African Journal of Advanced Pure and Applied Sciences (AJAPAS)
Online ISSN: 2957-644X
Volume 5, Issue 3, 2026
Page No: 75-89
Website: https://aaasjournals.com/index.php/ajapas/index
ISI 2025: 1.126 SJIFactor 2024: 6.752 معامل التأثير العربي: 1.62

Developing an AI-Based Digital Twin for Predictive Quality Control in 3D Printing Processes

Abdussalam Ali Ahmed [1*], Rafat S. A. Abumandil [2], Omar Ahemed Mohamed Edbeib [3]
[1,2,3] Mechanical and Industrial Engineering Department, Bani Waleed University, Libya

تطوير توأم رقمي قائم على الذكاء الاصطناعي لمراقبة الجودة التنبؤية في عمليات الطباعة ثلاثية الأبعاد

عبد السلام علي أحمد [1*]، رأفت سالم أبو منديل [2]، عمر أحمد محمد الدبيب [3]
[1، 2، 3] قسم الهندسة الميكانيكية والصناعية، جامعة بني وليد، ليبيا

[*]Corresponding author: abdussalam.a.ahmed@gmail.com

Received: April 18, 2026	Accepted: June 30, 2026	Published: July 15, 2026

Abstract
Additive manufacturing can produce complex parts, yet print quality often changes between runs. Small changes in temperature, speed, material, or layer settings may cause rough surfaces and weak parts. This paper develops an AI-based digital twin for predictive quality control in 3D printing. The twin joins machine data, process settings, quality models, and operator decisions within one synchronized system. It predicts surface roughness, tensile strength, and elongation before final inspection. A public material-extrusion dataset provides the practical test. The dataset contains 70 printing runs from an Ultimaker S5 printer. Nine process settings and three measured quality outcomes are available. One invalid record was removed before model training. Ridge regression, random forest, gradient boosting, and extra trees were evaluated. Repeated five-fold cross-validation reduced dependence on one favorable data split. Gradient boosting achieved the lowest average normalized error across the three outcomes. Its out-of-fold R^2 reached 0.90 for roughness, 0.53 for tensile strength, and 0.54 for elongation. The mean absolute errors were 21.49 µm, 4.50 MPa, and 0.39%, respectively. A replay test converted these predictions into simple quality alerts. The alert logic reached 79.1% balanced accuracy under illustrative acceptance limits. Results show useful predictive value, but they do not prove industrial readiness. Larger datasets, sensor streams, uncertainty limits, and printer-specific validation remain necessary. The proposed structure offers a practical path from offline machine learning toward safe, traceable, closed-loop quality control.

Keywords: additive manufacturing; 3D printing; digital twin; machine learning; predictive quality control; fused filament fabrication; process monitoring.

الملخص
يمكن للتصنيع الإضافي إنتاج أجزاء معقدة، ومع ذلك فإن جودة الطباعة غالباً ما تتغير بين دورة تشغيل وأخرى. فالأخطاء أو التغيرات الطفيفة في درجات الحرارة، أو السرعة، أو المادة، أو إعدادات الطبقات قد تتسبب في أسطح خشنة وأجزاء ضعيفة. تطور هذه الورقة البحثية توأماً رقمياً قائماً على الذكاء الاصطناعي لمراقبة الجودة التنبؤية في الطباعة ثلاثية الأبعاد. يدمج التوأم بين بيانات الآلة، وإعدادات العملية، ونماذج الجودة، وقرارات المشغل داخل نظام واحد متزامن؛ حيث يتنبأ بخشونة السطح، وقوة الشد، والاستطالة قبل إجراء الفحص النهائي.
وقد وفرت مجموعة بيانات عامة لبثق المواد اختباراً عملياً؛ حيث تحتوي مجموعة البيانات هذه على 70 دورة طباعة من طابعة من نوع(Ultimaker S5) ، وتتضمن تسع إعدادات للعملية وثلاث نتائج جودة مقاسة. تم استبعاد سجل واحد غير صالح قبل تدريب النماذج. جرى تقييم كل من انحدار ريدج(Ridge regression) ، والغابة العشوائية(Random forest) ، وتعزيز التدرج (Gradient boosting)، والأشجار الإضافية (Extra trees). كما استُخدم التحقق المتبادل المتكرر الخماسي-(Repeated five-fold cross-validation)لتقليل الاعتماد على تقسيم واحد ملائم للبيانات.
حقق نموذج "تعزيز التدرج (Gradient boosting) "أدنى متوسط خطأ معياري عبر النتائج الثلاثة؛ حيث بلغ معامل التحديد (R^2) خارج العينة 0.90 (Out-of-fold) للخشونة، و0.53 لقوة الشد، و0.54 للاستطالة. كما بلغت قيم متوسط الخطأ المطلق 21.49

ميكرومتر، و4.50 ميجاباسكال، و0.39% على التوالي. وحوّل اختبار إعادة التشغيل (Replay test) هذه التنبؤات إلى تنبيهات جودة بسيطة، حيث وصلت دقة منطق التنبيه المتوازنة إلى 79.1% تحت حدود قبول توضيحية.
تظهر النتائج قيمة تنبؤية مفيدة، لكنها لا تثبت الجاهزية الصناعية التامة؛ إذ لا تزال هناك حاجة إلى مجموعات بيانات أكبر، وتدفقات بيانات المستشعرات، وحدود عدم اليقين، والتحقق الخاص بكل طابعة. يقدم الهيكل المقترح مساراً عملياً للانتقال من التعلم الآلي غير المتصل بالإنترنت (Offline) نحو مراقبة جودة آمنة، وقابلة للتتبع، وذات حلقة مغلقة(Closed-loop).

الكلمات المفتاحية: التصنيع الإضافي؛ الطباعة ثلاثية الأبعاد؛ التوأم الرقمي؛ التعلم الآلي؛ مراقبة الجودة التنبؤية؛ تصنيع الخيوط المنصهرة؛ مراقبة العمليات.

Introduction

Additive manufacturing builds a part from digital geometry through successive material layers. The method supports complex forms, local production, and rapid design changes. It now serves aerospace, medical, energy, tooling, and consumer markets. Recent work also extends printing toward electronics and functional materials (Chen et al., 2025). These strengths increase the need for dependable quality control.

Print quality remains difficult to maintain across machines and production runs. Many parameters act together during each build. These include temperature, speed, layer height, feed rate, cooling, orientation, and material condition. Their effects may change with geometry and equipment wear. A setting that works for one part may fail on another. Traditional inspection usually finds defects after material and time have already been spent.

Current inspection methods include visual checks, microscopy, tensile testing, surface measurement, thermography, and X-ray computed tomography. These methods provide valuable evidence. However, many tests are slow, expensive, destructive, or unsuitable for every produced part. Predictive quality control addresses this timing problem. It estimates quality while the process can still be changed.

Artificial intelligence can learn links between process inputs and measured outcomes. Machine learning has supported parameter selection, defect recognition, property prediction, and process control (Wang et al., 2020). Image models can also detect printing errors during material deposition. Brion and Pattinson (2022) demonstrated broad error detection across several printers, geometries, colors, and lighting conditions. Their public dataset contains more than one million labeled images.

A predictive model alone is not a digital twin. A digital twin must remain linked with its physical counterpart. It should update its state when the printer state changes. It also needs identity, time, model version, and decision history. Scime et al. (2022) described a scalable platform that joins additive manufacturing data throughout the component lifecycle. Lu et al. (2024) further showed that twin quality depends on data, models, and computing services.

Several reviews describe digital twins for additive manufacturing. They discuss virtual models, sensors, synchronization, and process control (Ben Amor et al., 2024; Jyeniskhan et al., 2024). Yet many studies remain conceptual. Others focus on one sensor, one printer, or one process stage. Rojek et al. (2025) identified continuing problems with data scarcity and weak cross-machine generalization. Altun et al. (2025) reported similar barriers for intelligent control and industrial certification.

This paper connects these research areas through one practical design. It presents an AI-based digital twin for predictive quality control. The architecture covers data capture, state estimation, quality prediction, alerts, and safe feedback. A public material-extrusion dataset supports the experimental study. The experiment predicts roughness, tensile strength, and elongation from common print settings.

The study makes three contributions. First, it defines a traceable digital-twin structure for quality prediction. Second, it compares four models through repeated cross-validation. Third, it tests an alert rule through an out-of-fold production replay. The result is a reproducible proof of concept, not an industrial certification study.

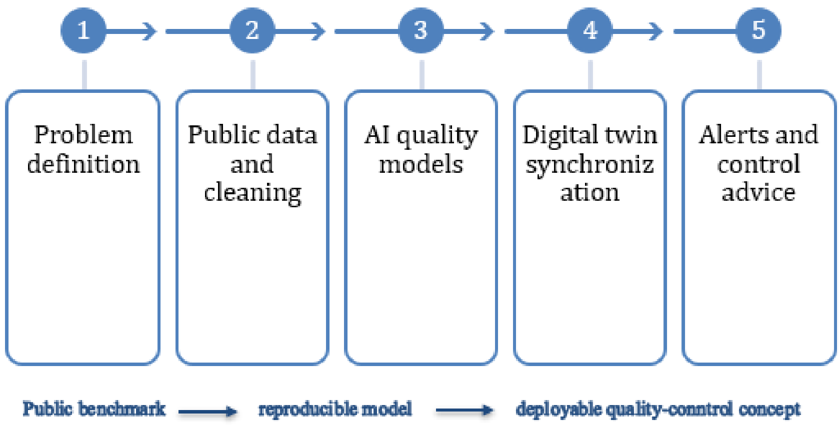

Figure 1 Research design used for the proposed digital twin and public-data experiment.

Figure 1 summarizes the research flow. The design starts with a clear quality problem and a public dataset. It then compares predictive models. The selected model becomes one service inside the digital twin. Its output supports warnings and parameter advice.
The remaining paper has seven sections. Section 2 reviews quality control, machine learning, and digital twins. Section 3 explains the proposed architecture. Section 4 describes the public dataset and experimental method. Section 5 presents results. Section 6 discusses meaning, risks, and industrial use. Section 7 gives a practical validation plan. Section 8 concludes the paper.

Related Research

Quality variation in 3D printing

Additive manufacturing includes material extrusion, powder bed fusion, binder jetting, vat photopolymerization, material jetting, and directed energy deposition. Each process uses different energy and material mechanisms. However, all processes convert digital instructions into physical layers. Any error within this chain can alter the final part.
Material extrusion forms roads of softened material through a moving nozzle. Quality depends on bonding, cooling, flow stability, and geometric placement. Low temperature may weaken bonding between adjacent roads. Excessive temperature can increase deformation or stringing. High speed may reduce deposition stability. Large layer height can increase roughness and change mechanical behavior.
Metal powder bed fusion has different signals and failure modes. Melt-pool temperature, plume behavior, recoating quality, spatter, and scan history become important. Thermal measurements can reveal unusual energy input or cooling. Powder-bed images can reveal streaks, incomplete spreading, and contamination. NIST AM-Bench provides controlled data for testing these relations (NIST, 2022).
The final quality state is therefore multi-dimensional. A part may have low roughness but inadequate strength. Another part may meet strength limits but miss dimensional tolerances. A useful twin should preserve separate outcomes. A single score may hide an unsafe trade-off.
Tran (2025) reviewed recent developments across printing technologies and materials. The review showed continuing growth in AI-supported parameter prediction. Chen et al. (2025) also noted the value of automation for complex printing workflows. These studies support a process-wide view of quality.

Artificial intelligence for quality prediction

Machine learning maps measured inputs to known outputs. In additive manufacturing, inputs may describe settings, images, acoustic signals, or thermal histories. Outputs may represent defect classes, dimensions, density, roughness, or mechanical properties. The model learns these links from past builds.
Linear models provide simple and readable relationships. They are useful when effects are nearly additive. However, printing processes often contain thresholds and interactions. Tree ensembles can represent these patterns without a fixed equation. Neural networks can learn complex image and time-series features. Their success usually requires larger and more varied datasets.
Wang et al. (2020) reviewed machine learning across additive manufacturing. The authors stressed both predictive value and interpretability problems. Kim et al. (2023) reviewed image-based fault monitoring. They found strong progress across material extrusion and metal processes. However, lighting, camera position, machine changes, and limited labels still affect transfer.
Brion and Pattinson (2022) developed a multi-head neural network for extrusion error detection and correction. Their system learned material flow, speed, and nozzle offset from images. The work also included closed-loop correction. It provides strong evidence for camera-based control when representative training data exist.
Bast et al. (2025) released another public image dataset for fused filament fabrication. It contains normal parts, extrusion defects, and layer shifts.
Tabular models remain useful when images are unavailable. Process settings exist in most printer logs or build files. Tatar (2025) compared several regression models for roughness, strength, and elongation. Gaussian process regression gave strong results on the studied dataset. Yet small tabular datasets can produce unstable accuracy estimates. Repeated cross-validation offers a safer evaluation than one train-test split.

Digital twins in additive manufacturing

A digital model becomes a twin through synchronized data exchange. The physical system sends measurements and state changes to the digital system. The digital system updates predictions and decisions. A return channel may send recommendations or approved control changes.
The literature uses several twin levels. A digital model has no automatic data link. A digital shadow receives automatic data from the physical process. A digital twin supports meaningful two-way exchange. This distinction matters because many monitoring dashboards are incorrectly called digital twins.
Scime et al. (2022) proposed a scalable platform for component-level twin data. Their design joins design intent, machine data, process measurements, and part inspection. This traceability supports later qualification. Ben Amor

et al. (2024) reviewed the steps needed for additive manufacturing twins. These steps include acquisition, modeling, calibration, validation, creation, and real-time use.
Zhang et al. (2020) organized additive manufacturing twins around models, data, and services. Gaikwad et al. (2020) joined thermal simulation, sensing, and analytics for fault detection. Corradini and Silvestri (2022) tested a twin for material-extrusion monitoring and quality assessment.
Jyeniskhan et al. (2024) reviewed 133 publications on digital twins and additive manufacturing. The study reported strong interest in machine learning and process monitoring. It also identified data, modeling, and implementation barriers. These findings suggest that architecture alone is insufficient. A twin also needs measured performance and operating limits.
Lu et al. (2024) proposed a quality evaluation framework for additive manufacturing twins. The framework tracks quality from data through models and computing services. Poor input quality can weaken model accuracy. Slow or unreliable services can then weaken decisions. This chain supports the validation approach used in this paper.

Research gap

Four gaps remain important. First, many published twins stop at visualization. They show printer motion or sensor values without predicting final quality. Second, many AI studies report prediction metrics without live state management. Third, acceptance rules often ignore uncertainty and model limits. Fourth, reproducible public demonstrations remain limited.
The proposed work addresses these gaps through a joined design. It keeps the predictive model inside a larger twin service. It records data quality, model version, and decision history. It also separates predictions from control authority. The public experiment allows other researchers to reproduce the results.

Table 1 Representative research related to AI quality control and additive manufacturing twins.

Study	Scope	Evidence	Main contribution	Remaining limit
Wang et al. (2020)	Broad AM review	Published studies	ML quality prediction	No synchronized twin test
Gaikwad et al. (2020)	Metal AM	Thermal simulation and sensing	Fault detection twin	Process-specific study
Scime et al. (2022)	Multiple AM systems	Lifecycle process records	Scalable twin platform	Limited predictive benchmark
Brion and Pattinson (2022)	Material extrusion	1,272,273 labeled images	Error detection and correction	Camera and compute needs
Corradini and Silvestri (2022)	Material extrusion	Printer monitoring data	Twin-based quality assessment	Limited transfer evidence
Ben Amor et al. (2024)	AM review	Published studies	Twin development hierarchy	Mainly conceptual
Lu et al. (2024)	AM quality framework	Data, model, and service quality	Tracks quality propagation	No public model benchmark
Present study	Material extrusion	69 cleaned public runs	Architecture, model test, alert replay	Small offline dataset

Proposed AI-Based Digital Twin

Design goals

The proposed twin should answer four practical questions. What is the current process state? What quality is expected? How certain is that prediction? What action is allowed? Each answer needs a clear data source and time stamp.
The design follows six principles. The twin must preserve part and layer identity. It must reject invalid or late data. It must expose model uncertainty. It must record every model version. It must separate advice from automatic control. It must allow later inspection of each decision.
The first version supports material extrusion. The architecture also fits other printing processes. Sensor fields and predictive models would change for powder bed fusion or directed energy deposition. The main data and decision layers would remain similar.

Physical system and sensing

The physical layer contains the printer, material, part, environment, and operator. Basic printer data include nozzle temperature, bed temperature, speed, fan level, and commanded position. Better systems add actual flow, motor current, vibration, chamber temperature, and camera streams.
Every observation needs a trusted time stamp. Each record also needs printer, build, part, layer, and material identifiers. These fields prevent accidental mixing between builds. Calibration status should travel with each sensor value. Missing calibration information should reduce confidence.

Data rates differ greatly across sensors. Controller settings may update once per second. Cameras can produce many frames each second. Acoustic sensors may produce thousands of samples. The ingestion service should preserve raw data before feature extraction. This choice allows later model improvement.

Data and state layers

The data layer validates units, ranges, timing, and identity. A temperature of 260 degrees may be possible for some materials. The same value may be invalid for another printer profile. Range checks must therefore depend on process context.

The state estimator combines recent measurements into one current process state. It can smooth noisy values and mark stale sensors. The state includes commanded settings and measured responses. Their difference can reveal drift. A rising difference between target and actual temperature may indicate heater problems.

The twin should retain both current and historical states. Current state supports fast decisions. Historical state supports investigation, retraining, and qualification. A simple database can store low-rate values. Object storage can hold images, thermal files, and large arrays.

Predictive model service

The model service receives a validated state vector. It returns separate predictions for each quality outcome. The current proof of concept predicts roughness, tensile strength, and elongation. A production twin may also predict dimensions, porosity, density, or defect probability.

Each output should include a model identifier and training-data version. It should also include a confidence measure. Confidence can come from prediction intervals, ensemble spread, or calibrated probabilities. A prediction without model context is difficult to audit.

The service should detect unfamiliar inputs. A model trained between 200 and 250 degrees should not silently accept 300 degrees. Distance checks can flag such requests. The twin can then request inspection or use a safer fallback model.

Decision and control layer

The decision layer compares predicted outcomes with product requirements. It can create three states: normal, warning, and stop review. A normal state means predictions remain within validated margins. A warning means one outcome approaches a limit. Stop review means risk exceeds an approved threshold.

Automatic control should begin with low-risk actions. Examples include reducing speed within an approved band or requesting operator inspection. High-impact changes need human approval. These include large temperature changes or geometry changes. The twin must never exceed machine and material limits.

The return channel should record the proposed action, approval, execution, and measured response. This record supports later learning. It also prevents the model from receiving credit for an action that was never applied.

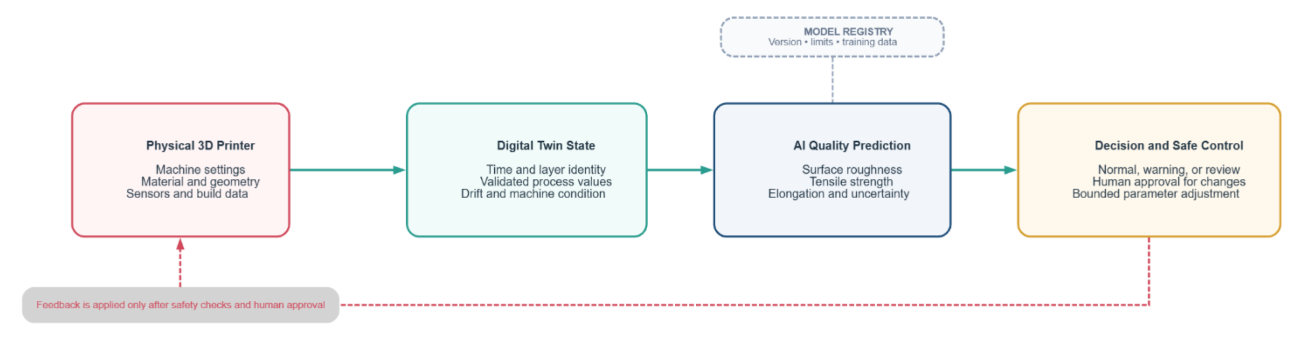

Figure 2 Proposed architecture for an AI-based 3D printing quality twin.

Figure 2 shows the full architecture. Data move from the physical printer toward the prediction service. The decision layer converts predictions into controlled actions. The red return path remains conditional. Safety checks and human authority must control its use.

Synchronization cycle

The twin operates through a repeated cycle. It first captures new process data. It then checks units, timing, identity, and ranges. Valid records update the current state. The model predicts quality from that state.

The decision service converts predictions into risk. It then recommends an approved response. The physical process continues or changes. New measurements verify the effect. This cycle repeats until the build ends.

Synchronization has a strict time requirement. A late warning may have no value. The allowed delay depends on the process. Material extrusion may tolerate seconds for some defects. Laser powder bed fusion may require much faster response.

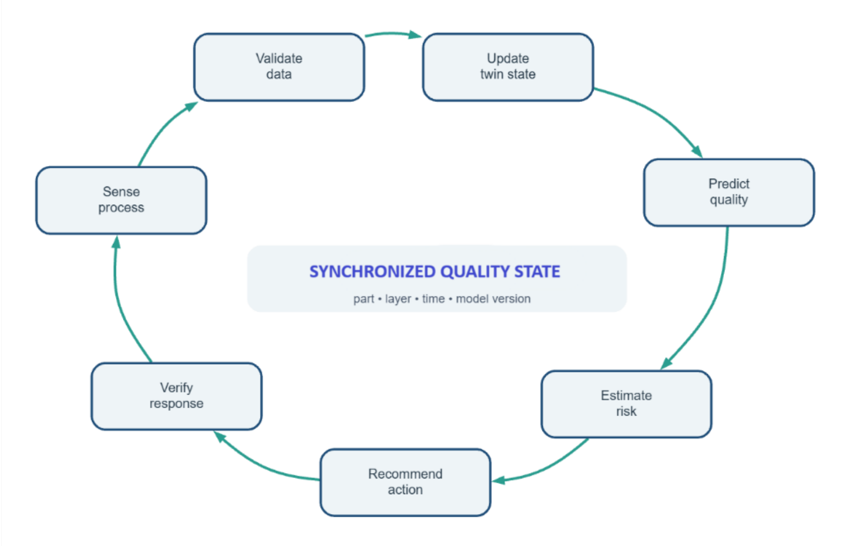

Figure 3 Continuous synchronization and decision cycle for predictive quality control.

Table 2 defines the minimum data carried by the twin. These fields support traceability across models, builds, and decisions.

Table 2 Minimum traceability fields for the proposed digital twin.

Field	Meaning	Update level	Quality purpose
event_time	Coordinated time stamp	Every event	Detect late or misordered data
printer_id	Unique machine identity	Build	Separate machine behavior
build_id	Unique production run	Build	Join all process records
part_layer_id	Part and layer identity	Layer	Align local process history
material_state	Grade, batch, moisture, condition	Build or update	Track material variation
command_state	Requested temperature, speed, flow	Controller cycle	Record process intent
measured_state	Actual sensor values and status	Sensor cycle	Update physical state
prediction	Quality value and model version	Inference cycle	Support traceable decisions
uncertainty	Interval or confidence value	Inference cycle	Limit unsafe confidence
action_state	Advice, approval, execution, response	Decision cycle	Close and audit feedback

Materials and Methods

Public dataset

The practical study used a public additive manufacturing dataset from APMonitor. The data can be downloaded without registration. It combines PLA and ABS experiments from an Ultimaker S5 printer. Mechanical testing used a Sincotec system with a 20 kN capacity (APMonitor, n.d.).

The file contains 70 printing runs. Nine input variables describe the process. They include layer height, wall thickness, infill density, pattern, temperatures, speed, material, and fan speed. Three output variables describe measured quality. These outputs are roughness, tensile strength, and elongation.

The public file supports teaching and model comparison. It does not contain continuous sensor streams. It also lacks geometry identifiers and replicate labels. Therefore, this experiment tests the predictive service only. It does not test live synchronization hardware.

One record contained negative roughness and extreme process values. Its bed temperature was 260 degrees Celsius. Its print speed was 360 millimeters per second. The record was removed through a written rule. No other record was removed.

The final sample contained 69 runs. Both materials and both infill patterns remained present. Table 3 summarizes the data. Figure 4 shows the measured outcome distributions.

Table 3 Variables and descriptive values after data cleaning.

Variable	**Role**	**Unit**	**Range or classes**	**Mean**	**SD**
Layer height	Input	mm	0.02–0.2	0.10	0.06
Wall thickness	Input	mm	1–12	5.50	2.94
Infill density	Input	%	10–100	54.09	27.80
Infill pattern	Input	category	grid, honeycomb	—	—
Nozzle temperature	Input	°C	200–250	221.88	14.93

Variable	Role	Unit	Range or classes	Mean	SD
Bed temperature	Input	°C	60–100	70.51	8.58
Print speed	Input	mm/s	40–120	64.64	28.92
Material	Input	category	abs, pla	—	—
Fan speed	Input	%	0–100	49.32	35.84
Roughness	Output	µm	21–368	157.45	94.88
Tensile strength	Output	MPa	4–38	19.75	9.21
Elongation	Output	%	0.4–3.3	1.63	0.75

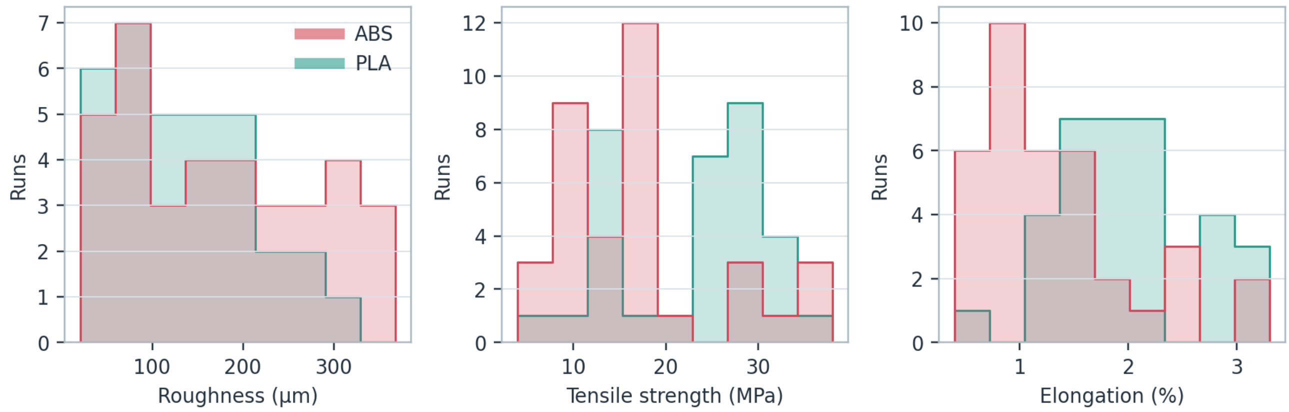

Figure 4 Distributions of the three measured quality outcomes by material.

Data preparation

Numerical inputs were standardized within each training fold. Standardization used the training mean and standard deviation. Categorical inputs used one-hot encoding. These inputs included material and infill pattern.

All transformations occurred inside each validation fold. This prevented information from the test fold entering model training. The same transformation pipeline was reused for every model. No synthetic rows were added.

The three outcomes were modeled separately. This choice kept errors visible for each quality property. It also allowed different relations between settings and outcomes. A shared multi-output model may be tested later.

Figure 5 shows simple correlations after category coding. Layer height has a strong positive relation with roughness. Tensile strength and elongation also move together. Correlation does not prove a direct cause. Several settings change together in this small design.

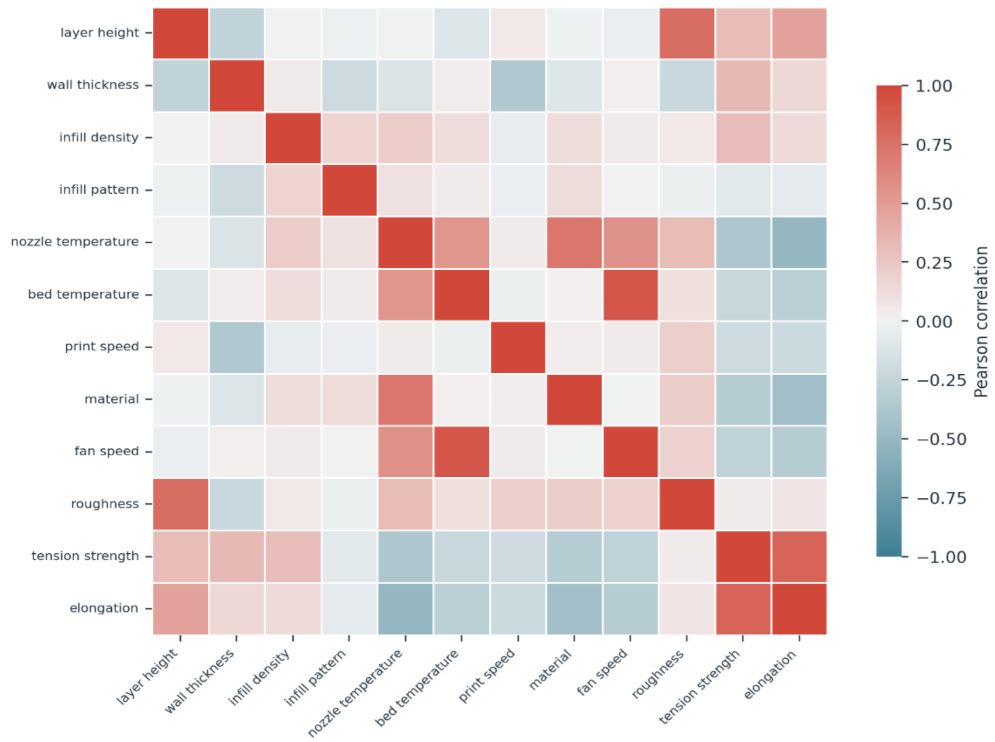

Figure 5 Correlation matrix for process settings and measured quality outcomes.

Predictive models

Four regression methods were compared. Ridge regression provided a regularized linear baseline. Random forest combined many bootstrapped decision trees. Extra trees increased randomization between trees. Gradient boosting built a sequence of small corrective trees.

The models used fixed settings before final evaluation. This reduced the risk of selecting settings from the same results. Gradient boosting used 180 trees, a 0.035 learning rate, and depth two. Random forest and extra trees used 500 trees. Both tree ensembles required at least two samples in each leaf.

Table 4 lists the main model settings. The settings favor stable behavior on a small dataset. They do not represent a full tuning study.

Table 4 Predictive models and fixed experimental settings.

Model	Type	Main settings	Reason for inclusion
Ridge regression	Linear	Alpha = 1.0	Readable baseline with shrinkage
Random forest	Tree ensemble	500 trees; leaf ≥ 2	Nonlinear interactions and stable averaging
Gradient boosting	Boosted trees	180 trees; rate 0.035; depth 2	Sequential correction with Huber loss
Extra trees	Randomized trees	500 trees; leaf ≥ 2	Stronger randomization between trees

Validation method

Repeated five-fold cross-validation estimated model performance. The data were split into five folds. Four folds trained the model, and one fold tested it. This process repeated until each fold served as the test set.

The full five-fold process was repeated ten times. Each repeat used a different controlled split. Therefore, every reported cross-validation mean used 50 test folds. This design reduces dependence on one fortunate split.

Three metrics were calculated. Mean absolute error gives an easy average error. Root mean squared error gives more weight to large errors. The coefficient of determination reports explained test variation.

Normalized RMSE supported comparison between different units. Each RMSE was divided by the observed outcome range. The result was expressed as a percentage. The best overall model had the lowest mean normalized RMSE.

Out-of-fold replay and alert rule

A second validation produced one honest prediction for every cleaned row. Ten-fold cross-validation generated these out-of-fold predictions. Each row was predicted by a model that never trained on that row.

The replay used three illustrative quality limits. Roughness needed to remain at or below 200 micrometers. Tensile strength needed to reach 20 MPa. Elongation needed to reach 1.5 percent. These limits are not standards. A real product needs engineering limits from its design and risk analysis.

A row passed only when all three conditions were met. The same rule was applied to measured and predicted outcomes. The comparison produced a confusion matrix, balanced accuracy, and pass-class F1 score. False passes received special attention because they create quality risk.

Reproducibility

The analysis used Python, pandas, NumPy, and scikit-learn. The random seed was fixed at 42. Raw data remained unchanged in a separate file. Cleaning, modeling, evaluation, and figure generation ran from one script.

The public data link appears in every data-based figure caption. The figures were created for this paper. No stock images or licensed image libraries were used. The same script can regenerate all reported metrics and charts.

Results

Descriptive results

The cleaned dataset contained 69 runs. Roughness ranged from 21 to 368 micrometers. Its mean was 157.45 micrometers. Tensile strength ranged from 4 to 38 MPa. Its mean was 19.75 MPa.

Elongation ranged from 0.4 to 3.3 percent. Its mean was 1.63 percent. The wide ranges support prediction testing. However, 69 rows remain a small sample for general industrial use.

PLA and ABS show overlapping outcome distributions. This overlap means material alone cannot explain quality. Process settings still matter. Figure 4 also shows a right-skewed roughness distribution. Tree methods can handle this shape without assuming normal residuals.

Model comparison
Gradient boosting achieved the lowest overall normalized RMSE. Its mean normalized error was 14.46 percent across the three outcomes. Extra trees followed at 14.94 percent. Random forest reached 15.16 percent. Ridge regression reached 18.90 percent.
The result suggests important nonlinear effects. Ridge regression performed worst for roughness and tensile strength. Tree ensembles better represented changing effects across settings. Figure 6 compares normalized error for every outcome.

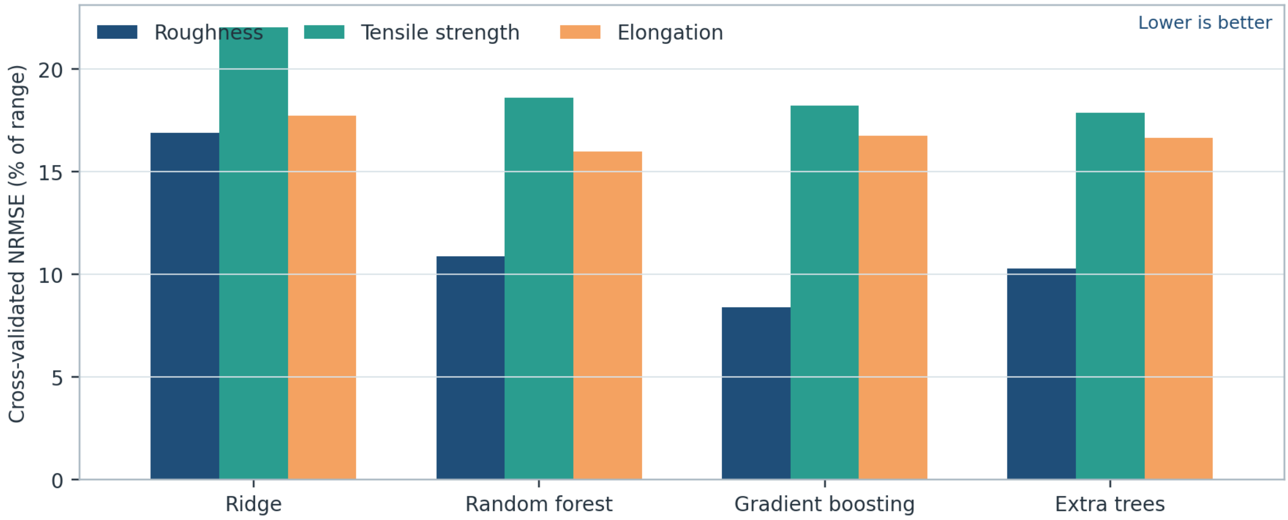

Figure 6 Repeated cross-validation error for the four regression models.

Gradient boosting gave the strongest roughness result. Mean cross-validated RMSE was 29.16 micrometers. Mean R^2 was 0.88 across the 50 folds. Tensile strength was harder to predict. Mean RMSE was 6.19 MPa, and mean R^2 was 0.47.
Elongation gave a mean RMSE of 0.49 percentage points. Mean R^2 reached 0.46. Random forest slightly improved elongation error. However, gradient boosting remained best across all three outcomes.
Table 5 reports the full repeated cross-validation results. Variation between folds remains meaningful. This variation reflects the small sample and changing test composition.

Table 5 Repeated five-fold cross-validation results across ten repeats.

Model	Outcome	MAE	RMSE	Mean R^2	NRMSE (%)
Ridge	Roughness	47.97	58.67	0.52	16.91
Ridge	Tensile strength	5.69	7.49	0.24	22.03
Ridge	Elongation	0.41	0.51	0.42	17.75
Random forest	Roughness	30.15	37.71	0.81	10.87
Random forest	Tensile strength	4.97	6.33	0.45	18.63
Random forest	Elongation	0.36	0.46	0.51	15.98
Gradient boosting	Roughness	21.85	29.16	0.88	8.40
Gradient boosting	Tensile strength	4.55	6.19	0.47	18.22
Gradient boosting	Elongation	0.38	0.49	0.46	16.76
Extra trees	Roughness	27.98	35.71	0.83	10.29
Extra trees	Tensile strength	4.59	6.08	0.49	17.88
Extra trees	Elongation	0.37	0.48	0.47	16.66

Out-of-fold predictions
The ten-fold replay produced one prediction for each cleaned row. Roughness achieved an out-of-fold R^2 of 0.90. Its mean absolute error was 21.49 micrometers. Tensile strength achieved an R^2 of 0.53. Its mean absolute error was 4.50 MPa.
Elongation achieved an out-of-fold R^2 of 0.54. Its mean absolute error was 0.39 percentage points. Figure 7 compares measured and predicted values. Predictions follow roughness closely. Strength and elongation show wider spread around the perfect-prediction line.

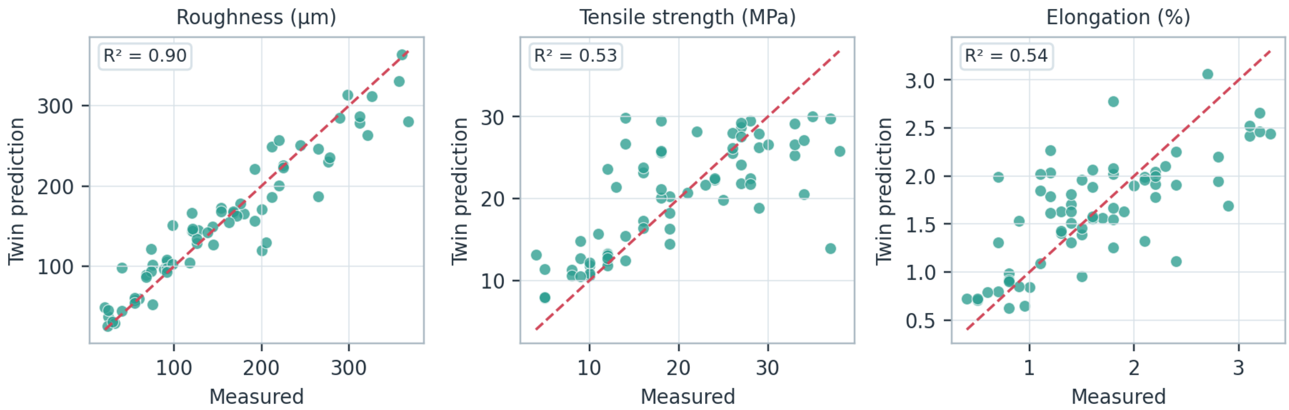

Figure 7 Out-of-fold predictions from the selected gradient boosting model.

The difference between roughness and mechanical outcomes is important. Roughness depends strongly on layer height within this dataset. Mechanical properties depend on several interacting settings. They may also depend on unrecorded factors.

Input importance

Permutation importance measured the loss of accuracy after shuffling each input. Results were normalized within each target and then averaged. Layer height had the largest mean importance. Nozzle temperature ranked second. Wall thickness ranked third.

These results fit known process behavior. Larger layers create a stronger staircase effect on surfaces. Nozzle temperature affects melt flow and bonding. Wall thickness changes load-bearing material and cooling behavior.

Infill pattern had low mean importance in this dataset. This result does not prove that pattern never matters. It only describes these runs and outcomes. Figure 8 shows the ranked values.

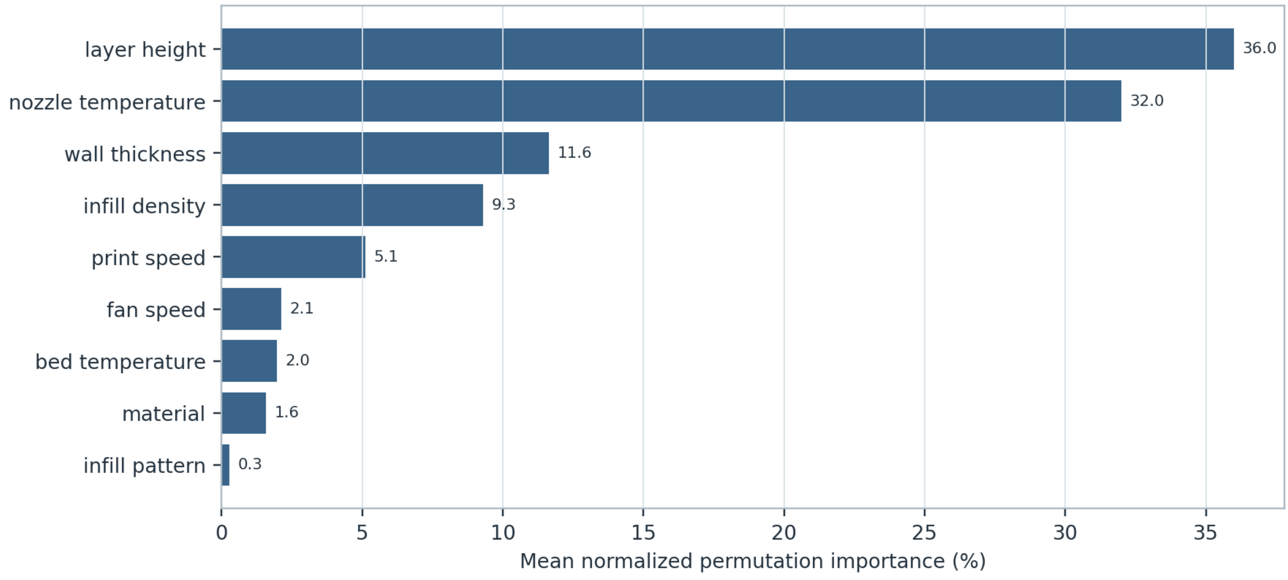

Figure 8 Mean normalized permutation importance across the three outcomes.

Predictive alert replay

Eighteen measured runs passed all three illustrative limits. The twin predicted 24 passes. Fourteen predictions were true passes. Four measured passes were missed.

Ten rows became false passes. Forty-one rows were correctly flagged for review. Balanced accuracy reached 79.1 percent. The pass-class F1 score reached 66.7 percent. The false-pass rate among measured failures was 19.6 percent.

Figure 9 shows the quality margin across the replayed run order. A positive value means all predicted limits were met. A negative value means at least one limit failed. The chart shows useful tracking, but several crossings remain incorrect.

Figure 9 Replayed production sequence using measured and predicted quality margins.

Figure 10 summarizes the final decisions. False passes are more serious than false alerts. Therefore, 79.1 percent balanced accuracy is not sufficient for autonomous release. The model should support inspection decisions, not replace them.

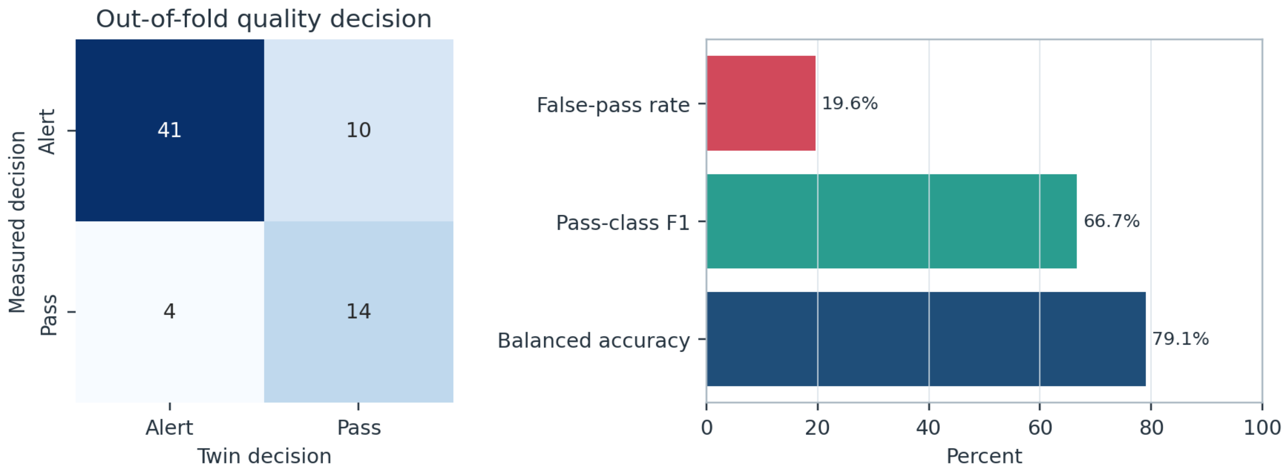

Figure 10 Alert confusion matrix and decision metrics from the out-of-fold replay.

Discussion

Main findings

The experiment shows that a small public dataset can support useful quality prediction. Roughness prediction was especially strong. Gradient boosting explained about 90 percent of out-of-fold roughness variation. This result agrees with the strong layer-height relation.

Mechanical outcomes were harder. Their out-of-fold R^2 values were near 0.54. These values can support screening and process learning. They are not strong enough for final certification.

The alert replay provides a clearer operational test than regression metrics alone. A model can have acceptable average error and still cross an acceptance limit incorrectly. Ten false passes occurred under the chosen limits. This finding supports cautious deployment.

Comparison with earlier work

Tatar (2025) reported very high accuracy for Gaussian process regression on a related open dataset. The present results are lower for strength and elongation. Several method differences explain this gap. This study removed one invalid row and used repeated validation. It also reported honest out-of-fold decisions.

Brion and Pattinson (2022) achieved general error recognition from a much larger image dataset. Their work used over one million labeled images. The present study uses only 69 tabular runs. The data scale explains the simpler models and lower information content.

ORNL has released larger additive manufacturing datasets for quality research. One recent release includes process data, thousands of powder-bed images, and thousands of tensile tests (ORNL, 2024). Such datasets can support stronger twin validation. They can also connect local process history with final properties.

The architecture aligns with Scime et al. (2022). Both approaches preserve component data across the manufacturing lifecycle. The present work adds a small predictive benchmark and an alert replay. It also applies the quality chain described by Lu et al. (2024).

Why the digital twin structure matters

An isolated model returns a number. A twin explains which process state produced that number. This difference improves trust and traceability. It also supports later investigation when a prediction fails.

Model version is especially important. Predictions may change after retraining. A quality record must therefore preserve the exact model used. Training-data identity and feature definitions are also required.

Time alignment is equally important. A camera frame must match the correct layer and tool position. Misaligned data can create false relations. The model may appear accurate offline but fail during live use.

The decision record closes the loop. It shows whether an operator accepted a recommendation. It also shows whether the printer applied it. Without this record, the system cannot learn action effects.

Data quality and generalization

The experiment exposed one invalid row. A negative roughness value has no physical meaning. The row also contained extreme temperatures and speed. Automatic range checks would catch this record before inference.

Other problems may remain hidden. The dataset lacks printer serial numbers, filament batch, moisture, room conditions, and geometry identity. These factors can change quality. Their absence limits causal interpretation.

Generalization across printers is a larger challenge. Two printers may use the same commanded temperature but produce different melt conditions. Sensor calibration, nozzle condition, and firmware can change the actual state. Cross-printer testing is therefore essential.

Rojek et al. (2025) identified data scarcity and weak transfer as major barriers. Altun et al. (2025) also noted limited generalization across machines and materials. The current findings support those concerns.

Uncertainty and safe decisions

Every quality prediction needs an uncertainty measure. Average error does not show risk for one build. A production twin should create prediction intervals for each outcome. It should widen those intervals for unfamiliar settings.

Acceptance should use conservative bounds. Strength can pass only when the lower bound exceeds its limit. Roughness can pass only when the upper bound remains below its limit. This approach reduces false passes.

False alerts also matter. Too many alerts can slow production and reduce operator trust. Thresholds should reflect product risk and inspection cost. A medical implant needs stricter rules than a visual prototype.

Human review remains necessary during early deployment. Operators should see the predicted outcome, uncertainty, reason, and allowed action. They should not receive an unexplained red warning.

Industrial deployment

A practical system can begin as a digital shadow. It receives printer data and makes predictions without changing the process. Engineers can compare predictions with inspection results for several weeks. This stage measures drift and false alarms.

The next stage allows advisory control. The twin suggests a change within an approved envelope. The operator accepts or rejects the change. Every response becomes part of the validation record.

Automatic control should follow only after controlled trials. The system must demonstrate stable latency, prediction accuracy, and safe fallback behavior. It must also survive sensor loss and network failure.

Table 6 lists the main deployment checks. These checks treat the twin as a quality system, not a dashboard.

Table 6 Validation gates for staged industrial deployment.

Gate	Measure	Release condition	Failure response
Data identity	Unique printer, build, part, and layer	No duplicate or orphan record	Stop inference
Sensor calibration	Current certificate and range check	All critical sensors valid	Use fallback or pause
Prediction error	Local out-of-fold or external test	Product-specific limit	Keep advisory mode
False-pass risk	Measured failures released by model	Risk-approved maximum	Require inspection
Uncertainty	Calibrated prediction coverage	Target coverage achieved	Widen review zone
Latency	Capture-to-decision delay	Below process deadline	Disable live action
Drift	Input and residual shift	Inside validated envelope	Retrain and revalidate

Gate	Measure	Release condition	Failure response
Fallback	Sensor, network, or model failure	Safe fixed behavior	Return to approved settings
Human factors	Alert clarity and operator response	Usable and auditable	Revise interface and rules

Limitations

The study has five important limits. The dataset is small. It contains only two materials and one printer type. It has no continuous sensor streams. It has no geometry or batch identifiers. The quality limits are illustrative.

Repeated cross-validation improves evaluation, but it cannot create missing variation. The model may fail on new printers or materials. Feature importance also describes association, not physical cause.

The experiment uses offline rows. It does not measure live latency or packet loss. It also does not apply control changes. Therefore, the architecture remains partly conceptual.

Practical Implementation and Validation Plan

Phase one: passive data capture

The first phase records printer data without changing production. Engineers define one build identifier and one time source. They also record machine, material, geometry, layer, and operator fields. Calibration records remain linked with every sensor.

The system should capture both commands and measured responses. A commanded nozzle temperature is not the actual temperature. Their difference may reveal drift. The same rule applies to speed, flow, and position.

Quality inspection remains unchanged during this phase. Each final measurement returns to the twin record. These labels support local model testing. They also reveal missing factors.

Phase two: shadow prediction

The validated model runs beside normal production. It predicts quality but does not alert operators. Engineers compare predictions with measured outcomes. They track error by machine, material, geometry, and time.

The model should pass three gates. Average error must meet the engineering target. False-pass risk must remain below the approved level. Prediction latency must remain below the process deadline.

Drift testing should compare current inputs with training inputs. A new filament grade may move several features. The system should mark this condition before making a confident claim.

Phase three: advisory alerts

The twin then displays reviewable alerts. Each alert shows the affected quality property. It also shows the predicted value and uncertainty band. The operator sees the evidence and allowed response.

Operators should record alert usefulness. They should also record false alerts and missed defects. This feedback supports threshold changes. It should not directly retrain the model without review.

The advisory stage should include controlled fault tests. Engineers can vary temperature, speed, or flow within safe limits. The twin should detect the resulting quality change. These tests measure sensitivity under known conditions.

Phase four: bounded control

Bounded control allows small automatic changes within approved limits. A safety controller checks every proposed action. The AI model cannot bypass this controller. The system returns to fixed settings after communication loss.

Each control rule needs a verified response model. Reducing speed may improve one defect but increase build time. Temperature changes may improve bonding but raise deformation risk. Multi-objective limits must guide the decision.

Automatic release remains a separate decision. It needs stronger evidence than automatic process adjustment. Final release may still require physical inspection. The required evidence depends on product risk and regulation.

Suggested future experiment

A stronger study should use at least three printers and three material batches. It should include repeated geometries and controlled faults. Camera, temperature, current, vibration, and flow data should share one clock.

The test should divide data by printer and build. Random row splitting can leak similar conditions between training and testing. Leave-one-printer-out validation gives a stronger generalization test.

The study should compare tabular, image, and fused models. A tabular model provides a low-cost baseline. An image model detects local deposition errors. A fused model combines both sources.

NIST AM-Bench and ORNL releases provide useful public reference data. Cambridge also provides a large extrusion image dataset (Brion & Pattinson, 2022). These sources can support external validation and benchmark comparisons.

Conclusion

This paper developed an AI-based digital twin for predictive quality control in 3D printing. The design joins process data, state management, quality models, decisions, and traceable feedback. It separates prediction from control authority.

A public material-extrusion dataset supported the practical experiment. The cleaned data contained 69 runs. Four regression methods predicted roughness, tensile strength, and elongation. Gradient boosting gave the lowest average normalized error.

Roughness prediction was strong. Mechanical properties were predicted with moderate accuracy. The alert replay reached 79.1 percent balanced accuracy. However, ten false passes prevented autonomous quality release.

The main lesson is simple. Useful prediction does not equal safe control. A production twin needs uncertainty, traceability, timing checks, drift detection, and human oversight. It also needs validation on local machines and materials.

Future work should use larger multi-printer datasets and synchronized sensor streams. It should test external generalization and bounded control. These steps can move the proposed twin from an offline proof toward dependable industrial quality support.

Data Availability

The experimental dataset is publicly available at https://apmonitor.com/pds/uploads/Main/manufacturing.txt.
The supporting case page is https://apmonitor.com/pds/index.php/Main/AdditiveManufacturing.
The Cambridge image dataset is available at https://doi.org/10.17863/CAM.84082.
NIST AM-Bench data links are available at https://www.nist.gov/ambench/direct-am-bench-data-links-and-referencing-guidance.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

1. Altun, F., Bayar, A., Hamzat, A. K., Asmatulu, R., Ali, Z., & Asmatulu, E. (2025). AI-driven innovations in 3D printing: Optimization, automation, and intelligent control. *Journal of* Manufacturing and Materials Processing, 9(10), 329. https://doi.org/10.3390/jmmp9100329
2. APMonitor. (n.d.). *Additive manufacturing: Machine learning for engineers*. Retrieved July 10, 2026, from https://apmonitor.com/pds/index.php/Main/AdditiveManufacturing
3. Bast, S., Scherer, K., Wahl, M., Naumann, S., Dartmann, G., & Didas, S. (2025). *Quality control in additive manufacturing: Image dataset for defect detection in fused filament* fabrication [Data set]. Zenodo. https://doi.org/10.5281/zenodo.14712897
4. Ben Amor, S., Elloumi, N., Eltaief, A., Louhichi, B., Alrasheedi, N. H., & Seibi, A. (2024). Digital twin implementation in additive manufacturing: A comprehensive review. *Processes, 12*(6), 1062. https://doi.org/10.3390/pr12061062
5. Brion, D. A. J., & Pattinson, S. W. (2022). Generalisable 3D printing error detection and correction via multi-head neural networks. *Nature Communications*, *13*, 4654. https://doi.org/10.1038/s41467-022-31985-y
6. Chen, J., Yuan, Y., Wang, Q., Wang, H., & Advincula, R. C. (2025). Bridging additive manufacturing and electronics printing in the age of AI. *Nanomaterials, 15*(11), 843. https://doi.org/10.3390/nano15110843
7. Corradini, F., & Silvestri, M. (2022). Design and testing of a digital twin for monitoring and quality assessment of material extrusion process. *Additive Manufacturing, 51*, 102633. https://doi.org/10.1016/j.addma.2022.102633
8. Gaikwad, A., Yavari, R., Montazeri, M., Cole, K., Bian, L., & Rao, P. (2020). Toward the digital twin of additive manufacturing. *IISE Transactions, 52*(11), 1204–1217. https://doi.org/10.1080/24725854.2019.1701753
9. Jyeniskhan, N., Shomenov, K., Ali, M. H., & Shehab, E. (2024). Exploring the integration of digital twin and additive manufacturing technologies. *International Journal of Lightweight Materials and Manufacture, 7*(6), 858–876. https://doi.org/10.1016/j.ijlmm.2024.06.004
10. Kim, R. G., Abisado, M., Villaverde, J., & Sampedro, G. A. (2023). A survey of image-based fault monitoring in additive manufacturing. *Sensors*, *23*(15), 6821. https://doi.org/10.3390/s23156821

11. Lu, Y., Xie, J., Safdar, M., Yang, Z., Elhambakhsh, F., Ko, H., Li, S., & Zhao, Y. (2024). An overarching quality evaluation framework for additive manufacturing digital twin. In *2024 IEEE 20th International Conference on Automation Science and Engineering* (pp. 676–683). IEEE. https://doi.org/10.1109/CASE59546.2024.10711607
12. National Institute of Standards and Technology. (2022). *Direct AM Bench data links and referencing guidance*. https://www.nist.gov/ambench/direct-am-bench-data-links-and-referencing-guidance
13. Oak Ridge National Laboratory. (2024, July 18). *Free 3D-printing datasets enable analysis, confidence in printed parts*. https://www.ornl.gov/news/free-3d-printing-datasets-enable-analysis-confidence-printed-parts
14. Rojek, I., Mikołajewski, D., Kempiński, M., Galas, K., & Piszcz, A. (2025). Emerging applications of machine learning in 3D printing. *Applied Sciences, 15*(4), 1781. https://doi.org/10.3390/app15041781
15. Scime, L., Singh, A., & Paquit, V. (2022). A scalable digital platform for the use of digital twins in additive manufacturing. *Manufacturing Letters, 31*, 28–32. https://doi.org/10.1016/j.mfglet.2021.05.007
16. Tatar, A. B. (2025). Predicting three-dimensional printing product quality with machine learning-based regression methods. *Firat University Journal of Experimental and Computational Engineering, 4*(1), 206–225. https://doi.org/10.62520/fujece.1604379
17. Tran, T. Q. (2025). Recent advances in 3D printing and additive manufacturing technology. *Applied Sciences, 15*(17), 9599. https://doi.org/10.3390/app15179599
18. Wang, C., Tan, X. P., Tor, S. B., & Lim, C. S. (2020). Machine learning in additive manufacturing: State-of-the-art and perspectives. *Additive Manufacturing, 36*, 101538. https://doi.org/10.1016/j.addma.2020.101538
19. Zhang, L., Chen, X., Zhou, W., Cheng, T., Chen, L., Guo, Z., Han, B., & Lu, L. (2020). Digital twins for additive manufacturing: A state-of-the-art review. *Applied Sciences, 10*(23), 8350. https://doi.org/10.3390/app10238350

