

Concept Paper

Not peer-reviewed version

AI-Enhanced Digital Twins as Decision-Support Layers for Precision Management of Microbial Communication: The Case of Cocoa Pulp Juice Fermentation

[Anthony Oppong Kyekyeku](#)*, Margaret Owusu*, John Edem Kongor*, Daniel Sitsofe Yabani*

Posted Date: 5 May 2026

doi: 10.20944/preprints202605.0102.v1

Keywords: cocoa pulp juice fermentation; quorum sensing; digital twin; bioprocess decision suite; transformer model; precision fermentation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Concept Paper

AI-Enhanced Digital Twins as Decision-Support Layers for Precision Management of Microbial Communication: The Case of Cocoa Pulp Juice Fermentation

Anthony Oppong Kyekyeku *, Margaret Owusu *, John Edem Kongor * and Daniel Sitsofe Yabani *

CSIR College of Science and Technology (CCST), Faculty of Bioresources and Applied Sciences, Department of Agro-Processing Technology and Food Biosciences, Accra, Ghana

* Correspondence: akyekyeku.stu@ccst.edu.gh (A.O.K.); mowusu@ccst.edu.gh (M.O.), johnedem.kongor@foodresearchgh.org (J.E.K.); sityab@yahoo.com (D.S.Y.)

Highlights

- QS signalling invisibly governs cocoa fermentation beyond bioreactor control.
- Digital twin is the QS-aware intelligence layer bioreactors currently lack.
- Enables in silico sensory predesign and batch-to-batch cocoa-wine consistency.
- Time-series transformers with SHAP explainability anchor the deployable AI layer.
- Three-tier framework scales QS decision support to full biomanufacturing.

Abstract

Fermentation quality is shaped not just by what microbes produce but by how they communicate. Cocoa pulp juice fermentation is governed by quorum sensing (QS) — bacterial autoinducers (AHLs, AI-2, peptides) and fungal mediators (farnesol, tyrosol) — that coordinate microbial succession and flavour precursor formation. Yet QS states are absent from operational fermentation control: conventional bioreactors act only on macroscopic variables (pH, temperature, dissolved oxygen) and cannot read the molecular language coordinating quality-determining transitions. This translational perspective proposes that a digital twin, implemented as a bioprocess decision suite, provides the missing intelligence layer. It is not a vessel replacement but a QS-aware reasoning system that translates communication states into explainable, audit-ready decisions, and enables in silico sensory predesign before a run begins. A key design constraint is polyphenol-mediated AHL quenching and acid-accelerated AI-2 degradation in the cocoa matrix, which we formalise as an architectural specification. Time-series transformer models with attention-based explainability anchor the near-term AI layer, with graph neural networks, Bayesian uncertainty quantification, and reinforcement learning forming a progressive maturation roadmap. A three-tier framework scales deployment from artisanal to biomanufacturing operations.

Keywords: cocoa pulp juice fermentation; quorum sensing; digital twin; bioprocess decision suite; transformer model; precision fermentation

1. Introduction

Fermentation quality is shaped not just by what microbes produce, but by how they communicate. In cocoa pulp juice fermentation — the conversion of the fermentable liquid fraction of *Theobroma cacao* pulp into cocoa-based wine and related beverages — the metabolic transitions that determine product quality are not the passive consequence of substrate depletion. They are actively orchestrated through density-dependent molecular communication: quorum sensing (QS)

autoinducers accumulate to threshold concentrations and trigger coordinated shifts in gene expression, metabolic activity, biofilm organisation, and stress response across yeasts, lactic acid bacteria (LAB), and acetic acid bacteria (AAB) (Schwan and Wheals, 2004; De Vuyst and Weckx, 2016; Rutherford and Bassler, 2012; Campos et al., 2025). By the time conventional fermentation control systems — whether in a laboratory bioreactor, an industrial stirred-tank, or a cooperative processing vessel — register a meaningful change in pH, ORP, or dissolved oxygen, the QS-driven community transitions responsible for that change have already occurred. The opportunity for early corrective or preventive action has passed.

Bioreactors represent the most sophisticated fermentation vessels currently available: engineered hardware systems with precise control of temperature, pH, dissolved oxygen, agitation, and nutrient feed. Yet the intelligence layer governing a bioreactor — its process logic, its control algorithms, its intervention triggers — is built entirely around macroscopic physicochemical setpoints. A bioreactor is an extraordinarily capable physical controller that operates in complete ignorance of the molecular communication occurring within it. It enforces conditions without reasoning about the biological conversation that is determining what those conditions should be and when they should change. This is not a design flaw; it is a structural consequence of building fermentation control systems around the assumption that the relevant process variables are all physicochemical and continuously measurable. QS communication signals violate that assumption: they are biological, they are intermittent or analytically inaccessible in real time, and they precede the physicochemical outcomes that the bioreactor monitors.

The bioprocess decision suite proposed in this perspective is designed to fill this gap. It is not a replacement for the fermentation vessel. It is the QS-aware intelligence layer that fermentation systems currently lack: a digital twin that reads the molecular communication of the microbial community, interprets it in the context of the desired product, predicts the trajectory toward the target sensory outcome, and translates its reasoning into governed, explainable, and audit-ready intervention decisions that any fermentation vessel — bioreactor, cooperative tank, or artisanal fermenter — can act upon. For existing bioreactors with adequate sensor arrays and digital control interfaces, the digital twin integrates as a software-defined intelligence upgrade requiring no hardware modification. For the longer term, the QS signal thresholds and state transition criteria of the four-state operational model constitute the functional specification for a next-generation QS-native fermentation platform — a bioreactor whose control logic is built around biological communication states rather than their downstream physicochemical consequences.

Two interlocking challenges must be resolved before this intelligence layer can operate reliably. The first is a matrix measurement problem: polyphenols in cocoa pulp juice substantially compromise AHL signal bioavailability and reduce AI-2 stability under acidic fermentation conditions, creating a measurement–control bottleneck that determines the sensing architecture, the role of proxy signals, and the uncertainty quantification requirements of the entire system (Section 4). The second is a decision architecture problem: no operational framework has previously existed for translating QS signal dynamics into the structured, explainable, biomanufacturing-compatible intervention decisions that governance-grade fermentation control requires (Mahanty, 2023). This perspective addresses both.

AI-enhanced digital twin architectures have demonstrated translational feasibility in food fermentation systems (Zhao et al., 2025; Abdurrahman and Ferrari, 2025). Time-series transformer models, with their attention-based explainability, are particularly well-suited to detecting QS-driven phase transitions in fermentation telemetry and producing the feature-attributed explanations that biomanufacturing audit requirements demand (Jin et al., 2020). The proposed architecture places transformer-based explainability at its operational core, with complementary AI components positioned as a progressive maturation roadmap. This perspective is structured as follows: Section 2 describes the article type and evidence synthesis approach; Section 3 examines microbial communication as a regulatory layer; Section 4 characterises signal quenching and matrix constraints as design specifications; Section 5 evaluates molecular and biosensing toolkits; Section 6 presents the

bioprocess decision suite architecture and its relationship to conventional fermentation vessels; and Section 7 develops the tiered translational implementation framework.

2. Conceptual Framework and Review Approach

2.1. Article Type and Methodological Note

This article is a translational perspective. Its objective is to map established QS mechanisms and matrix constraints into a deployable bioprocess decision suite that does not yet exist for cocoa pulp juice fermentation. As a translational perspective, it does not follow a systematic review methodology, and PRISMA reporting guidelines are therefore not applicable. PRISMA governs exhaustive, protocol-registered evidence synthesis aimed at answering a defined empirical question by pooling study outcomes. The present objective is categorically different: it requires integrating heterogeneous evidence from QS ecology, matrix measurement science, AI-enabled bioprocess control, sensory prediction science, and fermentation governance into a coherent, original architectural framework. This is the methodological purpose of a narrative evidence synthesis, which is selective and purposive by design — identifying, critically evaluating, and mapping the most relevant evidence bearing on each component of the proposed architecture. The proposed digital twin is a synthesis-based framework derived from analogous published systems; experimental validation in cocoa pulp juice fermentation contexts is identified as a research priority throughout.

2.2. Evidence Synthesis Workflow and Literature Scope

Literature was searched in PubMed, Scopus, and Web of Science through March 2026 using principal term combinations including: “quorum sensing AND cocoa fermentation”; “digital twin AND food fermentation”; “autoinducer-2 AND lactic acid bacteria”; “acyl-homoserine lactone AND fermentation monitoring”; “machine learning AND bioprocess control”; “transformer model AND fermentation”; “sensory prediction AND fermentation AI”; “bioreactor AND microbial communication”; and “AI-2 instability AND food matrix”. The synthesis proceeds through three steps: (1) mechanistic extraction of QS signals, producers, targets, and functional outcomes; (2) operational mapping of measurable variables and constraints as design specifications; and (3) decision translation to identify intervention points, sensory prediction linkages, explainability requirements, and biomanufacturing auditability standards for the proposed architecture.

3. Microbial Communication as a Regulatory Layer in Cocoa Pulp Juice Fermentation

3.1. Quorum Sensing Architectures Relevant to Cocoa Ecosystems

Quorum sensing operates through the synthesis, secretion, and population-level detection of diffusible autoinducers that accumulate in proportion to cell density and trigger coordinated changes in gene expression once threshold concentrations are exceeded (Rutherford and Bassler, 2012). Three QS architectures are active in cocoa pulp juice fermentation, each regulating distinct aspects of fermentation behaviour and product quality (Table 1). Critically, none of these regulatory events is detectable by any sensor in a conventional bioreactor. A bioreactor monitoring pH, dissolved oxygen, and temperature records the downstream physicochemical consequences of QS-driven transitions; it has no capability to observe the transitions themselves, or to intervene during the window between the QS event and its macroscopic expression.

Gram-negative AAB, which dominate the oxidative phase, communicate through N-acyl-L-homoserine lactones (AHLs). In *Acetobacter pasteurianus* and *Gluconobacter oxydans*, AHL signalling coordinates density-dependent activation of oxidative metabolism, biofilm formation at oxygen interfaces, and alcohol dehydrogenase expression governing the ethanol-to-acetic acid conversion (Churchill and Chen, 2011; Giaouris et al., 2015; Campos et al., 2025). The timing and extent of this

transition, governed by AHL accumulation above the quorum threshold, determines the acetic acid concentration and volatile acidity profile of the finished wine. A bioreactor equipped only with an ORP sensor will detect the consequences of this transition several hours after the QS events that caused it have already occurred.

LAB employ LuxS-mediated AI-2 signalling, confirmed in cocoa-associated *Lactiplantibacillus plantarum* and *Lactobacillus fermentum* (de Melo Pereira et al., 2012) at both pangenome and expression levels (Almeida et al., 2021; 2024), and peptide autoinducer circuits mediating bacteriocin production and spoilage exclusion (Kareb and Aider, 2020). The rate and trajectory of LAB-driven acidification, coordinated through AI-2 accumulation, shapes the organic acid profile and pH of the finished cocoa wine – primary sensory quality determinants that a bioreactor's pH controller responds to only after the trajectory has been established. Yeasts communicate through aromatic alcohol mediators (farnesol, tyrosol) that modulate ethanol yield, ester-alcohol balance, and the timing of the yeast-to-LAB community transition (Albuquerque and Casadevall, 2012; Chen and Fink, 2006). Broader evidence for yeast participation in AI-2 networks is provided by the demonstration that *Saccharomyces cerevisiae* produces an AI-2 mimic via the conserved enzyme Cff1p, with CFF1 homologs identified across diverse fungal genomes (Valastyan et al., 2021); whether *P. kudriavzevii*, a dominant early-phase yeast in cocoa pulp fermentation, similarly engages with bacterial AI-2 through analogous molecular machinery remains uninvestigated, and this cross-kingdom interaction is treated as biologically plausible but experimentally unresolved throughout.

Table 1. QS signal molecules, producers, targets, functional roles, and quality consequences in cocoa pulp juice fermentation. Signal classes were selected based on confirmed production or functional evidence in cocoa-associated microbial taxa reviewed through March 2026. "Invisible to bioreactor" column indicates whether the signal or its regulatory event is directly detectable by conventional bioreactor sensor suites; in all cases it is not – only downstream physicochemical consequences are measurable by standard instrumentation. Fermentation state designations refer to the four-state operational model in Section 6.3. AHL = acyl-homoserine lactone; AI-2 = autoinducer-2; DPD = (S)-4,5-dihydroxy-2,3-pentanedione; PKA = protein kinase A. Sources for tabulated signal classes, producers, and regulatory functions: AHL signalling in AAB – Churchill and Chen (2011), Blana et al. (2017), Wang et al. (2023); AI-2 and LuxS-coding gene expression in cocoa-associated LAB – de Melo Pereira et al. (2012), Almeida et al. (2021, 2024), Kareb and Aider (2020); peptide autoinducers and bacteriocin regulation – Kareb and Aider (2020); farnesol, tyrosol and fungal aromatic-alcohol QS – Chen and Fink (2006), Albuquerque and Casadevall (2012); oxylipins and fungal sporulation signalling – Albuquerque and Casadevall (2012); cross-kingdom and biofilm-mediated interactions – Giaouris et al. (2015), Valastyan et al. (2021); general QS framework – Rutherford and Bassler (2012).

Signal Class	Primary Producer(s)	Key Regulated Behaviours	Quality Consequence in Cocoa Wine	Invisible to Bioreactor's Standard Sensor Suite?	Fermentation State
AHLs (C8-HSL, C10-HSL)	<i>A. pasteurianus</i> , <i>G. oxydans</i>	Oxidative metabolism onset, biofilm at O ₂ interface, alcohol dehydrogenase upregulation	Acetic acid concentration; volatile acidity profile; over-acidification risk	Yes – ORP registers consequence hours later	III–IV
AI-2 (DPD)	<i>L. plantarum</i> , <i>L. fermentum</i>	Acid stress response, biofilm, substrate switching,	Lactic acid profile; pH trajectory; spoilage suppression	Yes – pH registers consequence after LAB expansion	II–III

Signal Class	Primary Producer(s)	Key Regulated Behaviours	Quality Consequence in Cocoa Wine	Invisible to Bioreactor's Standard Sensor Suite?	Fermentation State
		bacteriocin induction			
Peptide autoinducers	<i>L. plantarum</i> , <i>Leuconostoc</i> spp.	Bacteriocin production, competitive exclusion of spoilage taxa	Microbiological safety; spoilage-free shelf life	Yes — no physicochemical correlate exists	II
Farnesol	<i>S. cerevisiae</i> , <i>P. kudriavzevii</i>	Morphogenesis suppression, cross-kingdom competitive signalling	Ethanol yield; ester-alcohol balance; yeast-to-AAB transition timing	Yes — ethanol accumulation is the delayed consequence	I
Tyrosol	<i>S. cerevisiae</i> , <i>P. kudriavzevii</i>	Oxidative stress tolerance, metabolic rate modulation	LAB readiness; flavour precursor accumulation rate	Yes — no physicochemical correlate exists	I–II
Oxylipins	Minor fungal populations	Fungal sporulation; potential bacterial QS interference	Sporadic off-flavour risk; microbial community instability	Yes — not detectable by any standard bioreactor instrument	Sporadic

3.2. Cross-Kingdom Signalling, Emergent Community Behaviour, and Quality Governance

The regulatory events that determine cocoa wine quality are the product of a multi-kingdom molecular conversation that conventional fermentation control systems cannot hear. The yeast-dominant phase (State I) determines ethanol concentration and the ester-alcohol foundation of the aroma profile through farnesol- and tyrosol-mediated community coordination. The LAB-dominant phase (State II) determines organic acid balance and pH through AI-2-mediated coordination of acidification rate and spoilage exclusion. The AAB-oxidative phase (State III) determines volatile acidity and the final aroma complexity through AHL-mediated gating of oxidative metabolism. Each of these quality-determining transitions is governed by QS signal accumulation and threshold crossing — events that are biologically decisive but physically invisible to any current fermentation control system.

The failure mode that illustrates this most clearly is the QS-mediated gating of the yeast-to-AAB oxidative transition. When AHL concentrations exceed the quorum threshold, AAB initiate organised biofilm formation at oxygen interfaces and upregulate alcohol dehydrogenase expression as a coordinated population-level response (Giaouris et al., 2015). Farnesol from concurrent yeast populations may suppress premature AAB biofilm formation, synchronising oxidative metabolism with the completion of ethanol production (Albuquerque and Casadevall, 2012). When this gating fails — through polyphenol quenching of AHL signals, absence of QS-competent strains, or disruption by QS-quenching enzymes produced by competing organisms (Wang et al., 2023) — AAB oxidation becomes unsynchronised, acetic acid accumulates beyond the sensory target range, and the batch fails to meet the predesigned quality specification. A bioreactor responds to this failure only

after the ORP has risen and the acetic acid has already accumulated; the digital twin detects the gating failure through AHL kinetics and proxy-signal surrogate patterns before the transition completes, providing the pre-emptive intervention window that macroscopic monitoring cannot offer.

3.3. The Informational Lead-Time Advantage over Bioreactor Monitoring

QS events systematically precede the macroscopic physicochemical changes they cause. AHL accumulation in AAB populations begins to rise measurably before the ORP inflection marking organised oxidative metabolism. AI-2 peaks before the pH inflection signalling active lactic acid accumulation. In each case, the QS signal provides an informational lead-time advantage over the macroscopic indicator that a bioreactor would detect — a window during which the digital twin can apply a pre-emptive intervention to redirect the fermentation trajectory toward the predesigned sensory target before an outcome is locked in (Almeida et al., 2020). This lead-time advantage is the primary operational justification for the digital twin: it is the difference between preventive control and reactive correction, and it is only accessible if the intelligence layer is reading biological communication signals rather than waiting for their downstream physicochemical consequences.

4. Signal Quenching, Matrix Constraints, and the Measurement Gap

The matrix constraints characterised in this section are not analytical inconveniences; they are architectural specifications for both the digital twin and any future QS-native fermentation platform. They determine what can be measured, when, and with what confidence, which in turn determines the sensing layer design, the proxy-signal strategy, the uncertainty quantification requirements, and the conditions under which inference mode switching must occur. Figure 1 provides a structured summary.

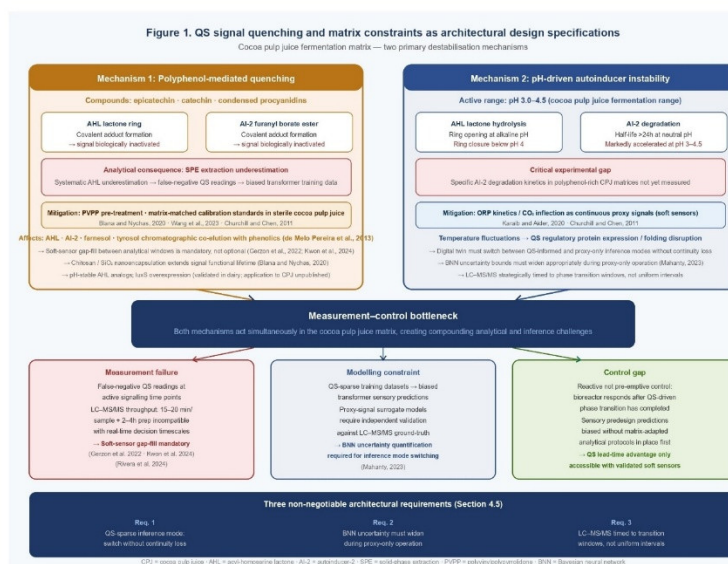


Figure 1. QS signal quenching and matrix constraints in cocoa pulp juice fermentation as architectural design specifications for the bioprocess decision suite and any future QS-native fermentation platform. The figure illustrates two primary destabilisation mechanisms in the cocoa pulp juice matrix (central panel): polyphenol-mediated covalent quenching of AHL and AI-2 signals, and pH-driven autoinducer instability at fermentation-range acidity (pH 3.0–4.5). These create a measurement–control bottleneck with architectural consequences for both the digital twin (QS-sparse training data; proxy-signal surrogate requirement; inference mode switching; BNN uncertainty quantification) and any future QS-native bioreactor (in-line biosensor validation requirement; controlled-release signal stabilisation; continuous proxy monitoring as primary control input). Mitigation

strategies (lower panel): PVPP pre-treatment; matrix-matched calibration; ORP/CO₂ proxy gap-fill; chitosan/SiO₂ encapsulation. The design specification box translates these constraints into the three non-negotiable architectural requirements stated in Section 4.5. CPJ = cocoa pulp juice; AHL = acyl-homoserine lactone; AI-2 = autoinducer-2; SPE = solid-phase extraction; PVPP = polyvinylpyrrolidone; BNN = Bayesian neural network.

4.1. Polyphenol-Mediated Quenching

The cocoa pulp juice matrix contains phenolic compounds — principally epicatechin, catechin, and condensed procyanidins — that form covalent adducts with the lactone ring of AHLs and the furanyl borate ester moiety of AI-2, rendering signal molecules biologically inactive and analytically inaccessible through standard SPE protocols (Churchill and Chen, 2011; Blana et al., 2017; Wang et al., 2023). This quenching effect means that naive application of standard QS analytical workflows to cocoa pulp juice samples will generate systematically biased data. Any digital twin or QS-native bioreactor control system trained or parameterised on such data will learn a distorted representation of QS state dynamics.

4.2. pH- and Temperature-Driven Instability

AHLs undergo acid-catalysed lactone ring closure below pH 4 and base-catalysed ring opening at alkaline pH (Churchill and Chen, 2011). AI-2 degradation is markedly accelerated at fermentation-range acidity (pH 3.0–4.5) relative to neutral conditions where half-life exceeds 24 hours (Kareb and Aider, 2020). The specific degradation kinetics in polyphenol-rich cocoa pulp juice matrices have not been directly measured and represent a critical experimental gap. Temperature fluctuations influence QS regulatory protein expression and folding. Farnesol and tyrosol present chromatographic co-elution challenges with matrix phenolics during LC–MS/MS analysis, requiring optimised gradient protocols and matrix-matched internal standards.

4.3. Throughput Constraints and the Proxy-Signal Imperative

A full LC–MS/MS QS panel requires 15–20 minutes of instrument time per sample plus 2–4 hours of sample preparation. For a 96-hour fermentation run sampled every 6 hours, this creates analytical throughput entirely incompatible with real-time decision timescales. This constraint is not merely an operational inconvenience for the digital twin; it is a fundamental specification for any future QS-native bioreactor as well. A QS-native bioreactor cannot rely on offline LC–MS/MS as its primary QS sensing modality. It requires validated in-line or near-line soft sensors — the established process analytical technology (PAT) term for data-driven models that provide continuous indirect estimates of process variables that cannot be measured directly in real time (Gerzon et al., 2022; Kwon et al., 2024; Rivera et al., 2024) — using ORP kinetics, CO₂ evolution inflections, and temperature drift profiles as proxy inputs for QS-driven state transitions. Developing and validating these surrogate models against periodic LC–MS/MS ground-truth data is therefore a prerequisite for both the digital twin and any future QS-native fermentation platform.

4.4. Mitigation Strategies and their Roles in the Architecture

PVPP pre-treatment reduces polyphenol-mediated extraction interference (Blana et al., 2017; Wang et al., 2023). Matrix-matched calibration standards in sterile cocoa pulp juice are an essential adaptation not yet systematically validated for this substrate. Chitosan-based and silica nanoparticle encapsulation extends signal functional lifetime by shielding autoinducers from polyphenol attack (Blana et al., 2017). pH-stable AHL analogs and *luxS* overexpression are validated in dairy fermentation contexts (Churchill and Chen, 2011; Kareb and Aider, 2020); direct application to cocoa pulp juice has not been published. Validated physicochemical proxy signals — ORP kinetics, CO₂ evolution inflections, temperature drift profiles — form the foundation of the continuous inference layer that operates between analytical sampling points in both the digital twin and any future QS-native bioreactor control system.

4.5. Architectural Specifications Derived from Matrix Constraints

Three non-negotiable architectural requirements follow directly from the matrix characterisation above: (1) the sensing layer must accommodate QS-sparse inputs, switching between QS-informed and proxy-only inference modes without loss of state estimation continuity; (2) uncertainty quantification must be calibrated to the measurement gaps introduced by matrix interference, producing confidence intervals that widen appropriately when operating on proxy signals alone; and (3) LC-MS/MS quantification must be strategically timed to the phase transition windows identified by continuous proxy-signal monitoring, concentrating analytical resources at the moments of highest informational value (Mahanty, 2023). These requirements apply equally to the digital twin architecture and to the design of any future QS-native fermentation platform.

5. Molecular and Biosensing Toolkits for QS-Guided Fermentation Control

5.1. The Measurement–Control Interface

Translating QS mechanisms into operational fermentation control requires both detection and engineering capability. In the context of the digital twin as a bioreactor intelligence layer, the detection toolkit provides the ground-truth QS measurements against which transformer model proxy-signal surrogates are validated, the training labels from which sensory prediction models are learned, and the real-time data inputs that update the digital twin's state estimate during a production run. The engineering toolkit provides the mechanisms by which digital twin intervention recommendations involving QS signal modulation are operationalised in the fermentation vessel — whether that vessel is a laboratory bioreactor, a cooperative-scale closed tank, or a production-scale stirred-tank fermentation system.

5.2. Analytical Platforms

5.2.1. LC-MS/MS as the Reference Standard and Transformer Training Label Source

LC-MS/MS provides simultaneous profiling of AHLs (C4–C14 chain lengths), AI-2 (as DPD or via DMPD derivatisation), farnesol, and tyrosol in a single targeted workflow, with critical cocoa pulp juice adaptations including pH adjustment before SPE, matrix-matched calibration standards in sterile cocoa pulp juice, and stable isotope-labelled internal standards. The LC-MS/MS dataset generated at strategic sampling windows, paired with QDA sensory data and GC-MS volatile profiling from the same fermentation runs, provides the ground-truth training corpus — the annotated pairing of QS signal kinetics with sensory outcomes — that enables the transformer model to learn *in silico* sensory pre-design capability. For any future QS-native bioreactor, LC-MS/MS serves the analogous function of generating the validated reference data against which in-line QS proxy sensors are calibrated.

5.2.2. Whole-Cell Biosensors and In-Line Sensing Prospects

Whole-cell reporter strains — *Vibrio harveyi* BB170 for AI-2, *Agrobacterium tumefaciens* NTL4 for medium-to-long-chain AHLs, *Chromobacterium violaceum* CV026 for short-chain AHLs — report biologically active signal fractions, enabling distinction of intact functional signals from inactive degradation products. Recent advances in miniaturised AI-integrated biosensing suggest that field-deployable and in-line biosensing approaches are approaching practical readiness for complex food fermentation environments (Jin et al., 2020; Niyigaba et al., 2025). For the digital twin, these systems function as functional screening and validation complements to LC-MS/MS reference quantification. For a future QS-native bioreactor, validated in-line whole-cell biosensors would provide the real-time QS signal monitoring capability that would make continuous QS-aware control possible without the analytical throughput constraint of LC-MS/MS. Validation in cocoa pulp juice matrices has not yet been published and is a prerequisite for both applications.

5.2.3. Multi-Omics Integration for Sensory–QS Model Linkage

PLSR applied to the combined QS time-series and GC–MS volatile profile dataset generates the quantitative sensory–QS coupling model: the mathematical relationship between QS signal kinetics and the volatile and organic acid composition that determines wine sensory descriptors. This coupling model is the engine of *in silico* sensory predesign, and its progressive refinement across annotated fermentation runs is the mechanism by which the digital twin learns to predict sensory outcomes with increasing precision. Metagenomics characterises the QS gene complement at each fermentation stage; metatranscriptomics resolves whether low signal concentrations reflect high degradation or genuinely low production — a distinction with direct implications for how the digital twin interprets its state estimates and for how a future QS-native bioreactor should respond to low-signal conditions. The integration of multi-omics data streams with fermentation process monitoring — linking genomics, transcriptomics, and metabolomics to process outcomes — has been demonstrated as a tractable approach for deciphering complex fermentation microbiome functionality and building data-driven fermentation control frameworks (Ferrocino et al., 2022).

5.3. Signal Engineering and its Role in the Bioreactor–Digital Twin System

pH-stable AHL analogs, *luxS* overexpression in LAB, and chitosan/silica nanoparticle encapsulation represent strategies for extending QS signal functional lifetime under matrix constraints (Churchill and Chen, 2011; Kareb and Aider, 2020; Blana et al., 2017). In the bioprocess decision suite, engineering interventions that modify QS signal profiles are detectable by the same LC–MS/MS methods that monitor endogenous signals, allowing the digital twin to verify that a recommended intervention has produced the expected change in QS state and to update its sensory trajectory prediction accordingly. In a future QS-native bioreactor, signal engineering strategies — particularly encapsulation for controlled release — could be implemented directly within the vessel to maintain QS signal concentrations above the analytical detection threshold throughout the fermentation run, enabling continuous in-line monitoring rather than periodic offline quantification. This would close the measurement gap that currently necessitates the proxy-signal strategy.

6. The Bioprocess Decision Suite: Architecture, Relationship to Fermentation Vessels, and Operational Purposes

6.1. The Digital Twin and the Bioreactor: A Precise Juxtaposition

Understanding what the digital twin is requires first being precise about what a bioreactor is and what it is not. A bioreactor is a physical vessel: an engineered hardware system that provides and regulates the physical and chemical environment within which a biological process occurs. Its sensors measure temperature, pH, dissolved oxygen, ORP, CO₂ evolution rate, and agitation. Its actuators adjust heating, cooling, acid/base addition, aeration, and agitation. Its control logic responds to deviations of measured variables from preset target values. A bioreactor enforces conditions; it does not reason about them. It is a sophisticated physical controller whose intelligence layer — the logic that determines what the conditions should be and when they should change — operates exclusively on macroscopic physicochemical signals.

The digital twin proposed in this perspective is not a bioreactor. It is not a vessel, not a hardware system, and not a replacement for fermentation infrastructure. It is the QS-aware reasoning layer that operates above and around the fermentation vessel — the decision intelligence that determines what the vessel should do, when, and why, based on the microbial communication states that the vessel itself cannot perceive.

The relationship between the digital twin and the fermentation vessel can be understood through three distinct framings, each relevant to a different stage of technological development and production context (Table 2).

Framing 1 – Intelligence upgrade for existing fermentation vessels. For any fermentation vessel – bioreactor, cooperative-scale closed tank, or pilot-scale fermenter – already equipped with pH, temperature, ORP, DO, and CO₂ sensors, and with a digital control interface, the digital twin integrates as a software-defined intelligence layer requiring no hardware modification. The vessel's existing sensor array provides the telemetry inputs to the digital twin's Module 1 sensing layer. The digital twin's Module 3 transformer model processes these inputs, detects QS-associated proxy-signal patterns indicative of approaching phase transitions, and generates intervention recommendations. These recommendations are communicated to the vessel's control system as updated setpoint instructions or operator alerts, which the vessel then enacts through its existing actuators. In this configuration, the digital twin transforms any adequately instrumented fermentation vessel from a reactive physicochemical controller into a QS-aware predictive control system. No new hardware is required; only the intelligence layer is added.

Framing 2 – QS-native bioreactor specification. Looking further forward, the four-state operational model and the QS signal thresholds that define each state boundary constitute the functional specification for a next-generation QS-native fermentation platform: a bioreactor whose control logic is built around biological communication states rather than their downstream physicochemical consequences. Such a platform would integrate validated in-line QS biosensors (Section 5.2.2) as primary sensing elements alongside conventional physicochemical sensors, with control logic parameterised on QS signal thresholds rather than pH or ORP setpoints alone. The digital twin framework is the intellectual blueprint for this platform: the four-state model provides the state classification logic; the transformer model provides the prediction engine; the SHAP attribution layer provides the interpretability required for regulatory approval of an automated bioreactor control system; and the decision log provides the parametric release record. A QS-native bioreactor designed around this specification would be a qualitatively different instrument from anything currently commercially available.

Framing 3 – Validation and calibration environment. The digital twin also functions as the simulation environment within which QS-native bioreactor control policies are developed and validated before physical deployment. The reinforcement learning component (maturation step 4, Section 6.3.3) trains optimal intervention policies within the digital twin simulation, discovers non-obvious control strategies by simulating thousands of fermentation scenarios, and validates those strategies against held-out experimental runs before any policy is enacted in a physical vessel. This simulation-first validation approach is a standard requirement of advanced biomanufacturing process development and is directly analogous to the virtual commissioning workflows used in pharmaceutical biomanufacturing (Mahanty, 2023).

Table 2. Digital twin as bioreactor intelligence layer: a structured comparison.

Dimension	Conventional Bioreactor / Fermentation Vessel	Digital Twin (Bioprocess Decision Suite)
What it is	Physical vessel: hardware system providing and regulating the fermentation environment	Software intelligence layer: computational reasoning system operating above the vessel
What it senses	pH, temperature, DO, ORP, CO ₂ , agitation – macroscopic physicochemical outcomes	QS signal kinetics (via LC–MS/MS + proxy surrogates), physicochemical telemetry, and historical run patterns
What it acts on	Deviations from physicochemical setpoints – reactive feedback control	QS-driven community state transitions – predictive, pre-emptive intervention

Dimension	Conventional Bioreactor / Fermentation Vessel	Digital Twin (Bioprocess Decision Suite)
Intelligence type	Setpoint-based PID control; no awareness of biological communication	Transformer-based temporal reasoning; attention-attributed explainability; BNN uncertainty quantification
Relationship to QS	Unaware of QS signals; monitors only their downstream physicochemical consequences	Reads QS proxy signals; interprets community communication states; predicts quality trajectory from QS kinetics
Sensory prediction	None — quality is assessed only after the batch is complete	In silico predesign: predicts sensory outcome distribution before the run is initiated
Batch consistency	Manual parameter replication; no learning between batches	Closed-loop learning: each run refines the process–QS–sensory coupling model
Governance	Process log; setpoint records; deviation reports	Full audit trail: QS state estimates, AI predictions, XAI attribution, operator decisions, observed outcomes
Near-term integration	Existing hardware; no modification required	Software add-on: integrates with existing bioreactor sensor and control interfaces
Long-term trajectory	Unchanged hardware paradigm	Functional specification for QS-native bioreactor with in-line QS sensing and QS-parameterised control logic

6.2. Purpose Architecture: Five Operational Functions of the Bioprocess Decision Suite

Purpose 1 — In silico product predesign. The digital twin enables the prediction of the sensory profile of the finished cocoa wine — aroma, flavour, acidity, volatile balance — before a fermentation run is initiated. A producer enters intended process parameters (inoculum composition, temperature profile, initial sugar concentration, oxygen exposure schedule) and receives a predicted sensory outcome distribution with confidence intervals. This predesign capability transforms fermentation from a reactive process — where quality is assessed after the fact — to a prospective, design-before-manufacture operation analogous to the model-informed product design workflows of advanced biomanufacturing. It also enables rapid product portfolio development: different sensory profiles can be explored in silico before committing raw materials, identifying the process parameterisation most likely to deliver each target profile.

Purpose 2 — Real-time state monitoring and predictive control. During a fermentation run, the digital twin continuously updates its estimate of the current QS-mediated fermentation state, forecasts the trajectory toward the predesigned sensory target, and generates explainable intervention recommendations when the trajectory deviates from the intended path. The QS lead-time advantage operationalised here enables pre-emptive intervention before macroscopic consequences are locked in — the capability that no conventional bioreactor or fermentation vessel currently possesses.

Purpose 3 — Batch-to-batch consistency and production scale-up. Each completed fermentation run contributes annotated data to the digital twin's training corpus, progressively refining the model's process–QS–sensory coupling. Over successive batches, the digital twin learns to reproduce the process conditions that reliably deliver a target sensory profile, systematically reducing flavour

variability. This cumulative learning is the mechanism through which consistent quality is achieved at scale: the digital twin encodes the process knowledge that enables a defined sensory outcome to be reproduced reliably across batches of varying substrate composition, inoculum age, and ambient conditions.

Purpose 4 – Biomanufacturing compliance and process validation. The digital twin generates the documentation required for biomanufacturing process validation: parametric release records linking each batch to the process conditions under which it was produced, deviation reports when a fermentation trajectory departs from the validated process window, and corrective action records documenting every intervention enacted and its observed outcome. The process parameters encoded in the four-state operational model – QS signal thresholds, physicochemical transition criteria, acceptable intervention windows – constitute the validated process specification against which each run is evaluated, implementing the digital twin as a process analytical technology (PAT) system within a biomanufacturing quality framework.

Purpose 5 – Food safety and audit-ready governance. Every intervention recommended – enacted or overridden – is logged alongside the QS state estimate, model outputs and uncertainty bounds, SHAP attribution summary, operator decision, operator ID, and observed outcome at the next measurement point. This structured decision log meets the documentation requirements of HACCP critical control point monitoring records, ISO 22000 food safety management system documentation, and GMP lot-level traceability obligations. The audit trail links microbial communication states, process decisions, and product quality outcomes in a single continuously maintained record, enabling retrospective quality investigations, regulatory submissions, and premium market certification to be grounded in process-level evidence.

6.3. Module Architecture: Four Interdependent Components

In Quality by Design (QbD) terminology (Gerzon et al., 2022), the bioprocess decision suite defines the design space of cocoa pulp juice fermentation: the four-state operational model maps critical process parameters (CPPs) – temperature, inoculum composition, oxygen exposure schedule, initial sugar concentration – to the critical quality attributes (CQAs) – ethanol concentration, organic acid profile, pH, volatile aroma composition, and sensory acceptability – that define acceptable product quality. This QbD framing is fundamental to the digital twin's biomanufacturing compliance purpose: each state boundary functions as a critical control point at which the digital twin evaluates whether the CPP trajectory will deliver the target CQA profile. The four interdependent modules implement this QbD logic in real time, as illustrated in Figure 2.

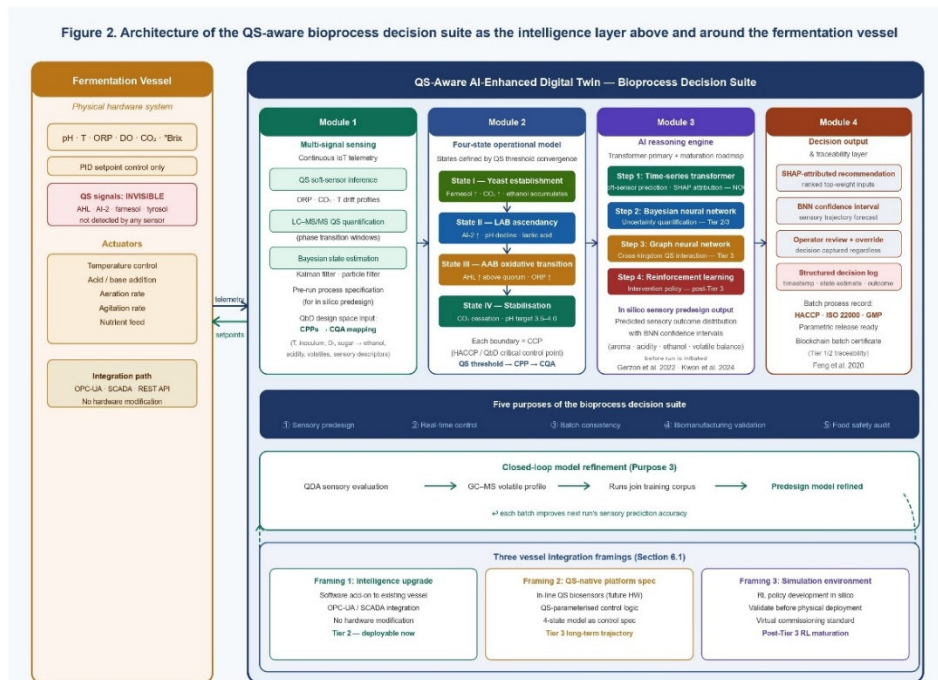


Figure 2. Architecture of the QS-aware bioprocess decision suite as the intelligence layer above and around the fermentation vessel. Left panel: the fermentation vessel (bioreactor or equivalent) with its standard sensor suite (pH, T, ORP, DO, CO₂, °Brix) and actuators (temperature control, acid/base addition, aeration, agitation). The vessel enforces physicochemical conditions but has no awareness of QS communication states occurring within it. Right panel: the four-module digital twin operating as the intelligence layer. Module 1 ingests vessel telemetry, strategic LC-MS/MS QS data, and proxy signals via Bayesian state estimation. Module 2 maps the integrated data to the four-state operational model (States I-IV) defined by QS signal threshold convergence. Module 3 applies the transformer-primary AI reasoning engine (Step 1: TST with attention-based explainability; Step 2: BNN uncertainty quantification; Step 3: GNN cross-kingdom interaction; Step 4: RL policy optimisation) to generate state forecasts and sensory trajectory predictions. Module 4 produces SHAP-attributed operator recommendations, records operator decisions, and generates the complete batch process record. Integration arrows: vessel telemetry flows to Module 1; Module 4 recommendations are communicated back to the vessel control interface as setpoint updates or operator alerts. For Tier 2 deployment: software-only integration, no hardware modification. For long-term QS-native platform: Module 3 RL policy and Module 2 QS thresholds constitute the bioreactor control specification with in-line QS biosensors replacing periodic LC-MS/MS. CPP = critical process parameter; CQA = critical quality attribute; QbD = Quality by Design; PAT = process analytical technology; CCP = critical control point; TST = time-series transformer; BNN = Bayesian neural network; GNN = graph neural network; RL = reinforcement learning.

6.3.1. Module 1 — Multi-Signal Sensing and Data Ingestion

Continuous IoT telemetry (temperature, pH, ORP, DO, CO₂, °Brix) feeds the digital twin in real time from any adequately instrumented fermentation vessel. Periodic LC-MS/MS QS quantification at phase transition windows supplements the telemetry stream. Validated soft-sensor models — data-driven models that provide continuous indirect estimates of QS state from ORP kinetics, CO₂ evolution inflections, and temperature drift profiles (Gerzon et al., 2022; Kwon et al., 2024) — serve as indirect indicators of QS-driven state transitions between analytical sampling points. Bayesian state estimation through Kalman filtering maintains probabilistic state estimates across irregular measurement intervals and temporary sensor unavailability (Mahanty, 2023). Pre-run process specifications (inoculum, temperature profile, planned duration) are ingested for in silico sensory pre-design computation.

6.3.2. Module 2 – Four-State Operational Model

Cocoa pulp juice fermentation is represented as a sequence of four operational states defined by convergence of QS signal thresholds and physicochemical criteria rather than clock time. Each state boundary constitutes both a QS-driven biological event and a HACCP-style critical control point at which the digital twin evaluates whether the trajectory is aligned with the predesigned sensory target. The four-state model is a conceptual framework grounded in QS mechanism knowledge and fermentation ecology evidence; its experimental validation is a research priority identified in the Future Perspectives section.

- State I – Yeast establishment: ethanol accumulation, CO₂ evolution, rising farnesol. Predesign role: ethanol concentration and ester-alcohol foundation of aroma profile established. Quality risk: insufficient ethanol; early contamination.
- State II – LAB ascendancy: AI-2 accumulation, progressive acidification, lactic acid production. Predesign role: organic acid balance and pH trajectory shaped. Quality risk: under- or over-acidification; insufficient spoilage exclusion.
- State III – AAB oxidative transition: AHL accumulation above quorum threshold, ORP rise, acetic acid production, flavour volatile emergence. Predesign role: volatile acidity and aroma complexity determined. Quality risk: over-acidification if QS-mediated gating fails.
- State IV – Stabilisation and endpoint: CO₂ cessation, target pH and volatile profile achieved. Quality decision: accept, continue maturation, or intervene.

6.3.3. Module 3 – Transformer-Primary AI Reasoning Engine with Progressive Maturation Roadmap

Transformer models – primary layer, deployable at Tier 2. Time-series transformer models address the core inferential challenge: identifying which time points and signal combinations in the fermentation telemetry record are most predictive of downstream sensory outcomes and state transitions (Kwon et al., 2024; Rivera et al., 2024). Multi-head attention mechanisms learn the temporal patterns in QS soft-sensor signals and physicochemical telemetry that precede phase transitions and correspond to specific sensory attribute trajectories. While attention weights do not constitute formal feature importance scores in all transformer configurations, they provide a directly interpretable record of which time points and signals the model weighted most heavily in generating each prediction – supporting SHAP-based operator attribution that identifies the top physicochemical and proxy-QS signals driving each recommendation in operator-readable terms. For the predesign function, the transformer generates *in silico* sensory predictions for proposed run parameterisations. Transformer models can be trained on Tier 2 telemetry datasets without requiring QS molecule measurements, making them the immediately deployable AI component of the decision suite.

Bayesian neural networks – maturation step 2, Tier 2/3 boundary. BNNs provide calibrated probability intervals on transformer predictions, producing the uncertainty bounds required for biomanufacturing parametric release decisions and confidence-graded sensory predesign outputs (Mahanty, 2023). They require multi-season annotated telemetry datasets and are deployable once Tier 2 data accumulation is sufficient.

Graph neural networks – maturation step 3, Tier 3. GNNs represent the fermentation ecosystem as a network of microbial taxa and metabolite classes with QS interaction-strength edges. They are architecturally suited to modelling the cross-kingdom interactions – AHL-mediated gating of AAB oxidation, farnesol-mediated synchronisation of the yeast-to-AAB transition – that most directly govern quality failure modes. GNN training requires annotated multi-omics interaction datasets that do not yet exist at scale for cocoa pulp juice; this is a Tier 3 objective contingent on the open dataset infrastructure identified in Future Perspectives.

Reinforcement learning – maturation step 4, post-Tier 3 validation. RL agents trained within the digital twin simulation environment discover optimal intervention policies that maximise a

defined reward function — for example, the probability of achieving the predesigned sensory target while minimising acetic acid exceedance risk. RL has demonstrated feasibility for multi-species microbial community control in bioreactor settings (Treloar et al., 2020). RL requires a validated simulation environment calibrated across multiple completed Tier 3 cycles; it is the final maturation step and the computational foundation for the fully autonomous QS-native bioreactor control system.

6.3.4. Module 4 — Decision Output and Traceability

Each AI recommendation is presented to the operator alongside a SHAP-attributed explanation, a BNN confidence interval on the sensory trajectory forecast, a digital twin simulation comparing projected outcomes with and without the recommended intervention, and a mandatory override capability with decision capture. Every decision event is logged in structured format: timestamp, QS state estimate, model outputs and uncertainty bounds, XAI attribution, operator decision, operator ID, and observed outcome at the next measurement point. This decision log constitutes the biomanufacturing process record, the HACCP CCP monitoring record, and the quality audit trail simultaneously. At run completion, the digital twin generates a batch process record linking the predesigned sensory target, the process trajectory, every decision event, and the final quality outcome — the primary artefact for internal process review, external certification, and progressive model refinement.

6.3.5. Worked Example: Predesign-to-Production Cycle with Bioreactor Integration

A producer specifies a target sensory profile for the next production batch: a fruity, moderately acidic cocoa wine with low volatile acidity, ethanol 10–12% v/v, and elevated 2-phenylethyl acetate. The digital twin's transformer model, trained on paired process–QS–sensory datasets, computes a recommended process parameterisation and returns a predicted sensory outcome distribution. The producer approves. The run is initiated in the production bioreactor; the digital twin integrates with the bioreactor's control interface and receives continuous telemetry. At 18 hours, the transformer detects an ORP slope pattern previously associated with imminent AI-2 accumulation above the State I–II threshold — the attention mechanism identifies ORP slope, CO₂ evolution rate, and current temperature as the three highest-weight predictor inputs. The digital twin recommends a 1°C temperature reduction to facilitate thermotolerant LAB establishment. The SHAP summary is presented to the operator; the operator confirms via the control interface; the bioreactor adjusts temperature; the intervention is logged. At run completion, QDA sensory scores and GC–MS volatile data are entered; the predicted-versus-actual sensory outcome comparison is generated; the batch record is closed; and the run joins the training corpus, incrementally refining the predesign model's predictive accuracy for subsequent batches.

7. Translational Implementation: From Artisanal Settings to QS-Native Biomanufacturing

7.1. Matching Decision Suite Capability to Production Context and AI Maturation Stage

The bioprocess decision suite is designed for phased deployment across production contexts that differ in fermentation vessel type, sensing infrastructure, data management capacity, AI maturation stage, and governance readiness. The three-tier framework maps each operational purpose and AI component to the tier at which the required data infrastructure becomes available, ensuring that the architecture is simultaneously ambitious in its long-term vision and honest about what can be deployed now. Each tier generates the data that enables the next: Tier 1 produces the annotated run dataset that trains the Tier 2 transformer; Tier 2 accumulates the multi-season dataset enabling BNN calibration; Tier 3 assembles the multi-omics corpus enabling GNN deployment. The progression from Tier 1 protocol-guided observation through Tier 3 full biomanufacturing deployment also traces

the progression from the digital twin as a conceptual decision aid to the digital twin as the operational intelligence layer of a QS-native fermentation platform. Table 3 summarises the full framework.

7.2. Tier 1 — Structured Observation and Data Collection (Artisanal and Smallholder Contexts)

Tier 1 operates without continuous electronic sensing or analytical laboratory access. A protocol-guided state checklist — paper-based or mobile application format — maps manually observed indicators (pH, temperature, °Brix, CO₂ evolution, colour, aroma) to the four operational states and provides pre-specified intervention recommendations for each state boundary. This translates the mechanistic state model into a practical decision aid accessible at minimal cost. Standardised fermentation run logs generate the longitudinal annotated dataset from which the Tier 2 transformer is trained. Blockchain-linked fermentation certificates provide traceable provenance documentation (Feng et al., 2020). At Tier 1, the digital twin's intelligence exists only in the protocol checklist: the analytical intelligence layer is human, the fermentation vessel is unmodified, and the data collected is the investment in future AI capability.

7.3. Tier 2 — Telemetry-Enabled Monitoring with Transformer Intelligence Layer (Cooperative and Semi-industrial Contexts)

Tier 2 adds continuous IoT telemetry, cloud-based data logging, and the first AI component: a time-series transformer classifier trained on annotated Tier 1 and early Tier 2 datasets. The transformer is integrated with the fermentation vessel's existing control interface — whether that vessel is a cooperative-scale closed tank or a small bioreactor — as a software layer requiring no hardware modification. The transformer delivers Purposes 1, 2, and 3 at proxy-signal resolution: in silico sensory trajectory prediction from process parameters and telemetry patterns; real-time state transition alerts with SHAP-attributed explanations; and progressive batch consistency improvement. A critical design acknowledgement for Tier 2 deployments in tropical smallholder and cooperative contexts is that challenges of high cost, absence of standardised implementation frameworks, and restricted access for small producers remain substantial barriers to smart fermentation technology adoption (Yee et al., 2025) and must be addressed through co-design with producer communities rather than technology push alone. The governance documentation framework — which transformer alerts require mandatory operator response, how decisions are logged, how the alert record maps onto HACCP CCP monitoring requirements — is established at Tier 2 and designed from the outset to be compatible with Tier 3 biomanufacturing process validation requirements.

7.4. Tier 3 — Full Bioprocess Decision Suite with QS-Informed Biomanufacturing Control and QS-Native Platform Specification

Tier 3 adds periodic LC-MS/MS QS quantification, the BNN uncertainty layer, and — as the multi-omics training corpus is assembled — the GNN cross-kingdom interaction component. All five purposes of the bioprocess decision suite are fully operational at Tier 3. The digital twin integrates with industrial bioreactor control systems through standard communication protocols (OPC-UA, SCADA integration), receiving continuous telemetry and communicating setpoint recommendations as a PAT implementation within the biomanufacturing quality system (Gerzon et al., 2022). The parametric process specification encoded in the four-state operational model — QS signal thresholds, physicochemical transition criteria, acceptable intervention windows, sensory target parameters — constitutes the master process record from which each batch is validated in accordance with QbD principles. Once the digital twin has been validated across sufficient Tier 3 production cycles and the RL agent has been trained and verified, the resulting system — transformer plus GNN plus BNN plus validated RL policy — provides the complete functional specification for a next-generation QS-native fermentation platform: a bioreactor whose control logic is built around QS signal states, whose

in-line biosensors provide continuous QS monitoring, and whose parametric release system is governed by the digital twin's inference and decision log.

7.5. Producer Data Sovereignty and Microbial Terroir

The digital twin does not seek to produce a single universal fermentation standard. The four-state operational model defines generic phase boundaries, but QS signal thresholds, telemetry profiles, and intervention parameters are calibrated to local substrate characteristics, indigenous microbial populations, and target sensory profiles. Fermentation records at every tier remain in the ownership of the producing entity, with data sharing governed by transparent agreements specifying how shared data may be used, who benefits from model improvements, and how intellectual property arising from local microbial diversity is protected. The accumulation of region-specific transformer models — each encoding the process knowledge for a distinct cocoa wine terroir — is an intended outcome of the framework, enabling the digital representation and reproducible production of distinct sensory profiles at biomanufacturing scale without erasing the regional distinctiveness that defines their market value.

Table 3. Three-tier implementation framework with AI maturation roadmap and vessel integration pathway for the QS-aware bioprocess decision suite. Tiers are defined by sensing infrastructure, data management capacity, AI maturation stage, fermentation vessel type, and governance readiness. “Vessel integration” column describes how the digital twin integrates with the fermentation vessel at each tier. Decision suite purposes (1–5) refer to Section 6.2. AI maturation steps (1–4) refer to Section 6.3.3. IoT = Internet of Things; TST = time-series transformer; BNN = Bayesian neural network; GNN = graph neural network; RL = reinforcement learning; PAT = process analytical technology; QDA = quantitative descriptive analysis; OPC-UA = OPC Unified Architecture (industrial control protocol).

Feature	Tier 1: Artisanal / Smallholder	Tier 2: Cooperative / Semi-Industrial	Tier 3: Industrial / Biomanufacturing
Fermentation vessel type	Open or closed artisanal tank; ceramic or food-grade vessel	Closed cooperative tank; small bioreactor; pilot-scale fermenter	Industrial stirred-tank bioreactor; commercial-scale closed fermentation system
Vessel integration	No integration — digital twin exists as protocol checklist and run log	Software add-on to existing vessel control interface; no hardware modification required	PAT integration via OPC-UA/SCADA; bioreactor telemetry as primary data feed; setpoint recommendations communicated to control system
Primary sensing	Manual pH, T, °Brix; visual/olfactory; standardised run logs	Continuous IoT telemetry: pH, T, ORP, CO ₂ , °Brix; cloud logging	Full telemetry + periodic LC-MS/MS QS quantification; GC-MS volatile profiling; QDA sensory data at run completion
AI maturation stage	None — human protocol intelligence; generates Tier 2 TST training data	Step 1: TST classifier + Kalman uncertainty; Step 2: BNN when multi-season dataset available	Steps 1–2: TST + BNN; Step 3: GNN when multi-omics corpus available; Step 4: RL post-validation → QS-native platform specification
Decision suite purposes	Purpose 3 (nascent): manual batch consistency tracking	Purposes 1–3 at proxy-signal resolution; Purpose 5 (partial)	All 5 purposes fully operational; biomanufacturing-grade process validation and food safety audit trail

Feature	Tier 1: Artisanal / Smallholder	Tier 2: Cooperative / Semi-Industrial	Tier 3: Industrial / Biomanufacturing
Sensory pre-design capability	Empirical: best-batch pattern identification from run log history	Transformer-based: predicted sensory trajectory from process parameters + telemetry	QS-informed: full sensory outcome distribution with BNN confidence intervals from QS-parameterised model
Governance / quality alignment	Blockchain batch certificates; cooperative quality protocol	TST alert governance policy; HACCP-compatible CCP monitoring records; cloud audit trail	Full PAT implementation; HACCP/ISO 22000/GMP process validation documentation; parametric release records
Long-term vessel trajectory	Unchanged vessel hardware; digital twin remains protocol-level	Vessel gains QS-aware intelligence layer through software integration	Digital twin provides functional specification for QS-native bioreactor with in-line QS sensing and QS-parameterised control logic

8. Conclusions

This perspective makes three specific contributions to the precision management, quality governance, and biomanufacturing-scale production of cocoa pulp juice fermentation. First, it establishes that polyphenol-mediated QS signal quenching and pH-driven autoinducer instability under acidic fermentation conditions are fundamental architectural specifications — not analytical inconveniences — for any QS-informed sensing, modelling, or control system in this substrate. These constraints apply equally to the digital twin architecture and to any future QS-native fermentation platform design.

Second, this perspective introduces the bioprocess decision suite — a QS-aware AI-enhanced digital twin designed to fulfil five distinct operational purposes: *in silico* product pre-design before a run is initiated; real-time QS-aware predictive control; batch-to-batch consistency management through closed-loop model refinement; biomanufacturing process validation documentation; and food safety audit-trail generation compatible with HACCP, ISO 22000, and GMP requirements. The digital twin is precisely distinguished from the fermentation vessel: it is not a bioreactor, not a replacement for fermentation infrastructure, but the QS-aware intelligence layer that fermentation systems have always lacked. For existing bioreactors with adequate sensor arrays, it functions as a software-defined intelligence upgrade requiring no hardware modification. For the longer term, the four-state operational model and QS signal threshold specifications constitute the functional blueprint for a next-generation QS-native fermentation platform (Treloar et al., 2020; Mahanty, 2023; Zhao et al., 2025).

Third, the three-tier translational implementation pathway maps each AI maturation step and vessel integration mode to the tier at which the required data infrastructure becomes available, creating a progressive pathway from artisanal protocol-guided observation to full biomanufacturing-scale QS-native fermentation control. The tiers are interdependent stages of a single knowledge-building programme rather than alternative configurations: each tier generates the annotated fermentation dataset that enables the next AI maturation step and the next level of vessel intelligence integration.

The four-state operational model and the full bioprocess decision suite remain conceptual frameworks whose experimental validation is the necessary next step. The hypothesis that convergent QS signal profiles define discrete fermentation phase transitions predictive of specific sensory outcomes must be tested empirically across multiple substrate origins, growing seasons, and inoculation conditions — the experimental programme that both validates the framework and generates the training corpus for transformer model deployment.

The broader significance of this work extends to other multi-kingdom fermentations in polyphenol-rich tropical matrices — coffee pulp, cacao mucilage, indigenous African fermented beverages — that share the same matrix challenges and the same commercial imperative to move from artisanal variability to biomanufacturing-scale consistency without losing the sensory distinctiveness that defines their market value. The bioprocess decision suite framework, the transformer-primary AI architecture, the explicit bioreactor-versus-intelligence-layer distinction, and the progressive AI maturation roadmap tied to a concrete data-building strategy are generalisable contributions to precision fermentation science wherever digital transformation and cultural authenticity must be reconciled (Zhao et al., 2025; Abdurrahman and Ferrari, 2025). Treating fermentation quality as shaped not just by what microbes produce, but by how they communicate, is the foundational reframing from which predesignable, reproducible, biomanufacturing-grade cocoa wine production can be built.

9. Future Perspectives

Standardised, matrix-adapted QS measurement protocols are the most immediate infrastructure priority: cross-laboratory validated SPE conditions, matrix-matched calibration standards in sterile cocoa pulp juice, and validated DMPD derivatisation for AI-2. These protocols produce the unbiased QS time-series data on which both the transformer sensory prediction model and any future QS-native bioreactor calibration depend.

Paired process-QS-sensory datasets — continuous telemetry, strategic LC-MS/MS QS quantification, GC-MS volatile profiling, and QDA sensory evaluation of the finished wine, generated simultaneously for every fermentation run — are the core training currency of the bioprocess decision suite. Building and maintaining these datasets under governance frameworks that protect producer data rights and enable equitable benefit-sharing is as much a governance challenge as a technical one. The QS-FERM database concept — an open-access platform aggregating global autoinducer data linked to fermentation outcomes — represents a particularly valuable infrastructure investment.

Experimental validation of the four-state operational model is the most important translational research priority, and it is simultaneously the primary data-generation activity for transformer model training. The experimental programme and the AI maturation roadmap share the same dataset requirements and should be co-designed accordingly.

In-line QS biosensor development and validation in cocoa pulp juice matrices represents the critical hardware component for transitioning from the digital twin as a software intelligence layer to the digital twin as the functional specification of a QS-native bioreactor. Validated in-line biosensors would close the analytical throughput gap that currently necessitates the proxy-signal strategy, enabling continuous QS-aware control without offline LC-MS/MS dependency.

Scale-up validation studies are needed to confirm that QS-process-sensory coupling relationships learned from pilot-scale runs transfer to production-scale fermentations, where mass transfer gradients and spatial QS signal heterogeneity may differ substantially. Multi-scale digital twin architectures modelling spatial QS signal gradients in production-scale vessels represent an important future development direction.

Quality management system integration warrants dedicated study: a formal process for incorporating QS state monitoring into HACCP critical control point designations and ISO 22000 food safety management plan documentation for fermented cocoa beverage production, developed in consultation with food safety regulatory authorities, would provide a replicable model for other multi-kingdom fermentation systems and directly accelerate biomanufacturing-scale adoption.

Author Contributions: Anthony Oppong Kyekyeku: Conceptualization, Methodology, Writing — original draft, Writing — review and editing, Visualization. Margaret Owusu: Supervision, Writing — review and editing. John Edem Kongor: Supervision, Writing — review and editing. Daniel Sitsofe Yabani: Writing — review and editing. All authors read and approved the final manuscript.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Consent for Publication: Not applicable.

Availability of Data and Materials: Not applicable. This is a translational perspective article; no primary data are presented.

Acknowledgments: The authors acknowledge all who provided guidance and support during the development of the research programme underpinning this perspective.

Competing Interests: The authors declare no competing interests.

Use of AI Tools: The authors used Anthropic's Claude AI assistant to support language polishing, structural review, and reference formatting during manuscript preparation. All AI-assisted outputs were critically reviewed, edited, and verified by the authors, who take full responsibility for the content and conclusions of this manuscript.

Abbreviations

AAB – Acetic Acid Bacteria; AHL – Acyl-Homoserine Lactone; AI-2 – Autoinducer-2; BNN – Bayesian Neural Network; CCP – Critical Control Point; CCST – CSIR College of Science and Technology; CPP – Critical Process Parameter; CQA – Critical Quality Attribute; DO – Dissolved Oxygen; DPD – 4,5-dihydroxy-2,3-pentanedione; GC – Gas Chromatography; GMP – Good Manufacturing Practice; GNN – Graph Neural Network; HACCP – Hazard Analysis and Critical Control Points; ISO – International Organization for Standardization; LAB – Lactic Acid Bacteria; LC – Liquid Chromatography; MS – Mass Spectrometry; OPC-UA – Open Platform Communications Unified Architecture; ORP – Oxidation-Reduction Potential; PAT – Process Analytical Technology; PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PVPP – Polyvinylpyrrolidone; QbD – Quality by Design; QDA – Quantitative Descriptive Analysis; QS – Quorum Sensing; RL – Reinforcement Learning; SCADA – Supervisory Control and Data Acquisition; SHAP – SHapley Additive exPlanations; SPE – Solid-Phase Extraction; TST – Time-Series Transformer; XAI – Explainable Artificial Intelligence.

References

- Abdurrahman O., Ferrari R., 2025. Digital twin applications in the food industry: a review. *Front. Sustain. Food Syst.* <https://doi.org/10.3389/fsufs.2025.1538375>
- Albuquerque P., Casadevall A., 2012. Quorum sensing in fungi: a review. *Med. Mycol.* 50, 337–345. <https://doi.org/10.3109/13693786.2011.652201>
- Almeida O.G.G., Pinto U.M., Matos C.B., Frazilio D.A., Braga V.F., von Zeska-Kress M.R., De Martinis E.C.P., 2020. Does quorum sensing play a role in microbial shifts along spontaneous fermentation of cocoa beans? An in silico perspective. *Food Res. Int.* 131, 109034. <https://doi.org/10.1016/j.foodres.2020.109034>
- Almeida O.G.G., Vitulo N., De Martinis E.C.P., Felis G.E., 2021. Pangenome analyses of LuxS-coding genes and enzymatic repertoires in cocoa-related lactic acid bacteria. *Genomics* 113, 1659–1670. <https://doi.org/10.1016/j.ygeno.2021.04.010>
- Almeida O.G.G., Pereira M.G., Bighetti-Trevisan R.L., Santos E.S., De Campos E.G., Felis G.E., Guimarães L.H.S., Polizeli M.L.T.M., De Martinis B.S., De Martinis E.C.P., 2024. Investigating luxS gene expression in lactobacilli along lab-scale cocoa fermentations. *Food Microbiol.* 119, 104429. <https://doi.org/10.1016/j.fm.2023.104429>
- Blana V.A., Lianou A., Nychas G.-J.E., 2017. Quorum sensing and microbial ecology of foods. In: Sant'Ana A.S. (Ed.), *Quantitative Microbiology in Food Processing: Modeling the Microbial Ecology*. Wiley, Chichester, pp. 600–613. <https://doi.org/10.1002/9781118823071.ch31>

- Campos S.D.M., Martínez-Burgos W.J., dos Reis G.A., Ocán-Torres D.Y., dos Santos Costa G., Rosas Vega F., Lima Serra J., Soccol C.R., 2025. The role of microbial dynamics, sensorial compounds, and producing regions in cocoa fermentation. *Microbiol. Res.* 16, 75. <https://doi.org/10.3390/microbiolres16040075>
- Chen H., Fink G.R., 2006. Feedback control of morphogenesis in fungi by aromatic alcohols. *Genes Dev.* 20, 1150–1161. <https://doi.org/10.1101/gad.1411806>
- Churchill M.E.A., Chen L., 2011. Structural basis of acyl-homoserine lactone-dependent signaling. *Chem. Rev.* 111, 68–85. <https://doi.org/10.1021/cr1000817>
- De Vuyst L., Weckx S., 2016. The cocoa bean fermentation process: from ecosystem analysis to starter culture development. *J. Appl. Microbiol.* 121, 5–17. <https://doi.org/10.1111/jam.13045>
- de Melo Pereira G.V., Miguel M.G.C.P., Ramos C.L., Schwan R.F., 2012. Microbiological and physicochemical characterization of small-scale cocoa fermentations and screening of yeast and bacterial strains to develop a defined starter culture. *Appl. Environ. Microbiol.* 78, 5395–5405. <https://doi.org/10.1128/AEM.01144-12>
- Feng H., Wang X., Duan Y., Zhang J., Zhang X., 2020. Applying blockchain technology to improve agri-food traceability: a review of development methods, benefits and challenges. *J. Clean. Prod.* 260, 121031. <https://doi.org/10.1016/j.jclepro.2020.121031>
- Ferrocino I., Rantsiou K., Cocolin L., 2022. Microbiome and -omics application in food industry. *Int. J. Food Microbiol.* 377, 109781. <https://doi.org/10.1016/j.ijfoodmicro.2022.109781>
- Gerzon G., Sheng Y., Kirkitadze M., 2022. Process analytical technologies – advances in bioprocess integration and future perspectives. *J. Pharm. Biomed. Anal.* 207, 114379. <https://doi.org/10.1016/j.jpba.2021.114379>
- Giaouris E., Heir E., Desvaux M., Hébraud M., Møretro T., Langsrud S., et al., 2015. Intra- and inter-species interactions within biofilms of important foodborne bacterial pathogens. *Front. Microbiol.* 6, 841. <https://doi.org/10.3389/fmicb.2015.00841>
- Jin X., Liu C., Xu T., Su L., Zhang X., 2020. Artificial intelligence biosensors: challenges and prospects. *Biosens. Bioelectron.* 165, 112412. <https://doi.org/10.1016/j.bios.2020.112412>
- Kareb O., Aïder M., 2020. Quorum sensing circuits in the communicating mechanisms of bacteria and its implication in the biosynthesis of bacteriocins by lactic acid bacteria: a review. *Probiotics Antimicrob. Proteins* 12, 5–17. <https://doi.org/10.1007/s12602-019-09555-4>
- Kwon H.J., Shiu J.H., Yamakawa C.K., Rivera E.C., 2024. Enhancing fermentation process monitoring through data-driven modeling and synthetic time series generation. *Bioengineering* 11, 803. <https://doi.org/10.3390/bioengineering11080803>
- Mahanty B., 2023. Hybrid modeling in bioprocess dynamics: structural variabilities, implementation strategies, and practical challenges. *Biotechnol. Bioeng.* 120, 2072–2091. <https://doi.org/10.1002/bit.28503>
- Niyigaba T., Küçüköz K., Kołożyn-Krajewska D., Królikowski T., Trzaskowska M., 2025. Advances in fermentation technology: a focus on health and safety. *Appl. Sci.* 15, 3001. <https://doi.org/10.3390/app15063001>
- Rivera E.C., Yamakawa C.K., Rossell C.E.V., Nolasco J., Kwon H.J., 2024. Prediction of intensified ethanol fermentation of sugarcane using a deep learning soft sensor and process analytical technology. *J. Chem. Technol. Biotechnol.* 99, 207–216. <https://doi.org/10.1002/jctb.7525>
- Rutherford S.T., Bassler B.L., 2012. Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harb. Perspect. Med.* 2, a012427. <https://doi.org/10.1101/cshperspect.a012427>
- Schwan R.F., Wheals A.E., 2004. The microbiology of cocoa fermentation and its role in chocolate quality. *Crit. Rev. Food Sci. Nutr.* 44, 205–221. <https://doi.org/10.1080/10408690490464104>
- Treloar N.J., Fedorec A.J.H., Ingalls B., Barnes C.P., 2020. Deep reinforcement learning for the control of microbial co-cultures in bioreactors. *PLoS Comput. Biol.* 16, e1007783. <https://doi.org/10.1371/journal.pcbi.1007783>
- Valastyan J.S., Kraml C.M., Pelczar I., Ferrante T., Bassler B.L., 2021. *Saccharomyces cerevisiae* requires CFF1 to produce 4-hydroxy-5-methylfuran-3(2H)-one, a mimic of the bacterial quorum-sensing autoinducer AI-2. *mBio* 12, e03303-20. <https://doi.org/10.1128/mBio.03303-20>
- Wang D., Fan F., Qin Y., et al., 2023. Quorum-quenching enzymes: promising bioresources and their opportunities and challenges as alternative bacteriostatic agents in food industry. *Compr. Rev. Food Sci. Food Saf.* 22, 1104–1127. <https://doi.org/10.1111/1541-4337.13104>

Yee C.S., Zahia-Azizan N.A., Abd Rahim M.H., Mohd Zaini N.A., Raja-Razali R.B., Ushidee-Radzi M.A., Ilham Z., Wan-Mohtar W.A.A.Q.I., 2025. Smart fermentation technologies: microbial process control in traditional fermented foods. *Fermentation* 11, 323. <https://doi.org/10.3390/fermentation11060323>

Zhao S.G., Jiao T., Adade S.Y.S.S., Wang Z., Ouyang Q., Chen Q., 2025. Digital twin for predicting and controlling food fermentation: a case study of kombucha fermentation. *J. Food Eng.* 393, 112467. <https://doi.org/10.1016/j.jfoodeng.2025.112467>

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