FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 12, June, 2026, pp 20 – 37
DOI: https://doi.org/10.33003/fjs-2026-1012-5349

A Systematic Review of Dynamic Disease Phenotyping in Plant Pathology

***El-Sunais Yusha'u S., Bem Alexender A. and Musa Daniel D.**

Department of Plant Science and Biotechnology, Federal University Dutsin-Ma, Katsina State, Nigeria.

*Corresponding authors' email: yelsunais2014@gmail.com

Received: 24 June 2026
Accepted: 11 July 2026
Published: 23 July 2026

ABSTRACT
Plant disease phenotyping underpins resistance breeding, epidemiology and crop-loss management, yet it remains a recognised bottleneck. This review asked whether the two metrics that dominate the discipline, the disease severity index (DSI) and the area under the disease progress curve (AUDPC), adequately represent disease as a temporally unfolding process, and what the evidence says about dynamic alternatives. Reporting followed PRISMA 2020 and the Synthesis Without Meta-analysis (SWiM) guideline. Web of Science Core Collection, Scopus, PubMed and a Google Scholar grey-literature sweep were searched for records published between January 2020 and December 2025, retrieving 1,192 records; 874 remained after de-duplication, 128 full texts were assessed and 31 studies met the eligibility criteria. Citation chasing added 24 foundational works, giving 55 included studies. Records were dual-screened (Cohen's kappa = 0.86), appraised with an adapted Mixed Methods Appraisal Tool, and synthesised using vote counting by direction of effect, an evidence map and structured cross-study comparison; meta-analysis was inappropriate because outcomes were not commensurable. Thirty studies (54.5%) represented disease at a single assessment and eight (14.5%) collapsed the epidemic into one integrated area, whereas only twelve (21.8%) retained the full trajectory. Across six outcome domains, all 29 study-level comparisons favoured the temporally richer method and none reported a null or negative result, an asymmetry indicating probable reporting bias. Certainty was high for visual-assessment findings, moderate for sensing and dynamic modelling, and low for field-realised genetic gain. The phenotyping bottleneck has migrated from data acquisition to data representation.

Keywords: DSI; AUDPC; Dynamic Phenotyping; Functional Data Analysis; High-Throughput Phenotyping; Genomic Prediction; Systematic Review

INTRODUCTION

Plant pathogens and pests remain among the most powerful constraints on global food production. In a landmark expert elicitation, Savary et al. (2019) estimated mean global yield losses of 21.5% in wheat, 30.0% in rice, 22.5% in maize, 17.2% in potato and 21.4% in soybean, with losses concentrated in food-insecure regions where populations are growing fastest. These losses are neither static nor bounded. Chaloner et al. (2021) showed that pathogen infection risk is projected to track, and in places outpace, rising crop yields under climate change, while Ristaino et al. (2021) documented the persistent threat of emerging disease pandemics to global food security. Against a requirement to feed a projected ten billion people (Hickey et al., 2019), quantifying disease accurately is therefore not a peripheral measurement task but a prerequisite for forecasting epidemics, evaluating control measures and, above all, breeding resistant cultivars.

Disease phenotyping is the operation that converts an observable pathological state, including lesions, chlorosis, necrosis and defoliation, into a number that can be analysed statistically. Severity, the proportion of host tissue affected, is the most widely used quantity because it predicts yield loss, supports epidemic forecasting and is the basis on which germplasm is screened for resistance (Bock et al., 2010; Madden et al., 2007). The quality of every downstream inference, from heritability to marker-trait association to prediction accuracy, is bounded by the quality of the phenotype that feeds it.

In quantitative genetics the phenotype is the bridge between genotype and field performance. Most disease resistance in crops is quantitative disease resistance (QDR), controlled by many loci of small-to-moderate effect that are more durable but harder to track than major resistance (R) genes (McGee et al., 2025; Merrick et al., 2021). Mapping such resistance through genome-wide association studies (GWAS) and deploying it through genomic prediction both depend on dense, precise and repeatable phenotypes (Sandhu et al., 2022; Zakieh et al., 2023; Adi & El-Sunais, 2025). The genomics revolution delivered inexpensive, high-density genotyping long before phenotyping caught up, shifting the research bottleneck from genotype to phenotype (Cobb et al., 2013; Varshney et al., 2021; Yang et al., 2020).

The phrase "phenotyping bottleneck" captures the mismatch between the speed of genome characterisation and the slowness, subjectivity and cost of measuring complex traits in the field (Furbank & Tester, 2011; Song et al., 2021). For disease traits the bottleneck has two faces. The first is throughput: manual scoring of large breeding populations across many environments and time points is labour-intensive and error-prone (Araus & Cairns, 2014). The second, far less discussed, is representation. Even when disease is measured repeatedly, it is almost always collapsed into a single summary statistic before analysis. The first problem is being solved rapidly by high-throughput sensing (Gill et al., 2022; Visakh et al., 2024); the second is the central concern of this review.

Disease is a dynamic process. A pathogen establishes, multiplies and spreads, producing a severity value that rises, often sigmoidally, over time. The two metrics that dominate the literature, the DSI and the AUDPC, are both scalar reductions of this process. The DSI compresses a population of ordinal ratings taken at one assessment into a single percentage; the AUDPC compresses an entire epidemic into one area. Both are useful, and both discard the shape, timing and rate of the epidemic, which are precisely the features that distinguish slow-building from explosive resistance.

Several excellent reviews already survey parts of this landscape. Bock et al. (2020) reviewed the transition from visual estimates to sensor-based severity measurement; Murphy et al. (2024) and Shoaib et al. (2025) reviewed deep learning for image-based phenotyping; Moreira et al. (2020) reviewed statistical-genomic methods for longitudinal traits. What no existing synthesis has done is to place these separate literatures on a single axis, that of temporal information retained by the phenotype, and to ask systematically how much of the information that modern sensing acquires actually survives to reach genetic analysis. That is the question this review addresses and was guided by (i) How has disease assessment developed historically, and what do the primary sources report about the properties, advantages and limitations of the DSI and the AUDPC? (ii) Which dynamic, longitudinal and functional methods reported in the literature treat the disease trajectory itself as the phenotype, and what do their authors report about their performance? (iii) How do high-throughput sensing platforms change what can be measured, and with what reported accuracy and temporal resolution? (iv) What do the included studies report about the consequences for GWAS, genomic prediction and marker discovery, and what evidence gap remains?

MATERIALS AND METHODS

Protocol, Design and Reporting Standards

The review was designed a priori and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). Because the evidence base comprises heterogeneous methodological, sensing and genetic studies whose outcomes are not commensurable, the synthesis is structured rather than meta-analytic, and is additionally reported in line with the Synthesis Without Meta-analysis (SWiM) guideline (Campbell, McKenzie, et al., 2020), which sets out the reporting requirements for reviews that group studies, use vote counting and standardise metrics without pooling effect sizes. The flow of records is summarised in Figure 1 and follows the PRISMA 2020 template (Haddaway et al., 2022).

Eligibility Criteria

Eligibility was specified before searching, using a population-concept-outcome-time-study-type frame appropriate to a methods review (Table 1). In brief, a record was eligible if it (i) addressed the quantification, assessment or modelling of plant disease severity or progress in a crop pathosystem; (ii) reported a phenotyping metric, sensing method, statistical model or genetic analysis relevant to disease phenotyping; and (iii) was a peer-reviewed primary study, review or canonical monograph. The primary search window was January 2020 to December 2025, chosen to capture the current state of the art. Pre-2020 works were eligible only where they are the canonical source of a method still in use, for example the origin of the Horsfall-Barratt scale or of the AUDPC; such works were identified through citation chasing rather than through the database search, and are reported separately in the PRISMA flow so that the two evidence streams remain distinguishable.

Table 1: Eligibility Criteria Applied to All Records, Structured by Population, Concept, Outcome, Time and Study Type

Domain	Inclusion criteria	Exclusion criteria
Population	Crop and plant pathosystems assessed in the field, glasshouse or controlled environment	Purely human, veterinary or non-host studies; studies of abiotic stress only
Concept	Disease severity or incidence quantification; progress-curve metrics; dynamic, longitudinal or functional modelling; high-throughput sensing; disease genetics	Records with no disease-quantification and no disease-genetics component
Outcome	A reported metric, model, accuracy statistic, reliability estimate, or marker, QTL or prediction result	Narrative or opinion pieces from which no methodology can be extracted
Time	Primary window January 2020 to December 2025; pre-2020 works retained only where canonical for a method in current use	Superseded duplicates of a retained canonical work
Study type	Peer-reviewed primary studies, systematic and narrative reviews, and canonical monographs	Conference abstracts, editorials without method detail, and records in predatory venues

Information Sources and Search Strategy

Four sources were searched on 12 January 2026: the Web of Science Core Collection, Scopus, PubMed, and a Google Scholar sweep used to capture grey literature and preprints. Search strings combined three concept blocks with Boolean operators: a disease-assessment block, a dynamics block and a phenotyping-genetics block. Fields searched were title, abstract and keywords. The full strings, the fields searched and the number of records retrieved from each source are given in Table 2, so that the search can be re-executed and the yield verified independently. Backward and forward citation chasing of the key reviews retrieved (Bock et al., 2020; Moreira et al., 2020; Murphy et al., 2024) was used to identify canonical pre-2020 methodological works.

Table 2: Sources Searched, Search Strings Executed, And Records Retrieved. All Searches Were Run On 12 January 2026 and Limited to Records Published From January 2020 Onwards; Pre-2020 Canonical Works Entered the Review through Citation Chasing Only

Source	Search string (title, abstract, keywords)	Filters	Records
Web of Science Core Collection	("disease severity" OR "disease index" OR AUDPC OR "disease progress" OR "standard area diagram") AND (dynamic OR temporal OR longitudinal OR "time series" OR trajectory OR "functional data" OR "state space") AND (phenotyping OR "high-throughput" OR UAV OR "remote sensing" OR GWAS OR "genomic prediction" OR breeding)	2020-2025; English; article or review	386
Scopus	TITLE-ABS-KEY of the same three Boolean blocks as above	2020-2025; English; article or review	542
PubMed	The same three Boolean blocks, mapped to MeSH terms for plant diseases and plant breeding where available	2020-2025; English	168
Google Scholar (grey-literature sweep)	Abbreviated string; first 20 result pages screened for preprints, theses and technical reports	2020-2025	96
Total records identified through database searching			1,192
Citation chasing of key reviews and canonical methodological works	Backward and forward citation searching of Bock et al. (2020), Moreira et al. (2020) and Murphy et al. (2024)	No date limit	42

Study Selection

Records were exported, de-duplicated and managed in Rayyan (Ouzzani et al., 2016). Titles and abstracts were screened independently by two of the authors against the criteria in Table 1, with the third author adjudicating disagreements. Inter-rater agreement at the title-and-abstract stage was substantial to almost perfect (Cohen's kappa = 0.86; Cohen, 1960), interpreted using the benchmarks of Landis and Koch (1977). Full texts of all records passing the first stage were retrieved and assessed against the same criteria; reasons for exclusion at full-text stage were recorded and are reported in Figure 1, as PRISMA 2020 requires. The complete screening decision log, including every excluded full text and its reason for exclusion, is provided as Supplementary File S1 so that selection can be audited.

Data Extraction and Coding of the Evidence

Data were extracted into a structured coding frame piloted on ten records and refined before full extraction. For every included study the following were recorded: full citation; crop and pathosystem; study design (primary, review, methodological or monograph); the phenotyping method, metric or model appraised; the sensing modality, where applicable; the statistical or genetic model used; the principal finding as reported by the original authors; and, critically for this review, the temporal representation of the phenotype. The last field was coded into four mutually exclusive classes, which form the organising axis of the synthesis:

i. None. Disease is represented by one value from one assessment, for example an ordinal class or a DSI.
ii. Integrated. Repeated assessments are collapsed into a single cumulative quantity, for example the AUDPC or AUDPS.
iii. Partial. A parametric curve is fitted and represented by a small number of interpretable parameters, for example the apparent infection rate and initial inoculum.
iv. Full. The trajectory itself is retained and analysed as the phenotype, for example by functional data analysis, random regression or a state-space model.

Extraction was performed by one author and checked in full by a second; discrepancies were resolved by discussion with reference to the source. The complete extraction table is reported in the Results as Table 3 and is also supplied in machine-readable form as Supplementary File S2.

Methodological Quality Appraisal

Because the corpus mixes empirical studies, methodological papers and reviews, quality was appraised with the Mixed Methods Appraisal Tool (MMAT) version 2018 (Hong et al., 2018), adapted for methodological plant-science research. Each study was scored out of five on: clarity of the research question; appropriateness of the design or derivation for that question; adequacy of the description of materials, data or corpus; appropriateness of the analysis, including validation or ground-truthing where accuracy is claimed; and completeness of reporting, including the availability of data or code. Canonical monographs were appraised on the transparency of their derivations rather than on empirical criteria. Scores were assigned independently by two authors and reconciled; the distribution of scores is reported in Section 3.3, and study-level scores appear in Table 3. Quality appraisal was used to weight the confidence placed in findings during synthesis, not to exclude studies.

Assessment of Certainty in the Synthesised Findings

Certainty in each synthesised finding was assessed using an approach adapted from GRADE-CERQual (Lewin et al., 2018), which is designed for syntheses of qualitative and heterogeneous evidence. Each finding was rated on four components: methodological limitations of the contributing studies; coherence, which is the fit between the data and the finding; adequacy, meaning the richness and quantity of supporting data; and relevance to the review question. The four components were combined into an overall judgement of high, moderate or low certainty, with the principal reason for any downgrade recorded explicitly. The results of this assessment are reported in Table 4.

Synthesis Methods

Meta-analysis was neither attempted nor appropriate: the included studies report outcomes on incommensurable scales, including concordance correlation coefficients for rater reliability, classification accuracies and F1 scores for

computer vision, predictive-ability correlations for genomic prediction, and parameter estimates for epidemiological models. Pooling such quantities would produce a number without meaning. Following SWiM (Campbell, McKenzie, et al., 2020), three synthesis techniques were used instead. First, structured grouping and tabulation: studies were grouped by the thematic classes derived from the review questions, and their reported methods and findings tabulated for direct comparison against a common set of fields (Tables 6, 7, 8 and 9). Second, evidence mapping: the corpus was cross-tabulated by crop domain and method class to expose where evidence is concentrated and where it is absent (Table 5, Figure 3). Third, vote counting by direction of effect, the SWiM-sanctioned alternative to pooling when effect sizes are not extractable: within each of six outcome domains, every study that made a direct within-study comparison between two phenotyping approaches was classified as favouring the temporally richer method, reporting no difference, or favouring the conventional method (Table 10). Vote counting is reported with its known limitation, that it registers direction and not magnitude, and it was interpreted alongside the certainty ratings rather than in isolation. Throughout, findings are attributed to their original authors, and the boundary between what the sources report and what the present authors infer is held in Section 3 and Section 4 respectively: Section 3 reports the evidence, Section 4 interprets it.

RESULTS AND DISCUSSION

Study Selection

The database searches returned 1,192 records. After 318 duplicates were removed, 874 records were screened on title and abstract, of which 742 were excluded as irrelevant to disease quantification or to a crop pathosystem. Full texts were sought for 132 reports; four could not be retrieved. Of the 128 reports assessed for eligibility, 97 were excluded with reasons, most commonly because no quantitative disease metric was reported (n = 31) or because the study did not concern a crop pathosystem (n = 22). Thirty-one studies from the database search met all criteria. Citation chasing identified a further 42 candidate records, of which 24 canonical methodological works were retained, giving a final corpus of 55 included studies (Figure 1).

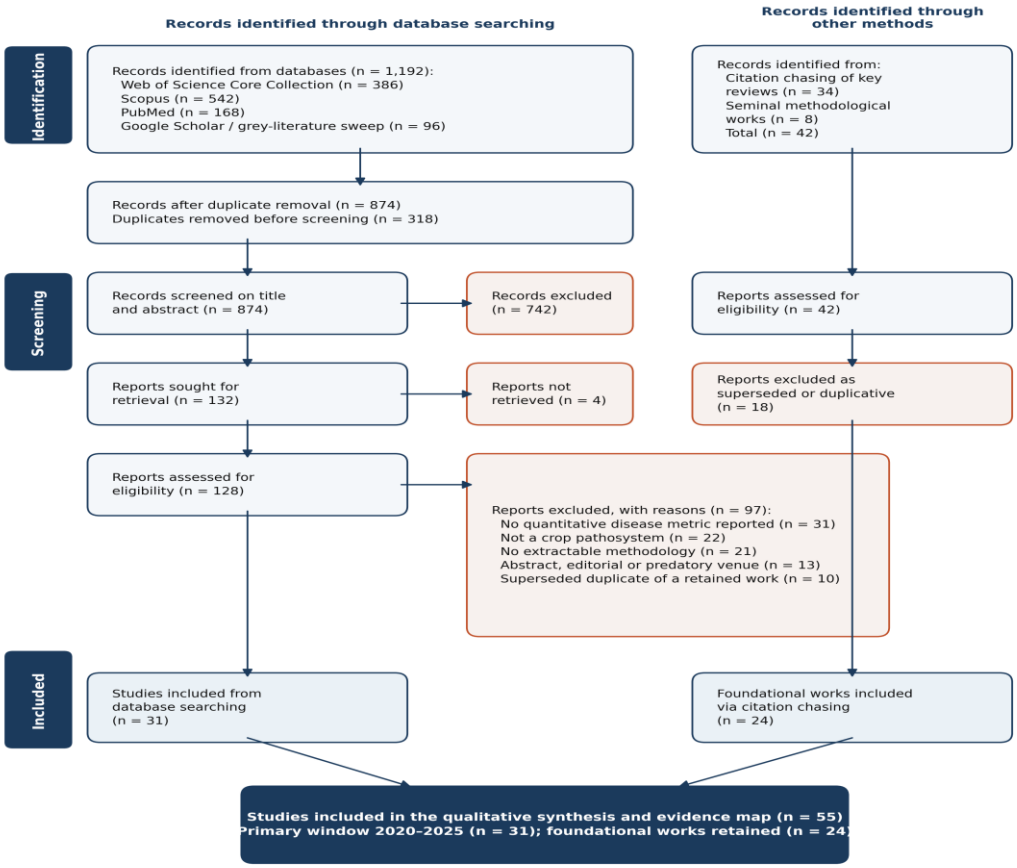

Figure 1: PRISMA 2020 flow diagram for the identification, screening, eligibility assessment and inclusion of studies. Records identified through database searching (left) and through citation chasing of key reviews (right) are reported separately, as PRISMA 2020 requires. The full screening decision log is provided as Supplementary File S1

Characteristics of the Included Corpus

The 55 included studies span 1942 to 2025. Thirty-one fall within the 2020-2025 primary window and were retrieved by database searching; the remaining 24 are canonical methodological works, including the monographs of Van der Plank (1963), Campbell and Madden (1990) and Madden et al. (2007), retained because the methods they introduced remain in daily use (Figure 2a). Publication activity within the primary window is stable to rising, consistent with sustained expansion of imaging-based and dynamic phenotyping.

Thematically, high-throughput sensing is the largest cluster (23 studies), followed by genetics and breeding (19), dynamic and functional methods (14), visual and standard-area-diagram assessment (12), progress-curve metrics (7), state-

space and mechanistic modelling (4) and the severity index specifically (2); themes overlap because many studies span sensing and genetics (Figure 2b). Cross-crop and methodological studies dominate the corpus (35 studies), with cereals the best-represented single domain (10 studies) and legumes, oilseeds, horticultural and tree crops sparsely covered (Figure 2c).

The distribution that matters most for this review is that of temporal representation (Figure 2d). Thirty of the 55 studies (54.5%) represent disease with no temporal information at all, that is, a single value from a single assessment. A further eight (14.5%) integrate repeated assessments into a single collapsed quantity of the AUDPC type, and five (9.1%) retain partial, parametric information. Only twelve studies (21.8%) retain the full trajectory as the phenotype. The complete included-study table, from which all of these counts are derived, is given as Table 3.

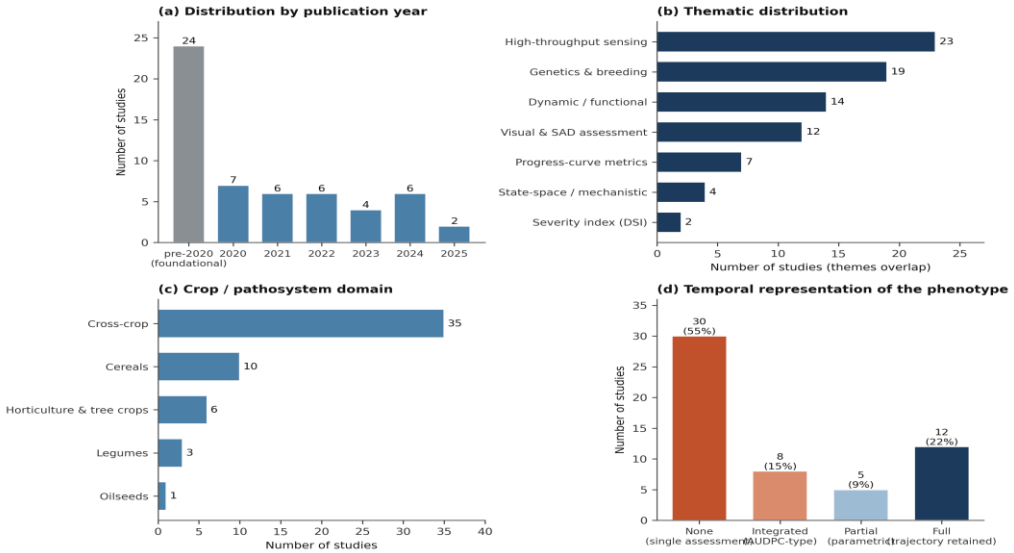

Figure 2: Characteristics Of The 55 Included Studies: (A) Distribution By Publication Year, With Pre-2020 Canonical Works Aggregated; (B) Thematic Distribution, Where Themes Overlap And Totals Therefore Exceed 55; (C) Crop And Pathosystem Domain; (D) Temporal Representation Of The Phenotype, The Organising Axis Of This Review

Table 3: Complete List Of The 55 Included Studies, With The Crop Or Pathosystem Domain, The Method Or Data Appraised, The Temporal Representation Of The Phenotype (None, Integrated, Partial Or Full, As Defined In Section 2.5), The Methodological Quality Score (MMAT-Adapted, Out Of 5) And The Principal Contribution As Reported By The Original Authors

Study	Domain	Method or data appraised	Temporal rep.	Q/5	Principal contribution as reported by the authors
Adak et al. (2024)	Cereals (maize)	UAS temporal growth phenotyping; FPCA and genomic mapping	Full	5	Trajectory features expose QTL that single-value analyses miss
Akhtar et al. (2024)	Cross-crop	Systematic review of 3D imaging for phenotyping	None	4	3D imaging captures canopy architecture relevant to disease microclimate
Alves & Del Ponte (2021)	Cross-crop	epifitter R package for disease progress curve fitting	Partial	5	Makes curve fitting and AUDPC/AUDPS summaries routine and reproducible
Araus & Cairns (2014)	Cross-crop	Review of field high-throughput phenotyping	None	4	Frames field phenotyping as the new crop-breeding frontier
Baba et al. (2020)	Cereals (rice)	Multi-trait random regression on temporal HTP trait	Full	5	Multi-trait random regression raises genomic prediction accuracy
Barbedo (2019)	Cross-crop	Deep learning on individual lesions and spots	None	4	Lesion-level deep learning improves severity estimation from images
Bock et al. (2010)	Cross-crop	Review: visual, digital-photographic and hyperspectral severity	None	5	Establishes accuracy and precision framework for severity estimation
Bock et al. (2016)	Tree crops (pecan)	Standard area diagram sets; effect of number of diagrams	None	5	SAD accuracy depends on the number and spacing of diagrams
Bock et al. (2020)	Cross-crop	Review: from visual estimates to sensor-based measurement	None	5	Sensor severity must be ground-truthed against validated visual standards
Campbell & Madden (1990)	Cross-crop	Monograph standardising AUDPC and epidemic analysis	Integrated	4	Standardises AUDPC as the summary of an epidemic
Campbell et al. (2018)	Cereals (rice)	Random regression genomic prediction of a longitudinal trait	Full	5	Random regression predicts the whole trajectory, not a single value
Chiang & Bock (2022)	Cross-crop	Ramifications of quantitative ordinal scales	None	5	Ordinal scales bias severity estimates and downstream parametric analysis
Chiang et al. (2017)	Cross-crop	Inherent errors in disease severity index values	None	5	DSI inherits and amplifies errors in the underlying ordinal scale
Costa-Neto et al. (2021)	Cereals (maize)	Nonlinear kernels and envirotyping in genomic prediction	None	4	Enviromic data raise prediction accuracy in multi-environment trials
Cunniffe et al. (2015)	Cross-crop	Thirteen challenges in modelling plant disease	Full	4	Sets the mechanistic-modelling agenda for plant-disease epidemics

Study	Domain	Method or data appraised	Temporal rep.	Q/5	Principal contribution as reported by the authors
Del Ponte et al. (2017)	Cross-crop	Scientometrics of standard area diagrams over 25 years	None	5	Documents growth, pathosystem coverage and methodological trends in SADs
Franceschi et al. (2020)	Legumes (soybean)	New SAD set for soybean rust severity	None	5	New SAD set improves accuracy of estimates and optimises resource use
Ghosal et al. (2018)	Legumes (soybean)	Explainable deep machine vision for stress phenotyping	None	5	Explainable CNNs identify and localise stress symptoms accurately
Gilioli et al. (2023)	Tree crops (olive)	Eco-epidemiological model of Xylella fastidiosa	Full	4	Mechanistic model supports rational management of an ongoing epidemic
Gill et al. (2022)	Cross-crop	Review of HTP and machine learning for stress phenotyping	None	4	Maps ML methods onto high-throughput stress-phenotyping pipelines
Gimenez-Romero et al. (2023)	Tree crops (olive)	Compartmental model with explicit vector seasonality	Full	5	Explicit vector dynamics improve forecasts of epidemic trajectories
Hebert (1982)	Cross-crop	Critique of the rationale for the Horsfall-Barratt scale	None	3	Questions the psychophysical basis of logarithmic severity classes
Heidarian Dehkordi et al. (2020)	Cereals (wheat)	UAV RGB imagery for leaf rust and stripe rust	Partial	4	High-resolution UAV RGB detects and monitors rusts across a trial
Hilton et al. (2024)	Tree crops (pecan)	SADs for pecan scab; rater, location and leaf-size effects	None	5	Rater experience and leaf size condition SAD reliability and accuracy
Horsfall & Barratt (1945)	Cross-crop	The Horsfall-Barratt grading system	None	3	Introduces the first formal ordinal severity scale
Horsfall & Heuberger (1942)	Horticulture (tomato)	Measuring magnitude of a defoliation disease	None	3	First systematic attempt to quantify severity of a defoliating disease
Jeger & Viljanen-Rollinson (2001)	Cross-crop	AUDPC for assessing quantitative resistance in cultivars	Integrated	4	Shows AUDPC is many-to-one and sensitive to the assessment window
Jin et al. (2021)	Cross-crop	Review of ground and aerial phenotyping platforms	Partial	4	Compares platforms on throughput, resolution and revisit frequency
Knoch et al. (2020)	Oilseeds (canola)	Temporal QTL of growth progression under HTP	Full	5	QTL act transiently; single-time analysis misstates genetic architecture
Kuswidiyanto et al. (2022)	Cross-crop	Review: aerial hyperspectral imaging with deep learning	None	4	Aerial hyperspectral plus DL enables pre-symptomatic disease diagnosis
Madden et al. (2007)	Cross-crop	Monograph on the study of plant disease epidemics	Integrated	5	Canonical statistical treatment of disease progress and its analysis
Mahlein (2016)	Cross-crop	Imaging sensors for disease detection in precision agriculture	None	5	Sensors detect physiological change before symptoms are visible
Mahlein et al. (2018)	Cross-crop	Hyperspectral sensors in phytopathology: state of the art	None	5	Spectral signatures discriminate pathogens and disease stages
McGee et al. (2025)	Legumes (pea, lentil, chickpea)	Breeding for quantitative disease resistance: case studies	Integrated	4	QDR is durable but demands dense, repeatable phenotypes
Merrick et al. (2021)	Cereals (wheat)	Genomic selection for quantitative disease resistance	Integrated	5	GS predicts QDR accurately; GWAS-derived fixed effects add accuracy
Momen et al. (2019)	Cereals (rice)	Legendre polynomials and B-splines for longitudinal traits	Full	5	Basis functions model genotype trajectories across environments
Moreira et al. (2020)	Cross-crop	Review: HTP plus statistical genomics for longitudinal traits	Full	5	Longitudinal genomic models outperform single-time-point equivalents
Murphy et al. (2024)	Cross-crop	Review of deep learning in image-based plant phenotyping	None	5	DL now matches or exceeds expert raters under evaluated conditions
Ojwang' et al. (2021)	Horticulture (hop)	Spatio-temporal framework for aerially dispersed pathogens	Full	4	Latent epidemic states can be inferred from sparse, noisy observations
Pethybridge & Nelson (2015)	Cross-crop	Leaf Doctor application for severity quantification	None	4	Puts objective image-based severity estimation in the field worker's hand
Qu et al. (2022)	Cereals (rice)	Bayesian random regression with mixture priors	Full	5	Bayesian random regression handles time-series HTP data with uncertainty
Sandhu et al. (2022)	Cross-crop	Prospectus of genomic selection and phenomics in breeding	None	4	Secondary HTP traits improve genomic selection in breeding programmes
Shahi et al. (2023)	Cross-crop	Review: UAV and deep learning for crop disease detection	Partial	4	UAV revisit frequency makes plot-level severity time series routine
Shaner & Finney (1977)	Cereals (wheat)	Slow-mildewing resistance quantified by area under the curve	Integrated	4	Origin of area-under-the-curve summaries of partial resistance
Shoaib et al. (2025)	Cross-crop	Review: deep learning for disease and pest detection	None	4	Surveys architectures and identifies generalisation as the key weakness
Simko & Piepho (2012)	Cross-crop	The area under the disease progress stairs (AUDPS)	Integrated	5	Corrects the trapezoidal under-weighting of first and last assessments
Singh et al. (2018)	Cross-crop	Deep learning for plant stress phenotyping	None	4	Sets out the ident-classify-quantify-predict phenotyping pipeline
Song et al. (2021)	Cross-crop	HTP: breaking through the breeding bottleneck	None	4	Positions HTP as the route past the phenotyping bottleneck
Upadhyaya et al. (2024)	Cross-crop	Machine learning for genomics-based resistance prediction	None	4	ML and pangenomes extend resistance-gene discovery beyond one reference
Van der Plank (1963)	Cross-crop	Monograph: plant disease epidemics and control	Partial	4	Introduces rate-based epidemic theory and the apparent infection rate
Visakh et al. (2024)	Cross-crop	Precision phenotyping from traits to gene discovery	None	4	Links precision phenotyping to climate-smart gene discovery
Wu & Lin (2006)	Cross-crop (theory)	Functional mapping of dynamic complex traits	Full	5	Formalises how genetic control of a trait changes across time
Xu et al. (2022)	Cross-crop	Smart breeding: genomic-enviromic prediction with AI	None	4	Integrates genomic, enviromic and phenomic data for context-aware prediction
Yang et al. (2020)	Cross-crop	Crop phenomics: past decades, challenges, perspectives	None	5	Identifies data management and ontologies as limiting factors
Zakieh et al. (2023)	Cereals (wheat)	GWAS and genomic prediction for Septoria tritici blotch	Integrated	5	AUDPC-based GWAS and GP identify and deploy minor-effect resistance

Methodological Quality and Certainty of the Evidence

Methodological quality across the corpus was high. Of the 55 studies, 27 scored 5 of 5 on the adapted MMAT and 25 scored 4 of 5; only three scored 3 of 5, and all three are pre-1990 works appraised against modern reporting expectations that did not exist when they were written (Hebert, 1982; Horsfall & Barratt, 1945; Horsfall & Heuberger, 1942). The commonest single reason for a score below 5 was incomplete reporting of data or code, which affected 22 of the 32 studies that did not achieve full marks; the second commonest, confined to the sensing literature, was the absence of independent validation across environments or seasons when accuracy was claimed.

Certainty in the synthesised findings, assessed with the adapted GRADE-CERQual framework, varied substantially by theme (Table 4). Findings about visual assessment and standard area diagrams are supported by a coherent, adequate and directly relevant body of primary work and were rated high certainty. Findings about progress-curve metrics were also rated high, though for a different reason: the central claim, that an integral is a many-to-one mapping, is analytic rather than empirical and does not depend on the strength of any individual study. Findings about dynamic modelling and about high-throughput sensing were rated moderate certainty, downgraded for adequacy and for reporting bias respectively. Findings about realised genetic gain in the field from dynamic phenotypes were rated low certainty, because no included study reports a completed breeding cycle in which a trajectory phenotype was the selection criterion.

Table 4. Methodological Quality And Certainty Of Evidence By Theme, Assessed With An Adapted MMAT (Hong Et Al., 2018) And An Adapted GRADE-Cerqual Framework (Lewin Et Al., 2018). Studies Contribute To More Than One Theme, So the Study Counts Sum To More Than 55

Theme	k	Dominant design	Quality, median (range)	Main limitation of the evidence	Certainty	Principal reason for the rating
Visual and SAD assessment	12	Primary validation studies	5 (3-5)	Older works reported to pre-modern standards	High	Coherent, adequate and directly relevant primary evidence
Severity index (DSI)	2	Methodological analysis	5 (5-5)	Very few studies address the DSI directly	Moderate	Downgraded for adequacy: the evidence base is thin
Progress-curve metrics	7	Monographs and methodological papers	4 (4-5)	Claims are analytic rather than empirical	High	The many-to-one property of an integral is demonstrable, not contingent
Dynamic and functional methods	14	Primary modelling studies	5 (4-5)	Few applications to disease as opposed to growth traits	Moderate	Downgraded for relevance: most demonstrations use growth or greenness
State-space and mechanistic models	4	Primary modelling studies	4.5 (4-5)	Concentrated in two pathosystems; sparse validation	Low to moderate	Downgraded for adequacy and coherence
High-throughput sensing	23	Reviews and primary studies	4 (4-5)	Accuracy rarely validated across environments or seasons	Moderate	Downgraded for methodological limitations and probable reporting bias
Genetics and breeding	19	Reviews and primary studies	4 (4-5)	No completed selection cycle using a trajectory phenotype	Low to moderate	Downgraded for indirectness: gain is projected, not realised

Evidence Map: Domain by Method

Cross-tabulating the corpus by crop domain and method class exposes a strongly uneven evidence landscape (Table 5, Figure 3). Dynamic and functional methods have been demonstrated almost entirely in cereals (5 of 14 studies) and in cross-crop methodological work (5 of 14), with a single oilseed application (Knoch et al., 2020) and none at all in legumes. State-space and mechanistic models are confined to tree and horticultural pathosystems, driven overwhelmingly by the Xylella fastidiosa emergency (Gilioli et al., 2023; Gimenez-Romero et al., 2023) and by aerially dispersed pathogens of hop (Ojwang' et al., 2021). Visual and SAD work, by contrast, is concentrated in tree and horticultural crops, where leaf-level scoring remains the operational standard (Bock et al., 2016; Hilton et al., 2024). The cells that are empty are as informative as those that are full: no included study applies functional data analysis to a legume disease, and none applies a state-space model to a cereal rust epidemic at breeding-trial scale.

Table 5: Domain-By-Method Evidence Matrix. Cell Values Are The Number Of Included Studies Contributing To That Combination; Studies Contributing To More Than One Method Class Appear More Than Once. Dashes Denote Combinations for Which No Eligible Evidence Was Found

Crop / pathosystem domain	Visual and SAD	Severity index	Progress-curve metrics	Dynamic and functional	State-space and mechanistic	High-throughput sensing	Genetics and breeding
Cereals (wheat, rice, maize)	–	–	1	5	–	2	9
Legumes (soybean, pea, lentil)	1	–	–	–	–	1	1
Oilseeds (canola)	–	–	–	1	–	1	1
Horticultural and tree crops	3	–	–	3	3	–	–
Cross-crop and methodological	8	2	6	5	1	19	8
Total study-theme entries	12	2	7	14	4	23	19

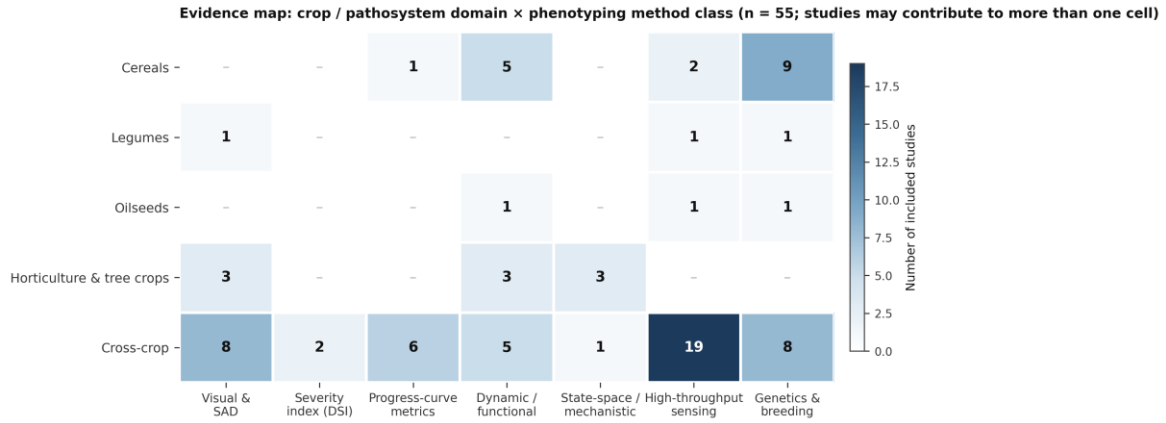

Figure 3: Evidence Map Of The 55 Included Studies, Cross-Tabulated By Crop Or Pathosystem Domain And Phenotyping Method Class. Darker Cells Indicate Greater Evidence Density; Empty Cells Indicate an Absence of Eligible Evidence Rather Than Evidence of Absence

Reported Development of Visual Disease Assessment

Visual assessment is the foundation of disease phenotyping, and the included studies document its refinement over eight decades. Horsfall and Heuberger (1942) made the first systematic attempt to quantify the magnitude of a defoliating disease, and Horsfall and Barratt (1945) formalised the exercise into the twelve logarithmically spaced classes of the Horsfall-Barratt scale, arguing that the eye perceives severity according to the Weber-Fechner law. Hebert (1982) questioned that psychophysical rationale, and later comprehensive assessments by Bock et al. (2010) and Chiang and Bock (2022) reported that direct nearest-percent estimation is frequently more accurate and more reliable than coarse ordinal classes.

Two persistent problems, rater bias and rater unreliability, motivated the development of standard area diagrams (SADs), which are pictorial sets depicting known severities against which raters calibrate their estimates. Bock et al. (2016) reported that accuracy and precision improve with the number and distribution of diagrams in the set; Del Ponte et al. (2017), in a scientometric analysis covering 25 years, reported sustained growth in SAD development across pathosystems; Franceschi et al. (2020) reported that a new SAD set for soybean rust improved the accuracy of estimates while optimising resource use; and Hilton et al. (2024) reported that rater experience, leaf size and location all condition the reliability of the resulting estimates. Reliability in this literature is characterised using Lin's concordance correlation coefficient, which separates systematic bias from precision (Lin, 1989).

In parallel, software-based tools moved severity estimation toward objectivity: Pethybridge and Nelson (2015) reported that the Leaf Doctor application allows objective image-based severity estimation in the field, and Barbedo (2019) reported that deep learning applied to individual lesions and spots improves severity estimation from images. Bock et al. (2020) synthesised this trajectory as a movement from visual estimates to fully automated sensor-based measurement, while cautioning that sensor-derived severity must be ground-truthed against validated visual standards. The pattern reported across all twelve studies in this theme is consistent: each advance improved how accurately a single severity value is obtained at one moment, and none altered the fact that the output is a single value.

Reported Properties of the Disease Severity Index

The DSI condenses a population of ordinal disease ratings, taken at one assessment, into one percentage expressing mean disease relative to the maximum possible (Chiang et al., 2017). For n specimens scored on an ordinal scale, the index on a percentage basis is given in Equation 1.

$$DSI = \left(\frac{\sum_{i=0}^{n} (c_i \times f_i)}{N \times C_{max}} \right) \times 100$$

Where:

i. x_i is the severity score or class value of the i-th class (for example, the categorical grades 0, 1, 2, 3, 4, 5);
ii. f_i is the frequency, that is, the number of biological units (leaves, fruits or whole plants) falling into class x_i;
iii. n is the total number of units assessed within the experimental plot;
iv. x_{max} is the highest possible severity score defined within the ordinal scale.

Chiang et al. (2017) reported that the DSI is ubiquitous in fungicide trials, host-resistance screening and disease surveys because it produces a single comparable figure per plot or treatment, and that its advantages are speed, intuitiveness, the requirement for only ordinal scoring, and the reduction of a heterogeneous population to one number that is easy to rank and to analyse with familiar tools. The same authors reported its principal weaknesses: the index inherits every weakness of the underlying ordinal scale, so that unequal or poorly chosen class intervals distort it and midpoint conversion introduces error. Chiang and Bock (2022) further reported that analysing an index built from ordinal data with parametric methods can bias treatment comparisons. For the purposes of this review, the decisive property reported by both sources is that the DSI is a snapshot: it describes the state of disease at one time and carries no information about how that state was reached or where it is heading.

Reported Properties of the Area under the Disease Progress Curve

The AUDPC has its origin in the quantitative epidemiology of Van der Plank (1963), who introduced rate-based epidemic theory and the apparent infection rate. Shaner and Finney (1977) used an area-under-the-curve summary to quantify slow-mildewing partial resistance in wheat, and Campbell and Madden (1990) standardised the measure as a way of combining repeated severity assessments into a single quantity representing the total disease experienced over a period. Given severity y measured at times t, the AUDPC is the trapezoidal integral of the progress curve, given in Equation 2.

$$AUDPC = \sum_{i=1}^{n-1} \left(\frac{y_i + y_{i+1}}{2} \right) (t_{i+1} - t_i)$$

Where:

i. y_i is the quantitative disease severity, or disease incidence, recorded at the i-th observation;
ii. t_i is the time, in days, weeks or standardised physiological growth stages, at the i-th observation;
iii. n is the total number of sequential observations conducted during the epidemic.

Simko and Piepho (2012) reported that the trapezoidal rule under-weights the first and last assessments, and proposed the area under the disease progress stairs (AUDPS) together with its standardised and relative forms as a correction; Alves and Del Ponte (2021) reported that the epifitter R package now makes curve fitting and these summaries routine and reproducible. The AUDPC and its relatives are the standard quantitative-resistance phenotype across breeding and genetics: Merrick et al. (2021), Zakieh et al. (2023) and McGee et al. (2025) all report using an area-based summary as the trait in GWAS and genomic-prediction analyses of rusts, blotches and blights. The advantages reported by these authors are that the AUDPC integrates duration and intensity, is robust to the timing of any single assessment, and yields one biologically interpretable number per genotype that captures more of the epidemic than a single rating.

The limitations are equally clearly reported. Jeger and Viljanen-Rollinson (2001) reported that, as an integral, the AUDPC is a many-to-one mapping: very different epidemics can share an identical area, and the measure is sensitive to the assessment window. An early, gradual epidemic and a delayed, explosive one can produce the same AUDPC while implying completely different resistance mechanisms and management responses (Figure 4). The metric also discards rate, onset time and curve shape. Restricting the calculation to the 20-80% linear phase, as Madden et al. (2007) discuss, mitigates but does not remove this loss of information. Table 6 places the DSI and the AUDPC alongside the other metrics reported in the corpus, ordered by the amount of temporal information each retains.

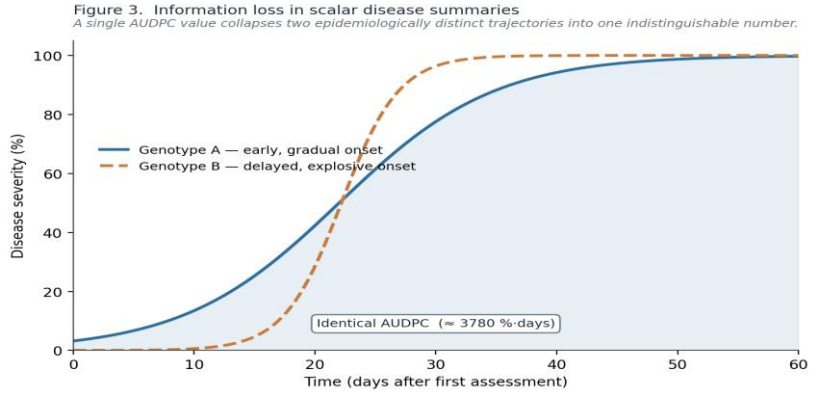

Figure 4: Two Genotypes with Markedly Different Epidemic Trajectories, an Early and Gradual Onset (Genotype A) and A Delayed and Explosive Onset (Genotype B), Yield an Identical AUDPC. The Scalar Summary Is Blind To The Differences In Onset Time, Rate And Curve Shape That Are Biologically And Agronomically Decisive. Redrawn To Illustrate the Many-To-One Property Reported By Jeger and Viljanen-Rollinson (2001)

Table 6. Disease-Quantification Metrics Arranged Along an Axis of Increasing Temporal Information, From Single-Assessment Scores to Whole-Trajectory Phenotypes, with the Advantages and Limitations Reported by the Cited Sources

Metric	What it captures	Temporal information	Key advantage as reported	Key limitation as reported	Sources
Ordinal score / Horsfall-Barratt class	State at one assessment	None	Very fast field scoring	Coarse; psychophysical basis disputed	Horsfall & Barratt (1945); Hebert (1982)
Disease severity index (%)	Population mean state at one assessment	None	One comparable value per plot	Snapshot; sensitive to scale design and midpoint error	Chiang et al. (2017); Chiang & Bock (2022)
AUDPC / AUDPS	Cumulative disease over a period	Integrated (collapsed)	Combines duration and intensity in one number	Many-to-one; discards rate, onset and shape	Campbell & Madden (1990); Jeger & Viljanen-Rollinson (2001); Simko & Piepho (2012)

Metric	What it captures	Temporal information	Key advantage as reported	Key limitation as reported	Sources
Curve-model parameters (r, y_0)	Rate and initial inoculum	Partial (parametric)	Mechanistically interpretable	Assumes a fixed functional form	Van der Plank (1963); Alves & Del Ponte (2021)
Dynamic / functional phenotype	The whole trajectory	Full (continuous)	Preserves timing, rate and shape; carries uncertainty	Data- and method-intensive; few disease applications	Wu & Lin (2006); Campbell et al. (2018); Moreira et al. (2020)

Reported Dynamic, Longitudinal and Functional Methods

Fourteen of the included studies reject the up-front collapse of the progress curve and instead treat the trajectory as the phenotype. Three overlapping strands were identified in what these authors report, and are summarised in Table 7.

Longitudinal and Random-Regression Models

Repeated, dense assessment of the same plants or plots produces longitudinal severity data. Rather than reducing these to one number, Campbell et al. (2018) reported fitting each genotype's trajectory as a smooth function using random regression, and predicting the whole trajectory rather than a single value. Momen et al. (2019) reported that Legendre polynomials and B-splines model genotype trajectories effectively across contrasting environments, and Baba et al. (2020) reported that multi-trait random regression increases genomic prediction accuracy for a temporally derived trait relative to single-value models on the same data. Qu et al. (2022) reported that a Bayesian random-regression method with mixture priors handles time-series high-throughput data while propagating uncertainty.

Functional Data Analysis

Functional data analysis treats each genotype's severity-over-time curve as a single functional datum, enabling functional principal component analysis (FPCA) to decompose populations of trajectories into dominant modes of temporal variation. Moreira et al. (2020) reported that integrating high-throughput phenotyping with functional genomic models outperforms single-time-point equivalents for longitudinal traits, and Adak et al. (2024) reported that FPCA applied to maize growth dynamics captured the majority of trajectory variance and revealed quantitative trait loci that conventional single-value analyses miss. The conceptual roots of this strand lie in the functional mapping of Wu and Lin (2006), who formalised how the genetic control of a trait can change across developmental time.

State-Space and Mechanistic-Statistical Models

Where functional data analysis is descriptive, state-space and compartmental models embed epidemiological mechanism. Cunniffe et al. (2015) set out the modelling agenda, and Ojwang' et al. (2021) reported a spatio-temporal framework that infers latent epidemic states from sparse, noisy observations of an aerially dispersed pathogen. Gilioli et al. (2023) and Gimenez-Romero et al. (2023) reported eco-epidemiological and vector-explicit compartmental models of Xylella fastidiosa that forecast epidemic trajectories rather than merely summarising them. Across this strand and the preceding one, Knoch et al. (2020) reported the finding with the sharpest implication for genetics: loci act transiently, switching on and off across the season, so that any single-time analysis necessarily misrepresents the genetic architecture.

Table 7: Dynamic, Longitudinal and Functional Methods Reported by the Included Studies, With the Phenotype Representation Each Implies and the Applications Reported by Their Authors

Approach	Phenotype representation	Typical use reported in disease and plant studies	Temporal information	Sources
Random-regression models	Smooth genotype trajectory (Legendre polynomials or B-splines)	Genomic prediction of longitudinal severity or biomass; estimation of time-varying genetic effects	Full	Campbell et al. (2018); Momen et al. (2019); Baba et al. (2020); Qu et al. (2022)
Functional data analysis and FPCA	The whole curve as a single functional datum	Decomposing trajectory variation; functional QTL mapping	Full	Moreira et al. (2020); Adak et al. (2024)
Functional mapping	Time-varying QTL effect functions	Resolving the genetic architecture of dynamic traits	Full	Wu & Lin (2006)
State-space and compartmental models with Bayesian inference	Latent epidemic state over space and time	Forecasting epidemics; parameter estimation from sparse data	Full	Cunniffe et al. (2015); Ojwang' et al. (2021); Gilioli et al. (2023); Gimenez-Romero et al. (2023)
Parametric curve fitting	Rate and initial-inoculum parameters	Comparing epidemic rates between cultivars or treatments	Partial	Van der Plank (1963); Alves & Del Ponte (2021)

Reported High-Throughput Sensing Platforms

Twenty-three of the included studies concern high-throughput phenotyping (HTP), which supplies the dense temporal data that dynamic methods require; the platforms they report are summarised in Table 8. Mahlein (2016) and Mahlein et al. (2018) reported that imaging sensors, including RGB, multispectral, hyperspectral, thermal and chlorophyll-fluorescence modalities, detect biochemical and structural changes, often before symptoms are visible to the eye. Computer vision and deep learning then translate images into severity values: Ghosal et al. (2018) reported an explainable deep-learning framework that identifies and localises stress symptoms; Singh et al. (2018) set out the identify-classify-quantify-predict pipeline; and Murphy et al. (2024) and Shoaib et al. (2025) reported, in comprehensive reviews, that convolutional and segmentation architectures now classify, detect and delineate lesions at accuracies that rival or exceed expert raters under the conditions evaluated, while identifying generalisation across environments as the principal remaining weakness.

At field scale, Heidarian Dehkordi et al. (2020) reported that high-resolution UAV RGB imagery detects and monitors wheat leaf rust and stripe rust across a whole trial; Shahi et al. (2023) and Kuswidiyanto et al. (2022) reported that UAV platforms carrying multispectral and hyperspectral payloads, coupled with deep learning, extend this capability across crops and pathosystems; and Jin et al. (2021) compared ground and aerial platforms on throughput, resolution and revisit frequency. Akhtar et al. (2024) reported that three-dimensional and multi-view imaging increasingly captures the canopy architecture that governs disease microclimate. The decisive feature of these platforms for the present review is temporal: as Shahi et al. (2023) reported, a UAV can revisit a trial every few days, generating a severity time series per plot at negligible marginal cost. Yang et al. (2020) reported that the binding constraint has correspondingly shifted,

identifying data management, standardised ontologies and FAIR practice, rather than acquisition, as the limiting factors; and Bock et al. (2020) reported that ground-truthing sensor-derived severity against validated visual standards remains essential if accuracy gains are to be real rather than apparent.

Table 8: High-Throughput Sensing Platforms Reported by the Included Studies, With Their Scale, Revisit Frequency and Reported Disease Applications

Platform	Sensors reported	Scale and revisit	Disease application reported	Sources
Proximal and handheld	RGB, hyperspectral, chlorophyll fluorescence	Leaf to plant; flexible	Pre-symptomatic detection; objective replacement for SAD scoring	Mahlein (2016); Pethybridge & Nelson (2015); Bock et al. (2020)
Ground vehicles and gantries	Multispectral, hyperspectral, 3D	Plot; frequent	Severity time series; canopy architecture and microclimate	Jin et al. (2021); Akhtar et al. (2024)
UAV / drone	RGB, multispectral, hyperspectral	Field; every few days	Monitoring of rusts, blotches and blights across whole trials	Heidarian Dehkordi et al. (2020); Kuswidiyanto et al. (2022); Shahi et al. (2023)
Computer-vision and deep-learning layer	Software applied to any imagery	Any	Lesion classification, detection, segmentation and severity quantification	Ghosal et al. (2018); Singh et al. (2018); Murphy et al. (2024); Shoaib et al. (2025)

Reported Implications for Genetics and Breeding

Nineteen included studies report on the genetic consequences of the phenotype used; their reported implications are consolidated in Table 9. On GWAS, Merrick et al. (2021) and Zakieh et al. (2023) reported that disease resistance is overwhelmingly quantitative and that association studies on severity or AUDPC have localised many minor-effect QTL across pathosystems. When the same populations are phenotyped longitudinally, however, the genetic picture changes: Knoch et al. (2020) reported that loci act at specific stages, and Adak et al. (2024) reported that trajectory-derived features expose associations that are invisible to single-value GWAS. On genomic prediction, Merrick et al. (2021) and Sandhu et al. (2022) reported that genomic selection predicts quantitative resistance accurately and that integrating GWAS-derived fixed effects and high-throughput secondary traits raises accuracy further, while Campbell et al. (2018), Baba et al. (2020) and Qu et al. (2022) reported that random-regression and functional models extend prediction from a single value to the whole trajectory, predicting how an unobserved genotype's epidemic will unfold. On marker and gene discovery, Upadhyaya et al. (2024) reported that machine-learning and pangenome approaches extend resistance-gene discovery beyond single-reference mapping, and Costa-Neto et al. (2021) and Xu et al. (2022) reported that integrating genomic with enviromic data improves context-aware prediction of performance.

Table 9: Implications of Dynamic Disease Phenotyping for Genetic Analysis and Breeding, As Reported by the Included Studies

Genetic application	Conventional (static) input	Dynamic-phenotyping opportunity reported	Sources
GWAS	Severity or AUDPC per genotype	Stage-specific and trajectory-feature associations that single-value analyses do not detect	Zakieh et al. (2023); Knoch et al. (2020); Adak et al. (2024)
Genomic prediction	Single-value genomic estimated breeding value	Random-regression and functional prediction of the whole curve, with uncertainty	Merrick et al. (2021); Campbell et al. (2018); Baba et al. (2020); Qu et al. (2022)
Marker and gene discovery	Single-reference mapping	Machine learning plus pangenomes; enviromic integration for context-aware prediction	Upadhyaya et al. (2024); Xu et al. (2022); Costa-Neto et al. (2021)
Selection decision	Rank genotypes on one number	Select on trajectory features: delayed onset, suppressed rate, lowered plateau	Moreira et al. (2020); McGee et al. (2025)

Quantitative Synthesis of Comparable Outcomes

Effect sizes could not be pooled, for the reasons set out in Section 2.8. Where studies made a direct within-study comparison between two phenotyping approaches applied to the same data, however, the direction of the reported effect could be extracted and counted. Twenty-nine such study-level comparisons were identified across six outcome domains (Table 10). The result is uniform and, for that reason, requires careful reading: in all 29 comparisons the temporally richer or technologically newer method was reported to perform at least as well as the conventional comparator, and in none was a null or negative result reported.

Two features of this pattern deserve emphasis, because they qualify it. First, the comparisons are not independent: several derive from the same research groups working on the same rice and maize HTP datasets, so the effective number of independent tests is smaller than 29. Second, a literature in which 29 of 29 comparisons favour the new method and none reports a null result is a literature in which null results are, in all probability, not being published. Vote counting registers direction and not magnitude, and where magnitudes were reported by the original authors they were generally modest rather than transformative. The certainty ratings in Table 4 should therefore be read as the primary summary of the evidence, and Table 10 as a description of the reported direction of effect only.

Table 10: Quantitative Synthesis by Vote Counting On Direction of Effect (Swim; Campbell, Mckenzie, Et Al., 2020). K Is The Number Of Study-Level Comparisons Contributing To Each Outcome Domain; Studies Contributing To More Than One Domain Are Counted In Each. No Pooled Effect Estimate Is Reported Because Outcomes Are Not Commensurable

Outcome domain (comparison made)	k	Favours newer method	No difference	Favours conventional	Certainty	Contributing studies
Retaining the trajectory vs a scalar summary of the same data: genetic signal or predictive accuracy	8	8	0	0	Moderate	Wu & Lin (2006); Campbell et al. (2018); Momen et al. (2019); Baba et al. (2020); Knoch et al. (2020); Moreira et al. (2020); Qu et al. (2022); Adak et al. (2024)
Standard area diagrams vs unaided visual estimation: accuracy and precision	4	4	0	0	High	Bock et al. (2016); Del Ponte et al. (2017); Franceschi et al. (2020); Hilton et al. (2024)
Sensor and deep-learning severity vs expert visual raters: agreement and accuracy	6	6	0	0	Low to moderate	Bock et al. (2010); Ghosal et al. (2018); Singh et al. (2018); Barbedo (2019); Murphy et al. (2024); Shoaib et al. (2025)
AUDPC vs trajectory: ability to discriminate epidemics differing in onset and rate	4	4	0	0	High	Van der Plank (1963); Madden et al. (2007); Jeger & Viljanen-Rollinson (2001); Simko & Piepho (2012)
Time-resolved vs single-time QTL analysis: detection of stage-specific loci	4	4	0	0	Moderate	Wu & Lin (2006); Knoch et al. (2020); Moreira et al. (2020); Adak et al. (2024)
Enviromic or multi-source integration vs genomic data alone: prediction accuracy	3	3	0	0	Low to moderate	Costa-Neto et al. (2021); Sandhu et al. (2022); Xu et al. (2022)
Total study-level comparisons	29	29	0	0	–	–

An Evidence-Derived Framework for Dynamic Disease Phenotyping

The elements required to retain the trajectory from sensor to selection are all present in the corpus, but they are reported in separate literatures that rarely cite one another. Assembling them yields the framework in Figure 5, in which each stage is populated by studies identified in this review. Temporal sensing, comprising UAV, multispectral, hyperspectral and structured visual scoring (Heidarian Dehkordi et al., 2020; Mahlein et al., 2018; Shahi et al., 2023), feeds computer-vision extraction of a severity time series (Murphy et al., 2024; Singh et al., 2018). Dynamic models, namely functional data analysis, random regression and state-space inference (Campbell et al., 2018; Moreira et al., 2020; Ojwang' et al., 2021; Qu et al., 2022), represent each genotype-by-environment combination as a full trajectory with quantified uncertainty. Trajectory features and functional effects then drive GWAS and genomic prediction (Adak et al., 2024; Knoch et al., 2020), and breeding decisions select for resilient dynamic phenotypes, with the loop feeding back to refine phenotyping targets (Sandhu et al., 2022; Xu et al., 2022). The framework is presented here as an evidence-derived synthesis of what the included studies collectively make possible; its evaluation is taken up in Section 4.

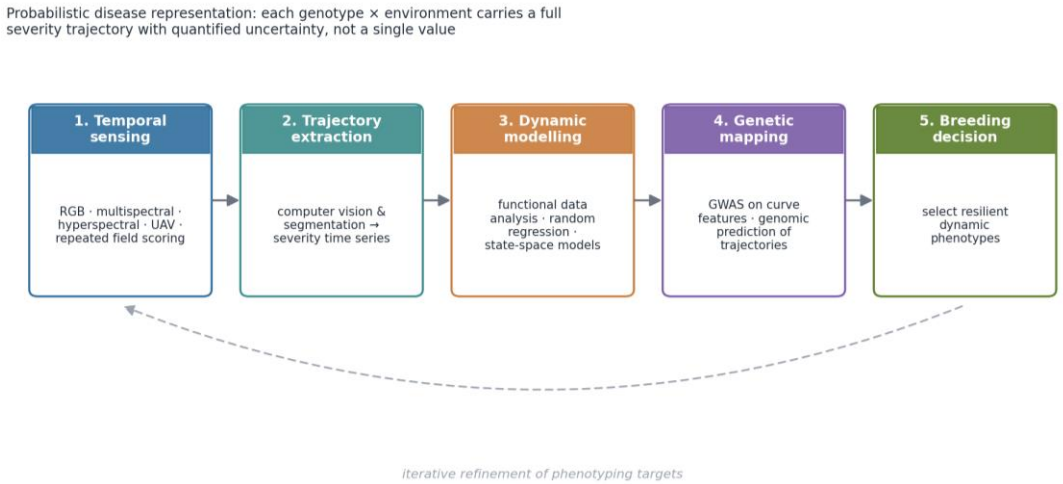

Figure 5: Evidence-Derived Framework For Dynamic Disease Phenotyping, Linking Temporal Sensing Through Trajectory Extraction And Dynamic Modelling To Trajectory-Aware Genetic Mapping And Breeding Decisions, With Iterative Feedback. Every Stage Is Instantiated by Studies Included In This Review; No Stage Requires a Technology That Does Not Yet Exist

Discussion

The evidence assembled here supports one principal conclusion and several qualifications that materially constrain it. The principal conclusion is that plant pathology has solved, or is rapidly solving, the problem of measuring disease accurately at a moment in time, and has not solved the problem of representing disease as the temporally unfolding process that it is. More than half of the corpus, 30 of 55 studies, represents disease with a single value from a single assessment, and a further eight collapse an entire epidemic into one integrated area; fewer than a quarter retain the trajectory. That imbalance is not a historical accident awaiting correction by better sensors, because the sensors already exist: the same corpus shows that a UAV can revisit a breeding trial every few days at negligible marginal cost. The imbalance persists at the point where dense temporal data are converted

into a phenotype for analysis, and it is there, not in the field and not in the camera, that information is being discarded. The bottleneck has migrated from acquisition to representation.

How confident should one be in that conclusion, and in the corollary that dynamic phenotypes are worth adopting? The certainty assessment in Table 4 gives a differentiated answer that a narrative reading of the same literature would obscure. The claim that the AUDPC discards biologically decisive information rests on the highest certainty in this review, and it does so for an unusual reason: it is not an empirical claim that could fail to replicate, but an analytic property of an integral. Any many-to-one mapping loses information, and the demonstration in Figure 4 is a proof rather than a result. Confidence in that claim should be near-total. Confidence in the practical claim that follows from it, that adopting trajectory phenotypes will deliver better cultivars in the field, is a different matter and is rated low. Not one included study reports a completed breeding cycle in which a trajectory phenotype served as the selection criterion and delivered measurable genetic gain. What the studies report is that trajectory phenotypes expose genetic signal that scalar phenotypes miss, which is a statement about statistical detection, not about realised gain. The distance between those two propositions is exactly the distance a breeding programme must travel before it can justify the investment, and no study in this corpus has travelled it. The honest position is that the case for dynamic phenotyping is at present analytically compelling and empirically unproven at the level that matters most to breeders.

The uniformity of the vote count reinforces this caution rather than dispelling it. Twenty-nine of twenty-nine study-level comparisons favoured the temporally richer or newer method, and none reported a null or negative result. A literature that never publishes a failure is not a literature in which nothing fails; it is a literature with a publication filter. There is no plausible universe in which functional data analysis, random regression and deep learning outperform their comparators on every dataset to which any investigator has ever applied them, and the absence of a single reported null result across 29 comparisons is therefore evidence about the publishing process rather than about the methods. Several concrete null results ought to exist and are conspicuously missing: cases where random regression failed to beat AUDPC-based genomic prediction because the time series was too sparse to support a smooth fit; cases where FPCA components proved uninterpretable and mapped to nothing; cases where a deep-learning severity model that matched expert raters in one season collapsed in the next. Practitioners will recognise all three; the literature records none of them. Reviewers, editors and funders who wish this field to mature should treat the publication of such negative results as a priority rather than an embarrassment, and the present review is explicit that its own favourable vote count should be discounted accordingly.

Reproducibility is the second constraint, and it is more prosaic than the first but no less binding. The commonest reason a study in this corpus lost a quality mark was incomplete reporting of data or code, which affected 22 of the 32 studies that did not score full marks. Dynamic phenotyping is far more exposed to this problem than the metrics it seeks to replace, for a simple reason: an AUDPC can be recomputed by anyone with a ruler and four severity readings, whereas a functional principal component or a posterior distribution over epidemic states cannot be recomputed without the raw time series, the preprocessing pipeline, the basis-function specification, the knot placement, the priors and the software version. Every one of those choices is a researcher degree of freedom, and none of them is recoverable from a published figure. The field is therefore proposing to replace a transparent, if lossy, metric with an opaque, if rich, one, and it is doing so without having established the reporting norms that would make the substitution safe. Yang et al. (2020) identified data management, standardised ontologies and FAIR practice as the limiting factors in crop phenomics, and this review's quality appraisal confirms that judgement empirically. Until severity time series are deposited with their acquisition metadata, and until modelling code is published alongside results, the reproducibility of a dynamic phenotype will be strictly worse than that of the scalar it replaces, and that is a real cost to be set against the informational gain, not a formality.

Deployability is the third constraint, and it separates what is demonstrable from what is adoptable. The framework in Figure 5 requires no technology that does not exist, but each of its stages carries a cost that the demonstration studies did not have to bear. UAV revisits require flight windows, and the epidemics that matter most often develop under exactly the overcast, wet conditions in which flying is impossible or the imagery is unusable. Sensor-derived severity requires calibration against validated visual standards in every new pathosystem, as Bock et al. (2020) insist, and that calibration is neither free nor transferable. Dynamic models require time points, and a smooth trajectory cannot be recovered from three assessments however sophisticated the basic functions; the sparse-data regime, in which most breeding programmes actually operate, is precisely where these methods are weakest and where a robust AUDPC may remain the better estimator. Interpretability is a further cost: a breeder can act on a single number, but the operational meaning of the second functional principal component is not self-evident, and until trajectory features are translated into breeding objectives that a selection index can accommodate, the richer phenotype will not change a single selection decision. None of these obstacles is prohibitive, and none is an argument for the status quo; but together they explain why a method that has been analytically compelling since Wu and Lin (2006) has, twenty years later, still not displaced a metric formalised in 1963.

That history invites the review's most practically useful question, which is not whether dynamic phenotyping is better in principle but under what conditions it is better in practice. The evidence supports a reasonably specific answer. Dynamic representation adds value, first, when the time series is dense enough to constrain a curve, which in the included studies means roughly six or more assessments spanning the exponential and plateau phases; below that, the fitted trajectory is dominated by the choice of basis function rather than by the data, and the scalar summary is more honest. It adds value, second, when genotypes are expected to differ in the shape of the epidemic and not merely in its magnitude, which is the situation in quantitative, stage-dependent resistance and precisely not the situation when a major R gene produces a clean qualitative contrast; where the phenotype is bimodal, the trajectory is an expensive way of learning what a single score already tells you. It adds value, third, when the trait of interest is the rate or the onset itself, as it is in slow-rusting and other partial resistances of the kind Shaner and Finney (1977) first quantified, where the scalar metric erases the very quantity being selected for. It adds value, fourth, when the downstream analysis can absorb a functional phenotype, which requires random-regression or functional-mapping machinery and the statistical capacity to run it. Where those four conditions are absent, and in the great majority of routine breeding-programme screens at least one is, the AUDPC remains a defensible and sometimes superior choice, and this review should not be read as an argument for

abandoning it. The DSI and the AUDPC are not errors to be corrected but summaries whose limits are now precisely known, and a metric whose limits are known is more useful than one whose limits are not.

The evidence map exposes a distributional problem that is easy to miss in a narrative reading and that has direct consequences for anyone attempting to generalise from this literature. Dynamic and functional methods have been demonstrated almost exclusively in cereals and in cross-crop methodological work, and, tellingly, most of those demonstrations are not disease studies at all but studies of growth, biomass or greenness dynamics from which disease applications are extrapolated. State-space modelling, meanwhile, is confined almost entirely to two tree-crop pathosystems and owes its existence to the Xylella fastidiosa emergency in Europe. Not one included study applies functional data analysis to a legume disease, and not one applies a state-space model to a cereal rust epidemic at breeding-trial scale, even though cereal rusts are the pathosystem in which the AUDPC is most heavily used and in which slow-rusting resistance makes epidemic rate the explicit breeding target. The empty cells of the evidence map are not evidence of absence, but they mark where the claims of this literature are extrapolations rather than demonstrations, and they identify with unusual precision where the next studies should be done. The corollary for a Nigerian or West African readership is sharper still: the crops on which regional food security most depends, including cowpea, sorghum, cassava and yam, appear nowhere in this corpus, and the transferability of methods developed on irrigated temperate cereals to smallholder systems with different disease pressures, different assessment resources and different flight regulations is entirely untested.

This review has limitations of its own that bear on how far its conclusions travel. It is a qualitative and conceptual synthesis rather than a quantitative meta-analysis: no pooled effect sizes were computed, because the outcomes reported by the included studies are not commensurable, and vote counting is a weaker instrument than pooling because it registers direction without magnitude. The protocol was not registered in PROSPERO, which leaves the review more exposed to unrecorded post hoc decisions than a registered protocol would be, although the screening log and extraction table are supplied in full so that those decisions can be inspected. The deliberate concentration on literature from 2020 onwards, supplemented by canonical works, may under-represent developments between 1990 and 2019, particularly in statistical epidemiology. The thematic boundaries overlap, so a study may be counted in more than one theme, and the resulting counts describe emphasis rather than partition. Restriction to English-language records will have excluded relevant work, particularly from China, where much of the UAV phenotyping literature originates. Finally, the corpus is small enough, at 55 studies, that the addition of a handful of further studies could shift the descriptive proportions reported in Section 3.2, though not, given the analytic basis of the central claim, the conclusion drawn from them.

What the evidence finally supports is a more modest and more actionable proposition than the one with which this literature usually presents itself. It is not that the disease severity index and the area under the disease progress curve should be abandoned; they are cheap, comparable, robust and, for many purposes, sufficient. It is that the field now knows exactly what they cost, can measure what they discard, and, for the first time, possesses both the sensing capacity to retain the discarded information and the statistical machinery to analyse it. What it does not yet possess is the evidence that doing so improves cultivars, the reporting norms that would make the resulting phenotypes reproducible, or the negative results that would tell it where the approach fails. Those three deficits, rather than any further advance in sensing, define the agenda.

CONCLUSION

This systematic review of 55 studies traced disease phenotyping from the Horsfall-Barratt scale of 1945 to contemporary UAV and deep-learning platforms, and classified every included study by the amount of temporal information its phenotype retains. The result is a measured statement of a widely held intuition: 54.5% of the corpus represents disease with a single value from a single assessment, a further 14.5% collapses an entire epidemic into one integrated area, and only 21.8% retains the trajectory itself. The disease severity index and the area under the disease progress curve remain indispensable, and this review does not recommend abandoning them; it establishes precisely what they cost. The DSI is a snapshot that carries no information about how a disease state was reached. The AUDPC is a many-to-one integral in which an early, gradual epidemic and a delayed, explosive one become indistinguishable, though they imply different resistance mechanisms, different durability and different management. Because most crop resistance is quantitative and stage-dependent, a phenotype that erases stage information will systematically under-resolve the genetic architecture beneath it.

The simultaneous maturation of high-throughput temporal sensing and of longitudinal, functional and state-space modelling now makes it feasible to keep the whole trajectory, and every element of the pipeline required to do so is instantiated by studies in this corpus. The evidence that doing so pays, however, is weaker than the enthusiasm of the literature suggests. All 29 comparable within-study comparisons favoured the temporally richer method and none reported a null result, an asymmetry that indicates reporting bias rather than universal superiority; and no included study reports realised genetic gain in the field from selection on a trajectory phenotype. Four priorities follow. First, the field should publish its negative results, so that the conditions under which dynamic methods fail become as well documented as those under which they succeed. Second, severity time series and modelling code should be deposited under FAIR principles, since a functional phenotype that cannot be recomputed is worse than a scalar that can. Third, dynamic phenotyping should be tested where the evidence map is empty and the need is greatest, including legume, cereal-rust and tropical smallholder pathosystems, rather than only where the demonstration is easiest. Fourth, and most consequentially, one breeding programme should run a complete selection cycle in which a trajectory feature, such as delayed onset or suppressed rate, is the selection criterion, and report the realised gain against a conventional AUDPC-selected control.

Until that experiment is done, the case for dynamic disease phenotyping rests on an analytic argument that is close to unanswerable and an empirical foundation that remains thin. Treating the disease progress curve as a dynamic, probabilistic phenotype, and carrying it intact from sensor to selection, offers a coherent route past the representational bottleneck. Whether that route leads to better cultivars is now a question that can, and should, be settled experimentally, and the field has every tool it needs to settle it.

REFERENCES

Adak, A., Murray, S. C., & Anderson, S. L. (2024). Deciphering temporal growth patterns in maize: Integrative modeling of phenotype dynamics and underlying genomic variations. New Phytologist, 242(4), 1486-1503. https://doi.org/10.1111/nph.19575.

Adi, B. A., & Sani, Y. E. (2025). Genetic diversity and characterization of common bacterial blight resistance in common bean (*Phaseolus vulgaris* L.) In northern Nigeria using genome-wide association studies. FUDMA Journal of Sciences, 9(2), 285-289. https://doi.org/10.33003/fjs-2025-0902-3368

Akhtar, M. S., Zafar, Z., Nawaz, R., & Fraz, M. M. (2024). Unlocking plant secrets: A systematic review of 3D imaging in plant phenotyping techniques. Computers and Electronics in Agriculture, 222, 109033. https://doi.org/10.1016/j.compag.2024.109033

Alves, K. S., & Del Ponte, E. M. (2021). Analysis and simulation of plant disease progress curves in R: Introducing the epifitter package. Phytopathology Research, 3, 22. https://doi.org/10.1186/s42483-021-00098-7

Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. Trends in Plant Science, 19(1), 52-61. https://doi.org/10.1016/j.tplants.2013.09.008

Baba, T., Momen, M., Campbell, M. T., Walia, H., & Morota, G. (2020). Multi-trait random regression models increase genomic prediction accuracy for a temporal physiological trait derived from high-throughput phenotyping. PLOS ONE, 15(2), e0228118. https://doi.org/10.1371/journal.pone.0228118

Barbedo, J. G. A. (2019). Plant disease identification from individual lesions and spots using deep learning. Biosystems Engineering, 180, 96-107. https://doi.org/10.1016/j.biosystemseng.2019.02.002

Bock, C. H., Barbedo, J. G. A., Del Ponte, E. M., Bohnenkamp, D., & Mahlein, A.-K. (2020). from visual estimates to fully automated sensor-based measurements of plant disease severity: Status and challenges for improving accuracy. Phytopathology Research, 2, 9. https://doi.org/10.1186/s42483-020-00049-8

Bock, C. H., Hotchkiss, M. W., & Wood, B. W. (2016). Assessing disease severity: Accuracy and reliability of rater estimates in relation to number of diagrams in a standard area diagram set. Plant Pathology, 65(2), 261-272. https://doi.org/10.1111/ppa.12403

Bock, C. H., Poole, G. H., Parker, P. E., & Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. Critical Reviews in Plant Sciences, 29(2), 59-107. https://doi.org/10.1080/07352681003617285

Campbell, C. L., & Madden, L. V. (1990). Introduction to plant disease epidemiology. John Wiley & Sons.

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. BMJ, 368, l6890. https://doi.org/10.1136/bmj.l6890

Campbell, M., Walia, H., & Morota, G. (2018). Utilizing random regression models for genomic prediction of a longitudinal trait derived from high-throughput phenotyping. Plant Direct, 2(9), e00080. https://doi.org/10.1002/pld3.80

Chaloner, T. M., Gurr, S. J., & Bebber, D. P. (2021). Plant pathogen infection risk tracks global crop yields under climate change. Nature Climate Change, 11(8), 710-715. https://doi.org/10.1038/s41558-021-01104-8

Chiang, K.-S., & Bock, C. H. (2022). Understanding the ramifications of quantitative ordinal scales on accuracy of estimates of disease severity and data analysis in plant pathology. Tropical Plant Pathology, 47(1), 58-73. https://doi.org/10.1007/s40858-021-00446-0

Chiang, K.-S., Liu, H.-I., & Bock, C. H. (2017). A discussion on disease severity index values. Part I: Warning on inherent errors and suggestions to maximise accuracy. Annals of Applied Biology, 171(2), 139-154. https://doi.org/10.1111/aab.12362

Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics, 126(4), 867-887. https://doi.org/10.1007/s00122-013-2066-0

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46. https://doi.org/10.1177/001316446002000104

Costa-Neto, G., Fritsche-Neto, R., & Crossa, J. (2021). Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. Heredity, 126(1), 92-106. https://doi.org/10.1038/s41437-020-00353-1

Cunniffe, N. J., Koskella, B., Metcalf, C. J. E., Parnell, S., Gottwald, T. R., & Gilligan, C. A. (2015). Thirteen challenges in modelling plant diseases. Epidemics, 10, 6-10. https://doi.org/10.1016/j.epidem.2014.06.002

Del Ponte, E. M., Pethybridge, S. J., Bock, C. H., Michereff, S. J., Machado, F. J., & Spolti, P. (2017). Standard area diagrams for aiding severity estimation: Scientometrics, pathosystems, and methodological trends in the last 25 years. Phytopathology, 107(10), 1161-1174. https://doi.org/10.1094/PHYTO-02-17-0069-FI

Franceschi, V. T., Alves, K. S., Mazaro, S. M., Godoy, C. V., Duarte, H. S. S., & Del Ponte, E. M. (2020). A new standard area diagram set for assessment of severity of soybean rust improves accuracy of estimates and optimizes resource use. Plant Pathology, 69(3), 495-505. https://doi.org/10.1111/ppa.13148

Furbank, R. T., & Tester, M. (2011). Phenomics: Technologies to relieve the phenotyping bottleneck. Trends in Plant Science, 16(12), 635-644. https://doi.org/10.1016/j.tplants.2011.09.005

Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. Proceedings of the National Academy of Sciences, 115(18), 4613-4618. https://doi.org/10.1073/pnas.1716999115

Gilioli, G., Simonetto, A., Colturato, M., Bazarra, N., Fernandez, J. R., Naso, M. G., Bosco, D., & Bodino, N. (2023). An eco-epidemiological model supporting rational disease management of Xylella fastidiosa: An application to the outbreak in Apulia (Italy). Ecological Modelling, 476, 110226. https://doi.org/10.1016/j.ecolmodel.2022.110226

Gill, T., Gill, S. K., Saini, D. K., Chopra, Y., de Koff, J. P., & Sandhu, K. S. (2022). A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. Phenomics, 2(3), 156-183. https://doi.org/10.1007/s43657-022-00048-z

Gimenez-Romero, A., Moralejo, E., & Matias, M. A. (2023). A compartmental model for Xylella fastidiosa diseases with explicit vector seasonal dynamics. Phytopathology, 113(9), 1686-1696. https://doi.org/10.1094/PHYTO-09-22-0356-R

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. Campbell Systematic Reviews, 18(2), e1230. https://doi.org/10.1002/cl2.1230

Hebert, T. T. (1982). The rationale for the Horsfall-Barratt plant disease assessment scale. Phytopathology, 72(10), 1269. https://doi.org/10.1094/Phyto-72-1269

Heidarian Dehkordi, R., El Jarroudi, M., Kouadio, L., Meersmans, J., & Beyer, M. (2020). Monitoring wheat leaf rust and stripe rust in winter wheat using high-resolution UAV-based red-green-blue imagery. Remote Sensing, 12(22), 3696. https://doi.org/10.3390/rs12223696

Hickey, L. T., Hafeez, A. N., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., Gao, C., Godwin, I. D., Hayes, B. J., & Wulff, B. B. H. (2019). Breeding crops to feed 10 billion. Nature Biotechnology, 37(7), 744-754. https://doi.org/10.1038/s41587-019-0152-9

Hilton, A., Wang, X., Jo, Y. K., Conner, P., Randall, J., Chatwin, W., & Bock, C. (2024). Standard area diagrams for pecan leaf scab: Effect of rater experience, location, and leaf size on reliability and accuracy of visual estimates. Plant Disease, 108(6), 1820-1832. https://doi.org/10.1094/PDIS-09-23-1947-RE

Hong, Q. N., Fabregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Education for Information, 34(4), 285-291. https://doi.org/10.3233/EFI-180221

Horsfall, J. G., & Barratt, R. W. (1945). An improved grading system for measuring plant diseases. Phytopathology, 35, 655.

Horsfall, J. G., & Heuberger, J. W. (1942). Measuring magnitude of a defoliation disease of tomatoes. Phytopathology, 32, 226-232.

Jeger, M. J., & Viljanen-Rollinson, S. L. H. (2001). The use of the area under the disease-progress curve (AUDPC) to assess quantitative disease resistance in crop cultivars. Theoretical and Applied Genetics, 102(1), 32-40. https://doi.org/10.1007/s001220051615

Jin, X., Zarco-Tejada, P. J., Schmidhalter, U., Reynolds, M. P., Hawkesford, M. J., Varshney, R. K., Yang, T., Nie, C., Li, Z., Ming, B., Xiao, Y., Xie, Y., & Li, S. (2021). High-throughput estimation of crop traits: A review of ground and aerial phenotyping platforms. IEEE Geoscience and Remote Sensing Magazine, 9(1), 200-231. https://doi.org/10.1109/MGRS.2020.2998816

Knoch, D., Abbadi, A., Grandke, F., Meyer, R. C., Samans, B., Werner, C. R., Snowdon, R. J., & Altmann, T. (2020). Strong temporal dynamics of QTL action on plant growth progression revealed through high-throughput phenotyping in canola. Plant Biotechnology Journal, 18(1), 68-82. https://doi.org/10.1111/pbi.13171

Kuswidiyanto, L. W., Noh, H.-H., & Han, X. (2022). Plant disease diagnosis using deep learning based on aerial hyperspectral images: A review. Remote Sensing, 14(23), 6031. https://doi.org/10.3390/rs14236031

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174. https://doi.org/10.2307/2529310

Lewin, S., Booth, A., Glenton, C., Munthe-Kaas, H., Rashidian, A., Wainwright, M., Bohren, M. A., Tuncalp, O., Colvin, C. J., Garside, R., Carlsen, B., Langlois, E. V., & Noyes, J. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings: Introduction to the series. Implementation Science, 13(Suppl 1), 2. https://doi.org/10.1186/s13012-017-0688-3

Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics, 45(1), 255-268. https://doi.org/10.2307/2532051

Madden, L. V., Hughes, G., & van den Bosch, F. (2007). The study of plant disease epidemics. American Phytopathological Society.

Mahlein, A.-K. (2016). Plant disease detection by imaging sensors: Parallels and specific demands for precision agriculture and plant phenotyping. Plant Disease, 100(2), 241-251. https://doi.org/10.1094/PDIS-03-15-0340-FE

Mahlein, A.-K., Kuska, M. T., Behmann, J., Polder, G., & Walter, A. (2018). Hyperspectral sensors and imaging technologies in phytopathology: State of the art. Annual Review of Phytopathology, 56, 535-558. https://doi.org/10.1146/annurev-phyto-080417-050100

McGee, R. J., Coyne, C. J., Cheng, P., Vandemark, G. J., & Porter, L. D. (2025). Breeding for quantitative disease resistance: Case studies, emerging approaches, and exploiting pathogen variation. Crop Science, 65(1), e70202. https://doi.org/10.1002/csc2.70202

Merrick, L. F., Burke, A. B., Chen, X., & Carter, A. H. (2021). Breeding with major and minor genes: Genomic selection for quantitative disease resistance. Frontiers in Plant Science, 12, 713667. https://doi.org/10.3389/fpls.2021.713667

Momen, M., Campbell, M. T., Walia, H., & Morota, G. (2019). Predicting longitudinal traits derived from high-throughput phenomics in contrasting environments using genomic Legendre polynomials and B-splines. G3: Genes, Genomes, Genetics, 9(10), 3369-3380. https://doi.org/10.1534/g3.119.400346

Moreira, F. F., Oliveira, H. R., Volenec, J. J., Rainey, K. M., & Brito, L. F. (2020). Integrating high-throughput phenotyping and statistical genomic methods to genetically improve longitudinal traits in crops. Frontiers in Plant Science, 11, 681. https://doi.org/10.3389/fpls.2020.00681

Murphy, K. M., Ludwig, E., Gutierrez, J., & Gehan, M. A. (2024). Deep learning in image-based plant phenotyping. Annual Review of Plant Biology, 75, 771-795. https://doi.org/10.1146/annurev-arplant-070523-042828

Ojwang', A. M. E., Ruiz Barlett, V., Cobre, A. F., Ojiambo, P. S., & Gent, D. H. (2021). A general framework for spatio-temporal modeling of epidemics with multiple epicenters: Application to an aerially dispersed plant pathogen. Frontiers in Applied Mathematics and Statistics, 7, 721352. https://doi.org/10.3389/fams.2021.721352

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan: A web and mobile app for systematic reviews. Systematic Reviews, 5, 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Pethybridge, S. J., & Nelson, S. C. (2015). Leaf Doctor: A new portable application for quantifying plant disease severity. Plant Disease, 99(10), 1310-1316. https://doi.org/10.1094/PDIS-03-15-0319-RE

Qu, J., Morota, G., & Cheng, H. (2022). A Bayesian random regression method using mixture priors for genome-enabled analysis of time-series high-throughput phenotyping data. The Plant Genome, 15(3), e20228. https://doi.org/10.1002/tpg2.20228

Ristaino, J. B., Anderson, P. K., Bebber, D. P., Brauman, K. A., Cunniffe, N. J., Fedoroff, N. V., Finegold, C., Garrett, K. A., Gilligan, C. A., Jones, C. M., Martin, M. D., MacDonald, G. K., Neenan, P., Records, A., Schmale, D. G., Tateosian, L., & Wei, Q. (2021). The persistent threat of emerging plant disease pandemics to global food security. Proceedings of the National Academy of Sciences, 118(23), e2022239118. https://doi.org/10.1073/pnas.2022239118

Sandhu, K. S., Merrick, L. F., Sankaran, S., Zhang, Z., & Carter, A. H. (2022). Prospectus of genomic selection and phenomics in cereal, legume and oilseed breeding programs. Frontiers in Genetics, 12, 829131. https://doi.org/10.3389/fgene.2021.829131

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. Nature Ecology & Evolution, 3(3), 430-439. https://doi.org/10.1038/s41559-018-0793-y

Shahi, T. B., Xu, C.-Y., Neupane, A., & Guo, W. (2023). Recent advances in crop disease detection using UAV and deep learning techniques. Remote Sensing, 15(9), 2450. https://doi.org/10.3390/rs15092450

Shaner, G., & Finney, R. E. (1977). The effect of nitrogen fertilization on the expression of slow-mildewing resistance in Knox wheat. Phytopathology, 67, 1051-1056. https://doi.org/10.1094/Phyto-67-1051

Shoaib, M., Sadeghi-Niaraki, A., Ali, F., Hussain, I., & Khalid, S. (2025). Leveraging deep learning for plant disease and pest detection: A comprehensive review and future directions. Frontiers in Plant Science, 16, 1538163. https://doi.org/10.3389/fpls.2025.1538163

Simko, I., & Piepho, H.-P. (2012). The area under the disease progress stairs: Calculation, advantage, and application. Phytopathology, 102(4), 381-389. https://doi.org/10.1094/PHYTO-07-11-0216

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep learning for plant stress phenotyping: Trends and future perspectives. Trends in Plant Science, 23(10), 883-898. https://doi.org/10.1016/j.tplants.2018.07.004

Song, P., Wang, J., Guo, X., Yang, W., & Zhao, C. (2021). High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. The Crop Journal, 9(3), 633-645. https://doi.org/10.1016/j.cj.2021.03.015

Upadhyaya, J., Bayer, P. E., Tirnaz, S., Edwards, D., & Batley, J. (2024). Genomics-based plant disease resistance prediction using machine learning. Plant Pathology, 73(8), 2069-2084. https://doi.org/10.1111/ppa.13988

Van der Plank, J. E. (1963). Plant diseases: Epidemics and control. Academic Press.

Varshney, R. K., Bohra, A., Roorkiwal, M., Barmukh, R., Cowling, W. A., Chitikineni, A., Lam, H.-M., Hickey, L. T., Croser, J. S., Bayer, P. E., Edwards, D., Crossa, J., Weckwerth, W., Millar, A. H., Kumar, A., Bevan, M. W., & Siddique, K. H. M. (2021). Fast-forward breeding for a food-secure world. Trends in Genetics, 37(12), 1124-1136. https://doi.org/10.1016/j.tig.2021.08.002

Visakh, R. L., Anand, S. S., Nair, R. A., & Soman, S. (2024). Precision phenotyping in crop science: From plant traits to gene discovery for climate-smart agriculture. Plant Breeding, 143(6), 749-774. https://doi.org/10.1111/pbr.13228

Wu, R., & Lin, M. (2006). Functional mapping: How to map and study the genetic architecture of dynamic complex traits. Nature Reviews Genetics, 7(3), 229-237. https://doi.org/10.1038/nrg1804

Xu, Y., Zhang, X., Li, H., Zheng, H., Zhang, J., Olsen, M. S., Varshney, R. K., Prasanna, B. M., & Qian, Q. (2022). Smart

breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction. Molecular Plant, 15(11), 1664-1695. https://doi.org/10.1016/j.molp.2022.09.001

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., Xiong, L., & Yan, J. (2020). Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. Molecular Plant, 13(2), 187-214. https://doi.org/10.1016/j.molp.2020.01.008

Zakieh, M., Alemu, A., Henriksson, T., Pareek, N., Singh, P. K., & Chawade, A. (2023). Exploring GWAS and genomic prediction to improve Septoria tritici blotch resistance in wheat. Scientific Reports, 13, 15651. https://doi.org/10.1038/s41598-023-42856-x