Journal of Economics
AND MANAGEMENT SCIENTIES
https://jems.ink
ISSN 2655-1934 (print), 2655-6685 (online)

Beyond Loss Cost: An AHP Framework for Prioritising Focused Improvement Initiatives in Total Productive Maintenance

Wina Indra Lavina[1], and Timotius FCW Sutrisno[2]*

[1,2] *Universitas Ciputra, Indonesia*

Journal of Economics and Management Scienties is licensed under a Creative Commons 4.0 International License.

ARTICLE HISTORY

Received: 14 October 26
Final Revision: 21 October 26
Accepted: 04 November 26
Online Publication: 31 December 26

KEYWORDS

Total Productive Maintenance, Focused Improvement, Analytic Hierarchy Process, Group Decision-Making, Improvement Prioritisation

A B S T R A C T

Focused Improvement (FI) teams in Total Productive Maintenance (TPM) must decide which loss-reduction initiatives to pursue first, yet TPM prescribes no method for this choice, and plants usually rank initiatives by loss cost. This study develops and tests a multi-criteria framework for this decision. An Analytic Hierarchy Process (AHP) hierarchy of four criteria and 16 sub-criteria, derived from the JIPM loss structure and the literature, was weighted by 15 division and department heads of an Indonesian pipe manufacturer, 12 of whom then compared five initiatives from the plant's 2026 improvement portfolio. Consistency, consensus, alternative aggregations, weight sensitivity, critical weights and a bootstrap over experts were analysed. Business impact received the highest weight (0.448), followed by sustainability (0.285), operational urgency (0.177) and implementation feasibility (0.090); customer impact, long-term benefit and quality outweighed cost reduction. The yield-loss initiative ranked first in 99.9% of bootstrap panels, and a quality initiative with the smallest realised saving rose from last to second place, so the AHP ranking agreed only weakly with the monetary ranking ($\rho = 0.40$). Loss cost should therefore inform, but not decide, the order of FI initiatives. The study extends AHP in TPM from ranking success factors to ranking FI initiatives and shows that managers value customer impact and risk above loss cost.

KATA KUNCI

Total Productive Maintenance, *Focused Improvement*, *Analytic Hierarchy Process*, Pengambilan Keputusan Kelompok, Prioritas Perbaikan

A B S T R A K

Tim *Focused Improvement* (FI) dalam *Total Productive Maintenance* (TPM) harus menentukan inisiatif pengurangan *loss* mana yang dikerjakan lebih dahulu, tetapi TPM tidak menetapkan metode untuk pilihan ini, dan pabrik umumnya mengurutkan inisiatif berdasarkan *loss cost*. Penelitian ini mengembangkan dan menguji kerangka kerja multikriteria untuk keputusan tersebut. Hierarki *Analytic Hierarchy Process* (AHP) yang terdiri atas empat kriteria dan 16 subkriteria, disusun dari struktur loss JIPM dan literatur, dibobot oleh 15 kepala divisi dan kepala departemen sebuah produsen pipa di Indonesia; 12 di antaranya kemudian membandingkan lima inisiatif dari portofolio perbaikan pabrik tahun 2026. Analisis mencakup konsistensi, konsensus, agregasi alternatif, sensitivitas bobot, bobot kritis, dan *bootstrap* atas para ahli. Dampak bisnis memperoleh bobot tertinggi (0,448), diikuti keberlanjutan (0,285), urgensi operasional (0,177), dan kelayakan implementasi (0,090); dampak pelanggan, manfaat jangka panjang, dan kualitas lebih berbobot daripada pengurangan biaya. Inisiatif *yield loss* menempati peringkat pertama pada 99,9% panel *bootstrap*, dan inisiatif kualitas dengan realisasi penghematan terkecil naik dari peringkat terakhir ke peringkat kedua, sehingga peringkat AHP hanya lemah selaras dengan peringkat berdasarkan nilai uang ($\rho = 0,40$). Karena itu, *loss cost* sebaiknya menjadi masukan, tetapi bukan penentu, urutan inisiatif FI. Penelitian ini memperluas penggunaan AHP dalam TPM dari pemeringkatan faktor keberhasilan ke pemeringkatan inisiatif FI dan menunjukkan bahwa manajer menilai dampak pelanggan dan risiko lebih tinggi daripada *loss cost*.

CORRESPONDING AUTHOR

windralavina@magister.ciputra.ac.id

DOI

10.37034/jems.v9i1.583

1. Introduction

1.1. Research Background

Manufacturing firms are now expected to produce more with less material and energy while also proving that their performance is sustainable. Industry 4.0 technologies combined with Total Productive Maintenance (TPM) have become a common response, because they link equipment performance directly to economic and environmental outcomes [1], [2]. Studies of sustainable manufacturing show the same shift: improvement is increasingly judged not only by cost, but also by waste, energy use and long-term effects [3], [4]. In water and infrastructure networks, PVC and HDPE pipes continue to replace metal pipes, chiefly because they weigh less, are quicker to lay and do not corrode [5]. Their producers therefore compete in a growing but

price-driven market in which material efficiency largely determines margins.

The same pressure is visible in Indonesia. The national economy expanded by 5.11% in 2025, and manufacturing, its largest sector, generated 19.07% of gross domestic product and grew by 5.30% in the same year [6], [7]. PVC imports rose by 49.39% to US$111.7 million, a surge that market analysts link to infrastructure projects [8], while the import price index for plastics (HS 39) climbed from 121.4 to 139.4 (2023 = 100) between the first and second quarters of 2026 [6]. For pipe producers, this means strong downstream demand but also exposure to the supply and price of raw material. Indonesian manufacturers have begun to adopt structured, multi-criteria approaches to improve efficiency and sustainability, for example in tyre, paving-block and agro-food production [9], [10], [11]. Evidence from the process industries, and from TPM-based plants in particular, remains scarce.

The case company, an Indonesian pipe manufacturer (hereafter Company X), shows how this pressure becomes a concrete decision problem. Its study plant (Plant A) has implemented JIPM-based TPM and a Manufacturing Execution System (MES), and converts its losses into money through a loss cost matrix. Two features stand out. First, the largest controllable loss, yield loss from overweight products and unrecyclable scrap, lies outside Overall Equipment Effectiveness (OEE), yet it has received only a small share of improvement themes, while rejects, a minor loss, have received the most. Second, initiatives are selected at an annual strategic deployment workshop through discussion, a Pareto ranking of loss cost and high/medium/low ratings of benefit and effort, without weighted criteria. Benefit classes have been applied inconsistently, effort ratings simply follow the improvement method, and the initiative portfolio covers only a small share of controllable losses. Plant A's OEE also remains below the 85% world-class benchmark [12], [13]. The decision is therefore important, recurrent and, at present, neither explicit nor auditable.

Three bodies of theory frame this problem. The first is the TPM theory of losses. Of the eight pillars on which TPM rests, Focused Improvement (FI) is the one charged with removing the causes of the 16 major losses [13], [14]. The standard asks FI teams to decide which losses to attack in each period, yet it supplies no method for that choice, and it places yield loss among the losses of production resources, outside OEE [13]. In World Class Manufacturing, cost deployment fills this gap by ranking losses according to their attributed cost and screening countermeasures by profitability, which turns money into the single yardstick of priority [15], [16].

The second is multi-criteria decision theory. Choosing improvement initiatives means trading off conflicting criteria, which is the domain of multi-criteria decision making (MCDM). The Analytic Hierarchy Process (AHP) structures such a decision as a hierarchy, derives priorities from pairwise judgements expressed on Saaty's 1–9 scale and checks whether those judgements hang together; a consistency ratio (CR) of 0.10 or less is the usual limit for acceptance [17]. The judgements of several experts can be pooled with the geometric mean [18], and few decision-support methods have been applied as widely [19].

The third is strategy–execution alignment. The *Balanced Scorecard* expresses strategy as a set of measures grouped into four perspectives [20], while *hoshin kanri* breaks the resulting objectives down into actions for every level of the organisation [21]. Together these theories imply that FI initiatives should be prioritised by explicit, weighted criteria that reflect both loss data and the company's targets, a link that is missing in the case plant.

Studies of improvement project selection supply structures that could be borrowed. AHP is the method most often used to choose Six Sigma projects [22], and hybrid models in which AHP fixes the weights and each alternative is scored on a common scale have ordered design projects and maintenance work, including across 695 Indonesian distribution substations [23], [24].

Within TPM, however, MCDM has been applied to other objects. AHP has ranked TPM success factors [25], and key performance indicators [26]; AHP–TOPSIS and DEMATEL-based ANP with ARAS have ranked drivers of TPM–Industry 4.0 integration [1], [2]. Studies of TPM implementation also note that poor selection of pilot equipment hinders success [27]. The data-driven framework of a study prioritises TPM pillars and actions by their impact on OEE, so losses outside OEE, such as yield loss, are not captured [28]. In Indonesia, multi-criteria methods support sustainability assessment rather than TPM decisions [4], [11].

Four gaps follow. First, TPM requires FI teams to prioritise, yet no study offers a structured method for prioritising FI initiatives themselves. Second, data-driven TPM prioritisation is anchored in OEE and misses the losses outside it that dominate process industries such as pipe manufacturing. Third, cost deployment, the most systematic loss-based method, judges initiatives by money alone. Fourth, empirical evidence from Indonesian process plants is limited, and existing studies rarely compare a multi-criteria ranking with the selection practice it would replace.

This study develops and tests an AHP-based framework for prioritising FI initiatives at Plant A. It uses a hierarchy of four criteria (business impact, operational urgency, implementation feasibility and sustainability) and 16 sub-criteria. It addresses three research questions. The first research question (RQ1) asks which criteria and sub-criteria are relevant for prioritising FI initiatives in pipe manufacturing. The second (RQ2) asks what weights internal experts assign to these

criteria and sub-criteria, and whether their judgements are consistent. The third (RQ3) asks how the resulting ranking of the plant's initiatives differs from the monetary logic of its current Pareto practice, and how robust that ranking is to changes in the weights and in the composition of the expert panel. The questions are answered with the judgements of a panel of 15 division and department heads, 12 of whom also evaluated five initiatives from the plant's 2026 improvement portfolio.

The study makes three contributions. Theoretically, it extends AHP in TPM from ranking success factors, KPIs and drivers to ranking the improvement initiatives that the FI pillar must choose, and it proposes a criterion set that separates the size of the benefit from urgency, feasibility and sustainability while bringing losses outside OEE into the assessment. Methodologically, it carries the AHP through to the evaluation of real initiatives by the same panel that set the weights, it separates the consistency of the judgements from the consensus among experts, and it tests the resulting ranking with three complementary checks: deterministic perturbation of the weights, a critical-weight analysis and a bootstrap over experts. Practically and contextually, it gives FI teams a transparent and auditable tool for the strategic deployment workshop and adds evidence from the Indonesian process industry.

1.2. Literature Review

1.2.1. TPM, Focused Improvement and the structure of losses

JIPM describes TPM as a company-wide effort to make the production system as efficient as it can be: every kind of loss is to be prevented over the whole life of the system, and all employees contribute through small groups that overlap across organisational levels [13]. Recent reviews show that the eight pillars of this model are being extended with digital tools, with MES platforms supporting continuous, data-based monitoring of losses [2], [14]. Within the model, the FI pillar (*kobetsu kaizen*) identifies losses, expresses them in monetary terms and turns them into improvement themes. The standard expects an FI team to master three abilities: to see which losses occur and how large they are, to decide, in line with management policy, which of them to address in a given period, and to remove them quickly [13]. The second ability is a prioritisation task, yet the standard does not specify how it should be carried out.

Equipment effectiveness is usually tracked through OEE, which multiplies the availability, performance and quality rates; plants that reach 85% or more are commonly regarded as world class [12], [13]. OEE nevertheless captures only part of the 16 major losses. JIPM groups them into one loss that limits the operating rate, seven that limit equipment efficiency, five that limit human work efficiency, and three that limit the use of production resources, namely energy, dies and tools, and yield [13]. Yield loss is defined by the difference between the weight of material fed into the process and the weight of good output; in plastics processing it also covers material used beyond the product standard [13]. In this study, overweight material and scrap that cannot be reprocessed are treated together as yield loss. Only the material above an achievable baseline, defined as the minimum specification plus a process capability margin, is regarded as controllable, because part of the overweight is a deliberate margin that protects quality when machine capability is limited.

Because losses are recorded in different units, such as hours, kilograms or kilowatt-hours, a loss cost matrix translates them into money to make them comparable. In pipe manufacturing, resin is the dominant cost item, which makes yield loss a direct material loss. A prioritisation that follows OEE impact alone therefore overlooks the largest losses of the plant [28]. Loss cost, however, measures only how large a problem is. It says nothing about how urgent the problem is, how feasible its removal is, or whether the result will last.

1.2.2. Prioritising improvement initiatives

Choosing improvement initiatives is a multi-criteria decision that balances expected benefit, urgency, resource needs and risk. On the shop floor, the most common approach is a Pareto ranking of loss cost, in which the largest losses are addressed first. The approach is simple and transparent, but it rests on one criterion. It ignores feasibility, because a large loss may require capital investment or a long project. It also misses risk, because a small loss, such as a dimensional defect, can damage customer trust, and it leaves the remaining judgement unstructured and difficult to audit.

Cost deployment, developed within World Class Manufacturing, is the most systematic form of this monetary logic. The method traces every resulting loss back to the loss that causes it, orders these causal losses by the cost assigned to them and selects countermeasures by their expected return within the budget [15]. When it is applied to whole production flows, it shows which types of loss account for most of the cost and thus where improvement should be concentrated [16]. Cost deployment makes prioritisation explicit and auditable, but it judges every initiative by one yardstick, money, and leaves urgency, customer risk and the durability of results outside the calculation.

TPM guidance itself points to further criteria. JIPM separates sporadic losses, which have a single cause and are removed by restoring basic conditions, from chronic losses, whose causes are many and interacting and which call for innovative countermeasures [13]. Automation should be justified by comparing the investment with the expected gain in efficiency, and improvement ideas are judged on effectiveness, feasibility and cost [13]. How these considerations should be weighted against each other is left to the team.

Studies of TPM implementation add that the choice of pilot equipment and the availability of trained people affect whether improvements succeed [25], [27].

Research on improvement project selection provides more structure. A study reviewed 59 studies on Six Sigma project selection and counted 111 methods; roughly a third of them were MCDM methods, and AHP appeared more often than any other [22]. Later studies ranked Lean Six Sigma projects with the Best-Worst Method [29], combined AHP weights with a normalised scoring model to choose Design for Six Sigma projects [23], and used AHP to weight criteria before ranking PFMEA improvement measures in a forging process [30]. In Indonesian manufacturing, structured methods such as Delphi, AHP and TOPSIS have been used to select sustainability indicators and waste-reduction measures, and simple, low-cost interventions were found to rank highest in one case [4], [11]. Outside manufacturing, AHP has also structured vendor selection in an Indonesian ICT system-integration company [31]. Together, these studies show that multi-criteria rankings can differ from the intuition that the largest problem deserves the first project.

1.2.3. Relative and absolute measurement in AHP

MCDM methods such as AHP, ANP, TOPSIS, DEMATEL, ARAS and the Best-Worst Method (BWM) rank alternatives against conflicting criteria, and AHP is among the most frequently used of them [19]. TOPSIS does not itself derive criterion weights, and BWM needs fewer comparisons [32], although its robustness relative to AHP is still debated [33]. AHP was chosen because its hierarchy fits criteria with sub-criteria, its consistency check gives experts direct feedback, and it is the method most familiar to improvement practitioners, which makes its results easier to accept in a strategic deployment workshop [22].

An AHP model arranges the goal, the criteria, the sub-criteria and the alternatives in levels. Experts judge pairs of elements on Saaty's 1–9 scale, each comparison matrix yields its weights through its principal eigenvector, and a matrix is normally accepted when its consistency ratio is 0.10 or lower [17]. When a study screened expert groups at $CR < 0.1$ and at $CR < 0.2$, the resulting priority orders hardly differed, so the looser limit can serve as a robustness check for small panels [3]. In group decisions, individual judgements can be combined with the geometric mean, which keeps the combined comparison matrix reciprocal [18]. Groups can be aggregated either at the level of judgements (AIJ) or at the level of the priority vectors derived from them (AIP), and AIP is common in applied work [34]. Both procedures can mislead when the experts' matrices are inconsistent or the experts disagree strongly, which makes it necessary to report consistency and consensus separately [35].

AHP can evaluate alternatives in two ways. In relative measurement, the alternatives themselves are compared in pairs under each criterion, so that every judgement is anchored in a direct comparison of concrete options. This mode is natural when the set of alternatives is small and the alternatives are well known to the experts, but the number of comparisons grows quickly: the more than 20 initiatives in the plant's 2026 portfolio would require at least 190 comparisons per criterion. In absolute measurement, AHP is used only to weight the criteria, and each alternative is rated against a fixed scale [23]. A study took this route, merging AHP weights with standardised condition scores into a health index that set the order of maintenance work [24]. Because rated alternatives do not depend on each other, adding or removing one does not alter the scores of the rest, which avoids the rank-reversal problem long discussed in the AHP literature [36]. Relative measurement, by contrast, produces a ranking that is valid for the set of alternatives evaluated and must be repeated when those set changes.

This study uses relative measurement for the initiatives of the current improvement portfolio, because the set is small and the experts know each initiative from the strategic deployment workshop. The sub-criteria serve two purposes: they define what each criterion covers when the initiatives are compared, and their weights show which aspects of an initiative the organisation values most. The second purpose also prepares the ground for absolute measurement once the framework is applied to the plant's full portfolio.

1.2.4. Prior studies and research gap

Table 1 summarises the studies closest to this research. They fall into four groups: MCDM applied to TPM, data-driven prioritisation within TPM, multi-criteria selection of improvement projects, and loss-based monetary prioritisation. Read column by column, Table 1 confirms the four gaps set out in the Introduction: MCDM in TPM has ranked factors, KPIs and drivers but not FI initiatives; data-driven TPM prioritisation stops at OEE; cost deployment relies on money alone; and no Indonesian process-plant study sets a multi-criteria ranking against the practice it would replace.

Table 1. Prior studies closest to this research

Study	Context and method	Main finding	Difference from this study
Singh and Gurtu (2022) [25]	AHP, TPM success factors	Top-management commitment is the most critical factor, followed by training	Ranks success factors, not initiatives
Toke and Kalpande (2023) [26]	AHP, TPM key performance indicators	Productivity and quality indicators receive the highest weights	Ranks KPIs, not initiatives
Samadhiya et al. (2024) [2]	AHP–TOPSIS, Industry 4.0–TPM drivers	Ranks drivers of integrated TPM	Ranks drivers, not initiatives
Fallahiarezoudar et al. (2025) [1]	DEMATEL-based ANP and ARAS, automotive firm	Ranks key drivers of Industry 4.0–TPM integration	Ranks drivers, not initiatives
Lucantoni et al. (2025) [28]	Association rules and network analysis of OEE data	Roadmap of TPM pillars and actions; OEE improved	Priorities follow OEE; yield loss outside OEE is not covered
Pakdil (2022) [22]	Review of Six Sigma project selection	AHP is the most frequently used of 111 methods	Not TPM; not linked to loss cost
Nakielski and Ginda (2026) [23]	AHP weights with normalised scoring, DFSS projects	Hybrid model works for many alternatives	Not production losses
Kastratović et al. (2026) [30]	AHP–SREM, PFMEA improvement measures	Technical, economic and operational criteria combined	Not TPM
Rosiani et al. (2026) [11]; Utama and Abirfatin (2023) [4]	Sustainable Lean Six Sigma, Delphi, AHP, TOPSIS, Indonesia	Low-cost measures can rank highest	Sustainability assessment, not TPM initiatives
Hermeindito et al. (2026) [31]	AHP, vendor selection, Indonesia	Price carries the largest weight	Vendor selection, not improvement initiatives
Bertolini and Castellano (2026) [15]; Hagström et al. (2023) [16]	Cost deployment in automotive and electronics production	Losses ranked by attributed cost; countermeasures by profitability	Single monetary yardstick; no weighting of non-monetary criteria

1.2.5. Proposed criteria and conceptual framework

The hierarchy combines the literature above, the plant's loss tree and the criteria used on its initiative form. An initial version with six criteria was consolidated into four criteria and 16 sub-criteria for the questionnaire, so that each pairwise matrix stays small and easy to judge, which can be seen on Table 2. The criteria are designed to avoid double counting. Each loss enters only one benefit sub-criterion: material and energy losses enter cost reduction (C1.1), machine-time losses enter productivity (C1.2), and rejects and complaints enter quality (C1.3). C1 and C2 also look in different directions. Business impact describes the result an initiative is expected to deliver, whereas operational urgency describes the size and pressure of the problem as it stands today; for example, the performance gap (C2.3) is the current distance between a KPI and its target, not the share of that gap a single initiative will close.

Table 2. Criteria, sub-criteria and their basis

Criterion	Sub-criteria	Meaning in this study	Basis
C1 Business impact	C1.1 Cost reduction; C1.2 Productivity; C1.3 Quality; C1.4 Customer impact	Expected benefit after implementation: money no longer lost on material, energy, maintenance and overtime; extra machine hours or output; fewer nonconforming products; better customer outcomes	[13]; [20]; [26]
C2 Operational urgency	C2.1 Problem frequency; C2.2 Problem severity; C2.3 Performance gap; C2.4 Risk exposure	Pressure of the problem today: how often it occurs, how serious each occurrence is, how far the related KPI lies from target, and the risk of leaving it unaddressed	[20]; [21]; [30];
C3 Implementation feasibility	C3.1 Investment need; C3.2 Technical feasibility; C3.3 Resource availability; C3.4 Complexity	How realistic implementation is with current means: capital required, in-house technical capability, availability of people, time and data, and coordination effort	[25]; [27]
C4 Sustainability	C4.1 Standardisation; C4.2 Replicability; C4.3 Monitoring; C4.4 Long-term benefit	Durability of the result: whether it can be built into standards, copied to other lines, tracked in the MES, and kept over time	[2]; [4]; [14];

Figure 1 shows the resulting four-level hierarchy. The criterion and sub-criterion weights are set by the expert panel, the initiatives are then compared in pairs under each criterion, and the overall priority of an initiative is the sum of its local priorities weighted by the criterion weights. Safety improvements and follow-up of customer complaints are mandatory in the case plant and are therefore handled outside the ranking.

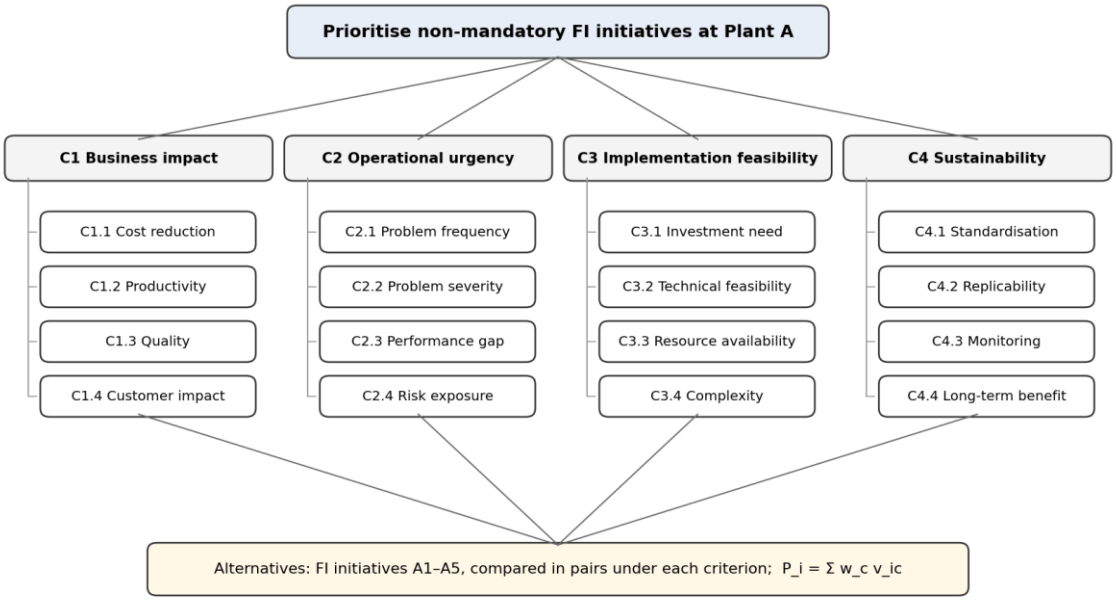


Figure 1. AHP hierarchy for prioritising FI initiatives at Plant A

2. Research Method

2.1. Research design

The study uses an embedded single-case design. A single case is appropriate because the aim is to build and test a decision framework in a setting where the problem is clearly visible and data are accessible, and the embedded units are the improvement initiatives assessed within the case. The approach is quantitative and decision-analytic: expert judgements are collected through structured pairwise comparisons and processed with AHP, first to weight the criteria and sub-criteria and then to evaluate initiatives from the plant's current improvement portfolio. Table 3 summarises the research procedure.

Table 3. Research procedure

Step	Activity	Output
1	Literature review and analysis of the plant's loss tree and initiative form	Initial set of criteria
2	Review of the criteria with practitioners and simplification of the hierarchy	Final hierarchy of 4 criteria and 16 sub-criteria
3	Pairwise-comparison questionnaire completed by the expert panel	Individual comparison matrices
4	Consistency check, aggregation and robustness checks for individual inconsistency	Group weights and global weights
5	Pairwise comparison of five initiatives under each criterion by the same panel	Local and overall priorities of the initiatives
6	Comparison with the monetary ranking; sensitivity, critical-weight and bootstrap analyses	Rank agreement and robustness of the ranking

2.2. Case setting and unit of analysis

The case is one pipe manufacturing plant (Plant A) of a large Indonesian pipe manufacturer, anonymised as Company X. The plant was selected because it operates JIPM-based TPM with a functioning FI pillar, records losses in an MES and converts them into loss cost, and selects its improvement initiatives at an annual strategic deployment workshop. The unit of analysis is the improvement initiative. The alternatives are five non-mandatory initiatives taken from the plant's 2026 initiative form, which can be seen on Table 4. The 2026 portfolio holds more than 20 initiatives, and the realised annual savings of the five selected equal about two-thirds (67%) of the planned annual saving of the whole portfolio, so the sample covers a large share of the money at stake. They were chosen to span the main loss categories of the plant (yield, maintenance cost, equipment breakdown, quality and availability), the three improvement routes in use (kaizen, horizontal expansion of proven solutions and small group activity) and the three functions that own improvement projects. Safety improvements and actions that follow customer complaints are excluded because they are carried out regardless of rank. The weights, in contrast, were elicited at division level, as described under Expert panel below. The criteria express the priorities of the pipe division, which apply to all of its plants, and the weights obtained are then applied to the initiatives of Plant A.

Table 4. Initiatives evaluated in Part C

Code	Initiative	Main loss addressed	KPI target (relative change)	Route	Owner	Realised annual saving (index, A1 = 100)
A1	Reduce yield loss from overweight on a high-volume pipe size	Yield loss (overweight)	Overweight −32%	Kaizen	Production	100.0
A2	Rationalise filter stock variants on three extrusion lines	Maintenance cost	Filter cost −90%	Horizontal expansion	Maintenance	27.6
A3	Eliminate breakdowns of the vacuum pumps of the calibration unit	Equipment breakdown	MTBF +40%; MTTR −60%	Kaizen	Maintenance	3.7
A4	Reduce quality rejects in the impact test on one extrusion line	Quality loss (rejects)	Reject rate −81%	Horizontal expansion	Quality assurance	3.3
A5	Shorten changeover time on one extrusion line	Availability loss (changeover)	Changeover time −40%	Small group activity	Production	6.7

2.3. Expert panel

AHP draws on informed judgement rather than on statistical sampling, so the panel was selected purposively. Two conditions applied: the expert held a position at department-head level or above in the pipe division, and the expert had led or approved improvement projects. Sixteen managers returned the questionnaire. One response could not be processed, because both sides of the scale had been marked in 29 of the 30 rows and the direction of preference could therefore not be determined. The analysis of Part B is based on the remaining 15 experts: four division heads and eleven department heads. They represent plant management, production, quality assurance, maintenance, engineering, continuous improvement, logistics, and demand and supply planning, and they work at Plant A, at four sister plants or in division-wide functions. Twelve of the 15 experts also completed Part C. Their profile is given in Supplementary Table S5.

2.4. Instrument and data collection

The questionnaire was completed by the experts on 7 and 8 October 2026. It contained a respondent profile (Part A), instructions with Saaty's scale and a worked example, definitions of every criterion and sub-criterion with shop-floor examples, and the pairwise comparisons among the criteria and sub-criteria (Part B). Each comparison was presented as one row with a 17-point scale running from 9 (left element much more important) through 1 (equal importance) to 9 (right element much more important). The instrument comprised 30 comparisons: 6 among the four criteria and 6 within each group of four sub-criteria. Before distribution, the definitions were checked with practitioners at the plant to remove ambiguity, in particular between cost reduction, productivity improvement and performance gap.

Part C asked the same experts to compare the five initiatives in pairs under each of the four criteria, which gave 10 comparisons per criterion and 40 in total. Twelve of the 15 experts completed Part C in October 2026. Seven of them revised one or more of their Part C matrices before the analysis, two of them more than once, and the final versions are used. A one-page fact sheet on each initiative accompanied Part C (Table 4 and Supplementary Table S2). It gave the loss addressed, the KPI baseline and target, the realised annual saving, the improvement route, the owner, the investment required, the departments involved and the plans for standardisation and replication, all taken from the plant's initiative records, so that every expert judged the same information. Experts were asked to read each criterion through its sub-criteria; for example, an initiative was to be judged stronger on business impact when it offered larger savings, more output, better quality or a greater benefit to customers.es are expressed as an index to protect company data. The savings are the realised annual savings recorded for each initiative; for A3, the midpoint of the reported range is used. They serve to reproduce the monetary ranking of current practice.

2.5. Data analysis

2.5.1. Weights and consistency

For each expert and each comparison matrix A of size n, the priority vector w is the normalised principal eigenvector of A. Consistency is assessed with the consistency index and the consistency ratio as can be seen on Equation (1), where $\lambda_{\max}$ is the principal eigenvalue and RI is the random index, which equals 0.90 for the 4 × 4 matrices of Part B and 1.12 for the 5 × 5 matrices of Part C. Matrices with CR up to 0.10 are accepted as consistent.

$$CI = \frac{\lambda_{\max} - n}{n-1}, CR = \frac{CI}{RI} \quad (1)$$

2.5.2. Aggregation

The judgements of the experts are combined element by element with the geometric mean. As can be seen on Equation (2), where K is the number of experts: 15 for Part B and 12 for Part C. Group weights are derived from each aggregated matrix in the same way as individual

weights, and the CR of every aggregated matrix is checked against the same threshold. The global weight of each sub-criterion is the product of its local weight and the weight of its parent criterion.

$$a_{ij}^{G} = \left(\prod_{k=1}^{K} a_{ij}^{(k)}\right)^{1/K} \quad (2)$$

Part B was collected within two days, and no revision round was held with experts whose matrices exceeded the threshold, so all 15 valid responses were kept in the base aggregation. To test whether individual inconsistency affects the results, the aggregation of Parts B and C was repeated twice: once with only the matrices that meet CR ≤ 0.10 and once with only those that meet CR ≤ 0.20, the more lenient threshold used by certain study [3]. The resulting weights and rankings are compared with the base results.

2.5.3. Consensus and aggregation method

Consistency and consensus are different properties: a group matrix can be consistent even when the experts disagree. Agreement among the experts on the order of the elements in each matrix was therefore measured with Kendall's coefficient of concordance *W*, computed from the ranks implied by each expert's priority vector and tested with its χ^2 approximation. As a check on the aggregation procedure, group priorities were also obtained by aggregating individual priorities (AIP), that is, as the normalised geometric mean of the experts' priority vectors, and compared with the AIJ results.

2.5.4. Evaluation of the initiatives

For each criterion c, the aggregated 5 × 5 matrix of Part C yields a vector of local priorities v_{ic} that sums to one across the initiatives. The overall priority of initiative i is on Equation (3), where w_c is the weight of criterion c from Part B. This is the distributive mode of AHP synthesis. Because the outcome of the distributive mode can depend on the set of alternatives, the ranking was also computed in the ideal mode, in which each local vector is divided by its largest element before weighting.

$$P_i = \sum_{c=1}^{4} w_c \; v_{ic} \quad (3)$$

2.5.5. Comparison with current practice

The AHP ranking is compared with a ranking of the same initiatives by realised annual saving, which reproduces the monetary logic of the plant's Pareto practice. Agreement is summarised with Spearman's rank correlation and examined initiative by initiative. With five alternatives, ρ is reported descriptively; no significance test is attempted, because a test based on five ranks has almost no power.

2.5.6. Sensitivity analysis

Each criterion weight in turn is raised and lowered by 10% and 20%, while the remaining three are scaled in proportion so that the total stays at one. For each of these 16 scenarios, the global weights of the sub-criteria and the overall priorities of the initiatives are recalculated. The stability of the sub-criteria is reported as Spearman's ρ with the base order and as the number of changes among the top five, and the stability of the initiatives as their ranks. Equal criterion weights serve as a reference scenario. Because only 12 of the 15 experts evaluated the initiatives, the ranking was also recomputed with criterion weights derived from these 12 experts alone (matched panel).

2.5.7. Critical weights

For the initiatives, the smallest relative change in a single criterion weight that alters the order of any two initiatives, and the smallest change that displaces the top-ranked initiative, were located by a fine grid search with the same proportional rescaling. These thresholds show how far the judgements of the panel would have to shift before the decision changes.

2.5.8. Bootstrap over experts

The results also depend on who sits on the panel. To estimate this source of uncertainty, 2,000 panels of 12 experts were drawn with replacement from the 12 experts who completed both parts. For each drawn panel, the aggregation of the criterion matrix and of the Part C matrices and the synthesis were repeated, and the share of panels in which each initiative held each rank was recorded.

The calculations were carried out in a spreadsheet and repeated independently in Python, which was also used for the critical-weight and bootstrap analyses. The spreadsheet uses the approximate method (normalised columns averaged by row), whereas this paper reports eigenvector weights; the two sets of weights differ by less than 0.01 for every element and give the same order of criteria.

2.6. Validity, reliability and ethics

Content validity of the hierarchy was supported by its grounding in the literature and the plant's own documents and by the practitioner review in Step 2. Reliability of the judgements was assessed through the consistency ratio of the individual and aggregated matrices, through the agreement among experts and through the robustness checks described under Data analysis. Company data are anonymised, monetary values are expressed as percentages or indices, and initiative names are replaced with generic descriptions. Participation was voluntary, responses were kept confidential, and the company approved the study. Because the first author is employed by Company X, she was not a member of the expert panel, the experts completed the questionnaire on their own, and the fact sheet was compiled from plant records rather than from the authors' assessment, to limit researcher influence.

3. Result and Discussion

3.1. Respondent profile

The panel of 15 experts (Supplementary Table S5) combines four division heads with eleven department heads from eight functions; eight experts come from quality assurance, production and maintenance, the functions that run TPM on the shop floor, and three work at Plant A. Experience at the company ranges from 3 to 22 years (median 10), and 14 experts had been involved in at least 20 improvement projects. Twelve experts (80%) completed Part C; the other three contribute to the criterion weights only.

3.2. Consistency and agreement of judgements

Table 5 summarises the consistency and agreement of the Part B judgements (individual values in Supplementary Table S1). Individual consistency was low: only 24 of the 75 individual matrices met CR ≤ 0.10 and 38 met CR ≤ 0.20, and only R4, R12 and R14 were consistent in all five matrices. Two patterns account for much of this. The criteria and C2 matrices of R10 (CR 2.72 and 2.03) suggest that the direction of the scale was reversed in some rows, and R3 and R8 show nearly identical CRs in every matrix (about 0.38–0.41), which points to a repeated answering pattern.

The aggregated matrices, by contrast, were all consistent, with CR values between 0.009 and 0.029. When judgements are combined with the geometric mean, individual deviations in opposite directions partly cancel out. Whether the inconsistent responses distort the group weights is examined below under Robustness of the weights.

Agreement among the experts is moderate on the criteria (Kendall's W = 0.45, $p < 0.001$): 8 of the 15 experts place business impact first, 5 place sustainability first, and 10 place implementation feasibility last. Within the criteria, agreement is weaker, with W between 0.18 and 0.34, although it is significant for every group ($p \leq 0.044$). The panel therefore shares a clear view of what matters most and a looser view of how each criterion should be broken down.

Table 5. Consistency and agreement of the Part B judgements by matrix

Matrix	CR of aggregated matrix	Individual matrices with CR ≤ 0.10	Individual matrices with CR ≤ 0.20	Kendall's W	χ^2 (df = 3)	p
Criteria	0.029	4 of 15	7 of 15	0.451	20.28	< 0.001
Sub-criteria of C1	0.013	6 of 15	8 of 15	0.180	8.10	0.044
Sub-criteria of C2	0.009	4 of 15	7 of 15	0.240	10.78	0.013
Sub-criteria of C3	0.014	5 of 15	8 of 15	0.337	15.16	0.002
Sub-criteria of C4	0.018	5 of 15	8 of 15	0.254	11.42	0.010

3.3. Criterion weights (RQ2)

Table 6 presents the aggregated criterion weights. Business impact received the highest weight (0.448), followed by sustainability (0.285), operational urgency (0.177) and implementation feasibility (0.090). The two benefit-oriented criteria together account for 73.3% of the weight, whereas feasibility carries about one-fifth of the weight of business impact.

Table 6. Aggregated criterion weights

Criterion	Weight	Rank
C1 Business impact	0.448	1
C2 Operational urgency	0.177	3
C3 Implementation feasibility	0.090	4
C4 Sustainability	0.285	2
λ_{max} = 4.078; CI = 0.026; CR = 0.029		

3.4. Sub-criterion weights (RQ1 and RQ2)

Table 7 lists the local and global weights of the 16 sub-criteria. The five sub-criteria with the highest global weights are customer impact (0.188), long-term benefit (0.128), quality (0.106), risk exposure (0.083) and cost reduction (0.080); together they represent 58.6% of the total weight.

Cost reduction (0.080) and productivity (0.074), which express the size of the benefit, each weigh more than twice as much as performance gap (0.030), which expresses the distance to a target; the panel thus kept apart the two aspects that the framework separates. Risk exposure dominates operational urgency (local weight 0.469), technical feasibility (0.472) dominates implementation feasibility, investment need has the lowest global weight of all (0.009), and long-term benefit (0.450) leads sustainability. All four within-group matrices were consistent.

In answer to RQ1, all 16 sub-criteria received a non-zero weight and were retained. Seven of them, however, each carry less than 4% of the total weight: problem frequency, problem severity, performance gap, investment need, resource availability, complexity and monitoring. These have little influence on the evaluation of initiatives.

Table 7. Local and global weights of sub-criteria

Sub-criterion	Local weight	Global weight	Global rank
C1 Business impact (CR = 0.013)			
C1.1 Cost reduction	0.179	0.080	5
C1.2 Productivity	0.165	0.074	6
C1.3 Quality	0.237	0.106	3
C1.4 Customer impact	0.419	0.188	1
C2 Operational urgency (CR = 0.009)			
C2.1 Problem frequency	0.169	0.030	13
C2.2 Problem severity	0.192	0.034	11
C2.3 Performance gap	0.169	0.030	12
C2.4 Risk exposure	0.469	0.083	4
C3 Implementation feasibility (CR = 0.014)			
C3.1 Investment need	0.102	0.009	16
C3.2 Technical feasibility	0.472	0.042	9
C3.3 Resource availability	0.226	0.020	14
C3.4 Complexity	0.201	0.018	15
C4 Sustainability (CR = 0.018)			
C4.1 Standardisation	0.231	0.066	7
C4.2 Replicability	0.195	0.056	8
C4.3 Monitoring	0.124	0.035	10
C4.4 Long-term benefit	0.450	0.128	2

3.5. Robustness of the weights (RQ2)

Across the 16 weight perturbations (Supplementary Table S4), the order of the criteria never changed, Spearman's ρ with the base order of the sub-criteria ranged from 0.974 to 1.000, and the top five changed by at most one element, always at fourth or fifth place. Equal criterion weights, by contrast, give a clearly different order (ρ = 0.547). The expert weights therefore matter, but the result does not hinge on their exact values.

Table 8 addresses the inconsistency reported above. With matrices of CR ≤ 0.20 only (seven or eight experts per matrix), the order of the criteria and the top five sub-criteria are unchanged (ρ = 0.891). With CR ≤ 0.10, only four experts remain for the criteria, and sustainability (0.362) moves ahead of business impact (0.304; ρ = 0.818). This reversal rests on one expert: three of the four rank business impact first, whereas R15 gives sustainability a weight of 0.640 and business impact 0.063. Without R15, business impact regains first place (0.423 against 0.252), whereas removing any of the other three leaves sustainability ahead. The aggregation procedure has no effect: aggregating individual priorities (AIP) gives the same order of all 16 sub-criteria (ρ = 1.000), and no local weight differs by more than 0.011. The robust findings are therefore the leading position of business impact, followed by sustainability, and the low weight of feasibility.

Table 8. Criterion weights under alternative aggregations

Aggregation	Experts per matrix	C1	C2	C3	C4	CR (criteria)	ρ with base order of sub-criteria	Top-five changes
All valid responses (base)	15	0.448	0.177	0.090	0.285	0.029	1.000	0
Matrices with CR ≤ 0.10	4–6	0.304	0.187	0.147	0.362	0.014	0.818	1
Matrices with CR ≤ 0.20	7–8	0.438	0.224	0.112	0.227	0.033	0.891	0
Aggregation of individual priorities (AIP)	15	0.443	0.178	0.092	0.286	–	1.000	0

3.6. Evaluation of the initiatives (RQ3)

Table 9 gives the local priorities of the five initiatives under each criterion, based on the 12 experts who completed Part C. A1 leads clearly on business impact (0.475). A4 leads on operational urgency (0.334), followed by A3 (0.247). A2 is judged the easiest to implement (0.348) and leads narrowly on sustainability (0.299, against 0.278 for A1). A3 receives the lowest priorities on business impact (0.098) and sustainability (0.115).

The Part C judgements were far more consistent than those of Part B (Supplementary Table S3): of the 48 individual matrices, 46 meet CR ≤ 0.10, and the aggregated matrices have CR values between 0.002 and 0.009. Agreement among the experts is moderate on business impact (W = 0.47) and operational urgency (W = 0.40), weaker on sustainability (W = 0.25), and not significant on implementation feasibility (W = 0.10, p = 0.29). The experts thus agree on which initiatives matter most to the business and are most urgent, but not on which are easiest to carry out.

Table 9. Local priorities of the initiatives under each criterion (12 experts)

Initiative	C1 Business impact	C2 Operational urgency	C3 Implementation feasibility	C4 Sustainability
A1 Yield loss reduction	0.475	0.189	0.132	0.278
A2 Filter-stock rationalisation	0.108	0.075	0.348	0.299
A3 Vacuum-pump breakdown elimination	0.098	0.247	0.212	0.115
A4 Quality reject reduction	0.159	0.334	0.172	0.177
A5 Changeover reduction	0.160	0.155	0.137	0.131
CR of aggregated matrix	0.007	0.006	0.002	0.009
Individual matrices with CR ≤ 0.10 (of 12)	11.000	12..000	12.000	11.000
Kendall's W (p)	0.472 (< 0.001)	0.404 (< 0.001)	0.104 (0.290)	0.251 (0.017)

3.7. Overall ranking and comparison with current practice (RQ3)

Table 10 combines the local priorities with the criterion weights. A1 ranks first with an overall priority of 0.337, of which 63% comes from business impact. A4 follows in second place (0.196) although its realised saving is the smallest of the five; its contribution through operational urgency (0.059) is the highest of all initiatives, and its contribution through business impact (0.071) is exceeded only by that of A1 and, marginally, A5. A2 takes third place (0.178), supported mainly by sustainability (0.085, the highest contribution of any initiative under this criterion). A5 (0.149) and A3 (0.140) close the ranking.

The AHP ranking agrees only weakly with the ranking by realised saving ($\rho = 0.40$). Both put A1 first, but A4 moves from last to second place, whereas A2, with the second-largest saving, drops to third because the experts gave it the second-lowest priority on business impact (0.108).

Table 10. Weighted contributions, overall priorities and ranks of the initiatives

Initiative	C1	C2	C3	C4	Overall priority P_i	AHP rank	Saving index	Saving rank
A1 Yield loss reduction	0.213	0.034	0.012	0.079	0.337	1	100.0	1
A4 Quality reject reduction	0.071	0.059	0.015	0.050	0.196	2	3.3	5
A2 Filter-stock rationalisation	0.048	0.013	0.031	0.085	0.178	3	27.6	2
A5 Changeover reduction	0.072	0.027	0.012	0.037	0.149	4	6.7	3
A3 Vacuum-pump breakdown elimination	0.044	0.044	0.019	0.033	0.140	5	3.7	4
Criterion weight w_c	0.448	0.177	0.090	0.285	1.000			

3.8. Robustness of the initiative ranking (RQ3*)*

Table 11 reports the ranks of the initiatives under the alternative scenarios. The ranking is identical in all 16 weight perturbations, in the ideal mode of synthesis, with the criterion weights of the 12 Part C experts and when only matrices with CR ≤ 0.20 are aggregated. A1 is first in every scenario. Two changes occur elsewhere. With equal criterion weights, A3 and A5 swap places. With only the most consistent matrices (CR ≤ 0.10), A2 and A4 swap second and third place and A3 and A5 swap the last two places. This change stems from the criterion weights of the four most consistent experts, which favour sustainability because of one expert's strong preference for it; filtering the Part C matrices alone leaves the base ranking unchanged.

The critical-weight analysis locates these boundaries. The order of A2 and A4 reverses when the weight of sustainability rises by 33% (to about 0.38), when operational urgency falls by 35% (to about 0.11) or when implementation feasibility rises by 95%. A3 and A5 swap when business impact falls by 21% or operational urgency rises by 41%. A1 loses first place only if business impact falls by 95%, or if the weight of operational urgency, feasibility or sustainability more than triples (to about 0.58, 0.48 and 0.92, respectively).

Table 11. Ranks of the initiatives under alternative scenarios

Scenario	A1	A2	A3	A4	A5
Base (AHP weights, distributive mode)	1	3	5	2	4
16 of 16 weight perturbations (±10%, ±20%)	1	3	5	2	4
Equal criterion weights	1	3	4	2	5
Ideal mode of synthesis	1	3	5	2	4
Criterion weights of the 12 Part C experts (matched panel)	1	3	5	2	4
Only matrices with CR ≤ 0.10	1	2	4	3	5
Only matrices with CR ≤ 0.20	1	3	5	2	4

The bootstrap over experts (Table 12) gives a similar picture. A1 is ranked first in 99.9% of the 2,000 resampled panels. A4 is second in 70.0% of the panels and A2 third in 57.2%, with the two initiatives swapping places in most of the remaining panels. A5 is fourth in 55.2% of the panels and A3 last in 70.5%. The complete order of all five initiatives is reproduced in 40.2% of the panels. In practical terms, the ranking supports a firm decision on A1, a well-supported but not certain preference for A4 over A2, and a consistent placement of A3 and, less firmly, A5 in the bottom two places.

Table 12. Share of bootstrap panels in which each initiative holds each rank (2,000 replications)

Initiative	Rank 1 (%)	Rank 2 (%)	Rank 3 (%)	Rank 4 (%)	Rank 5 (%)	Median rank	90% interval
A1 Yield loss reduction	99.9	0.1	0.0	0.0	0.0	1	1–1
A2 Filter-stock rationalisation	0.0	29.3	57.2	13.5	0.0	3	2–4
A3 Vacuum-pump breakdown elimination	0.0	0.0	1.1	28.4	70.5	5	4–5
A4 Quality reject reduction	0.1	70.0	26.2	2.9	0.9	2	2–3
A5 Changeover reduction	0.0	0.6	15.6	55.2	28.6	4	3–5

3.9. Discussion

3.9.1. Interpretation of the findings

The criterion weights show that the experts judge FI initiatives above all by what they deliver. Business impact alone carries 45% of the weight and, together with sustainability, 73%, whereas implementation feasibility receives only 9%; this contradicts the benefit–effort matrix used in the plant, where effort counts as much as benefit. A single dominant criterion is common in organisational AHP; price carried 58.5% of the weight in an Indonesian vendor-selection study [31]. The low weight of feasibility also qualifies the finding that simple, low-cost measures can rank highest [4], [11]. Senior managers appear to treat ease of implementation as a resourcing question after selection, not as a reason for selection.

The sub-criterion weights show what business impact means to the panel. The monetary sub-criteria, cost reduction and productivity, together carry 15.4% of the total weight, whereas customer impact and quality carry 29.4%. The panel thus partly endorses the monetary logic of cost deployment but places the customer and the risk carried by the business above it [15], [16]. This agrees with study which AHP study of TPM indicators weighted productivity and quality highest [26], and with the customer perspective of the Balanced Scorecard [20] and *hoshin kanri* [21]. The dominance of technical feasibility over investment need suggests that the experts regard know-how, rather than capital, as the binding constraint on FI work.

These preferences change actual decisions. The yield-loss initiative, which targets the largest controllable loss outside OEE, ranked first in almost every bootstrap panel, so a prioritisation anchored in OEE would overlook the most valuable work [28]. The quality initiative A4 rose to second place because of the urgency and customer risk it removes, whereas the maintenance initiative A2 fell to third. The weak agreement with the monetary ranking ($\rho = 0.40$) therefore reflects the panel's view that a saving on maintenance spending is worth less than the removal of a defect that customers can see.

Three features of the judgements qualify these conclusions. First, many experts were individually inconsistent in Part B, mostly because they used extreme scale values, yet the aggregated matrices were consistent, and business impact remained first under every check except the four-expert subset, in which one expert reversed the order. This supports reporting results at more than one consistency threshold when no revision round is possible [3]. Second, the Part C judgements were almost all consistent, which suggests that concrete initiatives described in a fact sheet are easier to judge than abstract criteria. Third, consensus was moderate on the criteria, weaker within them and absent for feasibility in Part C. Both AIJ and AIP can mislead when consensus or consistency is poor [35], so it matters that they agreed here; the weak consensus within criteria still calls for a facilitated discussion of the sub-criterion weights before use.

3.9.2. Theoretical implications

The study makes three contributions to theory. First, it moves the use of multi-criteria methods in TPM from ranking success factors, KPIs and drivers to ranking the initiatives that the FI pillar must choose, with a criterion set that separates the size of the benefit from urgency, feasibility and durability [2], [25], [26]. Second, it shows empirically that the managers responsible for these decisions value customer impact, quality and risk above the monetary measures on which cost deployment and Pareto practice rely, and that these preferences reorder a real portfolio; TPM prioritisation built on one yardstick, loss cost or OEE, captures only part of how practitioners value improvement. Third, it shows that consistency and consensus must be assessed separately in group AHP, because a consistent group matrix can hide inconsistent individuals and weak agreement; reporting both, with critical weights and a bootstrap, shows which parts of a group decision are firm.

3.9.3. Practical implications

For plants that run TPM, the results suggest four practices. Loss cost should inform the strategic deployment workshop without deciding the order on its own, and feasibility should be settled as a resourcing matter after selection. The weights should be set once a year, with a consensus round where agreement is weak and with forms that show experts their consistency ratio as they answer. Finally, rankings should be reported with their robustness, so that near-ties are resolved openly.

4. Conclusion

This study developed an AHP framework for prioritising FI initiatives in a TPM-based pipe manufacturing plant and tested it with 15 division and department heads, 12 of whom also evaluated five initiatives from the plant's 2026 improvement portfolio. For RQ1, the hierarchy of four criteria and 16 sub-criteria proved usable, although seven sub-criteria carry less than 4% of the weight each.

For RQ2, business impact received the highest weight (0.448), followed by sustainability (0.285), operational urgency (0.177) and implementation feasibility (0.090), and the aggregated weights were consistent and robust. For RQ3, the AHP ranking agreed only weakly with the monetary ranking ($\rho = 0.40$): the yield-loss initiative ranked first in 99.9% of bootstrap panels, and a quality initiative with the smallest realised saving rose from last to second place. The study has limitations that point to further research. It examines a single plant, the weights were set by a division-level panel whose members also work at other plants, many Part B matrices were inconsistent without a revision round, and three experts did not complete Part C. The evaluation covered five initiatives in relative mode; because they had already been implemented, the experts knew their realised savings, which may have influenced their judgements, and the savings capture only monetary effects. Future research should apply the framework ex ante and in absolute mode to the full portfolio, repeat the weighting with a revision round, compare AHP with the Best-Worst Method, extend the panel to other companies, and test whether higher-ranked initiatives deliver larger benefits.

References

[1] Fallahiarezoudar, E., Ahmadipourroudposht, M., Bagherian Rafi, M. H., & Ngadiman, N. H. A. (2025). A systematic approach of maintenance 4.0 towards a sustainable manufacturing policy: A case study on an automobile company. *Process Integration and Optimization for Sustainability, 9*(4), 1229–1251. https://doi.org/10.1007/s41660-025-00505-y

[2] Samadhiya, A., Agrawal, R., & Garza-Reyes, J. A. (2024). Integrating Industry 4.0 and total productive maintenance for global sustainability. *The TQM Journal, 36*(1), 24–50. https://doi.org/10.1108/TQM-05-2022-0164

[3] Nam, H., Cho, Y., Kim, J., Joo, Y., & Nam, H. (2025). An analytic hierarchy process for the sustainable development of small- and medium-sized enterprises in the manufacturing sector, South Korea. *Environment, Development and Sustainability, 27*(8), 20153–20180. https://doi.org/10.1007/s10668-024-05860-3

[4] Utama, D. M., & Abirfatin, M. (2023). Sustainable Lean Six-sigma: A new framework for improve sustainable manufacturing performance. *Cleaner Engineering and Technology, 17*. https://doi.org/10.1016/j.clet.2023.100700

[5] Chohan, I. M., Ahmad, A., Bheel, N., Najeh, T., & Almaliki, A. H. (2025). Sustainability assessment of different pipeline materials in freshwater supply systems. *Frontiers in Materials, 12*. https://doi.org/10.3389/fmats.2025.1566151

[6] Badan Pusat Statistik. (2026). *Ekonomi Indonesia tahun 2025 tumbuh 5,11 persen*. Retrieved from https://www.bps.go.id/assets/pressrelease/2026/02/05/2546/ekonomi-indonesia-tahun-2025-tumbuh-5-11-persen.html

[7] Ministry of Industry of the Republic of Indonesia. (2026). *Industri pengolahan mendominasi kinerja ekonomi nasional*. Retrieved from https://sippa.kemenperin.go.id/news/post/2026/02/20/37/industri-pengolahan-mendominasi-kinerja-ekonomi-nasional

[8] Tirta, E. B. E. (2026). *Impor plastik RI tembus Rp 46,3 triliun, siapa negara pemasoknya?* CNBC Indonesia. Retrieved from https://www.cnbcindonesia.com/research/20260406115504-128-724138/impor-plastik-ri-tembus-rp-463-triliun-siapa-negara-pemasoknya

[9] Dewi, S., Febrianti, R., & Utama, D. M. (2023). An integrated method for manufacturing sustainability assessment in tire industry: A case study in Indonesian. *International Journal of Sustainable Engineering, 16*(1), 1–12. https://doi.org/10.1080/19397038.2023.2276936

[10] Rosiani, T. Y., Utama, D. M., Ummudiyah, A. Y., & Amallynda, I. (2024). Sustainable manufacturing assessment using sustainable VSM and AHP involving workload and machine efficiency: A case study in Indonesian paving block production. *Circular Economy and Sustainability, 4*(3), 2049–2071. https://doi.org/10.1007/s43615-024-00366-6

[11] Rosiani, T. Y., Septian, R. I., & Masudin, I. (2026). Integrating Lean Six Sigma, Delphi, AHP and TOPSIS for waste minimization in sustainable manufacturing: A case of concrete paving block (CPB) production. *Circular Economy and Sustainability, 6*(3), Article 217. https://doi.org/10.1007/s43615-026-00947-7

[12] Di Luozzo, S., Starnoni, F., & Schiraldi, M. M. (2023). On the relationship between human factor and overall equipment effectiveness (OEE): An analysis through the adoption of analytic hierarchy process and ISO 22400. *International Journal of Engineering Business Management, 15*. https://doi.org/10.1177/18479790231188548

[13] Japan Institute of Plant Maintenance. (2016). *Monodzukuri test learning textbook*. JIPM.

[14] Mouhib, Z., Gallab, M., Merzouk, S., Soulhi, A., & Di Nardo, M. (2025). Total productive maintenance and Industry 4.0: A literature-based path toward a proposed standardized framework. *Applied System Innovation, 8*(4). https://doi.org/10.3390/asi8040098

[15] Bertolini, M., & Castellano, D. (2026). Quality cost deployment (QCD): A lean-inspired methodology for systematic reduction of total quality costs. *International Journal of Lean Six Sigma, 17*(8), 133–160. https://doi.org/10.1108/IJLSS-09-2025-0252

[16] Hagström, M. H., Bergsjö, D. H., Sathyanarayana, A., & Machado, C. (2023). Quantifying and visualising wastes and losses in automotive production flows (across multiple plants and organisations) for increased accuracy in improvement prioritisations. *International Journal of Product Development, 27*(3), 245–264. https://doi.org/10.1504/IJPD.2023.133062

[17] Salomon, V. A. P., & Gomes, L. F. A. M. (2024). Consistency improvement in the analytic hierarchy process. *Mathematics, 12*(6). https://doi.org/10.3390/math12060828

[18] Srđević, B., & Srđević, Z. (2026). Incomplete judgments in AHP: Transition-based approaches, aggregation strategies, and their impact on decision outcomes. *Algorithms, 19*(1). https://doi.org/10.3390/a19010002

[19] Madzík, P., & Falát, L. (2022). State-of-the-art on analytic hierarchy process in the last 40 years: Literature review based on latent Dirichlet allocation topic modelling. *PLOS ONE, 17*(5), Article e0268777. https://doi.org/10.1371/journal.pone.0268777

[20] Kumar, J., Prince, N., & Baker, H. K. (2022). Balanced scorecard: A systematic literature review and future research issues. *FIIB Business Review, 11*(2), 147–161. https://doi.org/10.1177/23197145211049625

[21] Wilson, R., Cudney, E. A., & Marley, R. J. (2024). Current status of Hoshin Kanri. *The TQM Journal, 36*(2), 460–483. https://doi.org/10.1108/TQM-07-2022-0216

[22] Pakdil, F. (2022). Six Sigma project prioritization and selection methods: A systematic literature review. *International Journal of Lean Six Sigma, 13*(2), 382–413. https://doi.org/10.1108/IJLSS-01-2021-0001

[23] Nakielski, M., & Ginda, G. (2026). AHP in design for Six Sigma project selection. *Sustainability, 18*(11). https://doi.org/10.3390/su18115258

[24] Prasojo, R. A., Zainal, M. R., Handani, G. P. C., Hakim, M. F., Wijaya, B. H., Ghoneim, S. S. M., Mahmoud, K., Lehtonen, M., & Darwish, M. M. F. (2025). Enhanced maintenance prioritization strategies for medium voltage distribution substations through precise health index assessment. *Sustainable Energy, Grids and Networks, 44*. https://doi.org/10.1016/j.segan.2025.101987

[25] Singh, R. K., & Gurtu, A. (2022). Prioritizing success factors for implementing total productive maintenance (TPM). *Journal of Quality in Maintenance Engineering, 28*(4), 810–830. https://doi.org/10.1108/JQME-09-2020-0098

[26] Toke, L. K., & Kalpande, S. D. (2023). An assessment of key performance indicators and its relationship for implementation of total productive maintenance in manufacturing sector. *International Journal on Interactive Design and Manufacturing, 17*(4), 1741–1753. https://doi.org/10.1007/s12008-023-01252-5

[27] Slavina, T., & Štefanić, N. (2024). Facing challenges of implementing total productive management and lean tools in manufacturing enterprises. *Systems, 12*(2), Article 52. https://doi.org/10.3390/systems12020052

[28] Lucantoni, L., Antomarioni, S., Ciarapica, F. E., & Bevilacqua, M. (2025). A data-driven framework for supporting the total productive maintenance strategy. *Expert Systems with Applications, 268*, Article 126283. https://doi.org/10.1016/j.eswa.2024.126283

[29] Swarnakar, V., Singh, A. R., Antony, J., Tiwari, A. K., & Garza-Reyes, J. A. (2023). Sustainable Lean Six Sigma project selection in manufacturing environments using best-worst method. *Total Quality Management & Business Excellence, 34*(7–8), 990–1014. https://doi.org/10.1080/14783363.2022.2139675

[30] Kastratović, N., Arsić, D., Komatina, N., Delić, M., & Marinković, D. (2026). Prioritization of process improvement measures in a forging process using an IPF-AHP-based SREM framework. *Journal of Manufacturing and Materials Processing, 10*(7), Article 250. https://doi.org/10.3390/jmmp10070250

[31] Hermeindito, H., Kristianto, L. S., & Christian, T. F. (2026). Vendor selection system in the system integration company PT. Visiniaga in Surabaya using the analytic hierarchy process (AHP) method. *Jurnal Manajemen Teknologi, 25*(1), 1–21. https://doi.org/10.12695/jmt.2026.25.1.1

[32] Ecer, F. (2024). A state-of-the-art review of the BWM method and future research agenda. *Technological and Economic Development of Economy, 30*(4), 1165–1204. https://doi.org/10.3846/tede.2024.20761

[33] Mazurek, J., Perzina, R., Strzałka, D., Kowal, B., Kuraś, P., Petrů Puhrová, B., & Rajs, R. (2024). Is the best–worst method path dependent? Evidence from an empirical study. *4OR, 22*(3), 387–409. https://doi.org/10.1007/s10288-023-00553-5

[34] Srđević, B., & Srđević, Z. (2024). Multi-model assessing and visualizing consistency and compatibility of experts in group decision-making. *Mathematics, 12*(11), https://doi.org/10.3390/math12111699

[35] Lin, C., Kou, G., Peng, Y., & Alsaadi, F. E. (2022). Aggregation of the nearest consistency matrices with the acceptable consensus in AHP-GDM. *Annals of Operations Research, 316*(1), 179–195. https://doi.org/10.1007/s10479-020-03572-1

[36] Tu, J., & Wu, Z. (2025). Analytic hierarchy process rank reversals: Causes and solutions. *Annals of Operations Research, 346*(2), 1785–1809. https://doi.org/10.1007/s10479-023-05278-6