StrideWise

MSBA Capstone Project

To be considered in the following categories:

- Business Impact & Decision Value
- Innovation on Technology or Approach
- Problem Framing & Insight

Simon Walsh

Executive Summary

The opportunity
Runners today can generate and access vast quantities of data across watches, apps, and platforms—yet many still struggle to translate that data into clear, timely decisions about what to prioritise in their training. This “information overload” is a well-documented barrier to effective decision-making (Eppler & Mengis, 2004; Bawden & Robinson, 2009). In parallel, running-related injuries and acute illnesses are common, disruptive events that reduce consistency, impair performance, and can lead to long-term dropout. Meta-analytic evidence suggests meaningful injury incidence across runner populations (Videbæk et al., 2015), and large cohort work shows that single-session distance “spikes” (e.g., a run exceeding the longest run in the prior 30 days) materially elevate overuse injury risk (Frandsen et al., 2025). These injuries can cut short a running career, and hence result in real negative lifetime health risks, as well as a direct economic loss from dealing with the actual costs of the injuries (e.g. from healthcare costs, and from work stoppages).

Our solution
StrideWise is a data analytics product that converts complex training signals into one simple, actionable outcome: a daily traffic-light status (Green / Amber / Red) that indicates whether a runner is on track or drifting into elevated risk. Under the hood, the system combines training exposure (volume, intensity, terrain, recent change), wearable-derived recovery signals (sleep, resting heart rate, HRV proxies), and optional subjective inputs (perceived fatigue, soreness, illness symptoms) to estimate short-horizon injury/illness risk and surface the most relevant drivers for that individual. The product is designed to be transparent and coach-like: it explains what changed, why it matters, and what to do next today—reducing data overload while preserving user agency.

How we overcome the key data constraint – access to data
High-quality longitudinal health and training datasets are difficult to access at scale due to privacy constraints and the cost of collection. StrideWise addresses this by using a clinically-informed synthetic data approach to model realistic training journeys, recovery dynamics, and plausible injury/illness onset patterns. The synthetic generator, SynthRun, is calibrated to reproduce real-world distributions and relationships (e.g., training progression, fatigue accumulation, recovery responses) and supports rapid iteration of modeling hypotheses before expanding to real-world trials. This approach aligns with the broader literature on synthetic data, which emphasises balancing fidelity, downstream utility, and privacy risk (Gonçalves et al., 2020; Adams et al., 2025; Vallevik et al., 2024). (See also Appendix 3 and Appendix 5 for more detail)

Progress to date
We have already built a working web-based MVP that demonstrates the end-to-end product experience: data ingestion and feature engineering, risk scoring, explainability, and a user-facing dashboard that communicates daily status and recent alerts. The MVP validates the core user journey (from raw signals to clear decisions) and provides a foundation for rapid product iteration, user testing, and investor-ready demonstrations. (See Appendix 1 for more detail)

Go-to-market, validation, and privacy-by-design plan
Our go-to-market plan is deliberately staged to de-risk both scientific validity and product-market fit. Our next step is to build a native mobile app MVP (iOS/Android) that mirrors the web experience while enabling frictionless daily use, push notifications, and device-integrated data capture. We will then run a small, informed pilot cohort to gather structured user feedback on usability, comprehension, and behavioural impact (e.g., do runners adjust training in response to Amber/Red signals).

In parallel, we will partner with an expert group—ideally a university sports science / sports medicine team—to formally review and validate the clinical plausibility of our synthetic training data generation as well as the assumptions within our injury/illness causation model. This partnership will help translate product hypotheses into a research-grade validation protocol, including appropriate outcome definitions, confounding considerations, and reporting standards.

Finally, we will begin collating rich real-world longitudinal data from the pilot cohort (potentially via a university-based study in conjunction with the partnership described above). The objectives are twofold: to empirically compare real and synthetic data to ensure the generator reproduces key distributions and relationships, and also to benchmark predictive performance and calibrate the model on real-world labels. This closes the loop between synthetic-first development and evidence-based deployment.

Privacy and security commitments

StrideWise is designed around privacy-by-design from the outset. During the validation phase, any collection of real health and training data will be conducted under informed consent, with appropriate ethics oversight where applicable, and in alignment with GDPR requirements for special category (health) data (Regulation (EU) 2016/679, Article 9). Data will be minimised to what is necessary, pseudonymised, encrypted in transit and at rest, and governed through clear retention and access controls. We will follow established regulatory guidance on anonymisation and pseudonymisation practices when sharing or analysing datasets (ICO, 2025; Data Protection Commission, 2025).

As we move from pilot to consumer-scale deployment, we will strengthen these provisions further. In particular, we will evaluate whether risk scoring and explainability can run fully on-device, reducing or eliminating the need to transmit sensitive health data off the user's phone. Where server-side processing is required, we will explore privacy-enhancing techniques (e.g., secure aggregation for model improvement, strict feature minimisation, and differential privacy where appropriate) and will give users clear, granular controls over what is collected and how it is used.

Figure 1. StrideWise UX Mockups

Company Overview

What is StrideWise

StrideWise is building a new category of running analytics: an injury- and illness-risk signal that is as simple as a traffic light, yet grounded in a rigorous understanding of training load, recovery, and behavioural decision-making. Where most platforms ask runners to interpret dozens of charts and specialist metrics, StrideWise converts the same complexity into a single daily recommendation with clear, coach-like reasoning: what changed, why it matters, and what a prudent adjustment looks like today.

The insight behind the company

Endurance training platforms have become extraordinarily powerful—but also increasingly demanding for end users. They often present multiple models of fitness, fatigue, readiness, and load (each with their own parameters), and expect users to reconcile conflicting signals across devices and dashboards. In practice, this creates information overload: too many metrics, too little clarity, and decisions that become reactive or inconsistent. Information overload is a recognised phenomenon in digital environments, where complexity can reduce comprehension and decision quality (Eppler & Mengis, 2004; Bawden & Robinson, 2009).

StrideWise's thesis is straightforward: most runners do not need more metrics—they need one reliable signal that protects training consistency—because consistency is what drives progress, and injuries and illnesses are what breaks it.

Why the status quo is hard for everyday runners

To illustrate the challenge, in Figure 2 below, we show examples of real-world training analytics interfaces. They are excellent tools for coaches and power users, but they require ordinary users to understand specialist terminology, select appropriate charts, and interpret multivariate data under uncertainty, and moreover, to do this every day.

Our product approach

StrideWise reduces the cognitive load of training decisions by separating two jobs that today are conflated:

- Analytics: robustly estimating short-horizon risk from training exposure and recovery signals.
- Communication: translating that estimate into a simple, trustworthy decision and explanation.

The user experience is intentionally minimal: a daily Green/Amber/Red status, a short explanation of the key drivers, and an actionable suggestion (e.g., keep plan, reduce intensity, or take recovery). This design is not "less capable", it is more focused: it optimises for behaviour change and adherence rather than dashboard exploration.

How StrideWise is defensible

Defensibility comes from how we combine domain knowledge with modern modelling practice and product execution:

- A clinically-informed risk model: features reflect established risk contributors such as abrupt training spikes and recovery disruption (Frandsen et al., 2025; Videbæk et al., 2015).
- Synthetic-first iteration: a calibrated synthetic data generator enables rapid experimentation while limiting privacy exposure (Gonçalves et al., 2020; Vallevik et al., 2024; Adams et al., 2025).
- Decision UX: we are not competing on chart count—we are competing on clarity, trust, and daily usefulness.

Over time, StrideWise will accumulate unique model and product assets: calibration data, validated protocols, and user interaction outcomes (how runners respond to guidance), forming a compounding advantage.

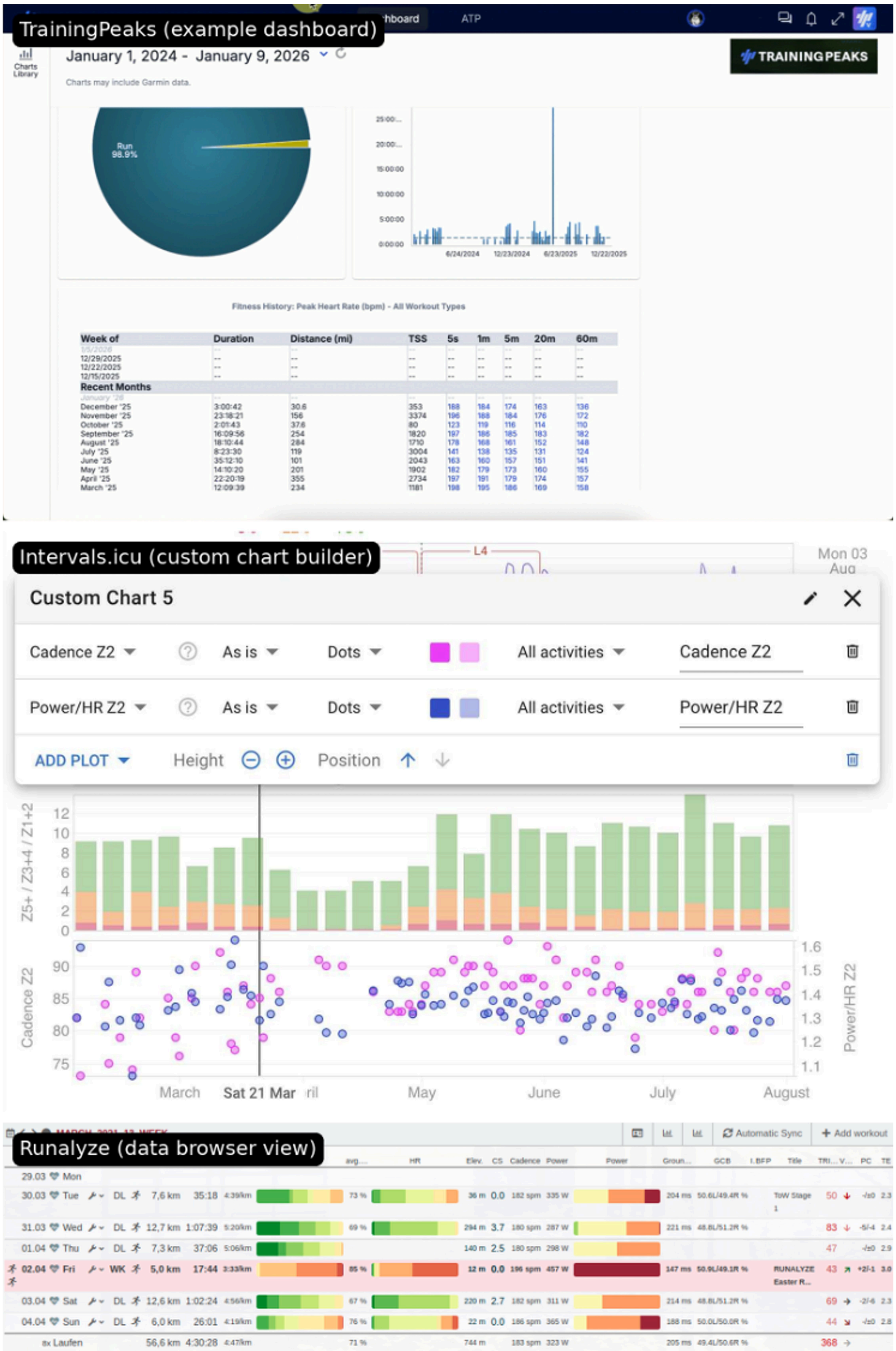

Figure 2. Examples from three of the main running/fitness platforms, showing their training analytics dashboards, which all require substantial interpretation by the end user. Note: Screenshots are used for illustrative purposes only; all product names and trademarks belong to their respective owners.

Business model

StrideWise is built for a scalable, software-first subscription model with optional licensing pathways:

- B2C subscription (initial focus): a consumer app with a free tier (basic status), and/or a free trail period, as well as paid tiers for deeper explanations, trend history, and coaching-style planning guardrails.
- B2B2C licensing (expansion): packaged offerings for coaches and clubs (team dashboards, cohort monitoring), and potential partnerships with health-adjacent organisations such as insurers or occupational wellness providers.
- API / embedded analytics (longer term): risk scoring and explainability modules that can be integrated into existing training ecosystems where appropriate.

This staged model allows us to prove consumer value and retention first, then expand to higher value channels where injury reduction and adherence have direct economic benefits.

Target customers and early wedge

Our first wedge is committed recreational runners: runners who train year-round, follow plans, and already use wearables, but who want fewer decisions and fewer setbacks. This cohort is large, digitally engaged, and has a clear willingness to pay for performance and health benefits. As trust is established, the product naturally extends to coached athletes, clubs, and organisations that benefit from improved consistency and reduced dropout.

Traction and current state

StrideWise has progressed beyond concept: we have a working web-based MVP that demonstrates the core user journey from data ingestion to daily risk status and explanation. The system architecture is designed to support rapid iteration and controlled validation as we expand to mobile and real-world pilots.

Strategic Analysis

External Analysis

Market opportunity and growth dynamics

StrideWise sits at the intersection of consumer wearables, fitness subscriptions, and the broader mobile health market. These adjacent markets are already large, and they are growing in ways that favour software that converts raw signals into daily guidance. Wearables are scaling rapidly: MarketsandMarkets forecasts the global wearable technology market to grow from US$84.53B in 2025 to US$176.77B by 2030 (MarketsandMarkets, 2025). On the software side, Grand View Research estimates the global fitness app market at US$12.12B in 2025, projecting growth to US$33.58B by 2033 (Grand View Research, 2025). The mobile health apps category shows similar momentum; Fortune Business Insights reports US$36.68B in 2024 with a forecast to US$88.70B by 2032 (Fortune Business Insights, 2025). For StrideWise, the strategic implication is clear: consumers increasingly accept recurring payments for health and performance guidance, and the addressable ecosystem of devices and data sources continues to expand.

It is also important that these are not just 'top-down' market forecasts. Shipment and engagement indicators suggest mainstream adoption: IDC reports 136.5 million wearable units shipped in Q2 2025, representing 9.6% year-on-year growth (IDC, 2025). At the same time, Strava's own reporting highlights the scale of digitally tracked running: in its 2025 Year in Sport materials, Strava references "over 180 million" users across 185 countries (Strava, 2025a), and in its Runna acquisition announcement it notes that "nearly 1 billion runs were recorded on Strava in 2024" (Strava, 2025b). This combination—more sensors, more tracked activity, and greater subscription willingness—creates a favourable environment for a focused, decision-first product.

Customer problem: data rich, insight poor

The incumbent ecosystem often assumes that users can interpret complex models of fitness, fatigue, readiness, and load across multiple dashboards. In practice, many recreational-to-serious runners face an 'information overload' problem: too many metrics, ambiguous or conflicting signals, and no clear translation into today's decision. Information overload is well established in the research literature as a barrier to comprehension and decision quality, particularly in data-dense digital settings. StrideWise's strategic wedge is therefore not 'more analytics', but a better interface between analytics and action: a single, trustworthy signal that reduces cognitive load while remaining explainable.

Target markets and reachable audiences

StrideWise's initial target is the large and growing segment of runners who already track training and recovery, but who do not have the time—or the desire—to interpret specialist terminology daily. This includes recreational runners training consistently (e.g., 2–6 sessions per week), runners preparing for events, and 'committed improvers' who invest in devices and subscriptions. From a distribution standpoint, the strategic advantage is that these audiences are already aggregated in platforms such as Strava and Garmin ecosystems, where integrations, communities, and content loops can compound adoption. Strava's reported footprint (over 180 million users across 185 countries) illustrates the reach available to a product that can establish credibility and trust (Strava, 2025a).

Competitive landscape and pricing benchmarks

The competitive environment is active and increasingly subscription driven. The most important strategic trend is that generalist platforms and device companies are moving 'up the stack' from passive tracking toward coaching, readiness, and personalised guidance. Strava's acquisition of Runna is emblematic: it is an explicit move into structured running training and coaching (Strava, 2025b). In parallel, Garmin introduced a paid tier—Garmin Connect+—bringing AI-driven insights and enhanced features to its previously free companion app (The Verge, 2025). These moves validate demand for guidance and insight, but they also raise the bar for differentiation.

StrideWise competes most directly on a specific dimension that is not yet well-served at consumer scale: running-specific, short-horizon injury and illness risk signalling. Most incumbents either provide broad readiness proxies; offer powerful but interpretation-heavy analytics environments; or focus on general recovery scoring. The strategic opportunity is to become the 'decision layer' that sits between data capture and training choice, simple enough for everyday runners, but clinically informed and explainable enough to earn trust.

Category	**Product**	**Typical primary promise**	**Published pricing (selected markets)**	**Strategic note for StrideWise**
Platform	Strava Subscription	Community + activity tracking + premium analysis	US: $11.99/month or $79.99/year (Strava, 2025c).	Price anchor for a broad 'fitness network' subscription; bundles now exist (e.g., Strava + Runna).
Device ecosystem	Garmin Connect+	AI summaries + enhanced dashboards on top of Garmin Connect	$6.99/month or $69.99/year (The Verge, 2025).	Signals a shift to monetising insights; creates head-to-head expectations on value-per-month.
Training analytics	TrainingPeaks Premium	Planning + analysis tools for structured training	$19.95/month or $134.99/year (TrainingPeaks, n.d.).	Demonstrates willingness-to-pay among serious runners, but requires high user interpretation.
Recovery wearable	ŌURA Membership	Readiness/sleep/recovery insights (ring + app)	€5.99/month or €69.99/year (ŌURA, n.d.); device pricing from $349 (ŌURA, n.d.-b).	Validates subscription acceptance for 'daily readiness' and shows appetite for simpler, wearable-led signals.
Recovery wearable	WHOOP Membership	Strain/recovery coaching (subscription-centric)	WHOOP One $25/month; WHOOP Peak $30/month; WHOOP Life $40/month (WHOOP, 2025).	Shows an upper bound for performance-focused subscriptions when coaching value is strong.
General fitness	Apple Fitness+	Guided workouts (broad fitness)	Ireland: €9.99/month (Apple, n.d.).	Relevant comparator for mainstream subscription pricing; not running-injury-specific.

Table 1: Stridewise competitor landscape

Figure 3 provides an illustrative positioning map. It is intended as a communication tool—not a scientific taxonomy—but it helps highlight StrideWise's intended market position: high injury/illness specificity with low required interpretation.

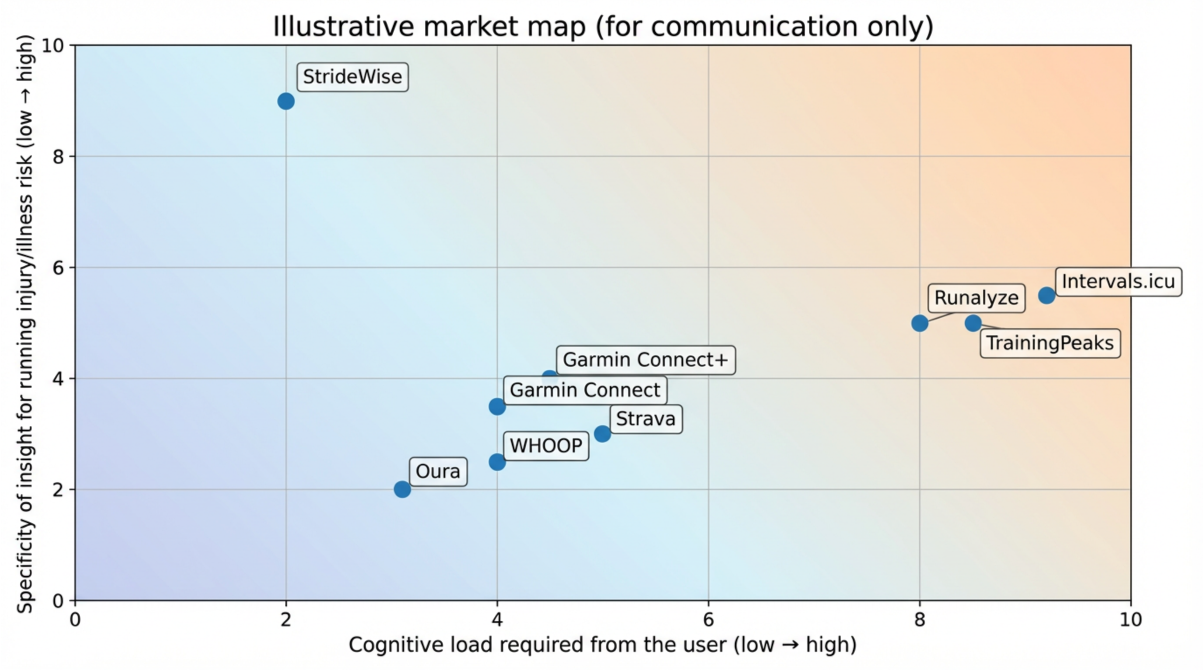

Figure 3. Illustrative market map.

Regulatory and data governance environment

Because StrideWise operates on training and recovery signals, it must treat privacy and governance as strategic enablers rather than 'compliance overhead'. In the EU, data concerning health is a special category of personal data under Article 9 GDPR, triggering additional requirements and safeguards (European Union, 2016; Data Protection Commission, n.d.). The governance strategy therefore needs to minimise data collection, maximise transparency, and implement privacy-by-design controls, especially during any real-world validation stage.

From a product perspective, two principles are particularly relevant. First, pseudonymisation can materially reduce risk when used correctly, and is recognised in GDPR-oriented guidance as an important safeguard (European Data Protection Board, 2025). Second, anonymisation is not a 'binary switch' but an identifiability risk management exercise; UK ICO guidance emphasises the need to understand techniques, strengths, weaknesses, and governance measures when producing and sharing anonymised data (ICO, 2025). These sources collectively support a synthetic-first development approach, and they also frame the requirements for any future real-data pilots.

Internal Analysis

Strengths and core competencies

StrideWise's core strength is its decision-first product design: it starts with the daily behaviour it wants to enable—better training choices under uncertainty—and works backward to a minimal set of signals, a calibrated risk score, and an explanation that a runner can understand. This design choice directly addresses the information-overload barrier that incumbents often amplify. A second strength is the synthetic-first development capability, which enables rapid iteration of hypotheses, simulation of edge cases, and privacy-preserving experimentation before large-scale real-world collection.

Key resources

The most strategically valuable assets today are: (1) the working web-based MVP (which proves the end-to-end flow from raw signals to a user-facing decision), (2) the synthetic data generator and modelling pipeline (which accelerate learning cycles), and (3) the emerging product narrative that positions StrideWise as an 'injury risk signal' rather than another analytics dashboard. These assets collectively reduce time-to-demonstration for investors and partners, while keeping optionality open for mobile, B2B2C, and research partnerships.

Sources of competitive advantage (what can become defensible)

Over time, defensibility is most likely to come from a combination of scientific credibility, product trust, and data network effects. Scientific credibility will be strengthened by partnering with an expert group (ideally a university sports science / sports medicine team) to validate model assumptions and evaluation protocols. Product trust will be strengthened by explainability that is both technically sound and psychologically reassuring, showing users 'what changed' and 'what to do today' without overclaiming certainty. Finally, responsibly collected longitudinal data from an informed pilot cohort can create a feedback loop for calibration and personalisation that generalist platforms may not prioritise for injury-specific outcomes.

Areas for improvement (near-term execution priorities)

From a strategic standpoint, the next phase is about de-risking two investor-critical questions: 'Is it scientifically valid?' and 'Will people actually use it daily?'. This points to three near-term priorities. First, build a native mobile MVP (iOS/Android) that mirrors the web experience but supports daily engagement patterns such as notifications and frictionless data capture. Second, run a small pilot to measure user comprehension and behavioural response to Amber/Red signals. Third, execute a structured validation plan with expert partners to assess the clinical plausibility of synthetic generation as well as the alignment between synthetic and real pilot data distributions, under strong governance controls.

Strengths Decision-first UX; synthetic-first iteration; MVP traction	**Opportunities** Subscription tailwinds; platform integrations; coach/club and insurer channels
Weaknesses Clinical validation not yet complete; requires trust-building and careful claims	**Threats** Platform 'move up the stack'; regulatory friction; model performance risk on real-world labels

Table 2. Strategic summary (compact SWOT).

Analytics Opportunity

Why this is an analytics opportunity now

StrideWise's core analytics opportunity is to translate runner-generated training and recovery data into a single, reliable, short-horizon risk signal that improves daily decisions. The substrate for this opportunity is now mainstream: runners log training digitally, wearables provide continuous recovery signals, and consumers increasingly accept subscription products that convert raw signals into practical guidance. The remaining gap is not a lack of data; it is the lack of a decision-first interpretation that is both explainable and credible.

The value at stake: injuries are common and disruptive

Running-related injuries (RRIs) are consistently reported as common across runner populations. A systematic review and meta-analysis by Videbæk et al. estimated injury incidence per 1,000 hours of running across runner types and highlighted meaningful rates in multiple settings (Videbæk et al., 2015). Beyond the immediate training interruption, RRIs carry economic and social costs. Prospective cohort work in novice runners has estimated direct and indirect costs per injury and per participant, demonstrating a measurable burden even in relatively short follow-up windows (Hespanhol Junior et al., 2016). A broader analysis of runners training for an event similarly discusses healthcare use, work impacts, and cost ranges associated with RRIs (Visser et al., 2021).

For StrideWise, the opportunity is not only to reduce injury incidence, but to protect the most valuable performance asset in endurance sport: training consistency. In consumer terms, the value proposition is fewer setbacks and more reliable progress, outcomes that are emotionally salient and commercially defensible.

Where existing analytics fall short

Most incumbent platforms provide extensive dashboards and advanced metrics, excellent for expert users, but cognitively demanding for everyday runners. This gap is well described by the information overload literature, which shows that increasing information volume and complexity can impair comprehension and decision quality (Eppler & Mengis, 2004; Bawden & Robinson, 2009). In practice, runners are asked to reconcile overlapping constructs such as fitness, fatigue, readiness, and load, and translate them into a decision under uncertainty, often without knowing which signal should dominate today.

StrideWise addresses this by treating 'prediction' and 'communication' as separate problems. The analytics objective is to estimate near-term risk with appropriate uncertainty; the product objective is to present that estimate as a simple decision aid that reduces interpretation burden while remaining explainable.

Evidence base for modelling: training exposure, recovery, and risk

A defensible injury/illness risk signal needs to be anchored in what is known, and honest about what is not. Training load change has long been treated as a key driver, but evidence has evolved over time. A systematic review by Camma Damsted et al. concluded that evidence linking sudden changes in training load to RRIs was limited at the time, reflecting heterogeneous definitions and study designs (Camma Damsted et al., 2018).

More recently, large-scale cohort evidence has sharpened the picture by focusing on specific, operationally meaningful 'high-risk sessions'. In an 18-month cohort of more than 5,000 runners, Frandsen et al. examined single-session spikes—runs that exceed the longest run in the prior 30 days—and reported materially elevated injury risk associated with these spikes (Frandsen et al., 2025). The strategic implication for StrideWise is that the most useful signal may not come from abstract weekly ratios alone, but from session-level deviations that are straightforward to detect and to explain.

Recovery state is the second pillar. Sleep duration has been linked to injury risk across athletic populations: Milewski et al. found that adolescent athletes sleeping less than eight hours per night were more likely to report injury (Milewski et al., 2014), and Watson et al. reported an association between increased sleep duration and reduced next-day injury risk in collegiate athletes (Watson et al., 2020). Wearable physiological signals also show promise as early indicators of maladaptation or illness. Düking et al. evaluated endurance training adapted on the basis of wearable-monitored HRV in a systematic review with meta-analysis (Düking et al., 2021). Case-level evidence shows meaningful HR/HRV changes during viral infection periods (Hottenrott et al., 2021), and observational work suggests that elevated resting heart rate may be informative in athlete health monitoring contexts (Pla et al., 2021).

Figure 4 summarises how these drivers translate into an explainable product output: the model estimates near-term risk, but the user receives a single daily decision accompanied by a concise explanation of the dominant drivers and a suggested adjustment.

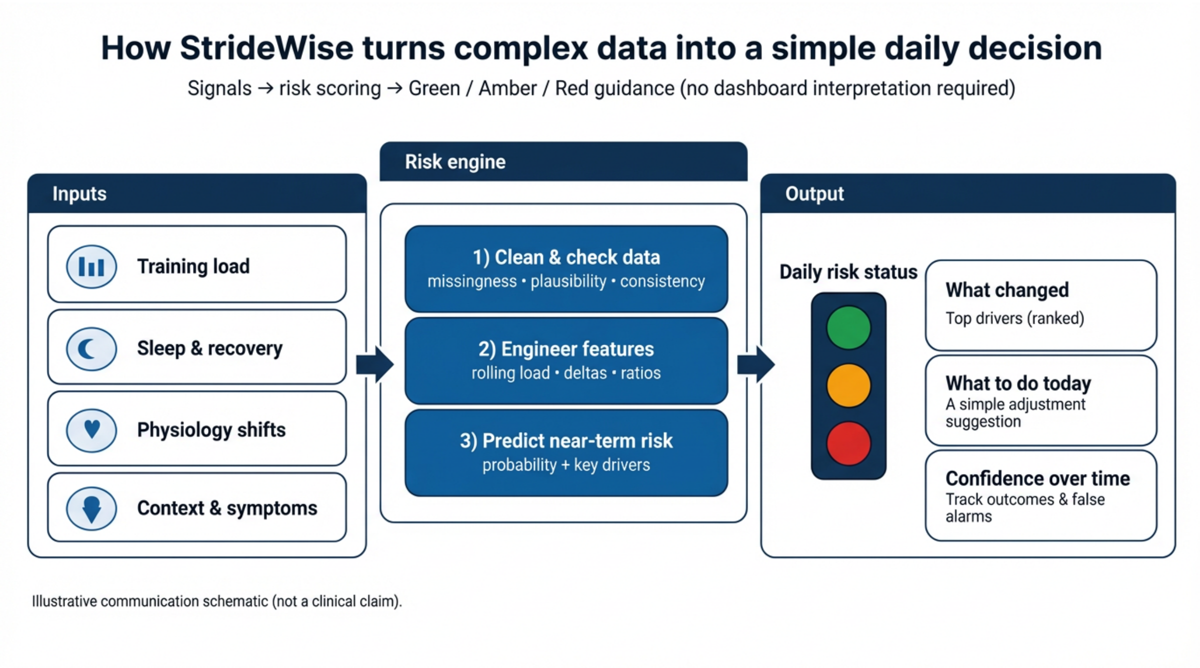

Figure 4. Risk driver schematic (illustrative): translating exposure and recovery drivers into an explainable daily decision (e.g., load spikes: Frandsen et al., 2025; injury incidence: Videbæk et al., 2015; sleep and injury: Milewski et al., 2014; Watson et al., 2020; HRV monitoring: Düking et al., 2021; illness physiology shifts: Hottenrott et al., 2021).

Data-to-decision: what StrideWise converts into value

The analytics opportunity becomes tangible when expressed as a conversion: transforming multi-source signals into one actionable decision. Figure 5 summarises the intended data-to-decision pipeline.

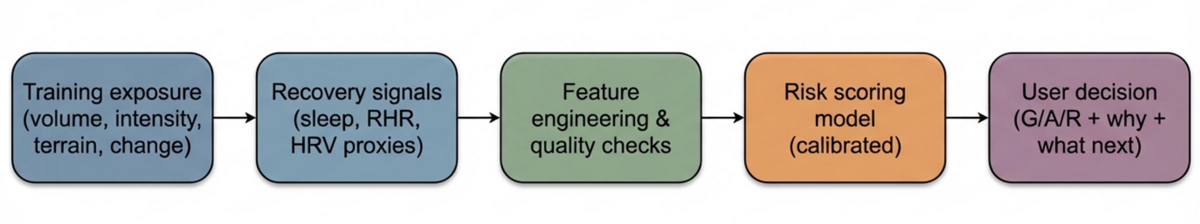

Figure 5. Data-to-decision pipeline (illustrative).

Why predictive analytics can work (and what 'good' looks like)

Injury prediction is difficult: outcomes are relatively rare, labels can be noisy, and causal pathways are multifactorial. This is precisely why the opportunity is meaningful: few products do it well. In running specifically, Lövdal et al. demonstrated that machine learning can be applied to detailed training logs to predict injuries, using a seven-year dataset of 74 high-level runners (Lövdal et al., 2021). The associated replication dataset is publicly available and provides a practical benchmark for external testing (Lövdal et al., 2021; Lövdal et al. dataset, DataverseNL).

More broadly, established prediction-model guidance emphasises the importance of transparent reporting, calibration, and validation. The TRIPOD statement sets expectations for how prediction models should be reported, including validation and performance disclosure (Collins et al., 2015). PROBAST provides a structured framework for assessing risk of bias and applicability in prediction model studies, reinforcing the need to avoid overfitting and to demonstrate generalisability (Wolff et al., 2019).

For StrideWise, 'good' therefore has three components: discrimination (ranking higher-risk days above lower-risk days), calibration (probabilities that mean what they say), and explainability that is stable enough to guide behaviour. Because injury/illness events are relatively rare, evaluation should prioritise metrics that reflect practical usefulness in the high-risk tail; precision–recall curves are often more informative than ROC curves in this setting (Saito & Rehmsmeier, 2015).

Validation approach: synthetic-first development, real-world benchmarking

StrideWise uses a synthetic-first approach to accelerate iteration while reducing unnecessary exposure of real health data during early development. However, credibility requires evidence on real-world data and disciplined evaluation. Figure 6 summarises a staged validation ladder that begins with synthetic calibration, then tests against an external public benchmark, and finally moves to a prospective pilot cohort.

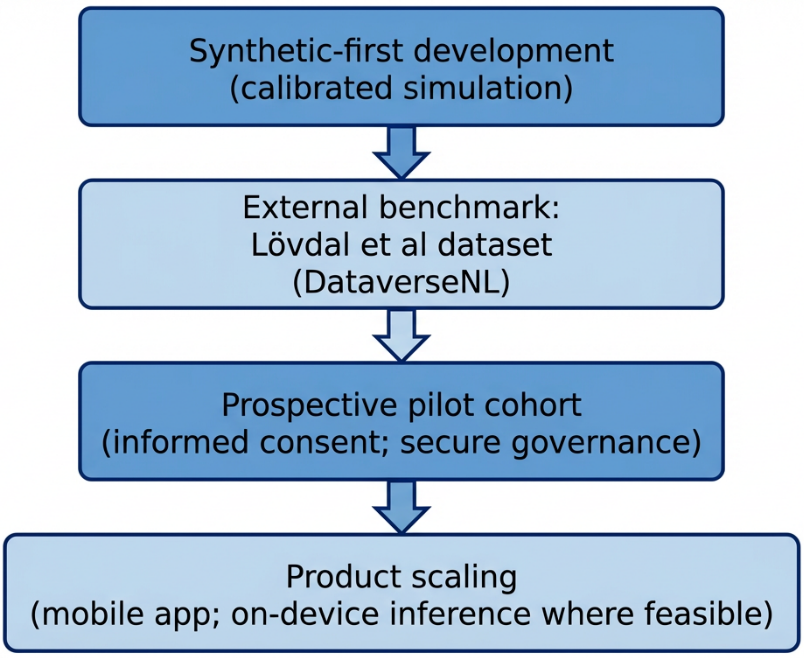

Figure 6. Staged validation ladder (illustrative).

Critically, when we reference external testing today, we refer to the Lövdal et al dataset. This dataset provides a concrete bridge between the academic injury prediction literature and pragmatic product evaluation, and it helps enforce honest reporting about what the model can and cannot claim at each stage.

To make those stage-specific claims explicit, Figure 7 sets out a practical scorecard: for each stage, it clarifies what we are trying to prove, what 'passing' looks like, and which claims become defensible only after those criteria are met. This keeps scientific credibility and product marketing aligned.

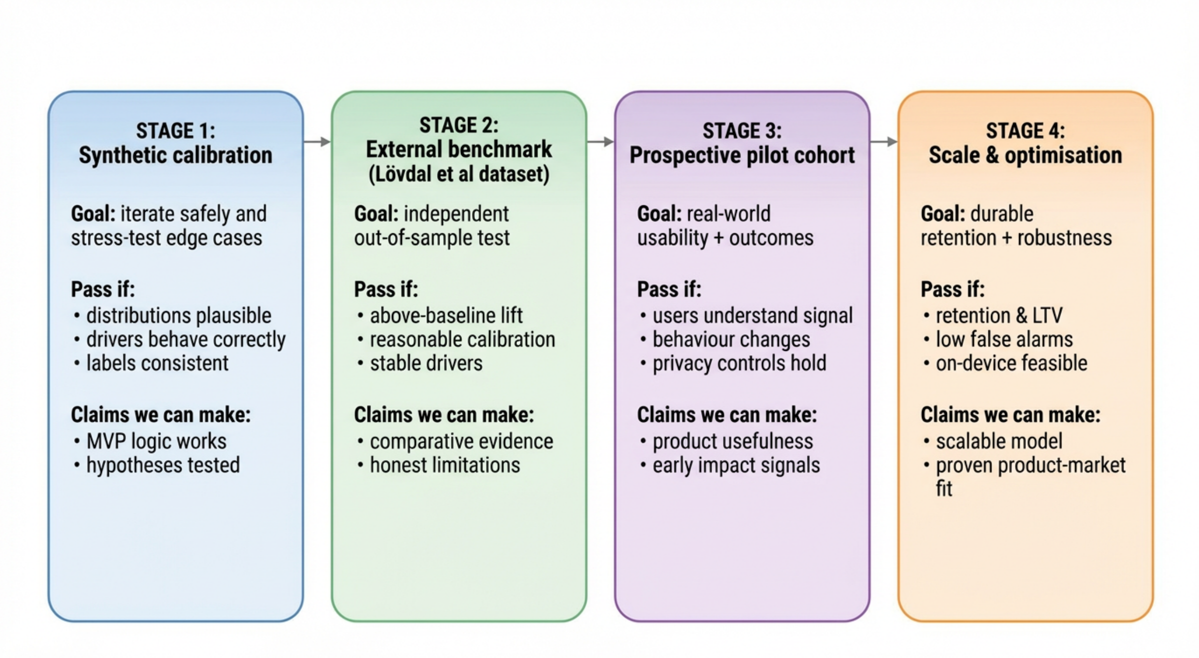

Figure 7. Validation scorecard with stage-gated pass criteria and claim boundaries (illustrative).

How we measure success

The success of this analytics opportunity should be measured with both scientific and product metrics. On the modelling side, performance should be assessed with metrics appropriate for rare outcomes (e.g., precision at top risk percentiles), calibration diagnostics, and stability across cohorts and time. On the product side, the most meaningful outcomes are behavioural: whether runners respond to Amber/Red signals by reducing avoidable high-risk spikes, and whether that response is associated with fewer training interruptions over time. The economic and quality-of-life literature on RRIs supports using healthcare utilisation, work impact, and training disruption as part of the value narrative (Hespanhol Junior et al., 2016; Visser et al., 2021).

Rationale

Rationale in brief

StrideWise is built around a simple proposition: most runners do not need more metrics, they need a trustworthy decision. The rationale for the product therefore sits at the intersection of human decision-making, the evolving evidence base for running injury risk, and privacy-aware product engineering. Rather than competing on the depth of dashboards, StrideWise competes on converting multi-source signals into a daily, explainable risk signal that a runner can act on immediately.

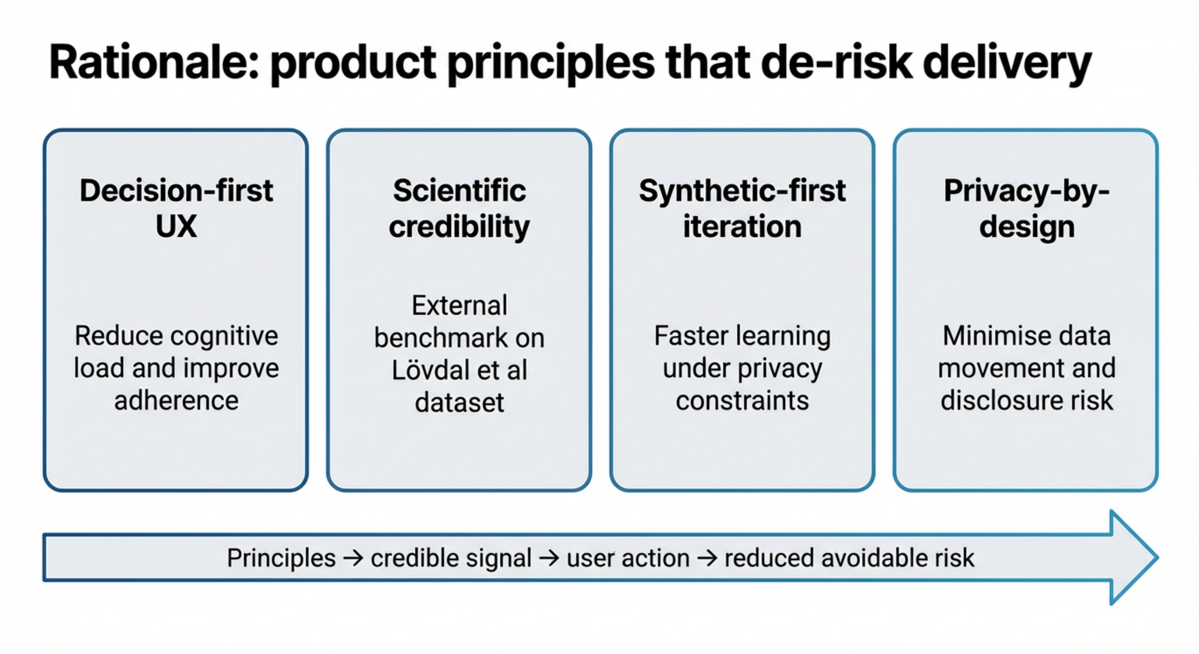

Figure 8. Product principles that shape StrideWise's rationale.

The problem we are solving

Recreational runners now have unprecedented access to training and health data through wearables and software platforms. The practical barrier is no longer data availability; it is actionable interpretation. Users are confronted with overlapping metrics and pseudo-scientific terminology without a clear decision framework for what to change today. As information volume and complexity increase, people's ability to process inputs and make good decisions can degrade, leading to uncertainty, delayed action, or suboptimal choices (Eppler & Mengis, 2004). The 'dark side' of abundant information can include anxiety, overinterpretation, and decision fatigue—especially when users are expected to synthesise many indicators independently (Bawden & Robinson, 2009).

In running, the stakes of poor interpretation are meaningful. Running-related injuries are common and disruptive, and although incidence varies by runner type and context, systematic review evidence shows substantial injury incidence across runner populations (Videbæk et al., 2015). In short: runners already have the signals, what is missing is a clear, reliable decision output that turns those signals into safer training choices.

The analytics solution being proposed

StrideWise proposes an injury-risk guidance layer that sits on top of existing wearable and training ecosystems and produces one primary output: a simple traffic-light risk signal (Green / Amber / Red) indicating whether the runner is at elevated near-term risk of injury or illness. The signal is paired with brief 'why' drivers and concrete, low-friction actions (for example: reduce intensity today, shorten the long run, take an easy day, prioritise sleep, or avoid hills).

This design is intentionally decision-first. It replaces dozens of competing indicators with a single interpretable decision aid, and it can be strengthened with a minimal daily check-in of two to three subjective items that capture early symptoms and perceived soreness or fatigue, signals that often precede measurable performance decline. The approach mirrors how best-in-class wearables already combine multiple physiological and training factors into a single readiness indicator. Garmin's Training Readiness, for example, is calculated from factors including sleep score, recovery time, HRV status, acute load, sleep history, and stress history (Garmin, n.d.). StrideWise's contribution is to make the output injury-specific, action-oriented, and calibrated to short-horizon risk.

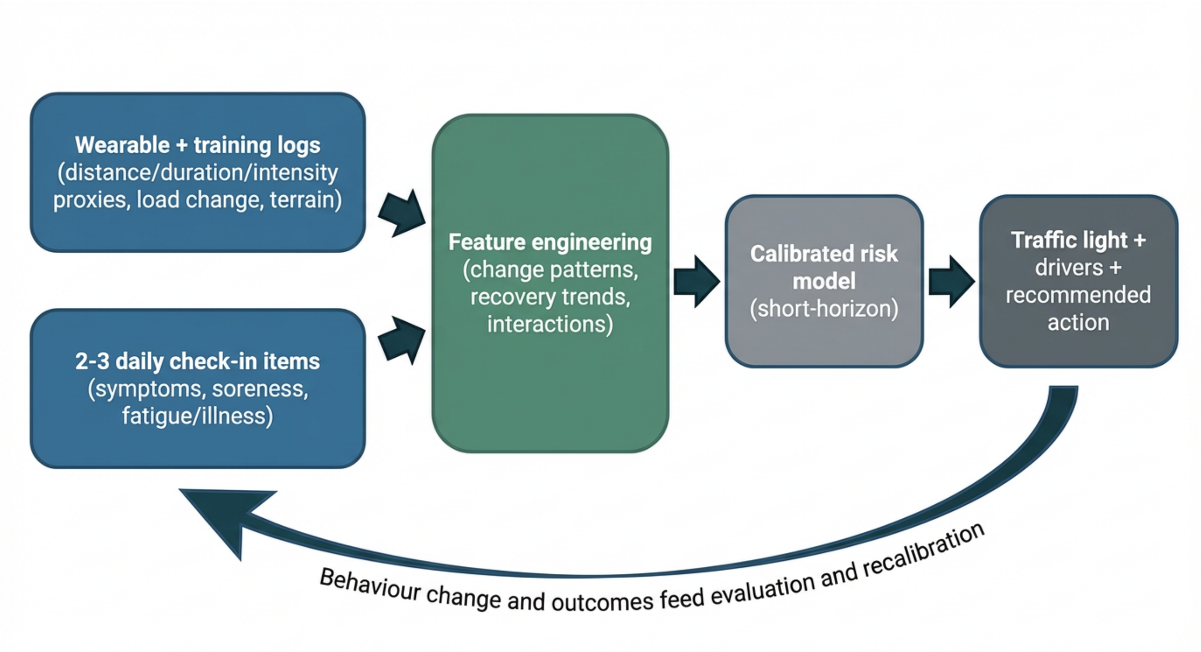

Figure 9. End-to-end decision loop: inputs → features → calibrated risk → action → evaluation (illustrative).

Why a synthetic dataset is central (and necessary)

When moving from product concept to model development, the core constraint is structural: large-scale, high-resolution wearable datasets suitable for injury prediction are difficult to access. Physiological and health-adjacent data is heavily regulated and sensitive, and in many jurisdictions 'data concerning health' is treated as a special category of personal data with strict conditions for processing and sharing (GDPR Article 9; Irish Data Protection Commission, n.d.).

This constraint makes the conventional approach—collect a large, richly labelled dataset and train a model—slow, expensive, and governance-heavy. StrideWise therefore adopts a synthetic-first strategy as an engineering and learning accelerator: it enables rapid iteration on pipelines and model design without exposing personal data, and it supports deliberate stress-testing of edge cases, missingness, and variable compliance. Importantly, synthetic is not 'automatically good'. Synthetic and anonymised datasets must be treated with explicit evaluation of identifiability risk and a clear 'release model', rather than assumed safe by default (ICO, n.d.).

This approach aligns with emerging best practice for synthetic data evaluation. Gonçalves et al. describe key trade-offs in synthetic patient data generation and evaluation, including disclosure risk and utility considerations (Gonçalves et al., 2020). Vallevik et al. propose a comprehensive quality assessment framework for synthetic tabular healthcare data, emphasising dimensions that support trustworthy use in downstream modelling (Vallevik et al., 2024).

Model behaviour: change and interaction matter more than absolute numbers

StrideWise is designed around the principle that injury risk is dynamic. The most informative signals are often not absolute levels, but changes and interactions, for example, a sudden spike in a long run when sleep has been poor, or a sustained upward drift in resting heart rate alongside rising load. Cohort evidence supports this emphasis on session-level deviations: in a study of more than 5,000 runners, Frandsen et al. found that single-session spikes—runs exceeding the longest run in the prior 30 days—were associated with materially increased injury risk (Frandsen et al., 2025).

Published machine learning work in runners further supports the idea that multivariate interactions can be informative. In a competitive runner setting, Lövdal et al. demonstrated meaningful discrimination using models that preserve short-horizon training dynamics and capture interactions across signals (Lövdal et al., 2021).

External benchmarking: the Lövdal et al dataset is a credibility anchor

Synthetic-first development must be paired with an external anchor to avoid closed-loop optimism, where a model appears strong only in the world it was built to simulate. StrideWise therefore treats the Lövdal et al dataset as a primary benchmark for testing generalisation. Lövdal et al. published both the injury prediction study and a publicly available replication dataset, providing an independent reference point for evaluation and calibration (Lövdal et al., 2021; Lövdal et al. dataset, DataverseNL).

Because injury events are relatively rare, evaluation should prioritise usefulness in the high-risk tail. Precision–recall analysis is often more informative than ROC analysis under class imbalance, which is typical in injury prediction settings (Saito & Rehmsmeier, 2015).

Privacy-by-design is a strategic advantage, not just compliance

Health-adjacent analytics must treat privacy and security as core product attributes. The GDPR introduces both heightened protections for health data and an explicit obligation to implement data protection by design and by default (GDPR Article 25; EDPB Guidelines 4/2019). In practice, this supports a ‘minimum necessary’ approach during validation: informed consent, purpose limitation, strict access control, encryption in transit and at rest, and auditable governance over any sensitive datasets.

As StrideWise evolves into a consumer mobile app, the same design lens can materially reduce risk. Where feasible, StrideWise can evaluate on-device inference so that sensitive health signals do not need to be transmitted off-device at all. Even where cloud processing is used, designing for minimal data movement keeps risk explicit and manageable.

Key assumptions and limitations

StrideWise is a decision-support product, not a clinical diagnostic system. It aims to reduce avoidable injuries by improving training decisions, not to predict every injury event; acute and traumatic injuries will always be less predictable because they are less tightly coupled to gradual exposure dynamics. This framing is consistent with the broader overuse-injury literature and with the inherent uncertainty in behavioural and physiological systems.

Synthetic data cannot establish clinical validity on its own. It accelerates development and creates a disciplined pathway to external testing, but ultimate validation requires prospective real-world evaluation and periodic recalibration as the product scales. This is why StrideWise treats the benchmark phase on the Lövdal et al dataset as necessary but not sufficient, and why the programme anticipates partnership-led validation studies as part of the go-to-market trajectory.

Stakeholders and Requirements

Purpose of this section

This section translates StrideWise's product vision into a concrete set of stakeholder needs and delivery requirements. The goal is to make explicit who the product must satisfy, what "good" looks like for each group, and which functional and non-functional requirements are necessary to deliver a credible, safe, and scalable injury-risk signal.

Stakeholder landscape

StrideWise sits at the intersection of consumer fitness, coaching practice, and health-adjacent data governance. The primary stakeholder is the runner, but the product must also earn trust and utility from coaches, research partners who support validation, and platform partners that enable data ingestion. Regulators and data protection authorities shape constraints around health data, and investors influence delivery priorities and evidence thresholds. Figure 10 provides an overview of these stakeholder relationships.

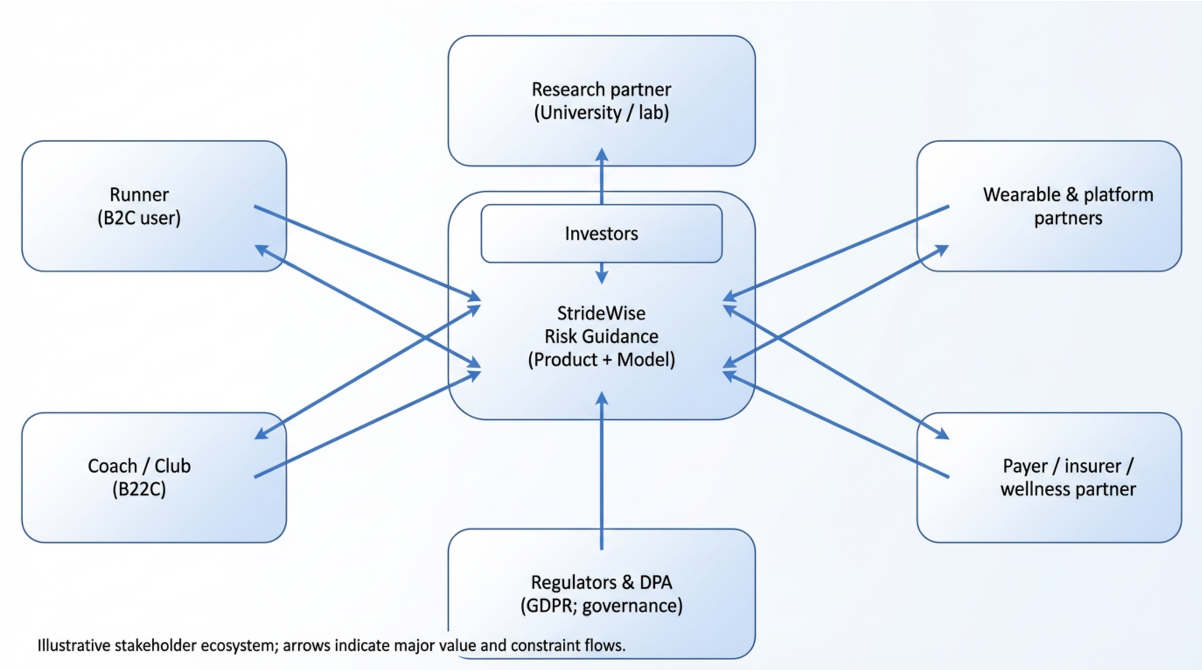

Figure 10. Stakeholder ecosystem for StrideWise (illustrative).

Runner requirements: trust, clarity, and controllable actions

For the runner, the product must reduce effort, not add to it. The runner-facing requirement is therefore a decision that is easy to understand, paired with a small amount of explanation that increases confidence without increasing cognitive burden. This direction is supported by the information overload literature, as previously discussed.

Operationally, runners also need the product to be fair and predictable: if a Red day appears, it must be defensible; if a runner follows guidance, they should not feel punished by persistent alarms. This translates into requirements for calibration, stable explanations, and a feedback loop that allows the model to improve as more diverse user patterns are observed (without turning the product into a research-grade data entry burden).

Coach and club requirements: transparency, accountability, and cohort-level insight

Coaches and clubs care about decision quality at both the individual and group levels. Their core requirement is not only a risk score, but a transparent reason for the score that can be communicated and justified to an athlete. They also value consistency across athletes—so that a cohort can be managed with a common language—and auditability when decisions are reviewed after an injury event.

This implies requirements for explainability that is stable across time, simple cohort views (e.g., a list of athletes with today's status), and mechanisms to tune alert thresholds to coaching context without breaking calibration. These needs align with a broader expectation in prediction-model best practice that models should be transparently reported and performance clearly communicated (Collins et al., 2015).

Research partner requirements: clinical credibility and reproducible evaluation

A university or expert research group partner is a key stakeholder because it provides a pathway to stronger clinical credibility. Their requirements include clear injury definitions, consistent outcome labelling, ethics approval where applicable, and reproducible evaluation. These requirements are aligned with the need to reduce bias and overclaiming in prediction modelling; PROBAST provides a structured way to assess risk of bias and applicability in prediction model studies, reinforcing the value of rigorous validation (Wolff et al., 2019).

Practically, this stakeholder also expects transparent documentation of feature construction, missing-data handling, and calibration procedures, as well as staged benchmarks on the Lövdal et al dataset before claims are extended to broader populations (Lövdal et al., 2021; Lövdal et al. dataset, DataverseNL).

Governance and regulatory requirements: health data, accountability, and privacy-by-design

StrideWise operates in a health-adjacent domain. Under the EU GDPR, data concerning health is a special category of personal data, and processing is generally prohibited unless specific conditions apply (GDPR Article 9). The Irish Data Protection Commission summarises these special category constraints and emphasises that health data requires heightened safeguards (Data Protection Commission, n.d.).

Beyond lawful basis, GDPR requires data protection by design and by default (GDPR Article 25). The European Data Protection Board's guidance on Article 25 clarifies expectations for technical and organisational measures that embed privacy into system design (EDPB, 2019/2020). In practice, this becomes a set of delivery requirements: data minimisation, purpose limitation, strong access control, encryption, and evidence of accountability. Governance expectations for anonymisation and disclosure risk are also well documented in regulator guidance (ICO, n.d.).

These requirements are not only compliance-driven. They are product-critical: the credibility of an injury-risk signal depends on user trust, and trust depends on both transparency and a security posture that is easy to explain to a non-technical audience.

Platform and integration requirements: reliable ingestion, quality controls, and portability

Because StrideWise aims to complement (not replace) existing platforms, integrations are a delivery prerequisite. The product must reliably ingest training exposure and recovery signals from multiple devices and logs, and it must handle missingness as a normal case rather than an exception. This drives requirements for robust data validation, schema normalisation across sources, and explicit quality flags that protect the model from silent corruption.

Portability is also a commercial requirement: a B2C subscription product needs to work across the heterogeneity of consumer devices, while future B2B2C distribution requires the ability to align to partner data availability. This implies a modular feature pipeline, clear versioning, and monitoring that can detect integration drift over time.

Product requirements: functional capabilities that support the decision

The functional requirements of StrideWise follow directly from the decision-first product promise. The system must produce a daily risk status, explain this status in plain language, recommend a concrete adjustment, and provide enough history to help users understand trends without reintroducing dashboard complexity. These capabilities must work under partial data availability and must fail safely when signals are missing or unreliable.

A secondary functional requirement is the ability to support learning without increasing burden. The product should support optional lightweight check-ins (e.g., symptoms, soreness) and should surface to the user how those inputs affect the signal, so that users perceive the system as responsive and fair. This also supports pilot evaluation in partnership settings.

Non-functional requirements: security, privacy, accessibility, and operational resilience

For investor-grade readiness, StrideWise must be secure and resilient by construction. For the mobile app trajectory, the OWASP Mobile Application Security Verification Standard (MASVS) provides a widely used reference for security requirements, including storage, cryptography, authentication, and network communication controls (OWASP MASVS, n.d.). At the organisational level, ISO/IEC 27001:2022 describes requirements for an information security management system (ISMS) that can support governance as the company scales (ISO, 2022).

Accessibility is also a practical requirement for consumer adoption. WCAG 2.2 is a W3C Recommendation and provides concrete guidance for making digital content more accessible, including input methods and target sizes that matter on mobile (W3C, 2023). ISO 9241-210:2019 provides requirements and recommendations for human-centred design processes, reinforcing the need to test the traffic-light experience with real users and iterate based on observed behaviour (ISO, 2019).

Finally, operational requirements include observability (monitoring model drift and data quality), incident response processes, and a clear audit trail for model versions and threshold updates. These requirements support trust, reduce operational risk, and simplify governance in research and partner contexts.

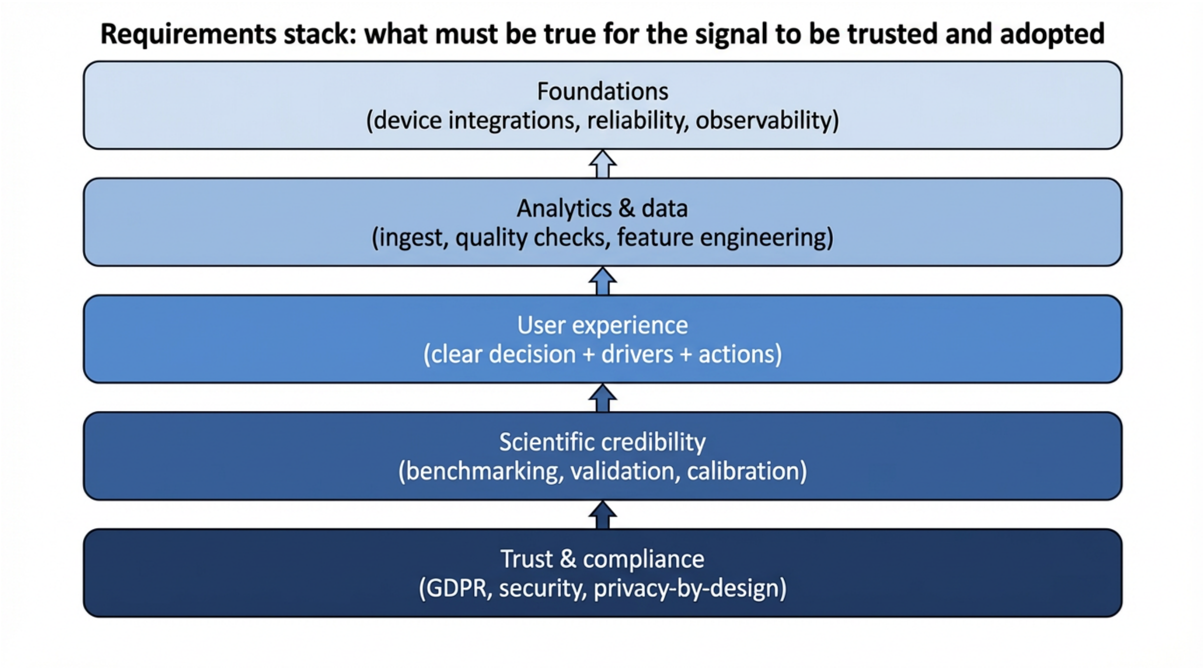

Figure 11. Requirements stack: layered conditions for a trusted and adopted risk signal (illustrative).

Implications for delivery planning

These stakeholder needs and requirements shape the delivery roadmap in two practical ways. First, they make validation and governance part of the core build, not an afterthought: the product cannot credibly scale without external benchmarking, documented evaluation procedures, and privacy-by-design controls. Second, they suggest that engineering effort should prioritise reliability and clarity over feature breadth, because the competitive edge is the quality of the decision and the trust around it.

In the pilot phase, requirements should be treated as acceptance criteria. A successful pilot is not only improved model performance; it is evidence that users understand the signal, that coaches and partners can interpret and justify it, and that the data governance approach is demonstrably safe and compliant.

Success Criteria

How we define success

StrideWise's success criteria are designed to be investor-grade and stage-gated. The early goal is not to "prove everything"; it is to demonstrate that the product can reliably convert messy, real-world signals into an understandable daily decision, then build credible evidence that users act on that decision and that outcomes improve. The criteria below are structured to avoid overclaiming and to align with accepted expectations for prediction models and health-adjacent data governance.

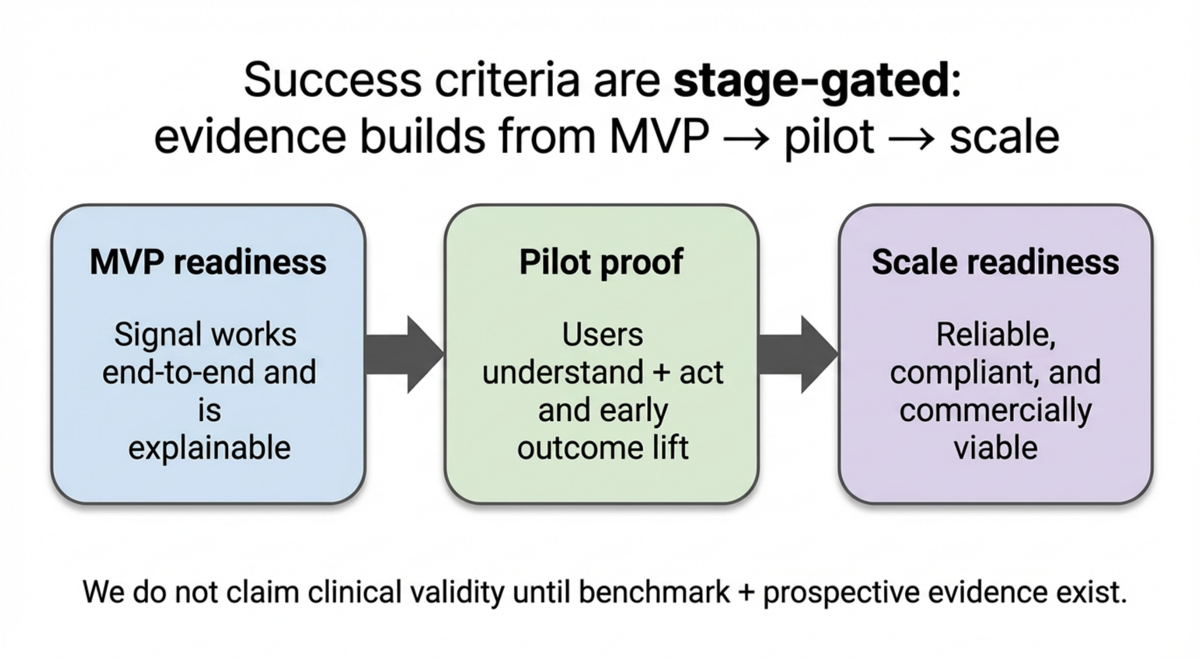

Figure 12. Stage-gated success scorecard (illustrative).

Stage-gated criteria

We evaluate success across four lenses—model evidence, user/product evidence, governance, and commercial traction—because a risk signal only creates value if it is both statistically credible and behaviourally usable. Table 3 summarises the concise acceptance criteria we use at each stage. Note clinical validation is a specific success criterion: we will partner with an expert institution (e.g., a university) to validate the clinical plausibility of the synthetic data generation assumptions and the injury causation logic. As real-world pilot data is gathered, we will test whether observed training, recovery, and outcome patterns are broadly consistent with the synthetic dataset, and update modelling assumptions accordingly.

Stage	Model evidence	User & product evidence	Governance & privacy	Commercial signal
MVP readiness	End-to-end pipeline runs reliably; meaningful lift over naïve baselines on internal tests; calibration	Users interpret Green/Amber/Red correctly in moderated tests; explanations are stable and action-oriented.	Data minimisation + clear consent flow for any pilot data; security controls documented; audit trail for model versions.	Clear willingness-to-pay hypothesis and packaging; early waitlist or partner interest to validate positioning.

	reviewed (not just ranking).			
Pilot proof	Benchmark performance on the Lövdal et al dataset is above baseline; precision in the high-risk tail is practically useful; calibration diagnostics stable. In addition, a clinical validation partnership is initiated and study protocol agreed, covering (a) validation of synthetic data realism and (b) validation of the injury causation model logic against expert review.	Users change behaviour in response to Amber/Red; low "alarm fatigue"; retention indicates ongoing value.	Ethics/governance in place (where applicable); DPIA-style risk assessment; incident response and access controls tested. Pilot governance supports clinical collaboration (ethics review where applicable), with documented data handling procedures appropriate for health-adjacent research.	Conversion and retention signals support a subscription path; coaches/club workflows show repeatable usage patterns.
Scale readiness	Model monitoring detects drift; periodic recalibration process; performance stable across cohorts/devices. Clinical validation is evidenced with results from real-world data collection, demonstrating that observed distributions and risk drivers are broadly consistent with synthetic assumptions and that performance	Self-serve onboarding, low support burden, and consistent engagement; clear pathways for B2C and B2B2C experiences.	Privacy-by-design strengthened (including on-device inference where feasible); security assurance aligned to standard controls.	Unit economics proveable (CAC/LTV), churn manageable, and partner distribution hypotheses validated. Evidence-backed claims (benchmark + clinical validation) support partner and enterprise conversations.

	generalises under prospective conditions.			

Table 3. Concise, stage-gated success criteria for StrideWise.

How we will measure model quality

Because injury and illness events are relatively rare, we focus on evaluation that reflects practical usefulness in the high-risk tail and avoids misleading "average-case" metrics. Precision–recall analysis is often more informative than ROC analysis for imbalanced classification problems (Saito & Rehmsmeier, 2015). We also assess calibration explicitly, because a risk signal must mean what it says if it is to support decisions; this focus is consistent with established prediction-model reporting expectations (Collins et al., 2015) and risk-of-bias guidance (Wolff et al., 2019).

How we will measure trust and governance readiness

StrideWise treats privacy and security as product attributes. GDPR makes health data a special category and requires data protection by design and by default (Regulation (EU) 2016/679, Article 25). For mobile delivery, we anchor security requirements to recognised baselines such as OWASP MASVS, and we align governance maturity to established ISMS expectations (ISO/IEC 27001:2022). For accessibility and usability at scale, we use WCAG 2.2 as a reference point for mobile interaction and readability requirements (W3C, 2023).

Scope and Delivery

Scope overview

StrideWise's scope is deliberately focused: we are building an injury-risk guidance layer that converts training and recovery signals into a daily, explainable decision (Green / Amber / Red) with clear recommended adjustments. The intent is to reduce avoidable risk by improving day-to-day training decisions, not to provide a medical diagnosis or replace clinical care. This positioning keeps the product credible and reduces the risk of overclaiming as evidence accumulates.

In scope for this delivery phase are: a web based MVP; robust ingestion and quality checks for training and recovery signals; a calibrated short-horizon risk model; user-facing explanations and recommended actions; and a staged validation pathway that includes benchmarking on the Lövdal et al dataset and a prospective pilot with an expert partner. Privacy-by-design requirements (GDPR Article 25) are treated as core product requirements rather than downstream compliance work (Regulation (EU) 2016/679; EDPB Guidelines 4/2019).

Out of scope for the near term are features that broaden breadth without strengthening the decision: comprehensive training plan generation, real-time coaching, and deep integration with every third-party platform at launch. We also avoid making clinical efficacy claims until benchmark and prospective pilot evidence exist; this reflects accepted expectations for prediction models and their reporting (Collins et al., 2015; Wolff et al., 2019).

StrideWise's current web-based MVP is intentionally narrow in scope and has been delivered to establish feasibility. It includes ingestion of core training and recovery signals, a short-horizon risk score translated into a simple Green/Amber/Red status, concise explanations of the dominant drivers, and a lightweight history view so users can understand recent trends without requiring dashboard-style interpretation.

Current MVP scope and completion status

We consider the current MVP scope complete for this delivery stage. The system runs reliably end-to-end, produces a daily decision with supporting drivers, and includes the tooling required to validate, iterate, and govern the next phase of evidence building.

The next delivery work should focus on portability (mobile), stronger external evidence (benchmark + prospective pilot), and security/privacy hardening suitable for broader distribution.

Finally, to ensure early external grounding, we have built a standalone validation harness aligned to the Lövdal et al. system and dataset. This benchmarking component provides a reproducible way to test generalisation on an independent public reference, compare against naïve baselines, and report performance using rare-event-appropriate metrics and calibration diagnostics. Together, the web MVP, synthetic generation and causation layer, core prediction model, and Lövdal-aligned validation harness form a complete 'MVP delivery package' that is finished and ready to progress into the next phases of benchmarking, clinical partnership, and pilot execution.

On top of this foundation, we have implemented the core prediction model that produces the daily risk signal. The model pipeline includes feature construction from exposure and recovery trends, quality checks to prevent silent data corruption, and calibration controls so that the traffic-light thresholds remain defensible and do not create 'alarm fatigue'. The intent at this stage is not to claim clinical validity, but to demonstrate a stable, explainable system that can be evaluated and improved under disciplined stage gates.

Behind the interface, the MVP delivery also includes the core data-science substrate needed to iterate credibly. We have built a synthetic data creation pipeline that generates longitudinal runner profiles and training blocks with realistic variability, alongside an injury causation and labelling layer that encodes how known risk drivers—such as session-level spikes, rapid load increases, and recovery degradation—translate into elevated near-term risk. This synthetic-first approach has allowed us to stress-test edge cases, missingness patterns, and model behaviour safely, while maintaining a clear separation between product development and any future handling of real health data.

All of the codebase, along with the supporting documentation, for the systems described above are available via Appendix 1.

Delivery approach

Delivery follows a stage-gated, evidence-driven approach. The guiding principle is "build–measure–learn": ship a minimal, usable product, measure behaviour and model performance, and iterate based on observed evidence rather than intuition (Ries, 2011). Practically, this is implemented through short development cycles with explicit acceptance criteria, consistent with modern agile delivery values (Agile Manifesto, 2001) and general project governance guidance that emphasises clear objectives, risk management, and stakeholder alignment (ISO 21502:2020).

Each stage has a narrow objective and a defined evidence bar. Model evaluation emphasises usefulness in the high-risk tail and calibration, rather than relying on headline discrimination metrics alone; precision–recall analysis is often more informative than ROC analysis under class imbalance typical of injury prediction (Saito & Rehmsmeier, 2015).

Scope by phase

Table 4 summarises the delivery scope at each phase, framed as the minimum set of outcomes required to justify moving forward. The table is intentionally concise: it functions as a shared contract between product, engineering, validation partners, and governance stakeholders.

Phase	Primary outcome	Key deliverables	Gate to progress
Now (web MVP)	Demonstrate end-to-end signal and UX	Working web MVP; core risk signal; early explanation patterns; baseline metrics instrumentation	Stable pipeline; coherent user understanding in quick tests; no critical governance gaps
Phase 1 (mobile VP/MVP)	Make the decision usable on-the-go	Mobile app; onboarding and consent flow; ingest + quality checks; history view; secure storage and logging	Retention signals and comprehension in user tests; security baseline met (e.g., OWASP MASVS controls)
Phase 2 (benchmark)	Demonstrate external generalisation	Evaluation report on the Lövdal et al dataset; calibration diagnostics; model card-style documentation; performance baselines	Above-baseline lift and stable calibration; clear limitations; readiness for prospective pilot
Phase 3 (clinical pilot)	Validate plausibility and usefulness prospectively	Partner protocol; ethics/governance package; secure data handling; pilot report (behaviour + early outcomes)	Evidence of user action and operational feasibility; partner sign-off on next-stage claims

Phase 4 (scale)	Harden for distribution and growth	Monitoring and drift detection; incident response; privacy hardening incl. on-device feasibility; subscription launch plan	Repeatable acquisition/retention; governance maturity aligned to scaling (e.g., ISO/IEC 27001 controls)

Table 4. Concise scope by delivery phase with stage gates.

Dependencies and delivery risks

The primary delivery dependencies are access to representative training and recovery signals (via user permissioned data), availability of the Lövdal et al dataset for benchmarking, and establishment of a clinical validation partnership to support the prospective pilot. The dominant risks are predictable: model overfitting and overclaiming, poor user comprehension leading to alarm fatigue, and governance gaps that block pilot execution. The stage-gated approach is designed to surface these risks early, before capital is committed to broad distribution.

Data Understanding

Purpose of data understanding

This section explains the datasets StrideWise uses today, what they contain, how they are structured, and the implications for modelling and validation. The focus is deliberately practical: the aim is to make transparent what the model can learn from, where uncertainty enters (missingness and label noise), and how we establish external credibility via benchmarking on the Lövdal et al dataset.

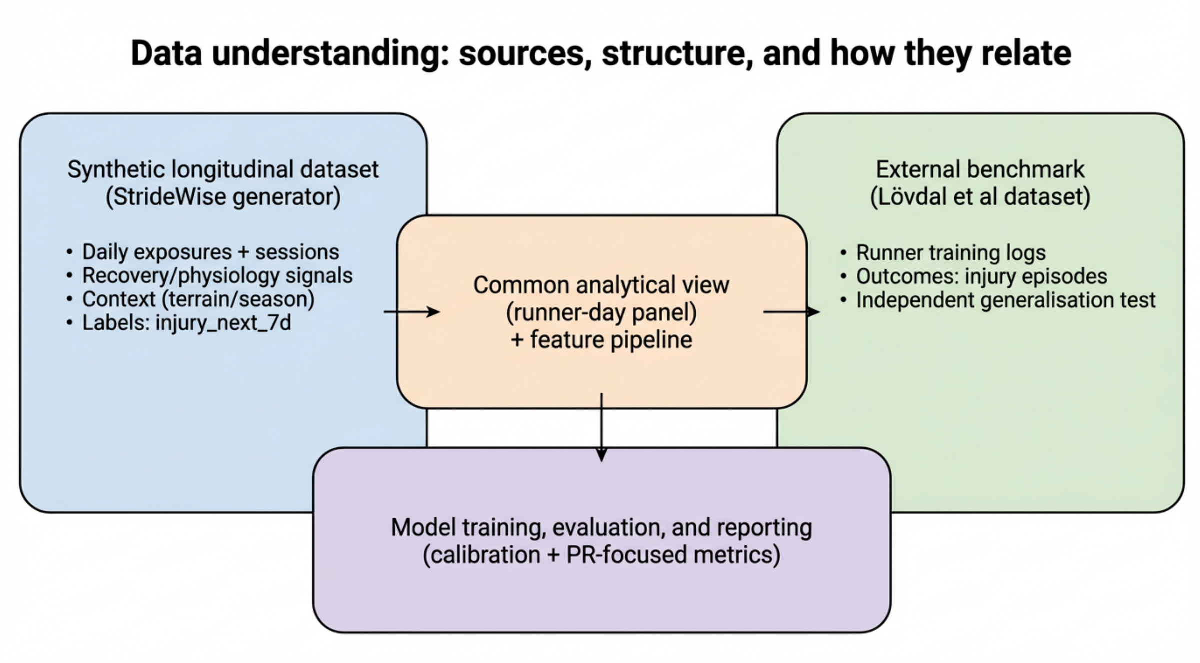

Figure 13. StrideWise data sources and how they relate (illustrative).

Data sources in scope

StrideWise works with two complementary sources of evidence. The first is a synthetic longitudinal dataset generated to mimic the structure and variability of real runner data while enabling fast iteration under privacy constraints. The second is an external benchmark—the Lövdal et al dataset—which anchors evaluation on an independent public reference used in published injury prediction work (Lövdal et al., 2021; Lövdal et al. dataset, DataverseNL).

These sources are intentionally kept distinct. Synthetic data is used to develop the end-to-end pipeline, stress-test rare edge cases, and prototype feature engineering and model behaviour. Benchmarking on the Lövdal et al dataset is used to test generalisation and to discipline claims. This division aligns with emerging guidance on synthetic data, which emphasises explicitly evaluating utility and disclosure risk rather than treating synthetic data as automatically safe or automatically valid (Gonçalves et al., 2020; Vallevik et al., 2024).

Common analytical unit: the runner-day panel

Across both sources, StrideWise uses a common analytical representation: a runner-day panel, optionally with session-level detail. The runner-day view aligns with how risk guidance is delivered (a daily decision) and how many meaningful drivers present (short-horizon changes, recovery trends). Session-level data can be rolled up into daily exposures (e.g., total duration, distance, intensity distribution proxies) and change features (e.g., today vs. prior 7/28 days).

Synthetic dataset: structure and key fields

The synthetic dataset is designed as a high-resolution longitudinal record of runners over time. It includes daily training exposure, recovery and physiology signals, and contextual modifiers such as terrain and seasonality. The generator produces both stable characteristics (e.g., baseline fitness/consistency profiles) and dynamic behaviour (e.g., training blocks, down-weeks, load spikes). The goal is to create a controlled environment in which the model can be tested against known causal assumptions, particularly whether risk increases under realistic combinations of load change and recovery degradation.

For transparency, the StrideWise synthetic dataset is persisted as three primary CSV files that mirror how consumer wearable ecosystems represent data: an activity table (session detail), a daily table (integrated day-level signals and labels), and a user table (stable traits). The full schema is extensive (including engineered features), so the tables below list the core fields that define structure and join keys. (Full details available in code via Appendix 1)

Synthetic file structure: activities.csv

Field (core)	Meaning / units	Notes
user_id, activity_id, start_time	Session identifiers	Primary key + link to user; date derived for daily rollup
session_type	Easy / long / tempo / intervals etc.	Used for sequence and density features
distance_km, duration_min, pace	Session exposure	Session-level load proxy
avg_hr	Average heart rate	Intensity proxy and realism check
elev_gain_m	Elevation gain	Terrain stress proxy
kms_z3_4, kms_z5_t1_t2, kms_sprinting	In-session intensity decomposition	Used for CC0 week mapping (session counts by zone)
cadence, stride_length, gct, vert_osc	Optional gait proxies	Only present when simulating supported devices

Table 5. activities.csv — session-level table — selected core fields (illustrative).

Synthetic file structure: daily.csv

Field (core)	Meaning / units	Notes
user_id, date	Runner-day key	Primary join key; daily panel used for prediction
session_type, sessions, has_double	Day structure	Counts and categorises sessions (incl. doubles)
km_total, duration_min, training_load	Exposure volume	Core load proxies at day level
kms_z3_4, kms_z5_t1_t2, kms_sprinting	Intensity decomposition (distance)	Aligns with Lövdal et al daily channels
strength_training, hours_alternative	Non-running exposure	Strength and cross-training volume
avg_hr, pace, elev_gain_m	Session/terrain proxies	Rolled up from activities or simulated at day level

rhr_bpm, hrv_ms	Physiology	Trend and deviation from baseline are informative
sleep_hours, sleep_efficiency, deep_sleep_hours	Sleep	Recovery signals; often missing in consumer data
stress_score, resp_rate, skin_temp	Additional physiology	Optional channels depending on device coverage
perceived_exertion, perceived_trainingSuccess, perceived_recovery	Subjective check-in	Minimal inputs used when available
device_worn, wear_7d_rate	Adherence	Explicit missingness modelling
missing_sleep, missing_hrv, missing_rhr (etc.)	Per-signal missingness flags	Prevents leakage and supports robustness
injury_onset, injury_ongoing, severity	State variables	Support label generation and analysis
injury_next_7d, illness_next_7d	Forward-looking labels	Primary supervised targets for risk prediction

Table 6. daily.csv — runner-day table — selected core fields (illustrative).

Synthetic file structure: users.csv

Field (core)	**Meaning / units**	**Notes**
user_id	Unique runner identifier	Primary key; joins to daily.user_id and activities.user_id
sex, age, height_cm, weight_kg	Demographics / anthropometrics	Stable traits; used for realism and normalisation
base_km_week, long_run_frac	Training background and tendencies	Shape baseline exposure patterns
vo2max, vo2max_adjusted	Fitness proxies	Influence pace/HR realism and capacity
rhr_base, hrv_base, hrmax	Heart-rate parameters	Support physiology realism and individual baselines
wear_rate, rest_day_frequency	Behaviour / adherence traits	Control missingness and recovery cadence
injury_proneness, injury_resilience	Risk trait parameters	Personalise risk response to similar loads

Table 7. users.csv — runner profile table — selected core fields (illustrative).

The labels in the daily file are forward-looking and operational: injury_next_7d indicates whether an injury/illness episode begins within the next seven days. In the synthetic setting, this label is generated by an injury causation layer that maps risk drivers to probability changes. This is intentionally transparent so that we can audit whether the model responds to the intended drivers and avoid "learning" artefacts that would not hold in real data.

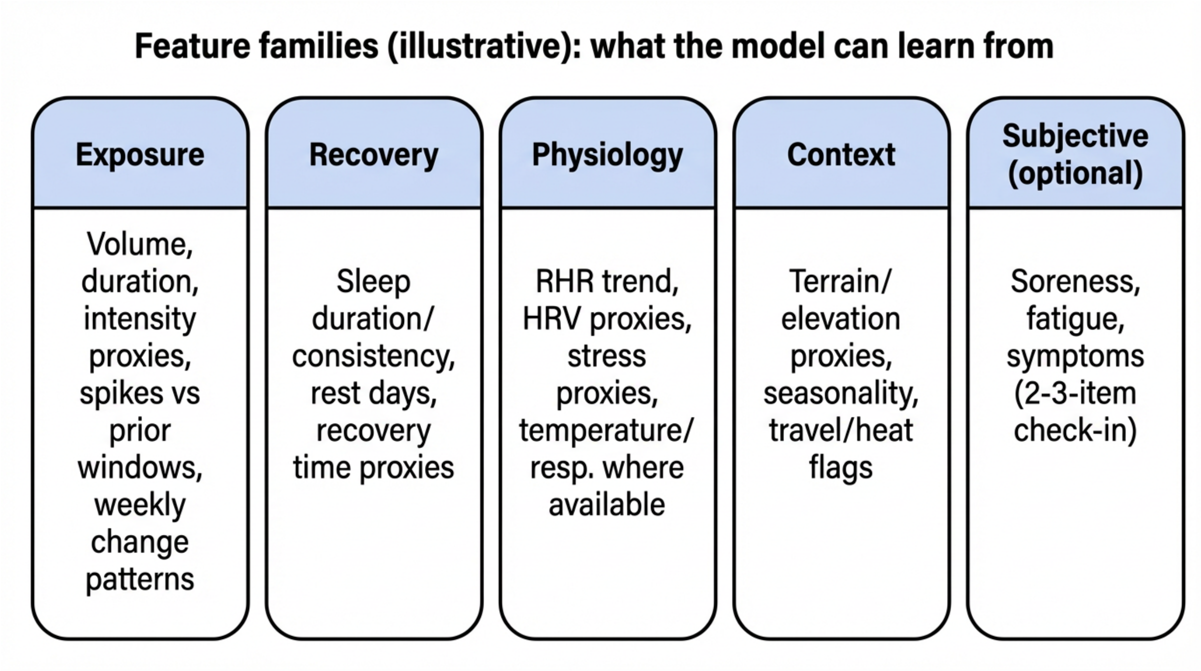

Figure 14. Feature families used to represent exposure, recovery, physiology and context (illustrative).

Lövdal et al dataset: role and constraints

The Lövdal et al dataset provides an external benchmark for running injury prediction. Lövdal et al. applied machine learning to competitive runners' training logs and published a replication dataset through DataverseNL (Lövdal et al., 2021; Lövdal et al. dataset, DataverseNL). For StrideWise, its primary role is not to replace a future prospective cohort, but to provide an independent dataset that can reveal whether the model generalises beyond synthetic assumptions and internal development conditions.

Like most real-world injury datasets, the Lövdal et al dataset has constraints that affect interpretation. Injury events are relatively rare, and labels reflect operational definitions that may differ from consumer self-report. The dataset is also specific to a competitive runner population, which may not fully represent recreational runners. These constraints are not weaknesses; they are exactly why the dataset is useful as a benchmark—because it forces StrideWise to be explicit about applicability and limitations.

To make benchmarking explicit, the Lövdal et al "day approach" represents each event as a 7-day window, with 10 daily features per day (70 values). Day 0 is the day immediately before the event; day 6 is seven days before. The table below summarises the daily feature schema and the direct mapping used in StrideWise's CC0-style translation.

Lövdal et al day approach: daily feature schema (7×10)

CC0 daily feature (paper)	StrideWise synthetic source field	Aggregation / rule (lag day d ∈ {0..6})	Notes
Number of sessions	daily.sessions	Value on date E−(d+1)	Synthetic includes doubles via has_double/double_*; sessions is the count used for CC0 equivalence.
Total distance	daily.km_total	Value on date E−(d+1)	Running km for the day.
Sum distance in Z3–Z4	daily.kms_z3_4	Value on date E−(d+1)	Matches CC0's "Z3–Z4" channel.
Sum distance in Z5, T1, T2	daily.kms_z5_t1_t2	Value on date E−(d+1)	Matches CC0's "Z5–T1–T2" channel.
Distance sprinting	daily.kms_sprinting	Value on date E−(d+1)	Separate sprint channel as in the paper.
Number of strength sessions	daily.strength_training	Value on date E−(d+1)	Stored as a per-day count/flag; CC0 expects count of strength sessions.
Hours alternative training	daily.hours_alternative	Value on date E−(d+1)	Cross-training hours (cycling, swimming, etc.).
Perceived exertion	daily.perceived_exertion	Value on date E−(d+1)	Scale aligned to [0,1] to reflect paper's ranges.
Perceived training success	daily.perceived_trainingSuccess	Value on date E−(d+1)	Same concept as CC0 subjective success.
Perceived recovery	daily.perceived_recovery	Value on date E−(d+1)	Same concept as CC0; how well athlete felt before the session.

Table 8. Lövdal et al "day approach" daily feature schema and its mapping to the StrideWise synthetic daily table (illustrative).

Data quality, missingness, and label noise

Wearable and training log data is inherently incomplete. Missingness can arise from device non-wear, recording failures, platform changes, and inconsistent subjective reporting. Importantly, missingness is rarely random: it can correlate with injury, illness, travel, or disengagement. The modelling approach must therefore treat missingness as a first-class signal, rather than simply dropping rows. The missing data literature distinguishes mechanisms such as missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR), and provides a framework for thinking about bias when missingness depends on unobserved factors (Rubin, 1976; Little & Rubin, 2002).

Label noise is also expected. In consumer settings, injuries may be self-reported, under-reported, or recorded with delay; in research settings, injury definitions vary. Prediction-model guidance stresses the importance of transparent reporting, validation, and avoiding overfitting—particularly when outcomes are rare and noisy (Collins et al., 2015; Wolff et al., 2019). StrideWise's staged approach (synthetic calibration → benchmark on Lövdal et al dataset → prospective pilot) is designed to manage these realities.

What we measure and how we report it

Because injury/illness is a rare-event problem, evaluation must reflect usefulness in the high-risk tail. Precision–recall analysis is often more informative than ROC analysis under class imbalance, and it aligns with the product question: when we flag a day as high-risk, how often is that flag meaningful (Saito & Rehmsmeier, 2015). We also report calibration diagnostics so that the probabilities and thresholds remain defensible and do not drift into unnecessary alarm behaviour.

Concise data inventory

Table 9 provides a concise inventory of the two data sources and how they are used. This inventory is a living artefact: as StrideWise begins prospective data collection under informed consent, a third source will be added and governed under an explicit data protection and research protocol.

Dataset	Granularity	What it contains	How it is used
StrideWise synthetic dataset	Runner-day (with session detail where needed)	Training exposure, recovery/physiology, context; forward-looking injury_next_7d label via causation layer	Pipeline development, feature engineering, model prototyping, stress-testing edge cases and missingness
Lövdal et al dataset (DataverseNL)	Training log with injury outcomes (published replication dataset)	Real runner training data and injury episodes from competitive runner context	Independent benchmark for generalisation; evaluation reporting and calibration discipline

Table 9. Data sources and their role in StrideWise's current evidence pathway.

Data Preparation

Purpose and scope

Data preparation in StrideWise is designed for two concurrent outcomes. The first is to enable robust modelling on smartwatch-grade “rich data” (training exposure, recovery and physiology, and lightweight subjective check-ins) under the kinds of missingness and inconsistency seen in consumer wearables. The second is to enable credible external benchmarking by translating the rich synthetic dataset into the Lövdal et al event-window representation, so we can evaluate behaviour in a published feature space rather than asserting realism by inspection.

Because these goals span multiple schemas (continuous user-day panels versus event-anchored vectors), preparation includes conventional cleaning (types, units, duplicates, and plausibility checks), plus structural alignment (window definitions, event construction, column ordering, and normalisation rules).

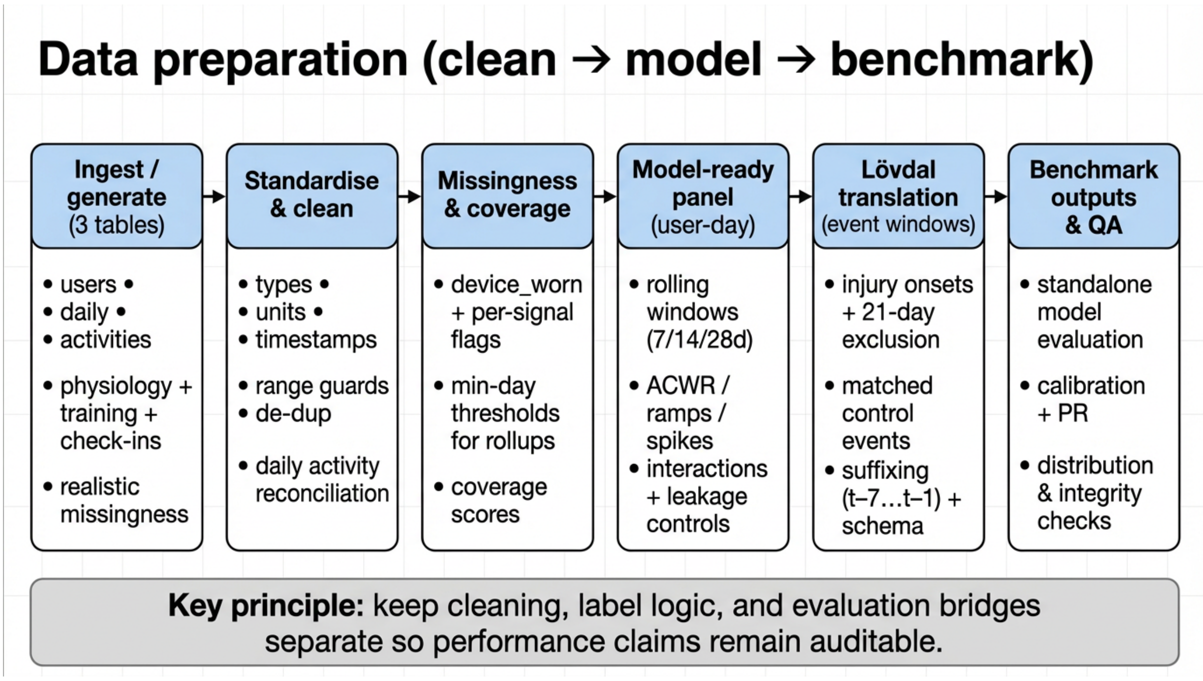

Figure 15. Expanded data preparation pipeline from ingestion through model-ready features and Lövdal-style benchmarking.

Datasets and how they connect

StrideWise uses two distinct dataset representations that are intentionally not merged at row level. The Lövdal et al dataset is an event-based table where each row represents an injury event (or matched control) with the preceding training history encoded as a fixed-length vector. Our StrideWise dataset is a continuous, longitudinal time series, stored as three linked tables (users, daily, activities), which more closely resembles how wearable ecosystems represent data.

To connect these worlds scientifically, we use a representation bridge rather than a direct join. Synthetic daily and activity records are transformed into a Lövdal-style event table via deterministic windowing, aggregation, naming, and ordering rules. This makes the comparison testable: models are compared under the same feature representation, rather than trying to force incompatible datasets into a single schema.

Dataset / table	Unit of analysis	Primary keys / join fields	Purpose in the pipeline
Lövdal et al dataset (event table)	Event-anchored rows (injury or control) with 7-day history vectors.	Athlete ID; event date/index; injury label; suffixed feature columns.	External benchmark representation for evaluation and replication.
users.csv (synthetic)	Runner-level stable profile (demographics, baseline capacity, risk traits).	user_id (PK).	Provides per-runner context, segmentation, and stable covariates.
daily.csv (synthetic)	Runner-day panel combining exposure, physiology, recovery, and labels.	(user_id, date) composite key.	Primary modelling table foundation; source for rolling features and label windows.
activities.csv (synthetic)	Session-level exposures (distance, duration, intensity proxies, elevation, etc.).	activity_id (PK); user_id (FK); start_time/date.	Drives daily rollups and intensity decompositions; enables reconciliation checks.
Lövdal-style translated table (synthetic → event vectors)	Event/control rows with 7-day feature windows (suffix convention).	Athlete ID; injury; Date; strict column order.	Allows indirect validation using the standalone Lövdal-style benchmark model.

Table 10. How datasets and tables connect across modelling and benchmarking.

Cleaning, standardisation, and integrity checks

Although the synthetic generator enforces many realism constraints, we still treat the resulting files as raw inputs. Preparation begins by normalising data types and harmonising time fields so that rolling windows and event extraction behave consistently. Distances, durations, and physiological measures are standardised to consistent units, and categorical labels (session types and identifiers) are canonicalised to avoid hidden duplicates.

We then apply plausibility guards to remove or neutralise impossible values before they can distort downstream aggregates. Negative distances or durations are treated as missing or removed when unrecoverable. Physiological channels are constrained to plausible ranges and extreme outliers are treated as missing rather than “winsorised” into a value that may look reasonable but is still artefactual. At the structural level, we enforce one daily record per athlete-date and run reconciliation checks where activity data exists, confirming that daily aggregates (e.g., total distance and session count) match the sum of session-level records.

Quality dimension	What we check / enforce	Why it matters
Types & units	Numeric enforcement for distances/durations/scores; consistent units (km, minutes, bpm); timestamp alignment.	Prevents silent coercion and incorrect windowing or aggregations.
Uniqueness & duplication	One record per (athlete, date) in daily panels; de-duplicate exact repeats in event tables.	Avoids double-counting exposure and inflating risk signals.

Plausibility / ranges	Impossible values set to missing; physiology constrained; exposure distributions checked by runner type.	Protects feature distributions and prevents unstable model behaviour.
Cross-table consistency	Daily totals reconcile with summed activities where available; session count derived if missing.	Ensures "what happened" is coherent across tables.
Label integrity	Forward-looking injury_next_7d windows; injury onsets isolated; exclusion windows for episodes.	Avoids leakage and ensures evaluation reflects onset prediction rather than episode persistence.

Table 11. Core cleaning and integrity controls applied during preparation.

Missing values and wearable realism

Missingness is treated as a first-class property rather than a nuisance. Consumer wearables are not worn every day and not all sensors record reliably even when a device is worn, so the synthetic generator explicitly models both device wear compliance and per-signal gaps. In practice this means that, alongside each physiological channel, we preserve missingness flags (e.g., HRV missing, sleep missing) and a device_worn indicator so that downstream feature engineering can be both gap-tolerant and informative.

Feature engineering is therefore designed to degrade gracefully when coverage is partial. Rolling windows are computed with minimum-day thresholds (for example, a 7-day HRV trend is only computed when sufficient HRV days exist in the window; otherwise the feature is set to missing and a companion 'missing-in-window' flag is activated). Where exposure can be represented through multiple signals, we use fallbacks: if distance is missing but duration is present, duration can act as the exposure proxy for acute load calculations. These rules preserve auditability and avoid over-imputation that can wash out the very patterns the model should learn.

Preparing the rich synthetic dataset for modelling

The modelling dataset is built as a single user-day panel derived from the daily table (with optional rollups from activities). Each row represents a day of training exposure and recovery state, paired with a forward-looking label (injury_next_7d) that indicates whether injury onset occurs within the following seven days. To prevent information leakage, predictors are computed exclusively from historical days, and the train/validation split is performed as a forward-time split per user to reflect deployment reality.

From this panel we construct rolling-window summaries (typically 7, 14, and 28 days) for exposure, intensity distributions, and physiology; spike constructs that quantify acute deviations from an individual's recent norm; and change features such as deltas and z-scores that describe 'how different this week is from the last month'. We also include ACWR-style ratios and ramp features to capture rapid changes in workload, and interaction terms that reflect how load, recovery, and individual susceptibility combine in practice. The result is a missingness-aware feature matrix that remains usable under realistic consumer data sparsity while still representing the dynamics required to detect rising injury risk.

Injury causation layer and label preparation

Labels are not assigned post-hoc; instead, injuries emerge from an explicit causation layer that maintains injury state over time (onset, ongoing, recovery/return) and exposes a forward-looking injury_next_7d outcome. This layer responds to combinations of drivers that are consistent with a load–capacity framing of running injuries, where risk increases when workload rises faster than the athlete's capacity to adapt, and where recovery signals modify that capacity. In practical terms, the causation layer uses workload change indicators (including spike behaviour and ACWR-style dynamics), fatigue accumulation, recovery deficits derived from physiology and sleep, and individual risk traits.

For preparation purposes, the key requirement is that labels remain interpretable and auditable. We therefore separate the causation calculation from the feature engineering pipeline, and we apply episode-aware filtering so that evaluation focuses on predicting onset rather than repeatedly scoring the same multi-day injury episode. This is also where exclusion windows are applied to avoid unrealistically dense injury clustering.

Transforming synthetic data into the Lövdal et al event-window format

To evaluate against a published benchmark representation, we convert the synthetic daily series into a Lövdal-style event dataset. This begins with event extraction. Injury events are anchored on onset days, and a 21-day exclusion window is applied so that days within the same episode do not create multiple near-duplicate events. We then sample control events from non-injury days, matching the injury event distribution to avoid an unrealistic control set.

For each event day E, we extract the preceding seven days (E−7 through E−1) and emit a fixed-length feature vector using the Lövdal suffix convention: the oldest day in the window has no suffix, and the most recent day carries suffix ".6". Daily measures are mapped deterministically (e.g., km_total to "total km", sessions to "nr. sessions", intensity decompositions to zone-kilometre fields, and subjective signals to perceived variables). Non-training days use a dedicated rest marker for perceived metrics, matching the benchmark convention, and perceived scales are normalised into the expected range. Finally, the converter enforces strict column ordering and completeness to ensure the translated table is structurally compatible with the standalone benchmark pipeline.

Data quality assessment and outputs

Data quality is assessed across completeness, consistency, plausibility, and label integrity. Completeness is characterised via coverage indicators (device wear and per-signal missingness). Consistency is checked via daily-to-activity reconciliation where session detail exists. Plausibility is monitored through range guards and distribution checks (including profile-dependent exposure expectations). Label integrity is verified by confirming forward-looking windows, enforcing episode-aware filtering, and maintaining the event-control construction rules used for benchmarking.

The preparation stage produces two principal analysis artefacts: a model-ready user-day feature table for the main StrideWise predictor, and a Lövdal-style translated event table for indirect validation. Keeping these outputs distinct helps ensure that claims about model performance, label logic, and dataset realism can be inspected independently, and that future real-world ingestion can slot into the same preparation interface.

Exploratory Data Analysis

Purpose of the EDA

Exploratory Data Analysis (EDA) is used to ensure that modelling outcomes reflect real signal rather than data artefacts. For StrideWise, this means confirming that training and recovery signals behave plausibly over time, missingness is realistic and explicitly represented, injury labels and event definitions are coherent, and our synthetic dataset remains externally grounded by comparing its Lövdal-style translation to the Lövdal et al benchmark representation.

EDA is therefore not a one-off report but a set of repeatable checks that run whenever the generator, the causation layer, or the feature pipeline changes. This mirrors the role of data profiling and integrity testing in production analytics systems, where data drift and silent quality regressions are a primary operational risk.

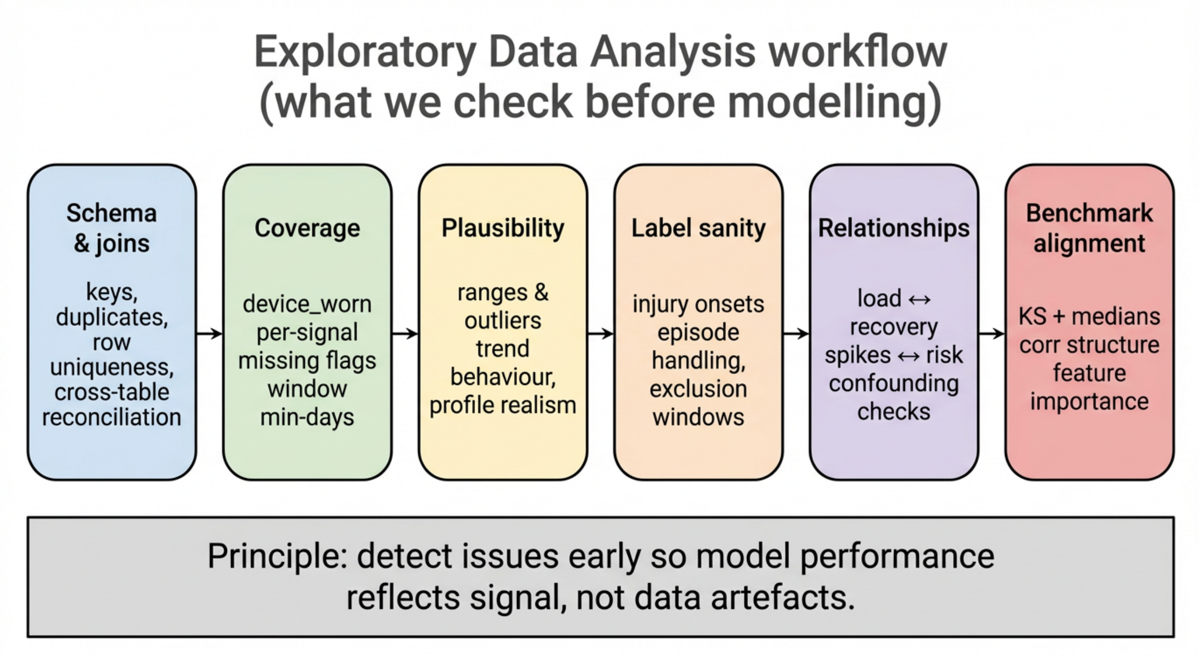

Figure 16. EDA workflow used to profile data quality and plausibility before training and evaluation.

Brief EDA overview by dataset

Tables 1 and 2 summarise the headline EDA facts for the two primary datasets used in this work. These summaries are intentionally brief and are complemented by the operational "sanity artefacts" described later in this section.

Synthetic dataset: EDA overview (StrideWise generator)

EDA dimension	Headline result	Interpretation / implication
Scale / coverage	3,000 runners; 540,000 user-days (2023-01-01 to 2023-06-29)	Sufficient breadth for stress-testing missingness, segment behaviour, and threshold policy.
Labels	injury_next_7d prevalence 10.38%; injury_onset prevalence 1.52%	Forward-looking label increases prevalence as expected; onset remains rare and is treated episode-aware in evaluation.

Exposure distribution	Mean distance 4.19 km/day (median 2.55; p95 13.33)	Plausible consumer training regime rather than an extreme synthetic artefact.
Load proxy	Mean training_load 5.32 (median 3.60; p95 15.83)	Supports ramp/spike features while retaining realistic dispersion across runner profiles.
Missingness realism	HRV missing 18.8%; RHR missing 15.3% (non-trivial sensor gaps)	Confirms the need for missingness-aware features and graceful degradation in the product.

Table 12. Headline EDA summary for the StrideWise synthetic dataset (as reported by the pipeline).

Lövdal et al dataset: EDA overview (benchmark)

EDA dimension	Headline result	Interpretation / implication
Time span / context	Training logs 2012–2019; competitive runner cohort (Dutch middle-/long-distance)	Strong longitudinal depth; applicability strongest to structured training populations.
Representation	Event-based table; each row is an event/control with a 7-day history vector (day approach)	Not directly compatible with wearable-rich daily modelling; requires translation bridge for indirect validation.
Class balance	After filtering: 42,183 healthy events vs 583 injury events (≈1.4% injury prevalence)	Rare-event regime; precision–recall and calibration are more informative than ROC alone.
Feature scope	Primarily training exposure + perceived measures; limited physiology vs consumer wearables	Benchmark is valuable for generalisation discipline, but cannot validate wearable-sensor features directly.
Event rules	Strict event filtering and episode handling in published pipeline	Requires careful matching in synthetic translation (onset anchoring, exclusion windows, control sampling).

Table 13. Headline EDA summary for the Lövdal et al benchmark dataset.

Note: We additionally run EDA on the Lövdal-format translated synthetic dataset to confirm that its event-table regime matches the benchmark setting (for example, the elite-250 translation produces 43,250 event rows with 1.50% injury prevalence), before running the standalone validation harness.

Structural profiling and join integrity

We begin by verifying that the dataset can be joined deterministically and that aggregation is coherent. For the synthetic dataset this means uniqueness of the (user_id, date) key in the daily table, referential integrity between activities and daily records, and stable runner identifiers in the user table. Where session detail exists, we reconcile daily totals (e.g., total distance and session count) against sums computed from the activities table to catch duplication, missing rows, or time-zone alignment issues.

Coverage and missingness analysis

Consumer wearable data is incomplete by default, so we treat missingness as a data feature rather than an error to “impute away”. EDA characterises both device wear behaviour (device_worn and short-window wear rates) and per-signal missingness (sleep, HRV, RHR, temperature, respiration). This provides two benefits: it prevents accidental leakage (e.g., a feature being present only on “good days”), and it ensures the prediction pipeline can degrade gracefully when users do not have full sensor coverage.

At feature level, rolling-window statistics are computed with minimum-day thresholds, and EDA reports the proportion of windows that are eligible for each feature family. This is a practical measure of whether features will be available often enough for a consumer product, and it highlights where fallbacks are needed (for example, using duration when distance is missing).

Plausibility checks for training, recovery, and physiology

We profile univariate distributions and time-series behaviour to confirm that the synthetic generator produces smartwatch-grade realism. Training exposures are checked for plausible ranges and structure (rest days, long runs, harder sessions, down-weeks), and intensity decompositions are verified to be consistent with the total distance. For recovery and physiology channels, we inspect both range plausibility and directional responsiveness, for example, whether a sequence of hard sessions is associated with temporary elevation in resting heart rate and degraded recovery signals, and whether recovery weeks return toward baseline.

These checks are conducted at both population level (are distributions plausible across runner profiles) and within-runner level (does an individual show coherent baselines, variability, and training blocks). Where anomalies are detected, the pipeline prefers to treat values as missing (paired with missingness flags) rather than coercing them into a plausible number that may still be artefactual.

Label and event sanity analysis

EDA verifies that injury labels behave coherently in time. In the continuous synthetic dataset, injuries are represented as state (onset, ongoing, recovery/return), and the modelling label injury_next_7d is computed as a forward-looking outcome relative to each day. EDA checks that onset events are not duplicated within the same episode and that evaluation windows do not inadvertently include post-injury behaviour that would inflate apparent predictive performance.

In the Lövdal-style translated dataset, we anchor rows on onset events and apply an exclusion window so that multiple rows are not produced from days within a single injury episode. Control events are sampled from non-injury days using matching logic so that the benchmark dataset reflects realistic exposure patterns rather than an artificially “easy” control set.

Relationship exploration (what should correlate with risk)

The goal of the product is not to maximise correlation, but EDA should confirm that directionality is sensible: risk should increase under combinations of rapid load escalation, high intensity density, and degraded recovery signals, particularly for athletes with lower baseline capacity or lower adherence. We therefore examine partial relationships between exposure spikes, recovery degradation, and injury labels, and we check that these relationships are not artefacts of missingness or label construction.

This relationship-oriented EDA step also informs feature engineering: it helps prioritise a small set of robust, interpretable drivers that are likely to remain available in consumer settings, rather than overfitting to narrow signals that disappear when a user changes device or stops providing check-ins.

Benchmark alignment checks (synthetic → Lövdal-style evaluation)

To ground the synthetic dataset externally, we run automated comparisons between the translated synthetic Lövdal-style event table and the real Lövdal et al representation. This includes distribution comparisons (e.g., Kolmogorov–Smirnov statistics and median tests), preservation of correlation structure, and whether feature importance patterns from a benchmark model trained on the public dataset are broadly consistent when the same model is applied to the translated synthetic data. These checks provide an indirect but disciplined answer to a central question: does the synthetic dataset produce the kinds of patterns that a published injury prediction framework expects?

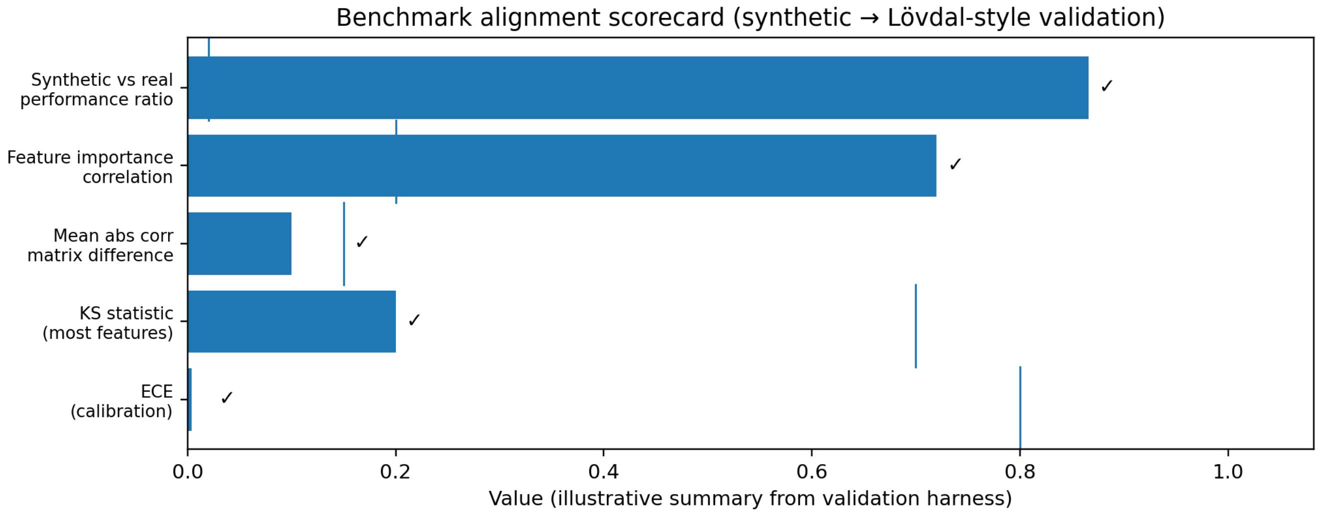

Figure 17. Summary alignment scorecard from the validation harness described in the system documentation (illustrative).

EDA artefacts produced and how they are used

A key design choice in StrideWise is to treat EDA as an operational layer with named outputs, not as a one-off narrative. Each run of the generator, feature pipeline, or translation step emits a small set of "sanity artefacts" that make it easy to detect regressions early. These artefacts are reviewed whenever parameters change (e.g., injury-causation weights, wear-rate assumptions, or feature filters) and form part of the same evidence chain as model metrics.

At dataset level, the large synthetic dataset produces machine-readable sanity summaries and trend checks (e.g., large_dataset/sanity_report.json, large_dataset/sanity_label_rate_by_decile_daily.csv, and large_dataset/sanity_injury_time_trend_weekly.csv). These files verify that basic distributions (coverage, volume, and label rates) remain stable across runs, and they provide lightweight time-trend monitoring to ensure injuries are not clustered into implausible bursts or seasonality artefacts.

At model level, the evaluation harness produces a consistent set of visual and tabular artefacts covering discrimination (ROC/PR), calibration, feature importance, and threshold policy. In the codebase this is generated via the script "/validation/create_main_model_visualizations.py", with outputs written to "main_model_visualizations/." These artefacts are the fastest way to confirm that performance changes are driven by real signal rather than pipeline changes such as leakage, altered window definitions, or shifted missingness handling.

For external grounding, the translation and indirect validation steps are also treated as first-class outputs. The conversion pipeline ("convert_synth_to_cc0.py"; with schema enforced by "cc0_feature_schema.json") emits a Lövdal-format event table that is compatible with the standalone benchmark model. The indirect validation run then generates an overall evaluation report capturing distribution similarity (e.g., KS statistics), correlation preservation, and model-behaviour checks such as the synthetic-to-real performance ratio.

Finally, a compact "EDA figures pack" is generated for human review, typically including a missingness summary, wear-rate distribution, injury prevalence by predicted-risk decile, injury time-trend plots, and spot-check distributions for key translated Lövdal-format fields (e.g., fig_01_missingness.png through fig_05_cc0_total_km_tminus1.png). This pack is particularly useful when sharing progress with non-technical stakeholders, because it makes the state of the data legible without exposing raw tables.

Artefact (examples)	What it confirms	How we use it
large_dataset/sanity_report.json	Row counts, basic ranges, missingness summaries, key label stats.	Quick regression check after any generator or preprocessing change.
large_dataset/sanity_label_rate_by_decile_daily.csv	Risk stratification behaves sensibly (higher predicted risk → higher label rate).	Detects label leakage or model collapse; supports threshold tuning.
large_dataset/sanity_injury_time_trend_weekly.csv	Injury onsets and label prevalence remain stable over time.	Flags unrealistic clustering and seasonality artefacts.
main_model_visualizations/ (ROC/PR, calibration, importance)	Discrimination, calibration, stability of driver signals.	Core "is the model still healthy?" pack for each training run.
Lövdal-format translation output + schema checks	Event-table construction and class-balance regime match the benchmark setting.	Ensures indirect validation is meaningful (not an artefact of formatting).
EDA figures pack (e.g., fig_01_missingness.png … fig_05_cc0_total_km_tminus1.png)	Human-readable validation of missingness, wear, trends, and translated-field plausibility.	Stakeholder communication and fast visual QA during iteration.

Table 14. Operational EDA artefacts emitted by the pipeline to detect regressions and ensure continued plausibility.

Concise EDA outputs

EDA produces a compact set of artefacts that are used repeatedly throughout development: a data profiling summary (coverage, ranges, and join integrity), a missingness and availability report by feature family, a label/event integrity report (onsets, episode clustering, and exclusion behaviour), and a benchmark alignment report that tracks distribution similarity, correlation preservation, and model-based sanity checks. Together these outputs allow rapid iteration while keeping assumptions transparent.

EDA check	What we look for	Why it matters
Join integrity	Stable user_id; unique daily keys; daily totals reconcile with activities where present.	Prevents silent duplication and invalid rollups that can dominate model behaviour.
Coverage & missingness	Wear rate plausible; per-signal missing flags present; rolling features only computed when windows have sufficient days.	Ensures the model remains usable under consumer sparsity and avoids leakage via "availability".
Plausibility	Ranges and trends align with expected physiology and training structure; intensity sums do not exceed totals.	Reduces artefactual patterns that inflate apparent performance.
Label/event integrity	Onsets separated from ongoing episodes; exclusion windows applied; control events sampled realistically.	Ensures performance claims reflect onset prediction, not repeated episode days.
Benchmark alignment	Distributions broadly similar; correlation structure preserved; importance patterns stable under the same representation.	Provides external grounding before prospective real-world data collection.

Table 15. EDA checks used to ensure modelling and evaluation remain robust and interpretable.

Methods and Frameworks

Approach overview

StrideWise combines two methodological threads in a single production-ready pipeline: an algorithmic simulation in Python to generate smartwatch-grade longitudinal training and recovery data under realistic missingness, and a supervised machine learning model to translate those histories into a simple, actionable injury-risk signal. The system is modular by design: synthetic generation, injury causation and labelling, feature engineering, model training, and benchmarking are separated so each layer can be inspected and improved without invalidating the others (StrideWise Development Team, 2025).

To keep the work legible to both technical and investor audiences, we map the end-to-end delivery to the CRISP–DM lifecycle (Chapman et al., 2000). This ensures the project remains anchored to a real user problem (reducing injury disruption) while maintaining disciplined controls on data leakage, imbalanced outcomes, and evaluation credibility.

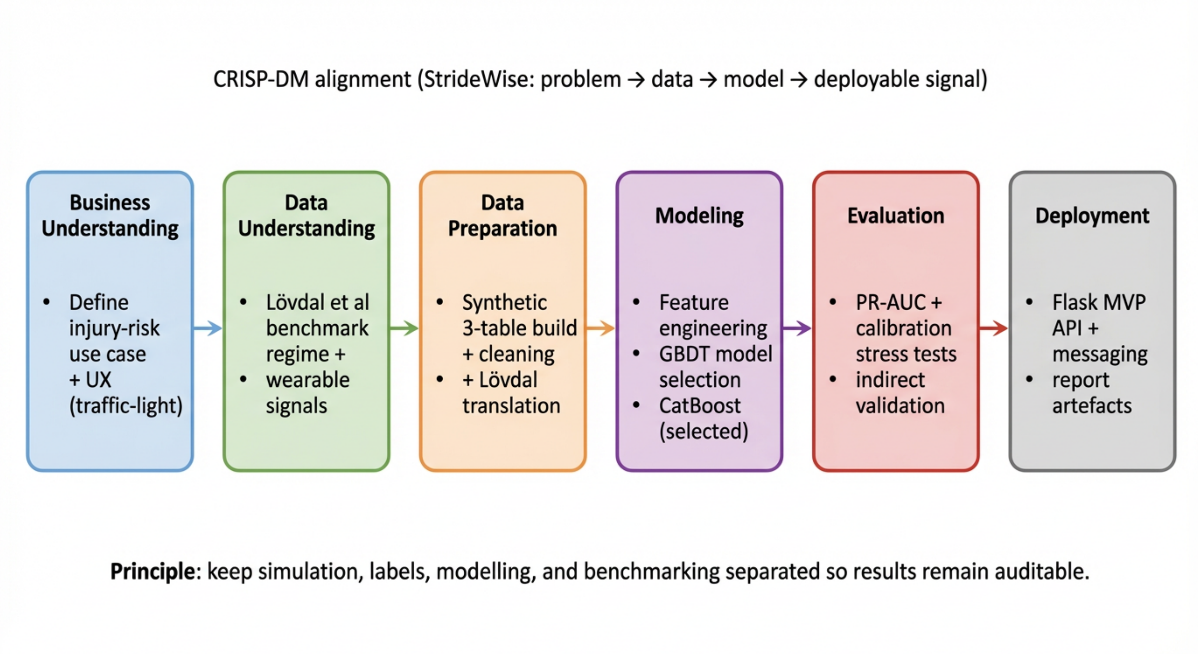

Figure 18. CRISP–DM mapping for StrideWise, linking problem framing, data work, modelling, evaluation, and the MVP deployment path.

CRISP–DM alignment

CRISP–DM provides a practical spine for the work: we begin with a concrete decision problem (should a runner back off today), define the necessary data representations and labels, prepare a model-ready dataset, train and evaluate models using metrics that match a rare-event setting, and package the output into a deployable product workflow (Chapman et al., 2000).

CRISP–DM stage	StrideWise activities	Methods / tools	Key outputs / artefacts
Business Understanding	Define the athlete decision: reduce load when risk rises; present a simple signal that avoids dashboard interpretation.	User-centred decision framing; probability-to-policy mapping for consumer clarity.	Risk messaging policy (Green/Orange/Red) and product success criteria.

Data Understanding	Profile the Lövdal et al benchmark regime and identify gaps versus smartwatch data; define the target wearable signal space.	Dataset profiling, schema analysis, class imbalance assessment, coverage analysis.	Benchmark dataset summary and translation requirements (event windows, suffixing, ordering) (Lövdal et al., 2021).
Data Preparation	Generate smartwatch-grade synthetic data; clean/standardise; build missingness-aware features; translate rich synthetic data into Lövdal-style format for evaluation.	Algorithmic simulation in Python; integrity checks; rolling-window engineering; deterministic event extraction and schema enforcement.	Model-ready user-day panel, Lövdal-format event table, sanity/trend artefacts and evaluation-ready splits.
Modeling	Train candidate models to predict injury_next_7d on the rich synthetic panel; tune and select the best generalising approach.	Gradient boosted decision trees (XGBoost, LightGBM, CatBoost), neural network exploration; early stopping; forward-time splits.	Production main model (CatBoost selected) and stable feature set (335 engineered features).
Evaluation	Evaluate discrimination and calibration; stress-test against leakage; validate synthetic realism indirectly by benchmarking in the Lövdal feature space.	PR-AUC, ROC-AUC, Brier, ECE; replication of paper approach; distribution similarity and “performance ratio”.	Model performance reports, calibration plots, benchmark alignment scorecard, and validation conclusion.
Deployment	Package inference into a product workflow; compute features, return risk band and guidance; support monitoring.	Python Flask MVP, REST endpoints, deterministic feature computation, artefact logging.	Web-based MVP prediction flow and repeatable reporting for iteration.

Table 16. CRISP–DM mapping, showing how technical work and product delivery remain tightly linked.

Synthetic data generation methods (Python simulation)

StrideWise’s synthetic dataset is generated algorithmically in Python and is designed to resemble how wearable platforms represent longitudinal training histories: a user table of stable traits, an activities table for session detail, and a daily table that integrates exposure, recovery/physiology, missingness flags, and labels (StrideWise Development Team, 2025). This structure supports both consumer decisioning (day-by-day risk) and research-style summarisation (rolling windows and event-based windows).

Injuries are produced through an explicit causation layer, modelled on real world research, that maintains injury state over time and exposes a forward-looking label (injury_next_7d). Risk is parameterised as a profile-specific baseline modified by interpretable drivers such as workload change (ACWR-style dynamics), fatigue accumulation, recovery deficits, sprinting exposure, and long-run spike behaviour (StrideWise Development Team, 2025). This keeps learning auditable and reduces the risk that a model simply learns label artefacts.

Physiological channels (e.g., HRV and resting heart rate) are simulated with lagged responses to training and recovery, plus natural noise; missingness and device wear are modelled explicitly (StrideWise Development Team, 2025). This matters operationally because consumer device coverage is incomplete and a deployable system must degrade gracefully rather than failing when signals are absent.

Feature engineering and model inputs

The main predictor is trained on a rich user-day panel derived from the synthetic daily table, optionally enriched with session rollups from activities. Feature engineering is dominated by time-aware constructs that reflect how injury risk emerges: short-term versus long-term workload, rapid ramp behaviour, cumulative fatigue, and recovery degradation.

In production, the system generates 335+ engineered features grouped into rolling-window summaries (7- and 28-day aggregates), ACWR-derived ratios and trajectories, ramp and spike features, interaction terms that capture profile-dependent sensitivity, temporal features, and recovery/fatigue state indices (StrideWise Development Team, 2025). A selection layer removes constant and highly redundant features to stabilise training and avoid overfitting on correlated signals (StrideWise Development Team, 2025).

Machine learning approaches and model selection

StrideWise uses supervised learning to predict injury_next_7d, aligning to a practical intervention window. Because the outcome is imbalanced, modelling is evaluated using metrics that reflect operational utility (precision–recall behaviour and calibration) alongside ROC-AUC (Saito & Rehmsmeier, 2015; Brier, 1950).

The codebase supports multiple model families and we tested gradient boosting variants (XGBoost and LightGBM), CatBoost, and exploratory neural networks (StrideWise Development Team, 2025). CatBoost provided the best balance of discrimination, generalisation, and stability under the available feature regime, achieving validation ROC-AUC of 0.7136, PR-AUC of 0.2427, and Brier score of 0.0973 in the documented configuration (StrideWise Development Team, 2025).

Benchmarking and indirect validation using the Lövdal et al framework

A central methodological challenge is that rich, labelled real wearable datasets are difficult to access at scale and typically involve sensitive health data. To keep validation credible, StrideWise uses an indirect approach grounded in a published injury prediction framework and dataset: we implement a standalone model in the Lövdal feature space, replicating as far as possible the approach outlined in their published paper, validate it against published performance, and then test whether synthetic data translated into the same feature space behaves similarly (Lövdal et al., 2021; StrideWise Development Team, 2025).

The standalone model replicates the day-approach method as a bagged ensemble of XGBoost models trained on bootstrap samples, with per-athlete z-scoring based on healthy baselines and strict event construction rules including a 21-day exclusion window (StrideWise Development Team, 2025). In the documented replication, this achieves AUC 0.7121 compared to the paper's 0.724, supporting implementation fidelity (Lövdal et al., 2021; StrideWise Development Team, 2025).

To evaluate synthetic realism under the same assumptions, we convert rich synthetic histories into a Lövdal-format event table by extracting onset events, sampling matched control events, and emitting a 7-day history vector using the published suffix convention and strict column ordering (StrideWise Development Team, 2025). Conversion checks verify schema compatibility and that injury prevalence in the translated table matches the benchmark regime (approximately 1.4–1.6%) (Lövdal et al., 2021; StrideWise Development Team, 2025).

Finally, the same benchmark model is applied to both the real and translated synthetic datasets, and an overall "performance ratio" is reported as a conservative alignment indicator. The system documentation reports 86.6% of the real benchmark performance, which is treated as supportive evidence (not clinical proof) that the synthetic data captures relevant structure (StrideWise Development Team, 2025).

Deployment framework and productisation

The system is packaged into a web-based MVP implemented in Flask, which loads the trained model, computes features from a runner's recent history, and returns a risk score and guidance message (StrideWise Development Team, 2025). This deployment pattern keeps the inference interface stable while allowing continued iteration on generator parameters and model retraining.

To make the output immediately actionable, the MVP maps predicted risk into simple bands (Green/Orange/Red) paired with training guidance. In parallel, the pipeline emits repeatable artefacts (sanity summaries, trend checks, and evaluation reports) so that any data or modelling change can be audited before release (StrideWise Development Team, 2025).

Solution Architecture and Proof of Concept

End-to-end architecture

StrideWise is built as a modular pipeline that turns longitudinal training and recovery histories into an extremely simple, daily injury-risk signal. The proof of concept intentionally separates simulation, labelling, modelling, benchmarking, and product delivery so that each layer can be tested independently and audited when parameters change (StrideWise Development Team, 2025).

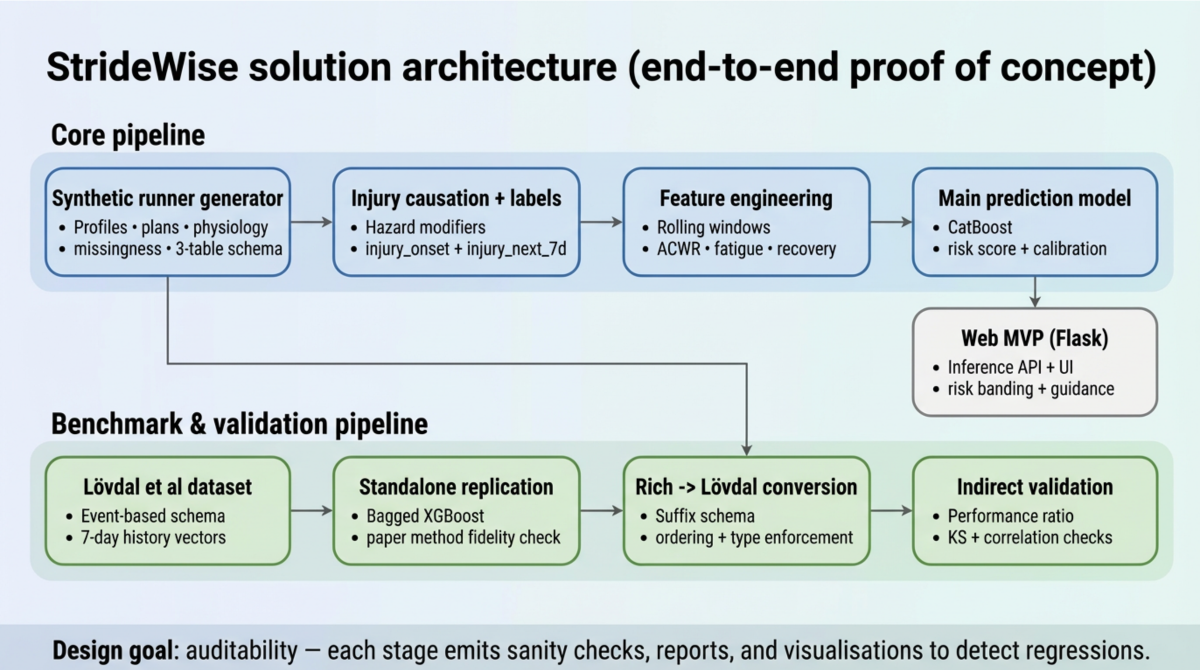

Figure 19. Solution architecture showing how synthetic generation, injury causation, modelling, benchmark replication, and the web MVP fit together.

Synthetic runner generation

The synthetic data generator creates individual runners rather than generic averages. Each synthetic runner is assigned a profile (novice, recreational, advanced, or elite) and a set of stable attributes that influence training exposure and recovery dynamics, including age, sex, anthropometrics, a normalised fitness score that informs VO_2max, baseline weekly volume, and individualised injury proneness and resilience parameters (StrideWise Development Team, 2025). The intent is to simulate a population with meaningful heterogeneity—so that the prediction model must learn patterns that generalise across different runner types.

Training histories are generated using a weekly periodisation structure (build, peak, recovery, and base phases) with realistic volume modulation and athlete-to-athlete variability. Day-to-day sessions are then instantiated with plausible session types—easy runs, tempo runs, interval/sprint sessions, and long runs—with weekend long runs and post-hard-session recovery days emerging naturally from the plan logic (StrideWise Development Team, 2025).

At the activity level, each session produces distance, duration, pace, intensity-zone distribution (including sprinting exposure), and supporting signals such as cadence and stride-mechanics proxies where relevant. Daily aggregation then yields a smartwatch-style day record that integrates exposure, load, subjective perceptions, and physiology. Training load is computed using a TRIMP-style formulation that combines session duration with an intensity factor to provide a unified load proxy for downstream modelling (Banister, 1991; StrideWise Development Team, 2025).

Science-informed physiological simulation

Physiological signals are simulated as lagged responses to training and recovery, rather than as independent random series. For example, HRV is generated around an athlete-specific baseline with a next-day suppression after high load and a rebound during recovery weeks, while resting heart rate increases with acute stress and can remain elevated under cumulative fatigue (StrideWise Development Team, 2025). Sleep duration and quality are generated with both training and stress effects, reflecting the practical reality that hard training can disrupt sleep while structured recovery tends to restore it.

Importantly, the dataset intentionally includes realistic missingness and device-wear behaviour. In a consumer context, the model must remain usable when signals such as HRV are intermittently absent; therefore, missingness and wear flags are treated as first-class properties of the data and are propagated into feature engineering and inference logic (StrideWise Development Team, 2025).

Injury causation and synthetic labelling

StrideWise labels injuries using an explicit causation layer rather than by post-hoc annotation. Each runner begins with a profile-dependent baseline daily hazard (e.g., novice 0.00624, recreational 0.00492, advanced 0.00115, elite 0.00138) that is then modified by interpretable drivers capturing the best-supported mechanisms in the running-injury literature: rapid workload increases, cumulative fatigue, insufficient recovery, and high mechanical stress exposure (StrideWise Development Team, 2025; van der Worp et al., 2015; Bertelsen et al., 2017).

The model combines multiplicative risk modifiers (for workload, fatigue, and recovery deficits) with additive "shock" risks for sprinting and long-run spikes. Workload risk is expressed through an acute-to-chronic workload ratio (ACWR) computed over 7-day versus 28-day windows, with sensitivity modulated by runner profile and fitness. Fatigue is represented as a persistent state with exponential decay, so that multiple hard days can accumulate risk even if any single day is not extreme. Recovery risk is driven by a deficit construct that blends sleep quality, HRV suppression, and RHR elevation. These design choices are aligned with how training stress and adaptation are commonly conceptualised, while remaining deliberately auditable and adjustable as clinical validation work progresses (Gabbett, 2016; StrideWise Development Team, 2025).

Labelling is forward-looking: the primary training label is injury_next_7d, which indicates whether an injury onset occurs within the subsequent seven days. This aligns the model objective to a practical intervention window, where the product can recommend reduced load before an athlete becomes injured (StrideWise Development Team, 2025).

Algorithmic risk calculation and worked example

The injury causation layer follows an explicit, auditable calculation that combines a profile-specific baseline hazard with multiplicative modifiers for training load, fatigue, recovery deficits, and individual susceptibility, then adds "absolute" shock risks for sprinting exposure and large long-run spikes. The probability is capped to avoid unrealistic extremes. This is implemented directly in code (see injury_causation_model.py) and is summarised below (StrideWise Development Team, 2025).

Daily injury probability (as implemented):

```
p_injury = base_hazard × (1 - resilience_factor)
          × load_risk_multiplier
          × fatigue_risk_multiplier
          × recovery_risk_multiplier
          × proneness_factor
          + sprinting_absolute_risk
          + spike_absolute_risk
```

The load multiplier is driven by acute-to-chronic workload ratio (ACWR) computed over 7-day versus 28-day windows. When ACWR exceeds a threshold (e.g., 0.95), the excess increases risk multiplicatively, scaled by a gain term (e.g., 1.2) and a profile-dependent sensitivity. Fatigue is represented as a persistent state that accumulates and decays according to “fatigue_t = 0.6 × fatigue_{t-1} + 0.4 × current_risk”, which captures the intuitive effect that multiple hard days raise risk even when any one day is not extreme. Recovery risk is derived from a simple deficit score that blends poor sleep, suppressed HRV, and elevated resting heart rate. Individual proneness acts as a global susceptibility multiplier, while resilience reduces the baseline hazard. Sprinting risk is additive and proportional to sprinting distance (eg sprinting_km × 0.1381, capped at 40%); long-run spikes add a discrete additive risk (small 4.33%, moderate 4.95%, large 6.19%). Finally, the overall probability is capped (typically 50%) to prevent implausible injury probabilities (StrideWise Development Team, 2025).

Worked example (advanced runner, high-risk day):

```
Base hazard: 0.00115 (0.115%)
Resilience: 0.8 → base = 0.00115 × 0.8 = 0.00092

Multiplicative factors:
- ACWR = 1.3, sensitivity = 1.0 → load_mult = 1.42
- Fatigue state = 0.6 → fatigue_mult = 1.36
- Recovery deficit = 0.4 → recovery_mult = 1.18
- Proneness = 0.7 → proneness_mult = 1.14

Combined multiplier: 1.42 × 1.36 × 1.18 × 1.14 = 2.60

Multiplicative risk: 0.00092 × 2.60 = 0.00239 (0.239%)

Absolute risks:
- Sprinting: 0.5 km × 0.1381 = 0.0691 (6.91%)
- Spike: Large spike = 0.06188 (6.19%)

Total: 0.00239 + 0.0691 + 0.06188 = 0.1334 (13.34%)
```

Injury generation mechanics and realistic patterns

Once the daily injury probability is computed, injury onset is generated by sampling from a Bernoulli draw. When an onset occurs, the system assigns a recovery duration (typically 7–42 days, with profile-dependent variation) and applies a 21-day exclusion window to prevent unrealistic clustering of injuries. This exclusion window is also enforced during the Lövdal-format conversion, ensuring that benchmark event construction remains consistent with the underlying synthetic injury episodes (StrideWise Development Team, 2025).

The resulting injury patterns are intentionally aligned to real-world intuition checks. Injuries should concentrate after sustained high-load periods, sprinting days should show elevated injury rates, and long-run spikes should act as visible risk events. Fatigue accumulation should create “late-stage” injuries after several hard sessions, while recovery weeks should reduce risk. The model also supports profile differences (for example, novices are more spike-sensitive, while elites can accumulate more sprinting exposure), which helps prevent the prediction engine from learning one-size-fits-all rules (StrideWise Development Team, 2025).

Feature engineering for model-ready training data

The prediction engine is trained on a rich user-day panel derived from the synthetic daily table and supported by roll-ups from activity records. Feature engineering emphasises time-aware representations of change: 7-day and 28-day rolling summaries, deltas and z-scores that measure short-term deviations from an athlete’s longer-term baseline, and engineered workload constructs including ACWR trajectory and the duration spent above risk thresholds (StrideWise Development Team, 2025).

To reflect the interaction between who the runner is and what they are doing, the pipeline also generates interaction terms (e.g., ACWR × fitness, proneness × workload, resilience × fatigue) as well as explicit state features such as recovery indices and persistent fatigue. The production pipeline yields 335+ features after removing constants and highly redundant predictors, balancing expressiveness with stability and overfitting control (StrideWise Development Team, 2025).

Model development and selection

StrideWise evaluated multiple machine learning approaches before selecting a production model. Gradient-boosted decision trees were prioritised because they handle heterogeneous feature types, non-linear interactions, and missingness patterns well, and they provide strong performance on tabular data without requiring deep feature normalisation (Chen & Guestrin, 2016).

In early experiments, XGBoost achieved promising discrimination but exhibited a large overfitting gap without careful regularisation. LightGBM controlled overfitting more effectively but underperformed on recall in this rare-event setting. Neural networks were explored but overfit heavily under the available feature regime (StrideWise Development Team, 2025). The selected production model is CatBoost, chosen for its robustness and its ability to maintain strong generalisation after tuning. In the documented configuration, the final CatBoost model achieved validation ROC-AUC 0.7136, PR-AUC 0.2427, and Brier score 0.0973 on 395,693 labelled user-days derived from 3,000 synthetic runners (StrideWise Development Team, 2025).

Training discipline is enforced through a forward-time split per athlete (preventing future leakage), early stopping, and calibration reporting. Because injury prediction is ultimately used to make a decision under uncertainty, calibration is treated as a first-class requirement alongside discrimination (Brier, 1950; StrideWise Development Team, 2025).

The results for our final CatBoost model show that it is definitely viable for both risk stratification and for decision support, at levels that are meaningful compared to baseline. The signal is clearly above chance, and demonstrates good discrimination, especially given that the prediction task is in an inherently noisy behavioural and physiological space.

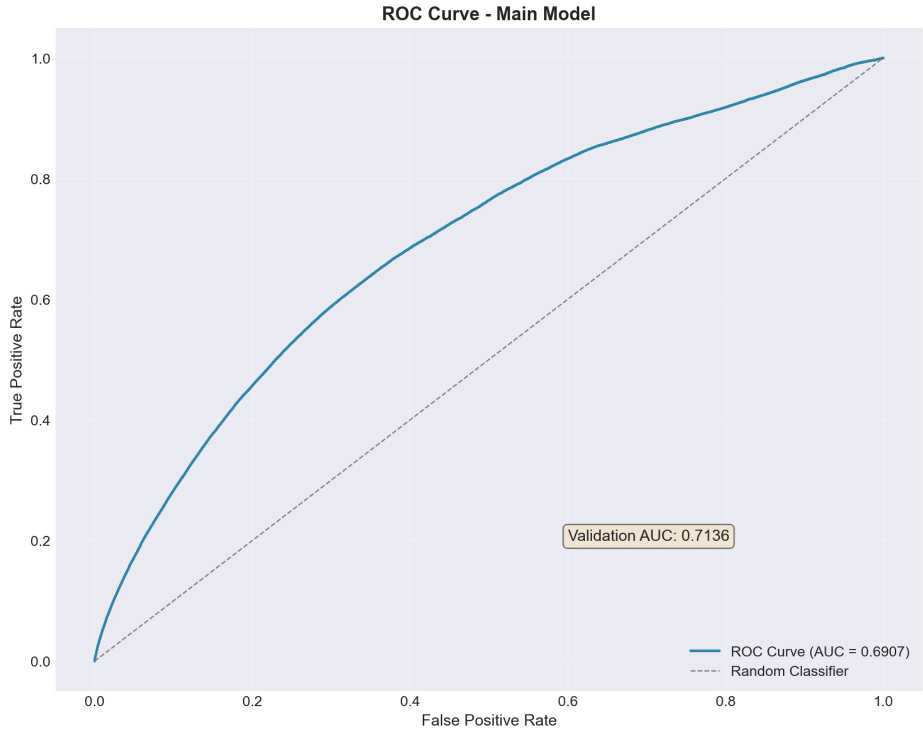

Figure 20. ROC AUC Graph for our main model (tested with cohort of 3,000 synthetic runners), which demonstrably confirms the model is identifying true injury risk (performance significantly above baseline)

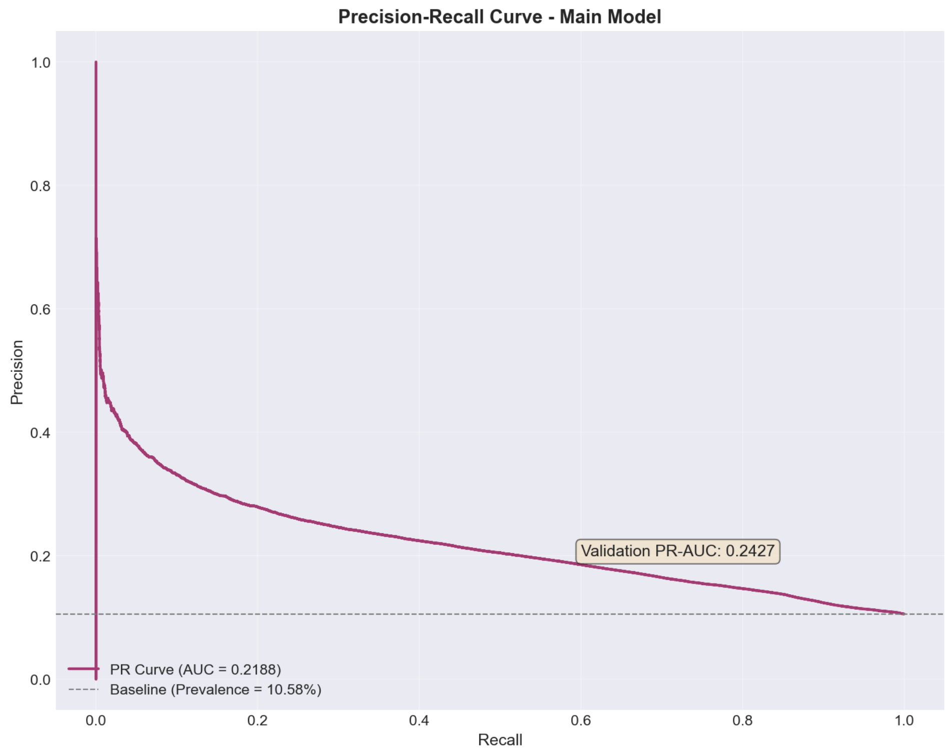

Figure 21. PR-AUC Graph for our main model (tested with cohort of 3,000 synthetic runners) demonstrating the model is good for triage – we have strong performance in the high risk zone (on the left), with usable performance as we widen coverage (i.e. a graceful decline)

Model exploration summary

Table 1 provides a concise summary of the model families evaluated and the rationale for the final selection (StrideWise Development Team, 2025).

Model family	Observed behaviour	Key issue(s)	Decision
XGBoost	Good raw discrimination; sensitive to tuning.	Large overfitting gap without strong regularisation.	Not selected.
LightGBM	Stable generalisation; fast training.	Lower overall performance in this setting.	Not selected.
Neural networks	Flexible; can model complex interactions.	Severe overfitting on tabular regime.	Not selected.
CatBoost	Best balance of discrimination and generalisation.	Requires tuning but robust once set.	Selected for production model.

Table 17. Summary of model families explored and the rationale for selecting CatBoost for production.

Standalone replication of the Lövdal et al model

To anchor evaluation in a published benchmark, we implemented a standalone replication of the "day approach" described by Lövdal et al. (2021). This method represents each candidate event as a 7-day history vector over a fixed set of features and trains a bagged ensemble of XGBoost models on bootstrap samples. The replication includes key procedural elements that materially affect performance, including per-athlete z-scoring using healthy baselines and event construction rules that enforce an exclusion window to avoid contaminating control examples (Lövdal et al., 2021; StrideWise Development Team, 2025).

In our implementation, the replicated model achieved AUC 0.7121 on the benchmark dataset compared to the published day-approach AUC of 0.724, providing evidence that the pipeline is faithful enough to be used as an external check on synthetic realism (Lövdal et al., 2021; StrideWise Development Team, 2025).

Format transformation and indirect validation of synthetic realism

Because the Lövdal et al benchmark is event-based and aggregated, it cannot be applied directly to the richer smartwatch-style synthetic tables. StrideWise therefore includes a deterministic conversion algorithm that transforms synthetic histories into the benchmark format by extracting injury onset events, sampling matched control events, and emitting a 7-day time series for each benchmark feature with the same suffixing and column ordering used in the published dataset (StrideWise Development Team, 2025).

This conversion necessarily discards physiology and other smartwatch-specific richness that is not represented in the benchmark. The goal is not to prove equivalence, but to test whether the key injury-relevant exposure signals (volume, intensity distribution, session structure) are sufficiently realistic that a benchmark model behaves similarly on converted synthetic data and on real data (StrideWise Development Team, 2025).

Using the standalone benchmark model, the converted synthetic dataset achieved ROC-AUC 0.6167 and PR-AUC 0.0216, with an injury prevalence of approximately 1.5%. Relative to the real benchmark performance, the system reports a performance ratio of 86.6%, alongside strong calibration alignment (ECE 0.0041) and supporting distribution and correlation checks (StrideWise Development Team, 2025). This provides conservative evidence that the synthetic generator encodes meaningful structure, increasing confidence that the main production model will behave sensibly when exposed to real rich wearable data, subject to future clinical validation. Realistically, we can state that our model is detecting a real signal in the data, but the strength of the signal is modest (likely due to the removal of many key features during the conversion process). We believe this demonstrates sufficient performance to move on to the next stage of development, involving clinical partners, and the subsequent use of real user data in place of the synthetic data. This will also allow us to iterate and improve the synthetic data creation algorithms as we identify more real word causal mechanisms with regard to the injury causation model, as well as to the correlations within the real word data itself (e.g between HRV and training intensity, etc.)

Web application MVP (functional proof of the end system)

The web-based MVP demonstrates the end-to-end system in a form that can be tested by non-technical stakeholders. Built with Flask, the application loads the trained model, generates (or accepts) a recent training history, computes features using the same pipeline as training, and returns a risk score mapped to a simple Green/Orange/Red band with guidance text (StrideWise Development Team, 2025).

For demonstration purposes, the MVP can generate a synthetic 90-day history and then allow a user to adjust their recent training preferences (e.g., sprinting, long runs, tempo exposure) and current health status (e.g., illness or pain flags). This makes it easy to validate the product logic—risk should rise when risk factors increase—without requiring access to any personal health data. The same MVP structure also provides a natural starting point for future app deployment, where real device data would be ingested instead of generated.

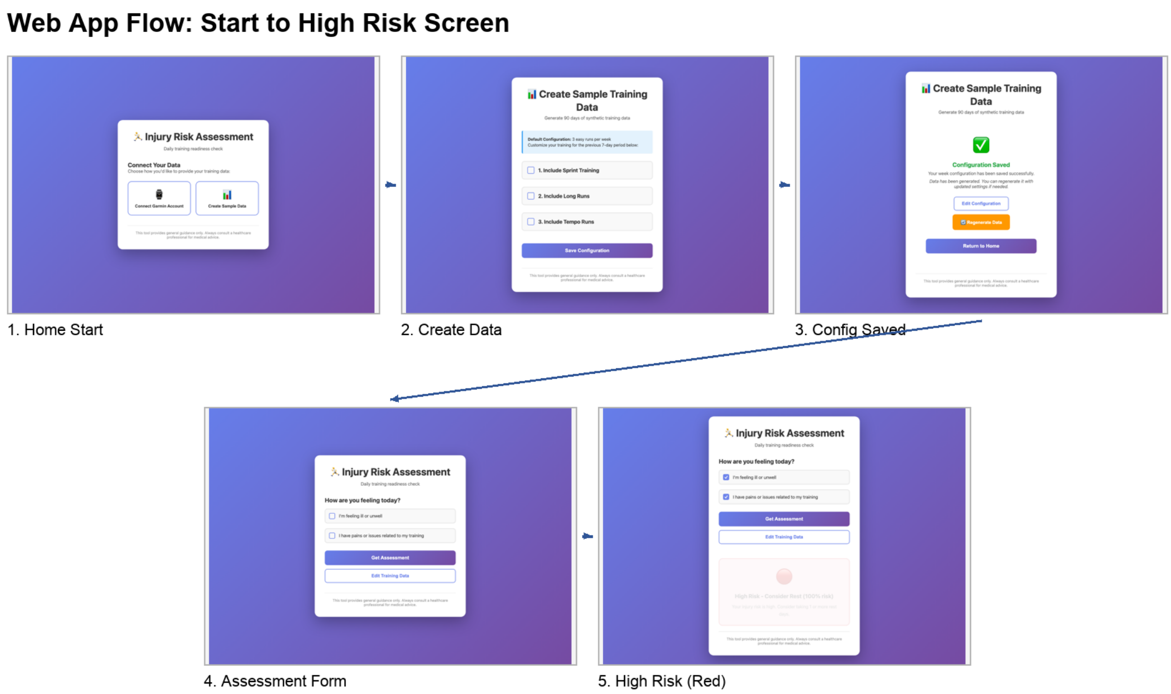

Figure 22. Screen flow for the WebApp

Codebase coverage and engineering artefacts

Beyond producing a working prediction model, StrideWise includes the engineering scaffolding needed to make results trustworthy and repeatable. The codebase therefore spans three categories: core pipeline code (generation, labelling, feature engineering, modelling, and inference), quality and audit code (sanity checks, leakage guards, benchmark alignment tests), and communication code (evaluation reports and visualisations). This breadth matters for an investor-facing proof of concept because it demonstrates that the system has been built as a productisable analytics engine rather than as a one-off notebook experiment (StrideWise Development Team, 2025).

In practice, each significant pipeline stage emits named artefacts that can be reviewed without reading raw tables. For example, the large synthetic dataset run produces dataset sanity summaries and trend files (e.g., large_dataset/sanity_report.json; large_dataset/sanity_label_rate_by_decile_daily.csv; large_dataset/sanity_injury_time_trend_weekly.csv), while model runs emit a standard pack of figures and metrics (ROC/PR curves, calibration plots, feature importance summaries, and threshold-policy reports) to "main_model_visualizations/". These outputs function as a lightweight "quality contract" that helps detect regressions quickly when generator parameters or feature definitions change (StrideWise Development Team, 2025).

Table 2 summarises the main code components and how they map to the overall architecture. File and folder names below reflect the structure described in the system documentation; the intent is to make the pipeline auditable by showing where each responsibility lives and which artefacts it produces.

Pipeline area	**Key modules / scripts (examples)**	**Purpose**	**Primary artefacts / outputs**
Synthetic runner generation	synthetic_runner_data_generator.py; generator config + profile logic	Create runners with heterogeneous profiles, training plans, physiology, and realistic missingness.	daily/user/activities tables; large_dataset/ exports; dataset-level sanity_report.json
Injury causation + labels	injury_causation_model.py (risk modifiers, episode logic)	Apply hazard-style drivers (load, fatigue, recovery deficits, sprinting, long-run spikes) to generate onset episodes and forward-looking labels.	injury_onset + injury_next_7d labels; onset timelines; sanity_injury_time_trend_weekly.csv
Feature engineering	feature_engineering_pipeline.py; rolling-window utilities	Compute 7/28-day aggregates, ACWR constructs, ramp/spike measures, recovery indices, and interaction terms; remove	Model-ready feature matrix; feature selection logs; schema snapshots

		constants/redundancy.	
Main model training	train_main_model.py (CatBoost); baseline model trainers (XGBoost/LightGBM)	Train and tune candidate ML models with forward-time splits; select production model balancing discrimination and calibration.	Trained model artifact; metrics summary; calibration and threshold policy reports
Model evaluation + visualisation	validation/create_main_model_visualizations.py; evaluation helpers	Generate consistent evaluation outputs: ROC/PR, calibration, feature importance, risk stratification plots, and summary tables.	main_model_visualizations/ figures; evaluation reports; threshold recommendations
Benchmark replication (Lövdal)	standalone_validation_model.py (bagged XGBoost); event construction helpers	Replicate the Lövdal day-approach model and validate against published benchmark performance.	Replication metrics; per-event predictions; benchmark evaluation report
Rich → Lövdal conversion	convert_synth_to_cc0.py; cc0_feature_schema.json	Deterministically translate smartwatch-style histories into the event-based Lövdal feature space (7-day suffix vectors, ordering, types).	Converted event table; schema validation logs; translated-field spot-check plots
Indirect validation harness	indirect_validation_runner.py; distribution/correlation comparators	Compare real vs translated-synthetic behaviour via distribution similarity, correlation preservation, and "performance ratio".	Alignment scorecard; KS/correlation summaries; performance ratio outputs
Web MVP (product proof)	flask_app/ (routes, templates); inference wrapper	Serve an end-to-end demo: generate/ingest history, compute features, score risk,	Working web UI; API responses; demo-friendly logs

		map to Green/Orange/Red guidance.	

Table 18. Codebase map showing how core pipeline code, quality controls, and reporting artefacts support an auditable architecture.

Proof-of-concept deliverables

Taken together, the proof of concept delivers: a production-ready synthetic runner generator with physiology and realistic missingness; an explicit injury causation and labelling layer that creates auditable associations between exposure and outcomes; a main prediction model trained on rich synthetic data achieving ROC-AUC 0.7136 with strong calibration; a replicated benchmark model validated against Lövdal et al. (2021); a schema-faithful conversion pipeline that enables indirect validation of synthetic realism (performance ratio 86.6%); and a web MVP that demonstrates the full product loop in an accessible format (StrideWise Development Team, 2025).

Evaluate Business Value: Cost Analysis

This section quantifies the business value of StrideWise in cost terms and sets out an investor-ready view of unit economics. It focuses on two questions: what is the economic cost of the problem we are trying to reduce (running-related injury), and what does it cost to deliver a scalable product that can meaningfully reduce that burden.

Cost of the problem and value pool

Running-related injuries (RRIs) are common enough to create a meaningful value pool, even when individual episodes are "small" in absolute monetary terms. In a large prospective cohort study of runners training for events, 46% of participants sustained an RRI during follow-up, and injured runners reported measurable limitations to daily activities, healthcare usage, and work absence (Visser et al., 2021).

Economic estimates vary by cohort definition and costing method, but multiple studies converge on a view that RRIs create both direct medical costs (e.g., physiotherapy and clinician visits) and indirect costs (primarily absenteeism and reduced productivity). Visser et al. estimate mean total costs of €74 per RRI in their cohort, while Hespanhol Junior et al. report higher estimates when separating direct and indirect components (€57.97 direct and €115.75 indirect per RRI, implying a total on the order of ~€174 per RRI) (Visser et al., 2021; Hespanhol Junior et al., 2016).

Study (population)	Direct cost estimate	Indirect cost estimate	Total cost estimate	Notes on interpretation
Visser et al. (2021) – runners training for events (Netherlands)	Included in total	Included in total	€74 per RRI (mean)	Large heterogeneous cohort; costs expressed as mean per RRI; work absenteeism reported in 5% of injured runners.
Hespanhol Junior et al. (2016) – runners training for an event	€57.97 per RRI (healthcare utilisation)	€115.75 per RRI (paid work absenteeism)	€173.72 per RRI (sum of means)	Prospective cohort; explicitly separates direct vs indirect; confidence intervals reported in paper.

Table 19. Evidence-based estimates of the direct and indirect economic burden of running-related injury.

StrideWise's value proposition can be expressed conservatively as an expected-value reduction in injury burden: Expected annual savings per subscriber ≈ (annual injury probability) × (injury reduction attributable to the product) × (economic cost per injury). In practice, the outcome we seek is not only fewer injuries, but fewer "avoidable" injuries driven by modifiable load, recovery, and progression factors—precisely the domains most amenable to data-driven alerting.

Annual injury probability	Injury reduction	Cost per injury (€)	Expected saving per subscriber / year (€)
40%	20%	74	5.92
40%	25%	74	7.40
40%	30%	74	8.88
40%	20%	174	13.92
40%	25%	174	17.40
40%	30%	174	20.88

Table 20. Illustrative expected-value savings per subscriber (40% annual injury probability; sensitivity to injury reduction and cost per injury).

These expected savings do not capture important non-monetary value drivers such as training continuity, race participation, wellbeing, and confidence. Indeed, for most runners, these indirect drivers carry a far more significant utility than the purely economic and notional savings (Russell, H.C., 2015). Nor do the saving capture the likely negative downstream health consequences from runners who cease the activity entirely post injury, as is common in both novice and recreational runners (Nelson, E.O., 2019). The savings noted also exclude downstream B2B2C value where a coach, club, or insurer can scale the benefit across a cohort and treat injury reduction as a measurable risk-management KPI.

Consumer pricing anchors and willingness to pay

The endurance training ecosystem provides clear reference points for subscription pricing. Leading platforms charge for analytics, planning, and health insights that remain complex for many recreational users. StrideWise is positioned differently—reducing complexity to a single interpretable signal—but it competes in the same ‘pay for insight’ budget line.

Product	Published pricing (examples)	What the user is paying for
Strava Subscription	US $11.99/month or $79.99/year (Strava pricing page)	Training analytics + athlete intelligence features; social platform premium tools.
TrainingPeaks Premium	$19.95/month or $134.99/year (TrainingPeaks pricing)	Structured planning, performance charts, and coaching workflow tools.
Oura Membership (app)	$5.99/month or $69.99/year (Oura membership)	Recovery and sleep insights tied to sensor data; ongoing analytics.
Garmin Connect+ (premium tier)	$6.99/month or $69.99/year (reported at launch)	AI insights and enhanced dashboards layered on top of free Garmin Connect.
Runalyze Premium	€5.50/month or €60/year (Runalyze pricing)	Advanced endurance analytics and reporting for serious athletes.
Intervals.icu supporter	Suggested $4/month (Intervals.icu homepage)	Support tier to sustain hosting + ongoing development for analytics/planning.

Table 21. Subscription pricing anchors in adjacent training and recovery products.

Taken together, these anchors suggest a well-established consumer willingness to pay in the ~€5–€20/month range for performance and recovery insight. A StrideWise subscription can therefore be positioned as a 'simplicity premium': the user is not paying for more charts, but for fewer decisions and a lower risk of preventable training interruption.

Delivery cost structure and gross margin considerations

From a cost perspective, StrideWise benefits from a software-first delivery model with a small marginal cost per additional user. The main fixed cost drivers are product engineering, model iteration, and the clinical validation and pilot studies required to substantiate injury-risk claims. Variable costs at scale are primarily distribution fees (App Store / Google Play), as well as compute and storage costs if inference is performed in the cloud.

For mobile subscriptions, platform fees are material. Apple's App Store Small Business Program reduces commission to 15% for qualifying developers, with standard rates applying above the threshold (Apple Developer, n.d.). Google Play publishes a similar 15% tier for the first $1M of annual revenue for eligible developers, with higher rates above that (Google Play Console Help, n.d.). These fees are a key input to net revenue per subscriber when subscriptions are sold through app stores, and they create a strong incentive to test alternative payment and fulfilment strategies where policy permits.

From a privacy and cost standpoint, an additional strategic lever is the deployment architecture. As StrideWise moves from a web MVP to a native mobile product, we will assess whether risk scoring can be executed fully on device, which would reduce cloud inference costs and limit the need to transmit health data off device. Where cloud components remain necessary (for example, cohort analytics, model monitoring, or optional cross-device sync), data will be processed under strict security controls appropriate to health-adjacent data.

Cost category (illustrative)	What it covers	Planning notes	Indicative cost weight
Product engineering (web → mobile)	Native app build, onboarding, integrations, UX iterations, analytics instrumentation	Front-loaded in Year 1; becomes incremental enhancement post-launch.	High (Year 1–2)
Data science and MLOps	Feature evolution, training runs, calibration/threshold policy, drift monitoring, evaluation reports	Ongoing; cost decreases per user with scale but remains a core competency.	Medium–High (ongoing)
Clinical validation partnership	Protocol design, ethics review support, recruitment coordination, statistical analysis, publication-quality reporting	Planned as a gated milestone before broader health-adjacent distribution.	Medium (milestone-gated)
Pilot cohort operations	Recruitment, consent, data collection tooling,	Cost scales with cohort size; early pilots can be small	Medium (pilot-dependent)

	support, retention communications	but ‘rich’ (high signal per participant).	
Cloud and tooling	Hosting, storage, CI/CD, monitoring, error tracking, dashboards	Highly elastic; can be minimized through on-device inference where feasible.	Low–Medium (elastic)
Customer support and compliance	Support workflows, data governance, DPIAs where applicable, security reviews	Increases with scale and with health-adjacent positioning; necessary for trust.	Medium (scales with users)

Table 22. Indicative cost categories for scaling from MVP to validated consumer product (planning view).

The most investor-relevant implication of this cost model is that the pathway to a durable defensible business is not dominated by commodity cloud spend; it is dominated by credibility spend (validation), product quality spend (native experience), and disciplined measurement (retention, injury outcomes, and model calibration). This aligns incentives: the work that reduces risk and increases trust also increases lifetime value and lowers churn.

Financial projections and capital requirements

We model three illustrative scenarios (Low, Base, High) over five years. Marketing spend is linked to subscriber growth via a blended acquisition cost (CAC) that combines paid acquisition and partner/community channels. Consistent with a UK-first launch, the marketing floor is intentionally modest in Year 1 and increases as the product expands across Europe and into the US. In the Base and High scenarios, we reach cumulative break-even in Year 4, reflecting the effect of maturing conversion/retention, and pulling forward a validation-led B2B2C motion that produces earlier cash inflows.

Validation-led go-to-market: B2B2C bridge

StrideWise is not positioned as a consumer subscription alone. Instead, we run a validation-led distribution strategy: we partner early with universities, clinics, coaches and clubs to validate the system in real cohorts, to acquire early users through trusted channels, as well as to generate paid pilot and licensing revenues that offset burn while the consumer product matures. This creates a flywheel—evidence builds trust, trust unlocks partners, partners reduce blended CAC—allowing a later, more efficient US launch. We explicitly account for incremental delivery and partner-enablement effort by applying a 15% cost allowance to incremental B2B2C revenue.

Assumption	Value / rationale
B2C pricing (planning)	€59/year blended (anchored to adjacent endurance subscriptions; see Table 3).
Free usage period	Two months free for all new users; model assumes net-new cohorts contribute circa 40% of annualised revenue in the acquisition year (conservative blend of trial + phased acquisition).
Platform fees	15% of subscription revenue (planning assumption), consistent with Apple’s Small Business Program for eligible developers (Apple Developer, n.d.).
Variable infra & support	3% of subscription revenue (cloud inference, telemetry, support; planning assumption).

Paid acquisition benchmark	Health & fitness eCPI benchmarks used as planning anchors for Europe and North America (Adjust, 2024).
Conversion calibration	Install→paid is a planning parameter; we anchor expectations to published Health & Fitness subscription conversion patterns (median and top-decile trial→paid rates) and refine using pilot funnel data (RevenueCat, 2025).
Partner/community economics	Partner acquisition modelled as a commission on first-year revenue (22% base). This sits within commonly reported SaaS affiliate ranges (e.g., 15–25% or 20–30% depending on program maturity) (Post Affiliate Pro, 2025; Rewardful, 2023).
B2B2C motion	Early paid pilots and cohort licenses (university validation, coaches/clubs, health-adjacent partners) are modelled explicitly and linked to the deal bridge in Tables 4B–4C.
Incremental B2B2C delivery cost	15% allowance on incremental B2B2C revenue (partner enablement, cohort reporting, integration support).
Exit multiple (illustrative)	Sensitivity shown at 3×–7× ARR, consistent with published commentary and indices used to contextualise private SaaS valuation ranges (SaaS Capital, 2025; Bessemer Venture Partners, n.d.).

Table 23. Key financial modelling assumptions (UK-first launch and validation-led B2B2C).

Package	Typical partner	Cohort size	Term	Pricing (planning)	What's delivered
Package A — University/clinical validation pilot	Sports science dept / clinic / lab	150–400 athletes	12–20 weeks	€40k–€100k pilot fee	Cohort onboarding, data governance, outcome tracking, interim + final report; supports publishable validation outputs.
Package B — Coach/club cohort license (annual)	Coaching org / club / training group	200–2,000 athletes	12 months	€5–€12 per athlete/year (≈€10k–€150k ACV)	Risk signal for athletes + cohort dashboard, admin tools, and partner reporting; reduces manual monitoring burden.
Package C — Health-adjacent feasibility pilot	Insurer innovation / wellness / MSK platform	500–3,000 athletes	3–6 months	€150k–€400k pilot fee	Feasibility + outcomes evaluation (engagement, risk reduction proxies), governance, and decision support for scale-up.

Table 24. B2B2C offer packaging and indicative pricing (planning ranges).

Year	Illustrative deal mix (planning)	Incremental revenue
Year 2	2× Package A (~€75k), 3× Package B (~€40k), 1× small feasibility add-on (~€30k)	€300k
Year 3	2× Package A (~€80k), 6× Package B (~€50k), 1× Package C feasibility (~€140k)	€600k
Year 4	Renew 6× Package B (~€50k), add 4× Package B (~€60k), 2× Package A (~€80k), 1× Package C expansion (~€200k)	€900k
Year 5	Renew 10× Package B (~€60k), 1× Package C rollout (~€300k)	€900k

Table 25. Deal-based bridge from pilots to annual licensing (Base scenario, incremental B2B2C revenue).

Phase	Share of net-new	eCPI (€)	Install→paid (rate)	Overhead factor	Paid share of net-new	Partner commission	Paid CAC (€/payer)	Partner cost (€/payer)	Marketing floor (€000)
Y1 (UK launch)	100%	0.82	3.2%	1.8×	55%	22%	46	13	60
Y2 H1 (UK focus)	45%	0.82	3.6%	1.8×	55%	22%	41	13	160
Y2 H2 (UK+EU + US prelaunch)	55%	0.90	3.9%	1.9×	60%	22%	44	13	160
Y3 (US launch)	100%	1.19	3.6%	2.0×	65%	22%	66	13	300
Y4 (Scale & optimise)	100%	1.10	4.5%	1.9×	65%	22%	46	13	400
Y5 (Scale & optimise)	100%	1.00	4.6%	1.8×	60%	22%	39	13	450

Table 26. Marketing and channel-mix assumptions by phase (used to convert growth into marketing spend).

Notes: eCPI benchmarks are anchored to published Health & Fitness benchmarks by region (Adjust, 2024). Trial→paid patterns in Health & Fitness are comparatively strong, but we treat install→paid as a planning parameter to be refined using pilot funnel data (RevenueCat, 2025). Partner commission is modelled as a first-year revenue share (Post Affiliate Pro, 2025; Rewardful, 2023).

Scenario projections (Years 1–5)

Tables 27,28 & 29 summarise cashflows for each scenario. Total costs include marketing (derived from growth), other operating expenses, variable platform/infra costs applied to subscription revenue, and an explicit allowance for incremental B2B2C delivery and partner enablement.

Year	Paying subs (end)	Revenue (€000)	Marketing (€000)	Other OPEX (€000)	Platform + infra (€000)	B2B delivery / sales (€000)	Total costs (€000)	Net cashflow (€000)	Cumulative (€000)
1	1500	35	60	540	6	0	606	-571	-571
2	7500	400	181	630	41	22	874	-474	-1,045
3	20000	1,088	594	765	133	45	1,537	-449	-1,495
4	35000	2,014	521	900	276	60	1,757	257	-1,238
5	50000	3,039	450	1,035	435	75	1,995	1,044	-194

Table 27. Low scenario cashflow summary (€, in thousands).

Year	Paying subs (end)	Revenue (€000)	Marketing (€000)	Other OPEX (€000)	Platform + infra (€000)	B2B delivery / sales (€000)	Total costs (€000)	Net cashflow (€000)	Cumulative (€000)
1	2000	57	62	600	8	0	671	-614	-614
2	12000	704	301	700	64	45	1,110	-406	-1,019
3	35000	2,001	1,093	850	225	90	2,258	-257	-1,277
4	60000	3,805	868	1,000	478	135	2,481	1,324	47
5	90000	5,498	860	1,150	765	135	2,910	2,588	2,635

Table 28. Base scenario cashflow summary (€, in thousands).

Year	Paying subs (end)	Revenue (€000)	Marketing (€000)	Other OPEX (€000)	Platform + infra (€000)	B2B delivery / sales (€000)	Total costs (€000)	Net cashflow (€000)	Cumulative (€000)
1	3500	108	109	672	15	0	796	-688	-688
2	22000	1,193	557	784	116	68	1,524	-331	-1,020
3	60000	3,395	1,806	952	395	135	3,288	107	-912
4	110000	6,420	1,737	1,120	850	180	3,886	2,534	1,621
5	170000	10,106	1,720	1,288	1,423	225	4,656	5,450	7,071

Table 29. High scenario cashflow summary (€, in thousands).

Table 30 reports peak funding need (maximum cumulative cash burn) and the year of cumulative break-even. Under the Base case, cumulative break-even occurs in Year 4; the Low case remains slightly loss-making at Year 5, highlighting sensitivity to adoption and partner conversion.

Scenario	Peak funding need (€000)	Break-even (cumulative)
Low	1,495	Beyond Year 5
Base	1,277	Year 4
High	1,020	Year 4

Table 30. Break-even and capital requirement summary (illustrative scenarios).

Return profile is typically evaluated using scenario-weighted cashflows and an exit assumption. Table 7 provides an illustrative IRR sensitivity for the Base case using an ARR multiple at Year 5; these figures are indicative only and exclude dilution, taxes, and execution risk (SaaS Capital, 2025; Damodaran, 2009; Bessemer Venture Partners, n.d.).

Exit multiple (× ARR)	Implied IRR (annual)
3	68%
5	85%
7	97%

Table 40. Illustrative IRR sensitivity (Base case) under exit valuation multiples.

Summary: business value versus delivery cost

In summary, published evidence indicates that running-related injuries produce measurable direct and indirect costs per episode, while the endurance analytics market demonstrates established consumer willingness to pay for insight. StrideWise's cost base is typical of a modern software subscription business, with material attention required for clinical validation and privacy-forward engineering. The resulting unit economics can be attractive because marginal delivery costs per additional subscriber are low, and because the product's core promise—preventing interruptions to training—directly supports retention.

Scale up the POC: Recommendations

StrideWise has already demonstrated the core end-to-end capability: a synthetic data generator, an evidence-informed injury causation and labeling layer, a trained prediction model, and a functional web MVP that translates complex wearable signals into a single daily traffic-light decision for runners. Scaling the proof of concept is therefore less about building ‘more charts’ and more about building a repeatable system that earns trust, learns from real-world cohorts, and grows sustainably, starting in one market (the UK) and expanding only as evidence and channel efficiency mature.

What “scalable” means for StrideWise

For StrideWise, scalable does not only mean handling more users. It means being able to move from a build-grade demonstration to a product that can be deployed to tens of thousands of runners while remaining operationally predictable, scientifically credible, commercially sustainable, and privacy-first. Operational scalability requires reliable ingestion of daily time-series data with controlled cloud cost and clear observability. Scientific scalability means performance remains valid as the product meets new runner segments, devices, and training contexts, with monitoring and recalibration as standard practice. Commercial scalability means growth does not depend on permanently loss-making acquisition, and the product experience avoids ‘alert fatigue’ by balancing sensitivity with trust. Governance scalability means consent, data minimisation and secure processing are built in from the start, recognising that health-adjacent data is often treated as special category personal data under GDPR/UK GDPR and therefore carries higher obligations.(Regulation (EU) 2016/679, 2016; UK Information Commissioner’s Office, 2024; Data Protection Commission, n.d.)

Recommended scale-up path aligned to the base case (UK-first, validation-led, B2B2C-enabled)

The recommended route is a staged plan that deliberately links growth to evidence. Year 1 focuses on a UK launch to reduce operational complexity and concentrate learning. In parallel, StrideWise partners with an expert group—ideally a university or clinical research collaborator—to run a structured validation programme. This partner-led work has two benefits: it accelerates clinical credibility and it creates an early B2B2C revenue line through paid pilots and cohort licences, which reduces capital intensity and supports the Year-4 break-even objective.

From mid-Year 2, the product expands beyond the UK into additional European markets through the same mechanisms—community channels, coaches/clubs, and validated cohorts—before a larger US launch in late Year 2 / Year 3. This sequencing is intentional: it avoids the common pattern of heavy paid acquisition before the product has accumulated enough trust, retention evidence, and calibration data. It also supports a blended acquisition strategy in which partner and community channels reduce dependence on paid user acquisition during the most expensive growth phase.

Clinical validation and ‘safe claims’ (a prerequisite for durable scale)

Because StrideWise outputs a risk signal related to injury, claims must be approached conservatively. Early scale-up should explicitly separate the wellness decision-support experience shipped to consumers and cohorts and the clinical validation workstream that evaluates outcomes rigorously. If, over time, StrideWise is positioned in a way that could be interpreted as a medical device, it is important to assess classification and regulatory obligations early, using appropriate UK guidance for software applications and medical device regulation.(Medicines and Healthcare products Regulatory Agency, n.d.-a; Medicines and Healthcare products Regulatory Agency, n.d.-b; Medicines and Healthcare products Regulatory Agency, 2024)

Validation activities should be designed and reported transparently. Where the system is evaluated as a predictive model, reporting should align with modern AI-aware guidance for prediction models (TRIPOD+AI) (Collins et al., 2024). Where an intervention trial is run (e.g., 'does the traffic-light signal reduce injuries or time-loss days?'), trial reporting and protocols should follow CONSORT-AI and SPIRIT-AI extensions (Liu et al., 2020; Rivera et al., 2020) In practice, this means agreeing definitions (injury/time-loss proxies), defining outcomes and follow-up windows, documenting cohort inclusion/exclusion, and pre-specifying how the model will be monitored and updated during the study.

What needs to happen to scale

Data scale-up should move from synthetic-first development to a controlled real-world learning loop. The first goal is not 'every integration', but reliable ingestion from one or two high-coverage sources, with clear handling of missingness and device heterogeneity. In early pilots, injury outcomes are often best captured via practical proxies such as time-loss days, training disruption weeks, or pain that prevents a planned session, with definitions standardised across cohorts.

Model scale-up should evolve from a single trained model artefact to a product-grade model system. This includes versioning, calibration tracking, drift monitoring, and explicit alert-policy controls (target alert rate, cool-down rules, and confidence indicators when data coverage is poor). Explanations should be stable and user-oriented (e.g., 'load spike', 'stacked intensity', 'poor recovery') rather than brittle feature lists, so that the experience remains trustworthy as the system is updated.

Product scale-up should be anchored in trust and retention. The primary interface must remain decision-first, with clear 'smallest adjustment' actions that reduce risk without pushing users toward extreme behaviour. A lightweight feedback loop ("was this helpful?" / "did you change training?" / "did symptoms emerge?") enables rapid learning while keeping user burden low.

Infrastructure scale-up should focus on reliability and cost control: a clean separation between ingestion, storage, feature computation, inference, and reporting; automated sanity checks; and monitoring of pipeline failures, alert rates, and data coverage. Governance scale-up should codify privacy-by-design, including explicit consent, data minimisation, and clear deletion/export controls. Health-adjacent data is typically treated as special category data under GDPR and UK GDPR, so the validation phase should adopt secure processing practices appropriate to that sensitivity.(Regulation (EU) 2016/679, 2016; UK Information Commissioner's Office, 2024; Data Protection Commission, n.d.)

Concrete roadmap (UK-first scale-up over 36 weeks)

The roadmap below keeps scope tight: deliver a production-ready mobile experience for the UK, run a partner-led validation pilot, and establish the B2B2C wedge for efficient acquisition and revenue.

Workstream	Weeks 1–12	Weeks 13–24	Weeks 13–36	Success signals
Product (UK)	Ship native app MVP from existing web design; onboarding within 24–48h to first signal.	Pilot-driven UX iteration; explanation templates + 'smallest adjustment' actions.	Release UK v1; refine retention and trust metrics; prepare EU localisation needs.	Activation ≥ target; week-4 retention improving; alert trust score trending up.
Clinical validation	Confirm partner; agree protocol, outcomes, consent; define injury/time-loss proxy.	Run 8–12 week cohort; collect outcomes + feedback; pre-specified evaluation.	Interim results + publishable reporting package; plan expanded cohorts.	Evidence pack produced; model calibration by segment; claims language validated.
Data & integrations	Implement one high-coverage ingestion path;	Stabilise daily pipeline; add QA/sanity	Expand ingestion only if needed; maintain	Coverage KPIs stable; low pipeline failure rate; reproducible histories.

	missingness flags as features.	artefacts and failure monitoring.	strict data minimisation.	
Model & MLOps	Model registry + versioning; baseline drift and calibration dashboards.	Threshold policy tuned to pilot alert rates; segment calibration begins.	Controlled retraining loop; governance for updates during studies.	Alert rate within policy; calibration stable; drift detected and handled.
GTM (B2B2C)	Package pilots (A/B/C) and materials; identify first UK cohorts (clubs/coaches).	Convert pilots to paid; define renewal terms; build basic cohort admin tooling.	Scale partners; prepare EU partner pipeline; plan US timing post-evidence.	Paid pilot conversion rate; renewals pipeline; blended CAC improving.

Table 41. Roadmap to initial full launch

Privacy and security considerations as scale accelerators (not constraints)

During validation and early cohorts, StrideWise should treat data governance as part of product quality: clear consent, documented lawful basis, minimal retention, and secure storage and access controls. Because health-adjacent information is typically classified as special category data, privacy-by-design reduces regulatory risk and improves partner confidence.(Regulation (EU) 2016/679, 2016; UK Information Commissioner's Office, 2024; Data Protection Commission, n.d.) As the consumer footprint grows, StrideWise should also evaluate whether some or all inference can be performed on-device to minimise the need to transmit sensitive signals off the user's phone. Even if cloud inference remains necessary for parts of the pipeline, designing for data minimisation and secure processing strengthens defensibility and reduces friction in partnership discussions.

In summary, the scale-up plan is intentionally staged: UK-first deployment to accelerate learning, partner-led clinical validation to establish credibility, and an early B2B2C wedge to reduce acquisition cost and fund growth. This approach aligns with the financial base case that targets cumulative break-even by Year 4, while building a repeatable evidence and distribution engine that can support later EU and US expansion.

Conclusion

StrideWise exists to solve a clear and persistent data analytics problem: runners are surrounded by complex dashboards and noisy signals, yet still lack a simple, reliable answer to the question that matters most, am I about to get injured if I keep training like this? Our approach is to translate rich wearable and training data into an extraordinarily simple outcome for the user: a clear injury-risk signal and practical guidance, without requiring the runner to interpret physiology, pseudo-scientific terminology, or dense performance charts.

We have already proven the feasibility of this vision by delivering an end-to-end working MVP. That MVP is rooted in a full analytics pipeline: synthetic data engineering to generate realistic longitudinal training and recovery records at scale; a research-informed injury causation engine to create meaningful labels; feature engineering to translate behaviour and physiology into model-ready predictors; and a machine learning prediction model that outputs actionable risk. We then built an independent validation pathway by implementing a comparable modelling approach to the Lövdal et al. dataset and creating a transformation process that allows our richer synthetic data to be expressed in the same structure, enabling direct evaluation and sanity checks against established real-world formats.

The next phase is structured to convert this technical foundation into market traction while strengthening credibility. We will launch UK-first to focus execution, partner with an expert group (ideally a university) to clinically validate both the synthetic data approach and the injury causation logic, and run a small, informed pilot to confirm that real-world signals align with our synthetic distributions. In parallel, we will take the MVP from web to app, and scale via a blended go-to-market motion: a consumer subscription as the core product, complemented by a B2B2C wedge through coaches, clubs, universities and health-adjacent partners. This not only accelerates distribution through trusted channels, but also creates an evidence-driven flywheel in which validation builds trust, trust unlocks partnerships, and partnerships reduce acquisition costs as we expand across Europe and into the US.

In short, StrideWise is a data-driven company with a functioning product and a credible plan to scale. The MVP demonstrates that we can take a complex analytics challenge, from data generation and engineering through modelling and validation, and turn it into a simple, valuable, user-facing experience. With validation, structured pilots, and a staged expansion strategy, StrideWise is well positioned to build a defensible injury-risk platform that improves training continuity for runners and creates a durable subscription and licensing business.

References

Adams, T., Birkenbihl, C., Otte, K., Ng, H. G., Rieling, J. A., Näher, A.-F., Sax, U., Prasser, F., & Fröhlich, H. (2025). On the fidelity versus privacy and utility trade-off of synthetic patient data. iScience, 28(5), 112382. https://doi.org/10.1016/j.isci.2025.112382

Adjust. (2024, July 30). Insights into what makes a good mobile app eCPI. Adjust. https://www.adjust.com/blog/ecpi-benchmarks/

Agile Alliance. (2001). Manifesto for Agile Software Development. https://agilemanifesto.org/

Apple Developer. (n.d.). App Store Small Business Program. Apple Developer. https://developer.apple.com/app-store/small-business-program/

Apple. (n.d.). Apple One (Ireland) – Apple Fitness+ monthly pricing. Apple. https://www.apple.com/ie/apple-one/

Banister, E. W. (1991). Modeling elite athletic performance. In J. D. MacDougall, H. A. Wenger, & H. J. Green (Eds.), Physiological Testing of the High-Performance Athlete (2nd ed.). Human Kinetics. Available via Google Books search: https://books.google.com/books?q=Physiological%20Testing%20of%20the%20High-Performance%20Athlete%202nd%20edition%20Human%20Kinetics%20Banister%201991

Bawden, D., & Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. Journal of Information Science, 35(2), 180–191. https://doi.org/10.1177/0165551508095781

Bertelsen, M. L., Hulme, A., Petersen, J., Brund, R. K., Sørensen, H., Finch, C. F., & Nielsen, R. O. (2017). A framework for the etiology of running-related injuries. Scandinavian Journal of Medicine & Science in Sports, 27(11), 1170–1180. https://doi.org/10.1111/sms.12883

Bessemer Venture Partners. (n.d.). BVP Nasdaq Emerging Cloud Index. https://cloudindex.bvp.com/

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Camma Damsted, C., Glad, S., Nielsen, R. Ø., & Sørensen, H. (2018). Is there evidence for an association between changes in training load and running-related injuries? A systematic review. International Journal of Sports Physical Therapy, 13(6), 931–942. PubMed search: https://pubmed.ncbi.nlm.nih.gov/?term=Is%20there%20evidence%20for%20an%20association%20between%20changes%20in%20training%20load%20and%20running-related%20injuries%20A%20systematic%20review%20Damsted%202018

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP–DM 1.0: Step-by-step data mining guide. SPSS. IBM SPSS Modeler documentation (CRISP-DM overview): https://www.ibm.com/docs/en/spss-modeler

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Collins, G. S., Moons, K. G. M., et al. (2024). TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use artificial intelligence. BMJ, 385, e078378. https://doi.org/10.1136/bmj-2023-078378

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Annals of Internal Medicine, 162(1), 55–63. https://doi.org/10.7326/M14-0697

Damodaran, A. (2009). Valuing young, start-up and growth companies: Estimation issues and valuation challenges. NYU Stern School of Business. https://pages.stern.nyu.edu/~adamodar/pdfiles/papers/younggrowth.pdf

Data Protection Commission (Ireland). (2025). Special Category Data (GDPR Article 9). https://www.dataprotection.ie/en/organisations/know-your-obligations/lawful-processing/special-category-data

Düking, P., Zinner, C., Trabelsi, K., Reed, J. L., Holmberg, H.-C., Kunz, P., & Sperlich, B. (2021). Monitoring and adapting endurance training on the basis of heart rate variability monitored by wearable technologies: A systematic review with meta-analysis. Journal of Science and Medicine in Sport, 24(11), 1180–1192. https://doi.org/10.1016/j.jsams.2021.04.012

EDPB. (2020). Guidelines 4/2019 on Article 25 Data Protection by Design and by Default (Version 2.0). European Data Protection Board. https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en

EDPB. (2019/2020). Guidelines 4/2019 on Article 25 Data Protection by Design and by Default (Final). European Data Protection Board. https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf

Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. The Information Society, 20(5), 325–344. https://doi.org/10.1080/01972240490507974

European Data Protection Board. (2025). Guidelines 01/2025 on Pseudonymisation. https://www.edpb.europa.eu/system/files/2025-01/edpb_guidelines_202501_pseudonymisation_en.pdf

European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation), Article 9 (processing of special categories of personal data). EUR-Lex. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

Fortune Business Insights. (2025, March 19). mHealth Apps Market to Grow at a CAGR of 11.8% by 2032. https://www.fortunebusinessinsights.com/press-release/mhealth-apps-market-9540

Frandsen, J. S. B., Hulme, A., Parner, E. T., Møller, M., Lindman, I., Abrahamson, J., Simonsen, N. S., Jacobsen, J. S., Ramskov, D., Skejø, S., Malisoux, L., Bertelsen, M. L., & Nielsen, R. O. (2025). How much running is too much? Identifying high-risk running sessions in a 5200-person cohort study. British Journal of Sports Medicine, 59(17), 1203–1210. https://doi.org/10.1136/bjsports-2024-109380

Gabbett, T. J. (2016). The training—injury prevention paradox: Should athletes be training smarter and harder? British Journal of Sports Medicine, 50(5), 273–280. https://doi.org/10.1136/bjsports-2015-095788

Garmin. (n.d.). Training Readiness (Owner's Manual). Garmin. https://www8.garmin.com/manuals/webhelp/GUID-626B0002-B026-4032-B7A5-7321047B8300/EN-US/GUID-C21BE0C8-A08E-4DA1-B6C6-2E0E2DDDB372.html

Gonçalves, A. R., Ray, P., Soper, B., Stevens, J. L., Coyle, L., Sales, A. P., et al. (2020). Generation and evaluation of synthetic patient data. BMC Medical Research Methodology, 20, 108. https://doi.org/10.1186/s12874-020-00977-1

Google Play Console Help. (n.d.). Service fees. Google Support. https://support.google.com/googleplay/android-developer/answer/112622

Grand View Research. (2025). Fitness App Market Size & Trends (2025–2033). https://www.grandviewresearch.com/industry-analysis/fitness-app-market

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. arXiv:1706.04599. https://arxiv.org/abs/1706.04599

Hespanhol Junior, L. C., Huisstede, B. M. A., Smits, D.-W., Kluitenberg, B., van der Worp, H., van Middelkoop, M., & Verhagen, E. A. L. M. (2016). The NLstart2run study: Economic burden of running-related injuries in novice runners participating in a novice running program. Journal of Science and Medicine in Sport, 19(10), 800–804. https://pubmed.ncbi.nlm.nih.gov/?term=NLstart2run%20study%20economic%20burden%20running-related%20injuries%20novice%20runners%202016%20Journal%20of%20Science%20and%20Medicine%20in%20Sport

Hespanhol Junior, L. C., van Mechelen, W., & Verhagen, E. (2016). The economic burden of running-related injuries in novice runners: A prospective cohort study. Journal of Science and Medicine in Sport, 19(10), 800–804. PubMed search: https://pubmed.ncbi.nlm.nih.gov/?term=economic%20burden%20of%20running-related%20injuries%20in%20novice%20runners%20Hespanhol%202016%20Journal%20of%20Science%20and%20Medicine%20in%20Sport

Hottenrott, L., Hoos, O., & Esperer, H.-D. (2021). Utilizing heart rate variability for coaching athletes during and after viral infection: A case report in an elite endurance athlete. Frontiers in Sports and Active Living, 3, 612782. https://doi.org/10.3389/fspor.2021.612782

ICO. (n.d.). Chapter 4: Accountability and governance (Anonymisation guidance). Information Commissioner’s Office. https://ico.org.uk/media2/migrated/4019713/chapter-4-anonymisation-guidance-accountability-and-governance.pdf

IDC. (2025, October 21). Wearable Devices Market Insights (Q2 2025 shipments). https://www.idc.com/promo/wearablevendor/

Information Commissioner’s Office (ICO). (2025). Anonymisation and pseudonymisation: Guidance (UK GDPR). https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/about-this-guidance/

International Organization for Standardization (ISO). (2020). ISO 21502:2020 — Project, programme and portfolio management — Guidance on project management. https://www.iso.org/standard/74947.html

Intervals.icu. (2026). Sports analytics and planning (example charts). https://intervals.icu/

ISO. (2019). ISO 9241-210:2019 — Ergonomics of human-system interaction — Human-centred design for interactive systems. International Organization for Standardization. https://www.iso.org/standard/77520.html

ISO. (2022). ISO/IEC 27001:2022 — Information security management systems. International Organization for Standardization. https://www.iso.org/standard/27001

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems (NeurIPS). https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree

Little, R. J. A., & Rubin, D. B. (2002). Statistical Analysis with Missing Data (2nd ed.). Wiley. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119013563

Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., & Denniston, A. K. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. BMJ, 370, m3164. https://www.bmj.com/content/370/bmj.m3164

Lövdal, S. S., Den Hartigh, R. J. R., & Azzopardi, G. (2021). Injury prediction in competitive runners with machine learning. International Journal of Sports Physiology and Performance, 16(10), 1522–1531. https://doi.org/10.1123/ijspp.2020-0518

Lövdal, S. S., den Hartigh, R. J. R., & Azzopardi, G. (2021). Replication Data for: Injury Prediction in Competitive Runners With Machine Learning. DataverseNL. https://doi.org/10.34894/UWU9PV

MarketsandMarkets. (2025). Wearable Technology Market – Global Forecast to 2030. https://www.marketsandmarkets.com/Market-Reports/wearable-electronics-market-983.html

Medicines and Healthcare products Regulatory Agency (MHRA). (2021). Software and AI as a Medical Device Change Programme. GOV.UK. https://www.gov.uk/government/publications/software-and-ai-as-a-medical-device-change-programme

Medicines and Healthcare products Regulatory Agency (MHRA). (2014). Medical devices: software applications (apps). GOV.UK. https://www.gov.uk/government/publications/medical-devices-software-applications-apps

Medicines and Healthcare products Regulatory Agency (MHRA). (2024). Regulating medical devices in the UK (overview guidance). GOV.UK. https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk

Milewski, M. D., Skaggs, D. L., Bishop, G. A., Pace, J. L., Ibrahim, D. A., Wren, T. A. L., & Barzdukas, A. (2014). Chronic lack of sleep is associated with increased sports injuries in adolescent athletes. Journal of Pediatric Orthopaedics, 34(2), 129–133. https://doi.org/10.1097/BPO.0000000000000151

Nelson, E. O., et al. (2019). Development of the University of Wisconsin Running Injury and Recovery Index. J Orthop Sports Phys Ther, 49(10), 751–760. https://doi.org/10.2519/jospt.2019.8868

Oura. (n.d.). Oura Membership. https://ouraring.com/membership/

OWASP. (n.d.). Mobile Application Security Verification Standard (MASVS). OWASP Mobile Application Security. https://mas.owasp.org/MASVS/

Pla, R., et al. (2021). Resting heart rate measurement in elite athletes during COVID-19 lockdown: A new approach to assess health and performance. Sustainability, 13(5), 2970. https://doi.org/10.3390/su13052970

Post Affiliate Pro. (2025, December 28). SaaS affiliate marketing commission rates and structures. Post Affiliate Pro. https://www.postaffiliatepro.com/blog/saas-affiliate-commission-rates/

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. arXiv:1706.09516. https://arxiv.org/abs/1706.09516

RevenueCat. (2025). State of Subscription Apps 2025. RevenueCat. https://www.revenuecat.com/state-of-subscription-apps-2025/

Rewardful. (2023, June 30). Affiliate commission guide for SaaS. Rewardful. https://www.rewardful.com/articles/affiliate-commission-explained

Ries, E. (2011). The Lean Startup. Crown Business. Publisher page: https://www.penguinrandomhouse.com/books/210088/the-lean-startup-by-eric-ries/

Rivera, S. C., Liu, X., Chan, A.-W., Denniston, A. K., & Calvert, M. J. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. BMJ, 370, m3210. https://www.bmj.com/content/370/bmj.m3210

Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Russell, H. C., & Wiese-Bjornstal, D. M. (2015). *Narratives of psychosocial response to microtrauma injury among long-distance runners*. Sports, 3(3), 159–177. https://doi.org/10.3390/sports3030159

Runalyze. (2026). Example training analytics dashboards. https://runalyze.com/

Runalyze. (n.d.). Pricing. https://runalyze.com/pricing/

SaaS Capital. (2025, January 24). 2025 Private SaaS Company Valuations (The SaaS Capital Index). SaaS Capital. https://www.saas-capital.com/blog-posts/private-saas-company-valuations-multiples/

Saito, T., & Rehmsmeier, M. (2015). The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Sports & Fitness Industry Association (SFIA). (2025, February 27). SFIA's Topline Participation Report shows 247.1 million Americans were active in 2024. https://sfia.org/resources/sfias-topline-participation-report-shows-247-1-million-americans-were-active-in-2024/

Strava. (2025a). Year in Sport 2025: What it means for brands. https://business.strava.com/resources/year-in-sport-brands-2025

Strava. (2025b, April 17). Strava to acquire Runna, a leading running training app. https://press.strava.com/articles/strava-to-acquire-runna-a-leading-running-training-app

Strava. (2025c, July 1). Pricing – Updating subscription prices. https://www.strava.com/pricing

StrideWise Development Team. (2025). COMPLETE_SYSTEM_DOCUMENTATION: Synthetic Runner Data Generation and Injury Prediction (Version 1.0). See Appendix 2.

StrideWise. (2026). MSBA Capstone Written Report (StrideWise internal report, Word manuscript).

The Verge. (2025, March 27). Garmin adds AI and a subscription tier to its app. https://www.theverge.com/news/636211/garmin-connect-plus-subscription-wearables

TrainingPeaks. (2026). Adding a Chart to your Dashboard (Help Center). https://help.trainingpeaks.com/hc/en-us/articles/115004963667-Adding-a-Chart-to-your-Dashboard

TrainingPeaks. (n.d.). Pricing for Athletes. https://www.trainingpeaks.com/pricing/for-athletes/

Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley. https://www.amazon.co.uk/Exploratory-Data-Analysis-Behavioral-Science/dp/0201076160

UK Information Commissioner's Office (ICO). (2024). Special category data (UK GDPR) – guidance. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/special-category-data/

Vallevik, V. B., Babic, A., et al. (2024). Can I trust my fake data – A comprehensive quality assessment framework for synthetic tabular data in healthcare. International Journal of Medical Informatics, 185, 105413. https://doi.org/10.1016/j.ijmedinf.2024.105413

van der Worp, M. P., ten Haaf, D. S. M., van Cingel, R., de Wijer, A., Nijhuis-van der Sanden, M. W. G., & Staal, J. B. (2015). Injuries in runners; a systematic review on risk factors and sex differences. PLOS ONE, 10(2), e0114937. https://doi.org/10.1371/journal.pone.0114937

Videbæk, S., Bueno, A. M., Nielsen, R. O., & Rasmussen, S. (2015). Incidence of running-related injuries per 1000 h of running in different types of runners: A systematic review and meta-analysis. Sports Medicine, 45(7), 1017–1026. https://doi.org/10.1007/s40279-015-0333-8

Visser, T. S. O. S., et al. (2021). The socio-economic impact of running-related injuries: A large prospective cohort study. Scandinavian Journal of Medicine & Science in Sports, 31(9), 1781–1792. https://doi.org/10.1111/sms.14016

W3C. (2023). WCAG 2.2 is a Web Standard "W3C Recommendation". Web Accessibility Initiative. https://www.w3.org/WAI/news/2023-10-05/wcag22rec/

Watson, A., et al. (2020). Decreased sleep is an independent predictor of in-season injury in collegiate athletes. Orthopaedic Journal of Sports Medicine, 8(11). https://doi.org/10.1177/2325967120964481

WHOOP. (2025, November 4). Membership Pricing. https://support.whoop.com/s/article/Membership-Pricing

Wolff, R. F., et al. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 170(1), 51–58. https://doi.org/10.7326/M18-1376

ŌURA. (n.d.). Membership Pricing. Ōura Support. https://support.ouraring.com/hc/en-us/articles/360025577993-Membership-Pricing

ŌURA. (n.d.-b). Oura Ring 4 – product pricing (from $349). https://ouraring.com/store/rings/oura-ring-4/black

Note

The author used the following AI systems to aid in the production of this paper, and the accompanying code base:

- ChatGPT: Research and experimentation – especially in regard to researching injury causation mechanisms.
- NotebookLM: Research and slide deck creation (specifically the deck as indicated in Appendix 3).
- Cursor (primarily utilising Claude): Research, experimentation, coding, code documentation.
- Runway (primarily utilising Nano Banana): Image creation (based on scanned drawings as created by the author).

All the code was run and tested on an Apple MacBook Pro using PyCharm as the primary IDE.

The author acknowledges accountability for the content in this document, and recognize that AI tool use disclosure does not excuse responsibility for errors, inaccuracies, or plagiarism.

Appendices

Appendix 1 – Code Repository

All the code (alongside supporting documentation) for this project is available at:

https://github.com/sgw2342/Stridewise/

Appendix 2 – Code Summary Documentation

The key summary document outlining & explaining the entire codebase and the underlying rationale, available at:

https://github.com/sgw2342/Stridewise/blob/1f50d302706eaee9b5e53be2b34a06925f8eabce/docs/COMPLETE_SYSTEM_DOCUMENTATION.md

Appendix 3 – SynthRun

Link to explanatory slide deck explaining our synthetic runner generation process (generated via NotebookLM, using codebase and code documentation as context):

https://github.com/sgw2342/Stridewise/blob/53fef7d4ad9cceb7a62c0a6102cca5f6a0fba74b/SynthRun_Generating_Realistic_Injury_Risk_Datasets.pdf

Appendix 4 – Model and actual dataset from Lövdal et al paper

We used only the Daily version of their dataset in our validation approach (the Weekly version being an aggregated view of the Daily dataset). That is the “day_approach_maskedID_timeseries” file in the folder.

https://github.com/sgw2342/Stridewise/tree/751b723ccb0c4343ade6e0e068cafe7047e50031/cc0_competitive_runners

Appendix 5 – SynthRun explanatory paper (academic journal format)

Further details on the SynthRun product – i.e. the code which produces our synthetic runner cohorts, incorporating the injury causation model. The paper was put into academic format via ChatGPT (based on original text as uploaded into context).

https://github.com/sgw2342/Stridewise/blob/main/SynthRun_Paper.pdf