three pillars is that limitations in reproducibility, usability, and validation—not algorithmic capability—now represent the dominant barriers to translation. In many contexts, simpler and well-validated tools outperform more complex models when deployed under real-world constraints. Consequently, investment in benchmarking, documentation, and user-centered design is likely to yield greater translational impact than incremental gains in model accuracy alone.

4.4. Usability, Training, and Human Capacity

Powerful tools will not be used to their full potential if interfaces are opaque or documentation is sparse. Teams with mixed skills—bench scientists, analysts, software engineers—benefit from graphical user interfaces for routine tasks, and from command-line interfaces and APIs for advanced use [73]. Interactive notebooks improve transparency but must be combined with environment capture to remain reproducible [65]. Investment in training (short courses, tutorials, example datasets) often yields disproportionate returns, especially for health services and small laboratories [74].

4.5. Integration Across Pillars

Interoperability is no longer “nice to have” but essential: bioinformatics outputs (e.g., pathogen genomes, differential expression signatures) routinely feed molecular modeling campaigns, while public health platforms contextualize candidate interventions through risk maps and forecasts [24,62,63]. Practical friction points include harmonization of identifiers (genes, proteins, compounds), alignment of coordinate systems and force fields, and linking tabular analyses with spatial layers [75]. “Data contracts”—explicit expectations about schema, units, and metadata—help teams move faster with fewer surprises throughout the lifecycle [65].

4.6. Data Governance, Privacy, and Security

Biomedical and surveillance data are inherently sensitive. Encryption at rest and in transit, role-based access control, auditable logs, and data minimization are baseline requirements [76]. When using cloud resources, organizations must plan for jurisdictional constraints and vendor lock-in by adopting portable workflows and clear exit strategies [77].

4.7. Limitations

As a review spanning diverse sub-disciplines, this article prioritizes widely used, well-documented tools and workflows rather than exhaustively cataloging every niche package. Heterogeneous evaluation practices across domains complicate direct “metric-to-metric” comparisons; where possible, the emphasis is on patterns and decision criteria rather than absolute performance values.

4.8. Critical Appraisal of Research Gaps

While software has moved from a supporting role to a central driver of discovery, several structural gaps prevent these tools from reaching their full translational potential. A primary standardization gap exists: although file formats such as FASTQ, BAM, and VCF are widely adopted, there is a lack of explicit “data contracts” or metadata schemas that govern seamless handoffs across different pillars [14,33,35]. For example, bioinformatics outputs often require manual curation before they can be ingested by molecular modeling software, leading to potential data loss or interpretation errors [11].

Furthermore, a significant benchmarking gap remains a dominant translational bottleneck. Performance metrics are frequently reported without access to the original pre-processing pipelines, reference databases, or exact parameter sets used, which complicates independent verification and limits reproducibility [12,62,65]. In epidemiology, forecast quality is often undermined by data latency and the lack of standardized reporting metrics across geographic regions [69].

4.9. The Usability-Performance Trade-Off

A recurring theme across the reviewed domains is the persistent usability gap [70]. Many high-performance tools rely on command-line interfaces (CLIs) that require significant computational expertise, creating a barrier for smaller laboratories or local health services with limited staff [28,38]. While graphical user interfaces (GUIs) and interactive dashboards have improved situational awareness in public health, they often lack the flexibility required for advanced research tasks [22,31,78]. Investment in user-centered design and comprehensive training materials is, therefore, as critical as incremental algorithmic improvements for ensuring the deployment of these tools in real-world settings [78].

4.10. Future Directions: Multimodal and Explainable Systems

Future progress in biomedical informatics depends on the development of integrated, multimodal pipelines [12,62]. Rather than functioning as isolated silos, next-generation software must natively link heterogeneous data sources—such as pathogen genomes, protein-ligand binding profiles, and spatial incidence maps—to provide a unified view of disease dynamics [13,14,58].

The role of machine learning is poised to expand, but its success depends on a shift toward explainable, interpretable AI [11,74]. For software to be trusted in clinical and public health decision-making, it must provide not only high accuracy but also calibrated uncertainty estimation and post hoc explanations suitable for human review [54]. Finally, ensuring the sustainability of open-source ecosystems is vital; the community’s reliance on these tools necessitates robust governance, stable funding, and responsible data stewardship to protect sensitive information while fostering innovation [30,55].

4.11. Domain-Specific Challenges

Each of the three pillars investigated in this review faces unique scientific hurdles that complicate the path from data to decision. In bioinformatics, the primary challenge is the ‘ $n = 1$ ’ problem in precision medicine, where the lack of large, diverse reference cohorts

for rare variants leads to significant uncertainty in interpretation [10,62,64]. Additionally, the sheer scale of multi-omics integration creates high-dimensional noise that requires more robust statistical regularization techniques to avoid false-positive discoveries [16,26,41,75].

In molecular modeling, the ‘force field challenge’ remains a dominant bottleneck; current mathematical models of atomic interactions often struggle to balance computational speed with the chemical accuracy required for predicting binding affinity in complex cellular environments [9,17,45]. Furthermore, the ‘sampling problem’—the difficulty of simulating biological processes that occur on timescales longer than microseconds—limits our ability to observe slow conformational changes critical for drug discovery [6,17,53].

In epidemiology and public health, the main challenge is data fragmentation and the ‘asynchrony’ of surveillance systems [31]. Variations in reporting standards between jurisdictions often result in inconsistent datasets that undermine the accuracy of regional and global forecasts [69]. Bridging these domain-specific gaps requires not only better algorithms but also increased international cooperation in data sharing and methodological transparency [31,69].

5. Conclusions

Software now constitutes core translational infrastructure in biomedicine, enabling the conversion of heterogeneous measurements into decisions at molecular, clinical, and population scales. This review demonstrates that, despite differences in data types and algorithms, bioinformatics, molecular modeling, and epidemiological software share common dependencies on standards, reproducibility practices, and transparent validation.

Near-term gains are most likely to come from disciplined engineering and stewardship: transparent benchmarking, clear “data contracts” and traceability, user-centered interfaces built atop robust command-line/API foundations, and secure, portable deployments. Machine learning enhances sensitivity, specificity, and speed, and is poised to remain a key enabler across multiple domains.

Future progress depends on integrated, multimodal pipelines (omics, imaging, electronic health records), standardized evaluation with open datasets, sustained investment in training and equitable access to compute, durable open-source ecosystems, and strong safeguards for security, privacy, and ethical data use. Under these conditions, software will continue to shorten the distance between measurement and decision, offering the clearest and most compelling value proposition for biomedical science and public health.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

3D	three-dimensional
API	application programming interface
BAM	Binary Alignment/Map (alignment file format)
BLAST	Basic Local Alignment Search Tool
CRAM	Compressed Reference-oriented Alignment Map
CWL	Common Workflow Language
Epi Info	epidemiological information software package for public health surveillance
FASTQ	text format for nucleotide sequences with quality scores
GPU	graphics processing unit
GROMACS	GRoningen Machine for Chemical Simulations (molecular dynamics package)
HPC	high-performance computing
mzML	mass spectrometry markup language
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
NPI	non-pharmaceutical intervention
PubMed	public interface to the MEDLINE biomedical literature database
QGIS	Quantum Geographic Information System (open-source GIS platform)
RNA	ribonucleic acid
VCF	Variant Call Format
WDL	Workflow Description Language
SANRA	Scale for the Assessment of Narrative Review Articles
PDB	Protein Data Bank

References

1. Roukos, V.; Misteli, T.; Schmidt, C.K. Descriptive No More: The Dawn of High-Throughput Microscopy. *Trends Cell Biol.* **2010**, *20*, 503–506. <https://doi.org/10.1016/j.tcb.2010.06.008>.
2. Buggenthin, F.; Marr, C.; Schwarzfischer, M.; Hoppe, P.S.; Hilsenbeck, O.; Schroeder, T.; Theis, F.J. An Automatic Method for Robust and Fast Cell Detection in Bright Field Images from High-Throughput Microscopy. *BMC Bioinform.* **2013**, *14*, 297. <https://doi.org/10.1186/1471-2105-14-297>.
3. Pegoraro, G.; Misteli, T. High-Throughput Imaging for the Discovery of Cellular Mechanisms of Disease. *Trends Genet.* **2017**, *33*, 604–615. <https://doi.org/10.1016/j.tig.2017.06.005>.
4. Berkhout, M.; Smit, K.; Versendaal, J. Decision Discovery Using Clinical Decision Support System Decision Log Data for Supporting the Nurse Decision-Making Process. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 100. <https://doi.org/10.1186/s12911-024-02486-3>.
5. Pastorino, R.; De Vito, C.; Migliara, G.; Glocker, K.; Binenbaum, I.; Ricciardi, W.; Boccia, S. Benefits and Challenges of Big Data in Healthcare: An Overview of the European Initiatives. *Eur. J. Public Health* **2019**, *29*, 23–27. <https://doi.org/10.1093/eurpub/ckz168>.
6. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of Machine Learning in Drug Discovery and Development. *Nat. Rev. Drug Discov.* **2019**, *18*, 463–477. <https://doi.org/10.1038/s41573-019-0024-5>.
7. Zitnik, M.; Nguyen, F.; Wang, B.; Leskovec, J.; Goldenberg, A.; Hoffman, M.M. Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. *Inf. Fusion* **2019**, *50*, 71–91. <https://doi.org/10.1016/j.inffus.2018.09.012>.
8. Sevimoglu, T.; Arga, K.Y. The Role of Protein Interaction Networks in Systems Biomedicine. *Comput. Struct. Biotechnol. J.* **2014**, *11*, 22–27. <https://doi.org/10.1016/j.csbj.2014.08.008>.
9. Dwivedi, S.; Purohit, P.; Misra, R.; Pareek, P.; Goel, A.; Khattri, S.; Pant, K.K.; Misra, S.; Sharma, P. Diseases and Molecular Diagnostics: A Step Closer to Precision Medicine. *Indian J. Clin. Biochem.* **2017**, *32*, 374–398. <https://doi.org/10.1007/s12291-017-0688-8>.
10. Clark, A.J.; Lillard, J.W. A Comprehensive Review of Bioinformatics Tools for Genomic Biomarker Discovery Driving Precision Oncology. *Genes* **2024**, *15*, 1036. <https://doi.org/10.3390/genes15081036>.

11. Jamialahmadi, H.; Khalili-Tanha, G.; Nazari, E.; Rezaei-Tavirani, M. Artificial Intelligence and Bioinformatics: A Journey from Traditional Techniques to Smart Approaches. *Gastroenterol. Hepatol. Bed Bench* **2024**, *17*, 241–252. <https://doi.org/10.22037/ghfbb.v17i3.2977>.
12. Rosati, D.; Palmieri, M.; Brunelli, G.; Morrione, A.; Iannelli, F.; Frullanti, E.; Giordano, A. Differential Gene Expression Analysis Pipelines and Bioinformatic Tools for the Identification of Specific Biomarkers: A Review. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 1154–1168. <https://doi.org/10.1016/j.csbj.2024.02.018>.
13. Larson, N.B.; Oberg, A.L.; Adjei, A.A.; Wang, L. A Clinician’s Guide to Bioinformatics for next-Generation Sequencing. *J. Thorac. Oncol.* **2023**, *18*, 143–157. <https://doi.org/10.1016/j.jtho.2022.11.006>.
14. Lubin, I.M.; Aziz, N.; Babb, L.J.; Ballinger, D.; Bisht, H.; Church, D.M.; Cordes, S.; Eilbeck, K.; Hyland, F.; Kalman, L.; et al. Principles and Recommendations for Standardizing the Use of the Next-Generation Sequencing Variant File in Clinical Settings. *J. Mol. Diagn.* **2017**, *19*, 417–426. <https://doi.org/10.1016/j.jmoldx.2016.12.001>.
15. Ismail, F.N.; Amarasoma, S. Mars: Simplifying Bioinformatics Workflows through a Containerized Approach to Tool Integration and Management. *Bioinform. Adv.* **2025**, *5*, vbaf074. <https://doi.org/10.1093/bioadv/vbaf074>.
16. Luo, Y.; Zhao, C.; Chen, F. Multiomics Research: Principles and Challenges in Integrated Analysis. *Biodes. Res.* **2024**, *6*, 0059. <https://doi.org/10.34133/bdr.0059>.
17. Ferreira, L.G.; dos Santos, R.N.; Oliva, G.; Andricopulo, A.D. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* **2015**, *20*, 13384–13421. <https://doi.org/10.3390/molecules200713384>.
18. O’Donoghue, S.I.; Goodsell, D.S.; Frangakis, A.S.; Jossinet, F.; Laskowski, R.A.; Nilges, M.; Saibil, H.R.; Schafferhans, A.; Wade, R.C.; Westhof, E.; et al. Visualization of Macromolecular Structures. *Nat. Methods* **2010**, *7*, S42. <https://doi.org/10.1038/nmeth.1427>.
19. Leelananda, S.P.; Lindert, S. Using NMR Chemical Shifts and Cryo-EM Density Restraints in Iterative Rosetta-MD Protein Structure Refinement. *J. Chem. Inf. Model.* **2020**, *60*, 2522–2532. <https://doi.org/10.1021/acs.jcim.9b00932>.
20. Deshpande, A.; Margevicius, K.; Generous, E.; Taylor-McCabe, K.; Castro, L.; Longo, J.; Priedhorsky, R. Tools and Apps to Enhance Situational Awareness for Global Disease Surveillance. *Online J. Public Health Inform.* **2014**, *6*, e111. <https://doi.org/10.5210/ojphi.v6i1.5183>.
21. Deo, V.; Ranganathan, P. Statistical Tools and Packages for Data Collection, Management, and Analysis—A Brief Guide for Health and Biomedical Researchers. *Perspect. Clin. Res.* **2024**, *15*, 209–212. https://doi.org/10.4103/picr.picr_160_24.
22. Shaw, N.T. Geographical Information Systems and Health: Current State and Future Directions. *Healthc. Inform. Res.* **2012**, *18*, 88–96. <https://doi.org/10.4258/hir.2012.18.2.88>.
23. Opiyo, S.O.; Nalunkuma, R.; Nanyonga, S.M.; Mugenyi, N.; Kanyike, A.M. Empowering Global AMR Research Community: Interactive GIS Dashboards for AMR Data Analysis and Informed Decision-Making. *Wellcome Open Res.* **2024**, *9*, 234. <https://doi.org/10.12688/wellcomeopenres.21010.4>.
24. Sintchenko, V.; Roper, M.P.V. Pathogen Genome Bioinformatics. *Methods Mol. Biol.* **2014**, *1168*, 173–193. https://doi.org/10.1007/978-1-4939-0847-9_10.
25. Sathasivam, S.; Rajendran, K.; Logeswaran, K.; Periasamy, K.; Sharmila, V.; Sangeetha, M. *Integration of Multi-Omics Data: Genomics, Proteomics, Metabolomics*; IGI Global: Hershey, PA, USA, 2025; pp. 149–184, ISBN 979-8-3693-9521-9.
26. Subramanian, I.; Verma, S.; Kumar, S.; Jere, A.; Anamika, K. Multi-Omics Data Integration, Interpretation, and Its Application. *Bioinform. Biol. Insights* **2020**, *14*, 1177932219899051. <https://doi.org/10.1177/1177932219899051>.
27. Putrama, I.M.; Martinek, P. Heterogeneous Data Integration: Challenges and Opportunities. *Data Brief* **2024**, *56*, 110853. <https://doi.org/10.1016/j.dib.2024.110853>.
28. Bessen, J.L.; Alexander, M.; Foroughi, O.; Brathwaite, R.; Baser, E.; Lee, L.C.; Perez, O.; Gustavsen, G. Perspectives on Reducing Barriers to the Adoption of Digital and Computational Pathology Technology by Clinical Labs. *Diagnostics* **2025**, *15*, 794. <https://doi.org/10.3390/diagnostics15070794>.
29. Weber, L.M.; Saelens, W.; Cannoodt, R.; Soneson, C.; Hapfelmeier, A.; Gardner, P.P.; Boulesteix, A.-L.; Saeys, Y.; Robinson, M.D. Essential Guidelines for Computational Method Benchmarking. *Genome Biol.* **2019**, *20*, 125. <https://doi.org/10.1186/s13059-019-1738-8>.
30. Ye, Y.; Barapatre, S.; Davis, M.K.; Elliston, K.O.; Davatzikos, C.; Fedorov, A.; Fillion-Robin, J.-C.; Foster, I.; Gilbertson, J.R.; Lasso, A.; et al. Open-Source Software Sustainability Models: Initial White Paper From the Informatics Technology for Cancer Research Sustainability and Industry Partnership Working Group. *J. Med. Internet Res.* **2021**, *23*, e20028. <https://doi.org/10.2196/20028>.
31. Scheirer, M.A.; Dearing, J.W. An Agenda for Research on the Sustainability of Public Health Programs. *Am. J. Public Health* **2011**, *101*, 2059–2067. <https://doi.org/10.2105/AJPH.2011.300193>.

32. Cotto, K.C.; Feng, Y.-Y.; Ramu, A.; Richters, M.; Freshour, S.L.; Skidmore, Z.L.; Xia, H.; McMichael, J.F.; Kunisaki, J.; Campbell, K.M.; et al. Integrated Analysis of Genomic and Transcriptomic Data for the Discovery of Splice-Associated Variants in Cancer. *Nat. Commun.* **2023**, *14*, 1589. <https://doi.org/10.1038/s41467-023-37266-6>.
33. Bravo, A.M.; Typas, A.; Veening, J.-W. 2FASTQ: A General-Purpose Sequence Search and Counting Program for FASTQ Files. *PeerJ* **2022**, *10*, e14041. <https://doi.org/10.7717/peerj.14041>.
34. Sadedin, S.P.; Oshlack, A. Bazam: A Rapid Method for Read Extraction and Realignment of High-Throughput Sequencing Data. *Genome Biol.* **2019**, *20*, 78. <https://doi.org/10.1186/s13059-019-1688-1>.
35. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
36. Althagafi, A.; Zhapa-Camacho, F.; Hoehndorf, R. Prioritizing Genomic Variants through Neuro-Symbolic, Knowledge-Enhanced Learning. *Bioinformatics* **2024**, *40*, btae301. <https://doi.org/10.1093/bioinformatics/btae301>.
37. Reimand, J.; Isser, R.; Voisin, V.; Kucera, M.; Tannus-Lopes, C.; Rostamianfar, A.; Wadi, L.; Meyer, M.; Wong, J.; Xu, C.; et al. Pathway Enrichment Analysis and Visualization of Omics Data Using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **2019**, *14*, 482–517. <https://doi.org/10.1038/s41596-018-0103-9>.
38. Schmid, S.; Jeevannavar, A.; Julian, T.R.; Tamminen, M. Portable BLAST-like Algorithm Library and Its Implementations for Command Line, Python, and R. *PLoS ONE* **2023**, *18*, e0289693. <https://doi.org/10.1371/journal.pone.0289693>.
39. Sievers, F.; Higgins, D.G. Clustal Omega for Making Accurate Alignments of Many Protein Sequences. *Protein Sci.* **2018**, *27*, 135–145. <https://doi.org/10.1002/pro.3290>.
40. Lloyd, G.R.; Jankevics, A.; Weber, R.J.M. Struct: An R/Bioconductor-Based Framework for Standardized Metabolomics Data Analysis and Beyond. *Bioinformatics* **2020**, *36*, 5551–5552. <https://doi.org/10.1093/bioinformatics/btaa1031>.
41. Le Piane, F.; Vozza, M.; Baldoni, M.; Mercuri, F. Integrating High-Performance Computing, Machine Learning, Data Management Workflows, and Infrastructures for Multiscale Simulations and Nanomaterials Technologies. *Beilstein J. Nanotechnol.* **2024**, *15*, 1498–1521. <https://doi.org/10.3762/bjnano.15.119>.
42. Ahmed, A.E.; Allen, J.M.; Bhat, T.; Burra, P.; Fliege, C.E.; Hart, S.N.; Heldenbrand, J.R.; Hudson, M.E.; Istanto, D.D.; Kalmbach, M.T.; et al. Design Considerations for Workflow Management Systems Use in Production Genomics Research and the Clinic. *Sci. Rep.* **2021**, *11*, 21680. <https://doi.org/10.1038/s41598-021-99288-8>.
43. Yu, Y.; Mai, Y.; Zheng, Y.; Shi, L. Assessing and Mitigating Batch Effects in Large-Scale Omics Studies. *Genome Biol.* **2024**, *25*, 254. <https://doi.org/10.1186/s13059-024-03401-9>.
44. Terefe, E.M.; Ghosh, A. Molecular Docking, Validation, Dynamics Simulations, and Pharmacokinetic Prediction of Phytochemicals Isolated From Croton Dichogamus Against the HIV-1 Reverse Transcriptase. *Bioinform. Biol. Insights* **2022**, *16*, 11779322221125605. <https://doi.org/10.1177/11779322221125605>.
45. Varghese, A.; Liu, J.; Patterson, T.A.; Hong, H. Integrating Molecular Dynamics, Molecular Docking, and Machine Learning for Predicting SARS-CoV-2 Papain-like Protease Binders. *Molecules* **2025**, *30*, 2985. <https://doi.org/10.3390/molecules30142985>.
46. Trott, O.; Olson, A.J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. <https://doi.org/10.1002/jcc.21334>.
47. Wang, J.; Watson, J.L.; Lisanza, S.L. Protein Design Using Structure-Prediction Networks: AlphaFold and RoseTTAFold as Protein Structure Foundation Models. *Cold Spring Harb. Perspect. Biol.* **2024**, *16*, a041472. <https://doi.org/10.1101/cshperspect.a041472>.
48. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
49. Baek, M.; McHugh, R.; Anishchenko, I.; Jiang, H.; Baker, D.; DiMaio, F. Accurate Prediction of Protein–Nucleic Acid Complexes Using RoseTTAFoldNA. *Nat. Methods* **2024**, *21*, 117–121. <https://doi.org/10.1038/s41592-023-02086-5>.
50. Pinzi, L.; Rastelli, G. Molecular Docking: Shifting Paradigms in Drug Discovery. *Int. J. Mol. Sci.* **2019**, *20*, 4331. <https://doi.org/10.3390/ijms20184331>.
51. Ma, Z.; Ajibade, A.; Zou, X. Docking Strategies for Predicting Protein-Ligand Interactions and Their Application to Structure-Based Drug Design. *Commun. Inf. Syst.* **2024**, *24*, 199–230. <https://doi.org/10.4310/cis.241021221101>.
52. Blanes-Mira, C.; Fernández-Aguado, P.; de Andrés-López, J.; Fernández-Carvajal, A.; Ferrer-Montiel, A.; Fernández-Ballester, G. Comprehensive Survey of Consensus Docking for High-Throughput Virtual Screening. *Molecules* **2022**, *28*, 175. <https://doi.org/10.3390/molecules28010175>.

53. Kamenik, A.S.; Linker, S.M.; Riniker, S. Enhanced Sampling without Borders: On Global Biasing Functions and How to Re-weight Them. *Phys. Chem. Chem. Phys.* **2021**, *24*, 1225–1236. <https://doi.org/10.1039/d1cp04809k>.
54. Sutton, R.T.; Pincock, D.; Baumgart, D.C.; Sadowski, D.C.; Fedorak, R.N.; Kroeker, K.I. An Overview of Clinical Decision Support Systems: Benefits, Risks, and Strategies for Success. *NPJ Digit. Med.* **2020**, *3*, 17. <https://doi.org/10.1038/s41746-020-0221-y>.
55. Ogwu, M.; Izah, S. Innovations in Disease Surveillance and Monitoring. In *Technological Innovations for Managing Tropical Diseases*. *Health Information Science*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 83–108, ISBN 978-3-031-82621-4.
56. Su, Y.; Yoon, S.S. Epi Info—Present and Future. *AMIA Annu. Symp. Proc.* **2003**, *2003*, 1023.
57. Graser, A.; Sutton, T.; Bernasocchi, M. The QGIS Project: Spatial without Compromise. *Patterns* **2025**, *6*, 101265. <https://doi.org/10.1016/j.patter.2025.101265>.
58. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-Time Tracking of Pathogen Evolution. *Bioinformatics* **2018**, *34*, 4121–4123. <https://doi.org/10.1093/bioinformatics/bty407>.
59. Kolivand, P.; Azari, S.; Bakhtiari, A.; Namdar, P.; Ayyoubzadeh, S.M.; Rajaie, S.; Ramezani, M. AI Applications in Disaster Governance with Health Approach: A Scoping Review. *Arch. Public Health* **2025**, *83*, 218. <https://doi.org/10.1186/s13690-025-01712-2>.
60. Syed, R.; Eden, R.; Makasi, T.; Chukwudi, I.; Mamudu, A.; Kamalpour, M.; Kapugama Geeganage, D.; Sadeghianasl, S.; Leemans, S.J.J.; Goel, K.; et al. Digital Health Data Quality Issues: Systematic Review. *J. Med. Internet Res.* **2023**, *25*, e42615. <https://doi.org/10.2196/42615>.
61. Vitorino, R. Transforming Clinical Research: The Power of High-Throughput Omics Integration. *Proteomes* **2024**, *12*, 25. <https://doi.org/10.3390/proteomes12030025>.
62. Wratten, L.; Wilm, A.; Göke, J. Reproducible, Scalable, and Shareable Analysis Pipelines with Bioinformatics Workflow Managers. *Nat. Methods* **2021**, *18*, 1161–1168. <https://doi.org/10.1038/s41592-021-01254-9>.
63. D'Auria, G.; Schneider, M.V.; Moya, A. Live Genomics for Pathogen Monitoring in Public Health. *Pathogens* **2014**, *3*, 93–108. <https://doi.org/10.3390/pathogens3010093>.
64. Zhang, S.; Liu, K.; Liu, Y.; Hu, X.; Gu, X. The Role and Application of Bioinformatics Techniques and Tools in Drug Discovery. *Front. Pharmacol.* **2025**, *16*, 1547131. <https://doi.org/10.3389/fphar.2025.1547131>.
65. Leipzig, J.; Nüst, D.; Hoyt, C.T.; Ram, K.; Greenberg, J. The Role of Metadata in Reproducible Computational Research. *Patterns* **2021**, *2*, 100322. <https://doi.org/10.1016/j.patter.2021.100322>.
66. Čuklina, J.; Lee, C.H.; Williams, E.G.; Sajic, T.; Collins, B.C.; Rodríguez Martínez, M.; Sharma, V.S.; Wendt, F.; Goetze, S.; Keele, G.R.; et al. Diagnostics and Correction of Batch Effects in Large-scale Proteomic Studies: A Tutorial. *Mol. Syst. Biol.* **2021**, *17*, e10240. <https://doi.org/10.15252/msb.202110240>.
67. Huang, L.; Kim, Y.; Balluff, B.; Cillero-Pastor, B. Quality Control Standards for Batch Effect Evaluation and Correction in Mass Spectrometry Imaging. *Anal. Chem.* **2025**, *97*, 10919–10928. <https://doi.org/10.1021/acs.analchem.5c02020>.
68. Welzel, C.; Brückner, S.; Brightwell, C.; Fenech, M.; Gilbert, S. A Transparent and Standardized Performance Measurement Platform Is Needed for On-Prescription Digital Health Apps to Enable Ongoing Performance Monitoring. *PLoS Digit. Health* **2024**, *3*, e0000656. <https://doi.org/10.1371/journal.pdig.0000656>.
69. Tomov, L.; Chervenkov, L.; Miteva, D.G.; Batselova, H.; Velikova, T. Applications of Time Series Analysis in Epidemiology: Literature Review and Our Experience during COVID-19 Pandemic. *World J. Clin. Cases* **2023**, *11*, 6974–6983. <https://doi.org/10.12998/wjcc.v11.i29.6974>.
70. Suetake, H.; Fukusato, T.; Igarashi, T.; Ohta, T. Workflow Sharing with Automated Metadata Validation and Test Execution to Improve the Reusability of Published Workflows. *Gigascience* **2023**, *12*, giad006. <https://doi.org/10.1093/gigascience/giad006>.
71. Josten, C.; Lordan, G. Automation and the Changing Nature of Work. *PLoS ONE* **2022**, *17*, e0266326. <https://doi.org/10.1371/journal.pone.0266326>.
72. Mehdiyev, N.; Majlatow, M.; Fettke, P. Interpretable and Explainable Machine Learning Methods for Predictive Process Monitoring: A Systematic Literature Review. *Artif. Intell. Rev.* **2025**, *58*, 378. <https://doi.org/10.1007/s10462-025-11399-0>.
73. Faulkner, M.; Wells, M. A Conversation with Research Software Engineers at the International Brain Laboratory. *Patterns* **2025**, *6*, 101315. <https://doi.org/10.1016/j.patter.2025.101315>.
74. Arora, A.; Alderman, J.E.; Palmer, J.; Ganapathi, S.; Laws, E.; McCradden, M.D.; Oakden-Rayner, L.; Pfohl, S.R.; Ghassemi, M.; McKay, F.; et al. The Value of Standards for Health Datasets in Artificial Intelligence-Based Applications. *Nat. Med.* **2023**, *29*, 2929–2938. <https://doi.org/10.1038/s41591-023-02608-w>.

75. Diamant, I.; Clarke, D.J.B.; Evangelista, J.E.; Lingam, N.; Ma'ayan, A. Harmonizome 3.0: Integrated Knowledge about Genes and Proteins from Diverse Multi-Omics Resources. *Nucleic Acids Res.* **2024**, *53*, D1016–D1028. <https://doi.org/10.1093/nar/gkae1080>.
76. Malin, B.A.; Emam, K.E.; O'Keefe, C.M. Biomedical Data Privacy: Problems, Perspectives, and Recent Advances. *J. Am. Med. Inform. Assoc.* **2013**, *20*, 2–6. <https://doi.org/10.1136/amiajnl-2012-001509>.
77. Salih, S.; Hamdan, M.; Abdelmaboud, A.; Abdelaziz, A.; Abdelsalam, S.; Althobaiti, M.M.; Cheikhrouhou, O.; Hamam, H.; Alo-taibi, F. Prioritising Organisational Factors Impacting Cloud ERP Adoption and the Critical Issues Related to Security, Usability, and Vendors: A Systematic Literature Review. *Sensors* **2021**, *21*, 8391. <https://doi.org/10.3390/s21248391>.
78. Lee, M.; Kang, D.; Joi, Y.; Yoon, J.; Kim, Y.; Kim, J.; Kang, M.; Oh, D.; Shin, S.-Y.; Cho, J. Graphical User Interface Design to Improve Understanding of the Patient-Reported Outcome Symptom Response. *PLoS ONE* **2023**, *18*, e0278465. <https://doi.org/10.1371/journal.pone.0278465>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.