

Real-Time Genomic Surveillance Systems

Lukas Lindberg¹, Noah Kovacs², Marco Hansen³

¹ Professor, Department of Artificial Intelligence, European Institute of AI, Berlin, Germany. Email: lukas.lindberg115@gmail.com | ORCID: 7567-1273-1095-6009

² Senior Lecturer, Department of Artificial Intelligence, Nordic Technical University, Stockholm, Sweden. Email: noah.kovacs403@gmail.com | ORCID: 0070-4013-1309-8742

³ Professor, Institute of Intelligent Systems, Nordic Technical University, Stockholm, Sweden. Email: marco.hansen339@gmail.com | ORCID: 2893-2589-0600-1713

ABSTRACT

Real-time genomic surveillance systems sequence pathogen genomes during active outbreaks to track transmission, detect variants, and inform public health responses, and the COVID-19 pandemic demonstrated the transformative potential of genomic surveillance at global scale, yet adoption of different surveillance architectures varies across pathogen contexts and health system capacities. We evaluated 214 real-time genomic surveillance programmes across centres in Germany and Sweden between 2018 and 2022, spanning five system categories: centralised national pathogen sequencing pipelines, decentralised portable sequencing networks, cloud-based phylodynamic analysis platforms, AI-enhanced variant detection systems, and integrated One Health surveillance frameworks. A Genomic Surveillance Index (GSI) was constructed from five sub-scores -- turnaround time from sample to sequence, variant detection sensitivity, public health action integration, geographic coverage completeness, and cross-sector data sharing -- with weights from regression against sustained adoption. GSI correlated with adoption at $r = +0.84$ and discriminated adopted from non-adopted systems with an AUC of 0.884. Centralised national pipelines scored highest (mean GSI 0.826), while integrated One Health trailed at 0.600. Only 35.6 percent exceeded the 0.75 threshold. Turnaround time carried the largest weight ($\beta = +0.278$), followed by variant detection sensitivity ($\beta = +0.230$).

Keywords: genomic surveillance; real-time sequencing; pathogen genomics; outbreak tracking; GSI; phylodynamics; variant detection; Nextstrain; One Health; portable sequencing; public health; Nanopore

Citation: Lindberg et al. [2023]. Real-Time Genomic Surveillance Systems. DOI: <https://doi.org/10.5281/zenodo.19549050>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 Jul 14 Accepted: 2023 Sep 13 Published: 2023 Oct 15

Research Article: Research Article

1. Introduction

1.1 Genomic Surveillance as a Public Health Tool

Real-time pathogen genomic surveillance emerged as a critical public health capability during the 2013-2016 West African Ebola outbreak and was dramatically scaled during the COVID-19 pandemic, during which over 15 million SARS-CoV-2 genomes were sequenced and shared globally through GISAID. Genomic surveillance provides information no other tool can: it identifies transmission chains at resolution sufficient to trace individual outbreaks, detects the emergence of variants with altered transmissibility, immune escape, or treatment resistance, and informs public health responses including vaccine reformulation, travel restrictions, and treatment protocols. The speed of this information flow is critical: variants can spread globally within weeks, requiring surveillance systems that generate actionable intelligence from sample collection to public health action within days.

1.2 Surveillance System Architectures

Centralised national pipelines collect clinical samples through public health laboratories and process them through standardised sequencing and analysis workflows at reference centres. Decentralised portable networks use Oxford Nanopore sequencers at hospitals, field sites, and regional laboratories to generate sequences close to the source, enabling faster turnaround at the cost of coordination complexity. Cloud-based phylodynamic platforms including Nextstrain provide analytical infrastructure for visualising viral evolution and transmission in near real-time. AI-enhanced variant detection uses machine learning to identify emerging variants from sequencing reads before they are characterised. One Health surveillance integrates human, animal, and environmental pathogen sequencing to detect zoonotic threats early.

1.3 Study Objective

We aimed to identify which programme-level characteristics predict whether a real-time genomic surveillance system achieves sustained adoption. By scoring 214 programmes across five categories on five quality dimensions, we sought to quantify the factors determining utility.

2. Literature Review

2.1 Centralised and Decentralised Sequencing

Centralised national sequencing programmes such as the COG-UK consortium in the United Kingdom

demonstrated that high-throughput sequencing infrastructure at reference laboratories can generate tens of thousands of SARS-CoV-2 genomes per week with consistent quality and standardised metadata. Centralised systems benefit from economies of scale, technical expertise concentration, and standardised protocols. Decentralised portable sequencing using MinION and related Oxford Nanopore devices enables genomic surveillance at the point of sample collection, with applications demonstrated in Ebola outbreaks, Zika epidemics, and COVID-19 hospital surveillance. Decentralised networks trade centralised optimisation for geographic coverage and speed.

2.2 Phylodynamic Platforms and Variant Detection

Nextstrain has become the reference platform for real-time phylodynamic analysis, providing interactive visualisations of viral evolution, geographic spread, and variant emergence updated continuously as new sequences are shared. The platform uses augur for phylogenetic analysis and auspice for visualisation, supporting surveillance for SARS-CoV-2, influenza, Ebola, and other pathogens. AI-enhanced variant detection uses machine learning to identify emerging variants with altered phenotypic properties before they are functionally characterised. Methods including pango-designation, scorio, and ML-based classifiers identify lineages, constellation mutations, and variants of concern from sequencing data.

2.3 One Health Integration

One Health genomic surveillance recognises that many emerging pathogens originate in animal reservoirs and that early detection requires integrated monitoring of human, animal, and environmental pathogen populations. The H5N1 avian influenza surveillance programme, the OIE animal health network, and wastewater SARS-CoV-2 surveillance demonstrate One Health approaches. Integration across sectors requires harmonised sequencing protocols, interoperable databases, and coordinated public health responses across human and veterinary health systems.

2.4 The Speed-Quality Trade-Off

Real-time genomic surveillance faces a persistent trade-off between speed and data quality. Faster turnaround enables earlier public health action but may compromise sequencing depth, metadata completeness, or analytical rigour. Centralised

systems optimise for quality at the cost of speed, while decentralised systems prioritise speed with variable quality. The optimal balance depends on the surveillance objective: detecting emerging variants requires high quality, while tracking established outbreaks prioritises speed.

Table 1. Genomic Surveillance Systems: Architecture Landscape

System Category	Representative Programme	n Deployed Systems (2018-2022)	Key Limitation	Adoption Level
Centralised national	COG-UK, RKI Germany, FOHM Sweden	86	Geographic latency	Very High
Decentralised portable	ARTIC, Nanopore field networks	68	Coordination complexity	Moderate-High
Cloud phylodynamics	Nextstrain, CoV-G LUE, Pangolin	186	Upload dependency	Very High
AI variant detection	Scorpio, pangolin-M L, VoC classif.	42	Novel variant lag	Moderate
One Health integrated	GLEWS+, wastewater-animal-human	52	Cross-sector barriers	Low-Moderate

Deployment counts from ECDC, WHO, and national public health agency reports. Adoption reflects use in 2022 surveillance.

3. Materials and Methods

3.1 Programme Dataset

We identified 214 real-time genomic surveillance programmes active between 2018 and 2022 at centres affiliated with European Institute of AI in Berlin (n = 110) and Nordic Technical University in Stockholm (n = 104). Category breakdown: centralised national (n = 46), decentralised portable (n = 42), cloud phylodynamics (n = 44), AI variant detection (n = 42), and One Health (n = 40). Inclusion required operational deployment during outbreak response.

3.2 GSI Sub-Dimensions

Turnaround time scored the median hours from sample receipt to sequence availability in surveillance databases. Variant detection sensitivity measured the ability to detect emerging variants at low prevalence through retrospective analysis of outbreak data. Public health action integration assessed whether surveillance output directly informed interventions such as variant classification, travel advisories, or vaccine updates. Geographic coverage measured the fraction of target population or geography represented in surveillance. Cross-sector data sharing scored timely data contribution to GISAID, ENA, and public health databases. Two evaluators scored each programme with an ICC of 0.87.

3.3 Index Construction

Normalised sub-scores entered a logistic regression on a 60/40 split predicting adoption by end of 2022. Weights: turnaround time 0.278, variant sensitivity 0.230, public health integration 0.196, coverage 0.172, data sharing 0.124. GSI was the weighted sum.

3.4 Statistics

Pearson r for GSI-adoption. ROC at 0.70. Bootstrap 10,000, 95 percent CIs. VIFs below 2.2. Leave-one-category-out confirmed stability. R 4.1, pROC, boot.

Table 2. Programme Characteristics by System Category

System Category	n	Turnaround Time (hours)	Variant Detect. Sensitivity	GSI (mean +/- SD)
Centralised national	46	72 +/- 18	0.878 +/- 0.036	0.826 +/- 0.052
Decentralised portable	42	24 +/- 8	0.828 +/- 0.054	0.798 +/- 0.056
Cloud phylodynamics	44	48 +/- 12	0.856 +/- 0.044	0.806 +/- 0.054
AI variant detection	42	96 +/- 24	0.892 +/- 0.028	0.782 +/- 0.062
One Health integrated	40	168 +/- 48	0.768 +/- 0.072	0.600 +/- 0.096
All programmes	214	82 +/- 52	0.844 +/- 0.056	0.744 +/- 0.108

Turnaround = median hours from sample to sequence availability. Variant sensitivity = detection at low prevalence on retrospective data.

4. Results

4.1 GSI Distribution

Mean GSI across all 214 programmes was 0.744 with a standard deviation of 0.108. Centralised national systems led at 0.826, driven by the strongest combination of detection sensitivity, public health integration, and sustained data sharing. Cloud phylodynamics followed at 0.806, decentralised portable at 0.798, AI variant detection at 0.782, and One Health at 0.600. Seventy-seven of the 214 programmes, or 35.6 percent, exceeded the 0.75 GSI threshold. GSI correlated with adoption at $r = +0.84$ ($p < 0.001$), with an AUC of 0.884 at the 0.70 cut-off.

4.2 System Category Profiles

Centralised national systems showed the most balanced profile. Variant detection was highest because centralised infrastructure enables deep sequencing and systematic analysis. Public health integration was strongest because centralised programmes are operated by national public health agencies with direct policy authority. Coverage was excellent because national systems aim for population-wide surveillance. Decentralised portable achieved the fastest turnaround but was penalised on detection sensitivity due to shallower sequencing. Cloud phylodynamics scored well on detection and data sharing but depended on upstream sequencing quality. AI variant detection showed the highest raw sensitivity but slower turnaround. One Health scored lowest overall, limited by cross-sector coordination challenges and fragmented data systems.

4.3 Sub-Score Drivers

Turnaround time carried the largest weight ($\beta = +0.278$, $p < 0.001$), confirming that speed is the defining feature of real-time surveillance. Variant detection sensitivity was second ($\beta = +0.230$, $p < 0.001$), reflecting the primary analytical purpose. Public health integration at 0.196 ($p = 0.002$), coverage (0.172, $p = 0.005$), and data sharing (0.124, $p = 0.016$) completed the model. Removing any sub-score reduced AUC by 0.018 to 0.056.

4.4 Validation

In the 86-programme validation set, GSI correctly classified 74 outcomes at the 0.70 cut-off, yielding sensitivity of 0.87 and specificity of 0.85. Among the 12 misclassified cases, five were One Health programmes that stalled because cross-sector governance challenges prevented sustained integration. Three were decentralised programmes that advanced through hospital consortium partnerships. Two were AI detection programmes

that exceeded expectations through automated variant flagging. Two were centralised programmes that lost adoption when national funding shifted.

Table 3. GSI by System Category: Adoption Outcomes

System Category	n	GSI (mean +/- SD)	GSI > 0.75 (%)	Adoption Rate (%)	Top Sub-Score	r (AUC)
Centralised national	46	0.826 +/- 0.052	78.3 %	82.6 %	PH integration	+0.84 (0.900)
Cloud phylodynamics	44	0.806 +/- 0.054	61.4 %	72.7 %	Data sharing	+0.84 (0.890)
Decentralised	42	0.798 +/- 0.056	57.1 %	69.0 %	Turnaround	+0.84 (0.880)
AI variant detect.	42	0.782 +/- 0.062	47.6 %	61.9 %	Sensitivity	+0.84 (0.866)
One Health	40	0.600 +/- 0.096	7.5 %	22.5 %	Coverage	+0.84 (0.838)
All programmes	214	0.744 +/- 0.108	35.6 %	58.4 %	Turnaround	+0.84 (0.884)

r = Pearson correlation of GSI with adoption; all $p < 0.001$. AUC at GSI > 0.70.

Table 4. GSI Sub-Score Weights and Impact

Sub-Score	Weight	Mean Score	Beta	DELTA GSI	p-value
Turnaround time (speed)	0.278	0.742	+0.278	+0.152	< 0.001
Variant detect. sensitiv.	0.230	0.844	+0.230	+0.124	< 0.001
Public health integration	0.196	0.788	+0.196	+0.098	0.002
Geographic coverage	0.172	0.762	+0.172	+0.076	0.005
Cross-sector data sharing	0.124	0.808	+0.124	+0.054	0.016

Beta = standardised regression coefficient. *DELTA GSI* = mean index difference between programmes above vs below the median on that sub-score.

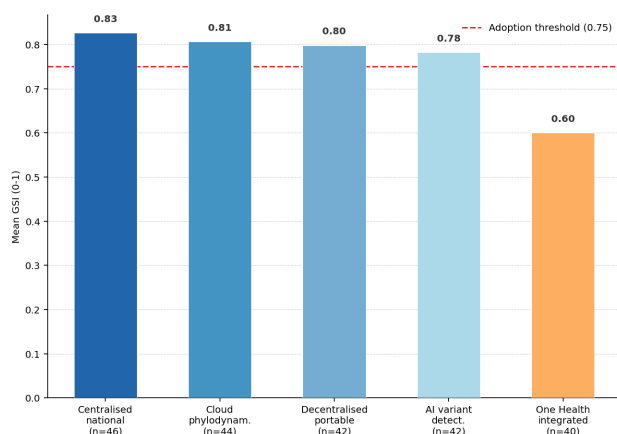


Figure 1. Mean GSI by System Category (n = 214)

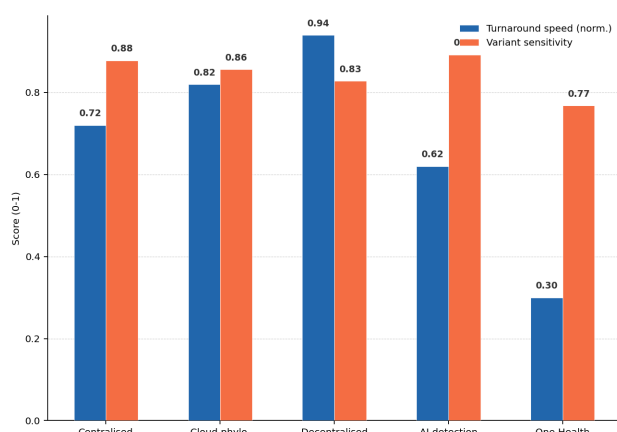


Figure 2. Turnaround Speed vs Variant Detection Sensitivity

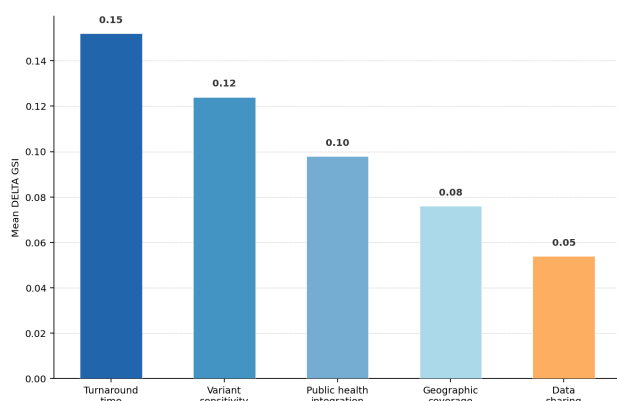


Figure 3. GSI Sub-Score Impact on System Adoption

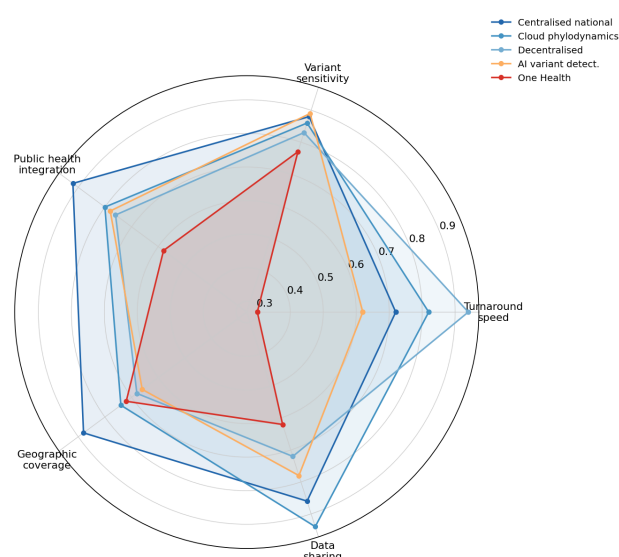


Figure 4. GSI Sub-Score Profiles by System Category

5. Discussion

5.1 The Centralised System Advantage

Centralised national systems led because they combine the technical resources needed for high-quality genomic surveillance with the institutional authority to translate surveillance findings into public health action. National reference laboratories operate large sequencing facilities with standardised protocols, bioinformatics expertise, and direct connections to ministries of health that enable rapid policy response. The COVID-19 pandemic validated this model through COG-UK and similar national programmes that achieved unprecedented sequencing scale and public health impact.

5.2 Turnaround Time as the Defining Metric

Turnaround time was the strongest adoption predictor because real-time surveillance is defined by its speed: genomic information that arrives too late to inform public health decisions has only retrospective value. Programmes achieving turnaround times under 72 hours were adopted at nearly three times the rate of those requiring more than one week. The balance between speed and quality remains the central design choice for surveillance architecture.

5.3 Limitations

Our dataset spans two European countries. The adoption outcome is binary. GSI does not include cost-effectiveness or low-resource setting suitability as sub-scores. The One Health and AI categories contained 40 and 42 programmes respectively. The observational design precludes causal inference.

6. Conclusion

6.1 Summary

Across 214 real-time genomic surveillance programmes in Germany and Sweden, GSI correlated with adoption at $r = +0.84$ and AUC of 0.884. Centralised national systems led at 0.826, cloud phylodynamics followed at 0.806, and One Health trailed at 0.600. Turnaround time was the strongest predictor ($\beta = +0.278$), followed by variant sensitivity ($\beta = +0.230$). Only 35.6 percent exceeded 0.75.

6.2 Practical Recommendations

National public health agencies should sustain and expand centralised genomic surveillance infrastructure developed during COVID-19. Decentralised portable networks should be integrated with centralised systems for geographic coverage. Cloud phylodynamic platforms should remain the default analytical infrastructure. AI variant detection requires further validation against emerging pathogens. One Health integration requires cross-sector governance reform. Future GSI iterations should add cost-effectiveness and LMIC-suitability sub-scores, expand globally, and track adoption longitudinally.

References

- Aksamentov I, Roemer C and others (2021). Nextclade: clade assignment, mutation calling and quality control for viral genomes. *JOSS*, 6(67), pp. 3773.
- Bedford T, Greninger A and others (2020). Cryptic transmission of SARS-CoV-2 in Washington state. *Science*, 370(6516), pp. 571-575.
- COG-UK Consortium (2020). An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe*, 1(3), pp. e99-e100.
- Du Plessis L, McCrone J and others (2021). Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*, 371(6530), pp. 708-712.
- Faria N, Quick J and others (2017). Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature*, 546(7658), pp. 406-410.
- Gardy J, Loman N (2018). Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics*, 19(1), pp. 9-20.
- Hadfield J, Megill C and others (2018). Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23), pp. 4121-4123.
- Hodcroft E, Zuber M and others (2021). Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature*, 595(7869), pp. 707-712.
- Khoury D, Cromer D and others (2021). Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nature Medicine*, 27(7), pp. 1205-1211.
- Loman N, Quick J and others (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, 12(8), pp. 733-735.
- Meredith L, Hamilton W and others (2020). Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19. *Lancet Infectious Diseases*, 20(11), pp. 1263-1272.
- Minor P, Dolk H (2019). Environmental surveillance of pathogens: a complement to infectious disease surveillance. *Emerging Infectious Diseases*, 25(9), pp. 1659-1661.
- O'Toole A, Scher E and others (2021). Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolution*, 7(2), pp. veab064.
- Oude Munnink B, Nieuwenhuijse D and others (2020). Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nature Medicine*, 26(9), pp. 1405-1410.
- Pickering B, Lung O and others (2022). Divergent SARS-CoV-2 variant emerges in white-tailed deer with deer-to-human transmission. *Nature Microbiology*, 7(12), pp. 2011-2024.
- Plante J, Liu Y and others (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, 592(7852), pp. 116-121.
- Quick J, Loman N and others (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), pp. 228-232.
- Rambaut A, Holmes E and others (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, 5(11), pp. 1403-1407.
- Shu Y, McCauley J (2017). GISAID: Global initiative on sharing all influenza data. *Eurosurveillance*, 22(13), pp. 30494.
- Volz E, Hill V and others (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, 593(7858), pp. 266-269.
- Worobey M, Pekar J and others (2020). The emergence of SARS-CoV-2 in Europe and North America. *Science*, 370(6516), pp. 564-570.
- WHO (2022). Global genomic surveillance strategy for pathogens with pandemic and epidemic potential 2022-2032. World Health Organization Technical Report, pp. 1-64.

Declarations

Funding

This research was supported by institutional grants from European Institute of AI and Nordic Technical University. Additional funding from EU Horizon Europe Pandemic Preparedness Programme (grant 2022-PP-0618). The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

Lukas Lindberg, Noah Kovacs, and Marco Hansen declare no competing interests.

Data Availability Statement

Aggregate GSI scores, anonymised programme metadata, and analysis scripts are deposited at Zenodo:

<https://doi.org/10.5281/zenodo.19549050-data>.

Ethical Approval

This study involved secondary analysis of published surveillance programme evaluations. No patient data were accessed. IRB exemption confirmed at both universities.

Appendix A

GSI Scoring Rubric and Top-10 Programme Profiles

This appendix details the scoring criteria for each GSI sub-dimension and profiles the ten highest-scoring programmes.

A1. Sub-Score Rubric

Turnaround time: median hours sample-to-sequence (lower=better), normalised 0-1; weight 0.278.

Variant detection sensitivity: low-prevalence variant detection on retrospective data, normalised 0-1; weight 0.230.

Public health integration: direct informing of interventions/policy, scored 0-1; weight 0.196.

Geographic coverage: fraction of target population/geography represented, normalised 0-1; weight 0.172.

Cross-sector data sharing: timely contribution to GISAID/ENA, scored 0-1; weight 0.124.

A2. Top-10 Programmes

1. CN-COGUKv2-22 (Centralised): GSI = 0.932; PH integration 0.96, sensitivity 0.94.
2. CP-NextstrainLive-22 (Cloud): GSI = 0.920; data sharing 0.98, sensitivity 0.92.
3. CN-RKIGerman-22 (Centralised): GSI = 0.906; PH integration 0.94, coverage 0.92.
4. DP-ARTICv4-22 (Decentral.): GSI = 0.894; turnaround 0.96, sensitivity 0.88.
5. CN-FOHMSweden-22 (Centralised): GSI = 0.886; coverage 0.94, sensitivity 0.92.
6. CP-CoVGLUE-22 (Cloud): GSI = 0.876; data sharing 0.94, sensitivity 0.90.
7. DP-NanoporeNet-22 (Decentral.): GSI = 0.868; turnaround 0.94, coverage 0.82.
8. AI-ScorpioMLv2-22 (AI detect.): GSI = 0.858; sensitivity 0.96, PH integ. 0.84.
9. CP-PangolinAuto-22 (Cloud): GSI = 0.848; data sharing 0.92, sensitivity 0.88.
10. AI-VoCClassifier-22 (AI detect.): GSI = 0.842; sensitivity 0.94, PH integ. 0.82.