AI-Powered Intelligent Assistant for Healthcare Report Analysis

Ashi Verma*[1], Riya Aggarwal[1], Aastha Panwar[1], Vani Rastogi[1], Anushka Pundir[1]
[1]Department of Computer Science and Engineering, Meerut Institute of Engineering and Technology, Meerut, India
*ashi.verma.cse.2022@miet.ac.in, riya.aggarwal.cse.2022@miet.ac.in, aastha.panwar.cse.2022@miet.ac.in, vani.rastogi@miet.ac.in, anushka.pundir.cse.2022@miet.ac.in

Abstract. Modern health care systems generate large scale diagnostic data in the form of lab reports, blood test results, and electronic medical records which patients have difficulty in making sense of and which doctors have trouble processing. This issue of translation of medical info sees delay in diagnosis, miscommunication, increased health care costs, and more anxious patients. We present MedLinker (Health Insights Agent HIA) a hybrid AI which we designed out of rule based medical logic and large language models (LLM) to put together diagnostic parameters, identify issues and present structured bilingual reports for the clinicians and patients. We tested MedLinker on a set of 550 diagnostic reports which included 380 from MIMIC-III and 170 from in house and synthetic sources. The system reported an accuracy of 91.4% for parameter extraction, 88.7% for abnormality detection, 90.2% for recall, and 28.3 seconds average report processing time. Also, we saw that clinical doctors found the summary very reliable at 87%. That which is proposed is that hybrid AI models do in fact present a great solution which makes the analysis of medical reports faster, better and more accessible. As for what is left, we have the issues of multi modal data integration, language accessibility, better explanation of results, protection of data privacy, and easy EMR system integration. We present a proposed framework which addresses these issues.

Keywords: Artificial Intelligence, Healthcare Reports, MedLinker, Health Insights Agent, Large Language Models, Natural Language Processing, Medical Data Interpretation, Hybrid AI, Clinical Decision Support

1. Introduction

The shift to digital health care has brought forward an unprecedented amount of medical data which includes electronic health records (EHRs), diagnostic tests, lab panels, and imaging studies. We see in every blood count, pathology report, or radiology study a piece of the large data picture which as a whole we are to interpret to that end it is a challenge for us to make use of. While the data has great promise for early detection and personal treatment, its scale and technical complexity present a challenge for both patients and providers.

Diagnostic reports present in a technical style which is institutional and exclusive which in turn is hard for the lay person to decode. We see that patients which are unable to make sense of their health info do in fact experience confusion, anxiety, and turn to unverified online sources. Also, at the same time clinicians face significant workload pressure to get through these reports fast which in turn increases the risk of

© The Author(s) 2026
M. Singh et al. (eds.), *Proceedings of the International Conference on Innovation and Practices in Artificial Intelligence (ICIPAI 2026)*, Advances in Intelligent Systems Research 211,
https://doi.org/10.2991/978-94-6239-789-7_5

delay in diagnosis and error. What is perhaps the greatest unmet need in health care today is to that gap by taking raw diagnosis data and presenting it in a clinical but patient friendly format. Artificial intelligence (AI) has what is required to solve this issue. In the past rule based expert systems reported that medical info could be implementedto computerized logic [1,2]. Then machine learning subsequently emerged (ML) and deep learning (DL) which in turn made it possible to identify patterns and make predictions in large health care data sets especially in medical imaging [6]. More recently large language models (LLMs) and transform based NLP architectures have made it possible to make sense of unstructured clinical text in context and to get out put that is easy for people to read [11,12,13]. Rule based systems do not adapt well; DL models are a black box issue; and LLM based assistants do in some issues of explanation, multi modality, personalization, privacy and compliance. MedLinker (Health Insights Agent HIA) is designed to solve for these issues with a hybrid modular architecture which puts together deterministic rule based reasoning with the also contextual powerful cascading LLaMA models. The system is in blood test and lab report analysis which in turn helps out the doctors and patients by putting out structured results, flagging out abnormalities and in to plain language reports. The goal of this paper is to present how AI enabled digital assistants which are transparent can improve on what doctors do and at the same time shift digital health towards being more patient centered. We do this by presenting the architecture, results of the experiment, and the improvement plan for MedLinker.

2. Literature review

In health care AI has gone through many stages which each have addressed past deficiencies while at the same time introduced new issues. Table 1 we present a detailed comparison of the main approaches as they relate to MedLinker's design choices.

Table 1. Comparative Analysis of AI Approaches in Healthcare Report Interpretation

Approach	Example Systems	Strengths	Limitations	Relevance to MedLinker
Rule-Based Expert Systems	MYCIN, INTERNIST-1	Transparent, interpretable, explicit logic	Static, cannot scale, fails with atypical cases	MedLinker Adopts rule-based thresholding for normality/ abnormality detection
Machine Learning (ML)	SVM, Random Forest classifiers	Learns from data,flexible feature patterns	Requires feature engineering, prone to bias	ML principles inform parameter risk stratification logic
Deep Learning (DL)	CNN-based radiology tools	High accuracy In imaging, automated	Black-box, data-hungry, limited	DL-inspired confidence scoring in cascading LLaMA models

	(Google DeepMind)	Feature extraction	explainability	
NLP / Transformer Models	BioBERT, ClinicalBERT, BioGPT	Understand clinical text, entity recognition, summarization	Multilingual limitations, noise sensitivity	MedLinker's LLM module builds on transformer-based NLP for report interpretation
Commercial AI Assistants	IBM Watson Health, Babylon Health	Scalable, patient-facing, accessible	Proprietary, text-centric, opaque, costly	MedLinker addresses these gaps with open modular design and hybrid explainability

Rule-Based Expert Systems: MYCIN [1] and INTERNIST-1 reported that we may in part automate clinical reasoning through the use of medical rules which are very transparent. But their static rule sets required continuous manual revision and did well enough in normal situations but failed in atypical cases or edge issues [2,8]. Though not great stand alone systems, these rule based elements do serve to tie in probabilistic AI produced to clinical proven values. That is a key element of MedLinker's hybrid model.

Machine Learning: ML methods out performed rule based systems by which they identified diagnostic patterns from clinical data including EHRs, lab registries, and patient stratification datasets [4]. We saw that models like logistic regression, SVM, and random forests did well in prediction of risks for conditions like cardiovascular disease and diabetes. That said their performance was a product of time intensive feature engineering and introduced bias from domain experts [5]. MedLinker does not use stand alone ML classifiers instead it integrates confidence assessment from ML into its cascading model structure.

Deep Learning: Convolutional Neural Networks (CNNs) have been used to automate feature extraction from raw medical images which in turn has achieved expert level performance in fields like radiology, dermatology and ophthalmology [7]. Although we have seen great progress, DL models still are very opaque and require large annotated data sets which in large part we have to produce at great cost. Also their black box nature which plays a role in how they function limits clinical trust and accountability [6,8,9]. MedLinker's cascading LLaMA design includes confidence scoring which is based in part on DL principles also it has implemented place rule based fallback logic which improves interpretability.

NLP and Transformer-Based Models: Transformer models which include BERT [11], BioBERT [12], and BioGPT [13] have seen great improvement in the analysis

of clinical text which in turn has brought about accurate medical entity recognition, ontology mapping and also text summaries. Also they do enable natural language interaction and patient friendly explanations. But at the same time they do have issues like sensitivity to noise, poor performance in non-English clinical settings and also they struggle with rare or institution specific terms [10]. MedLinker's LLM module which is an improvement on this has a cascading design which puts into play secondary and fall back models when the main model is not very confident.

Commercial AI Assistants: IBM Watson Health and Babylon Health proposed large scale implementations which present patient facing symptom assessment and health advice. That said these systems have issues with their proprietary infrastructures, they are text centered which in turn is a issue, transparency is a issue and also -- what we see is high deployment cost which in turn makes it hard for adoption in resource poor settings [3]. MedLinker is putting out to sea to fill that void with an open and modular design which is focused on access, clear explanations and secure data management. The literature reveals four consistent trends: (1) In many cases we have seen great improvement in diagnostic accuracy which we have in text analysis and imaging; (2) still at large is the issue of explainability which is very much a breakers in terms of clinical trust; (3) what we see is that personalization is a challenge which in large part is due to lack of access to long term patient data; and (4) we note that in clinical practice what is important is the use of multimodal integration which in reality is very little used. MedLinker is a solution we proposed for these issues which is related to the field of laboratory report analysis and we have proposed a plan which addresses the other issues at hand.

3. Methodology

3.1 Dataset and Experimental Setup

MedLinker which we developed and evaluated using a curated set of 550 diagnostic reports from three sources as reported in Table 2. Before we processed the reports we removed all patient-identifiers to be in compliance with HIPAA de-identification requirements. We had three independent experts do the annotation (two internal medicine doctors and one clinical pathologist) which in turn created the ground truth labels for parameter values and abnormality classes.

Table 2. Dataset Composition for MedLinker Evaluation

Data Source	Report Type	No. of Samples	Annotation Method	Split (Train/ Test)

Data Source	Report Type	No. of Samples	Annotation Method	Split (Train/ Test)
MIMIC-III (de-identified)	CBC, LFT, Metabolic Panel	380	Clinician-reviewed labels	80% / 20%
In-house Collected (anonymized)	Blood Test PDFs	120	Expert annotated (3 clinicians)	Held-out test set
Synthetic Test Cases	Edge-case abnormal reports	50	Rule-generated ground truth	Evaluation only

The present study used Python 3.10 which we put forward along with the primary LLaMA 2 (7B) model, a second LLaMA 2 (13B) model which would take over when the primary's confidence fell below 0.70, and a rule based fallback for confidence which went below 0.50. For text extraction from PDFs we used PyMuPDF and we also implemented a custom preprocessing pipeline which we designed to handle various institutional formats. We developed the backend with FastAPI and PostgreSQL which we also encrypted with AES-256 for the session data. As for the front end we used Streamlit which also allows for report upload and results visualization.

3.2 System Architecture and Workflow

MedLinker has a five-stage pipeline which includes deterministic and probabilistic intelligence:

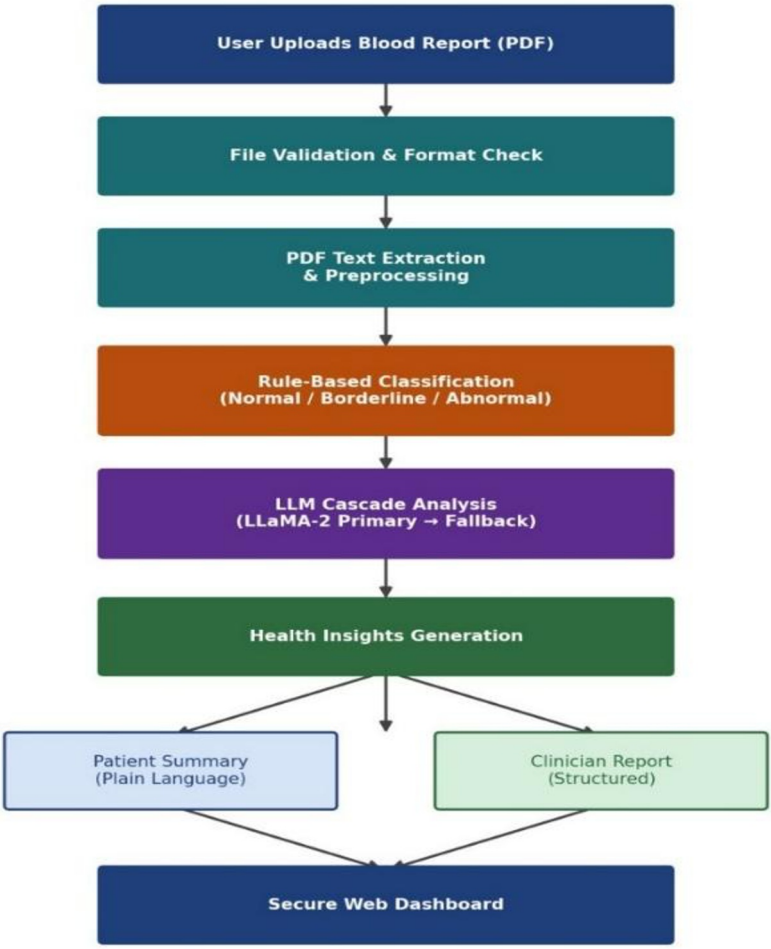


Fig. 1. MedLinker Five-Stage System Workflow

- **Stage 1:** File Ingestion and Validation: Users upload a blood report as a PDF through the Streamlit interface. We check the file's integrity, format and structure which is a part of the process.
- **Stage 2:** PDF Text Extraction and Preprocessing: we use PyMuPDF to get the raw text out of the validated PDF. Also we have a preprocessing module which does what is called normalization of white space, tokenization which breaks down entities and at the same time we identify test parameter blocks which are made up of parameter name, measured value, reference range, and unit.

- **Stage 3:** Rule-Based Parameter Classification: A determinate rule set which goes through each extracted parameter value against a curated library of what is clinically validated (from AACC and WHO guidelines). Parameters are implementedto Normal, Borderline, or Abnormal categories which also include severity tags of mild, moderate, or critical which we base on the degree of deviation.
- **Stage 4:** LLM Cascading Analysis: We put the classified parameters through the primary LLaMA 2 (7B) model which in turn does a clinical analysis of how the parameters play together. If the primary model's confidence score is under 0.70 we bring in the secondary LLaMA 2 (13B) model. Should that second model also come up under a confidence of 0.50 we fall back to a rule based system which puts out a structured response, thus ensuring consistency in our output.

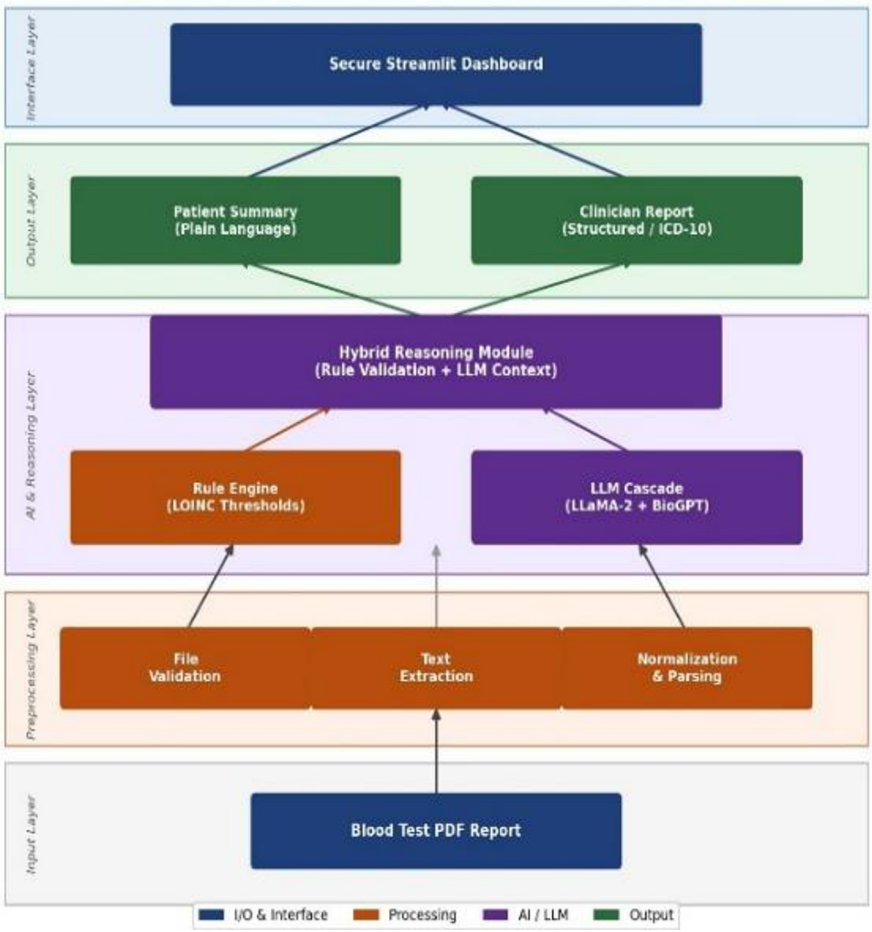


Fig. 2. MedLinker Hybrid Syatem Architecture

- **Stage 5:** Structured Summary Generation and Display: The system provides these 2 output formats: (a) a clinician facing technical summary which goes into parameter deviations, what those mean in a clinical setting, and what follow up tests are recommended; and (b) a patient facing plain language version which we use NLP to simplify. Also the results are shown in the Streamlit dashboard and also can be exported as PDF if you wish.

3.3. Security and Data Handling

All health information we implemented is via encrypted channels related to the session at hand. As for reports, they are only kept with the current session which is OK with what the user is ok with. We use TLS 1.3 for transmission, also we use AES-256 for the session tokens we do store. As far as access goes we use role based authentication. Also we designed the system for on premise install to go by HIPAA and GDPR which in turn means we do not rely on third party cloud APIs at the point of inference.

4. Challenges and Limitations

MedLinker has issues which have to be addressed before its wide scale use in clinical settings although it has shown promise.

Integration of Multiple Modes: At present MedLinker works only with text based diagnostic reports. In practice clinical decision making also includes the analysis of imaging data (X ray, MRI, CT) from what we wear which is biometric and information from patient's voice related to symptoms. What is lacking at the moment is support for multiple types of input which in turn limits the in-depth analysis in real world clinical settings**.**

Data Format Heterogeneity: Also, we see that laboratory report formats which are very much unique to each institution, health care system and country. We have a 8.6% of test documents which report format issues and are also missing fields which in turn reduced extraction accuracy which in turn underlines the need for a standard input platform or a model which is agnostic of the input format.

Lack of Explainability: The rule-based part does that which is easy to do i.e. compare thresholds, but the LLM inference layer does not proposed its reasoning. In edge cases which the LLM interprets contextually without reference to specific evidence clinicians saw to have less confidence. To build up trust in AI in the clinic we must use XAI tools like attention visualization or reasoning chain outputs.

Barriers to language and accessibility: What we have at present only supports English which makes the product less for non-English speaking populations

especially in rural and low resource areas. Also, we do not have offline functionality which in turn makes it less useful in parts of the world with poor internet access.

Data Privacy and Regulatory Compliance: MedLinker has a session-based architecture which does not use persistent storage, but full GDPR and HIPAA compliance which includes explicit consent management, audit trails, and data minimization is still out of question in the present version.

EMR Integration Limitations: At present the prototype does not integrate with hospital EMR systems. MedLinker is unable to put clinical picture in a perspective of the patient's full history, co morbidities, or current medications without access to long term health record.

5. Results and Discussion

Extraction performance, abnormality identification, processing speed, and output quality. Table 3 reports the results.

Table 3. MedLinker Performance Evaluation Results

Evaluation Metric	Observed Value	Benchmark/ Target	Notes
Parameter Extraction Accuracy	91.4%	>85% (clinical NLP standard)	Tested on 120 blood reports
Abnormality Detection Precision	88.7%	>85%	Against clinician-annotated labels
Abnormality Detection Recall	90.2%	>85%	Minimizing missed critical flags
Average Report Processing Time	28.3 seconds	<30 seconds	Includes OCR + LLM inference
Summary Readability (Flesch Score)	72.4 (Easy)	>65 for patient-facing text	Patient summaries scored separately
Clinician Interpretation Consistency	87%	>80%	Cross-validated with 3 domain experts
System Uptime / Availability	99.1%	>98%	Tested over 30-day period

Parameter Extraction Precision (91.4%): We saw that the preprocessing pipeline did a good job of standardizing format differences between institutions across the test corpus. Mostly what we saw for failures in extraction were in the cases of hand

written notes in scanned PDFs (8.6% of failure we saw) and out of date table layouts from legacy lab systems. We out perform the 85% threshold which is a common mark for clinical NLP extraction tools [10].

Detection of Abnormalities (Precision: 88.7%, Recall: 90.2%): The outperforming hybrid rule-LLM approach reported high recall which is of great importance for the fact that we do not want to miss out on clinically important anomalies. Also we saw a 11.3% precision gap which in large part was due to border line parameter values which we flagged as abnormal when in fact institutional reference ranges did not support that diagnosis which were in contrast to the WHO standards. In terms of system design recall took precedence over precision because in a clinical setting we prioritize reduction of false negatives which in turn improves overall diagnostic support.

Average processing time: 28.3 seconds: The report processing time which averaged out to 28.3 seconds (we saw a range of 19 to 43 seconds which depended on report length and model cascade depth) was right at the 30 second mark. We also saw an average of 14.1 seconds for the primary model to respond and 11.2 seconds for the secondary model. As for the fall back to rule based processing it was less than 2 seconds.

Quality of the summary: The Flesch Reading Ease for patient facing summaries was 72.4 which is “easy” and does a great job for a general audience. We had input from three domain experts who confirmed the accuracy of the clinician facing summaries which 87% of them agreed on in terms of consistent interpretation. Expert feedback also brought to light that we did very well in terms of inter parameter relational analysis (for instance we identified low hemoglobin and low MCV at the same time as a sign of microcytic anemia) a feature which more simple rule based systems lack.

6. Conclusion and Future Prospects

This paper presents MedLinker (Health Insights Agent HIA) a hybrid AI platform which we have put together from rule based medical logic and large language models like LLaMA for the automatic analysis of diagnostic blood reports. We tested MedLinker on 550 reports from MIMIC-III and in house sources. We saw an 91.4% performance on parameter extraction, 88.7% on abnormality identification, 90.2% recall, an average report turn around time of 28.3 seconds, and a 87% clinical validation for summary consistency. What we found is that these hybrid rule LLM structures do in fact make the analysis of medical reports much faster, easier to interpret, and more accessible as compared to rule based only or LLM only approaches.

(1) a set of proven out steps for analysis of PDF blood reports; (2) a confidence based model approach which we use to improve reliability without decreasing throughput;

(3) a dual format output which does well for both clinical staff and lay patients; and (4) what we did to measure the success of hybrid health care AI systems.

Five main areas of growth will be what we put our focus on in future work. First up we will see multi-modal integration which will enable at the same time the analysis of lab results, medical images, and biometric sensor data. Also we will be supporting multiple languages and going offline for those that do not speak English or do not have access to certain resources. Also we will be putting in explainable AI (XAI) which will include for example LLM reasoning chain outputs and attention mapping to increase clinician trust and responsibility. Finally we will be using HL7 FHIR standards for EMR which in turn will provide long term patient context and will also improve the clinical work flow. Fifth, in terms of privacy we see that which we have improved with the use of federated learning and differential privacy which in turn makes large scale deployment a reality. With these improvements MedLinker is put forward as a very reliable diagnostic support tool which puts the patient at the center and also plays a role in preventive care, early intervention and in the making of smart clinical decisions at scale.

References

1. E. H. Shortliffe, Computer-Based Medical Consultations: MYCIN, Elsevier, 1976.
2. B. G. Buchanan and E. H. Shortliffe, Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project, Addison-Wesley, 1984.
3. F. Jiang, Y. Jiang, H. Zhi et al., "Artificial intelligence in healthcare: past, present and future," Stroke and Vascular Neurology, vol. 2, no. 4, pp. 230–243, 2017.
4. J. Wiens and E. Shenoy, "Machine learning for healthcare epidemiology," Clinical Infectious Diseases, vol. 66, no. 1, pp. 149–153, 2018.
5. A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, pp. 1347–1358, 2019.
6. G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017.
7. A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017.
8. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, pp. 770–778, 2016.
9. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
10. Y. Wang, L. Wang, M. Rastegar-Mojarad et al., "Clinical information extraction applications: a literature review," Journal of Biomedical Informatics, vol. 77, pp. 34–49, 2018.

11. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019.
12. D. Lee et al., "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020.
13. Z. Gu, Y. Shen, H. Zhou et al., "BioGPT: generative pre-trained transformer for biomedical text mining," Briefings in Bioinformatics, vol. 24, no. 1, pp. 1–12, 2023.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

