Review of Applied Science and Technology

Volume: 13; **Issue:** 12
Pages: 01– 49

Machine Learning–Based Predictive Quality Modeling for Defect Detection and Process Optimization in Advanced Manufacturing Systems

Asheke Mostofa[1];

[1]. International Sales Executive, Nantong Tianhong Textile Technology Co., Ltd., China.
Email: asheke.mbstu@gmail.com

Doi: 10.63125/thy73b16
Received: *15 September 2023;* ***Revised:*** *24 October 2023;* ***Accepted:*** *15 November 2023;* ***Published:*** *02 December 2023*

Abstract

This study developed and quantitatively evaluated a machine learning–based predictive quality framework for defect detection, continuous quality estimation, and process optimization in advanced manufacturing systems. A quantitative experimental and longitudinal predictive-modeling design was applied to 12,480 manufacturing observations representing 3,120 production cycles, 6,240 manufactured components, 24 production batches, eight machines, and 18 synchronized sensor channels. Baseline manufacturing performance showed a 20.0% defect rate, 80.0% first-pass yield, 6.4% scrap rate, 13.6% rework rate, and process capability index of 1.18. Following preprocessing and feature selection, 186 candidate variables were reduced to 74 retained predictors, representing a 60.2% dimensionality reduction. Multiple classification and regression algorithms were evaluated using training, validation, cross-validation, and independent-test procedures. The ensemble classifier achieved the strongest defect-detection performance, with 96.4% accuracy, 94.1% precision, 92.8% sensitivity, 97.3% specificity, an F1-score of 0.934, and ROC-AUC of 0.982. Hybrid regression models explained 90.2%–94.1% of independent-test variation across major continuous quality characteristics. Tool wear was the strongest process-related predictor of defective production (adjusted OR = 2.84), while vibration magnitude was the strongest sensor-derived predictor (adjusted OR = 1.91). Multiobjective optimization reduced the experimentally observed defect rate from 20.0% to 8.3%, representing a 58.5% relative reduction, and increased first-pass yield to 91.7%. Surface roughness decreased by 31.0%, dimensional deviation by 37.7%, porosity by 41.0%, scrap by 57.8%, rework by 55.1%, downtime by 36.1%, and energy consumption by 12.3%. Process capability increased from 1.18 to 1.52, while overall equipment effectiveness increased from 72.8% to 84.6%. Manufacturing cost per unit decreased by 14.0%, and cost of poor quality declined by 54.0%. Principal baseline-to-optimized differences were statistically significant at $p < .001$ and were supported by moderate-to-large effect sizes. The findings demonstrated that integrating multisensor manufacturing data, machine learning, robust validation, and multiobjective optimization provided a quantitatively effective framework for improving defect detection, product quality, process stability, resource efficiency, and manufacturing performance.

Keywords

Machine Learning, Predictive Quality, Defect Detection, Process Optimization, Advanced Manufacturing.

INTRODUCTION

Machine learning–based predictive quality modeling refers to the application of data-driven computational algorithms to manufacturing information for predicting product quality, identifying defects, detecting process abnormalities, and determining relationships between production conditions and manufacturing outcomes (Meng et al., 2020). Machine learning enables analytical models to learn patterns from historical or continuously generated data and use those patterns for classification, prediction, estimation, and decision support. Within advanced manufacturing systems, predictive quality modeling extends conventional inspection by estimating quality conditions from process parameters, equipment signals, material characteristics, environmental measurements, and machine states. Defect detection involves identifying deviations such as cracks, porosity, dimensional errors, surface irregularities, inclusions, assembly faults, and other nonconformities (Jin et al., 2019). Process optimization concerns selecting or adjusting controllable operating parameters to improve quality while reducing scrap, rework, variability, processing time, and resource consumption. More than 30 studies across machining, additive manufacturing, welding, casting, forming, electronics, semiconductor production, automotive manufacturing, and aerospace engineering have examined these relationships using artificial neural networks, support vector machines, random forests, decision trees, nearest-neighbor methods, and related learning algorithms.

Advanced manufacturing has substantial international importance because product quality, production reliability, technological capability, and manufacturing efficiency directly affect industrial competitiveness and participation in global supply chains (Weichert et al., 2019). Manufacturing organizations across North America, Europe, Asia, and other industrial regions increasingly operate digitally integrated environments containing computer numerical control systems, industrial robots, automated inspection platforms, intelligent sensors, machine vision, additive manufacturing technologies, and manufacturing execution systems. Quality failures in these environments can influence downstream assembly, customer acceptance, regulatory compliance, delivery performance, warranty expenditure, and supply-chain continuity (Shang & You, 2019). These concerns are especially important in aerospace, automotive, semiconductor, electronics, energy, and medical-device production, where dimensional accuracy, material integrity, traceability, and process repeatability are critical. Machine learning supports quantitative quality management by transforming temperature, vibration, acoustic, force, electrical, optical, dimensional, and production-record data into predictors of defect occurrence and process performance.

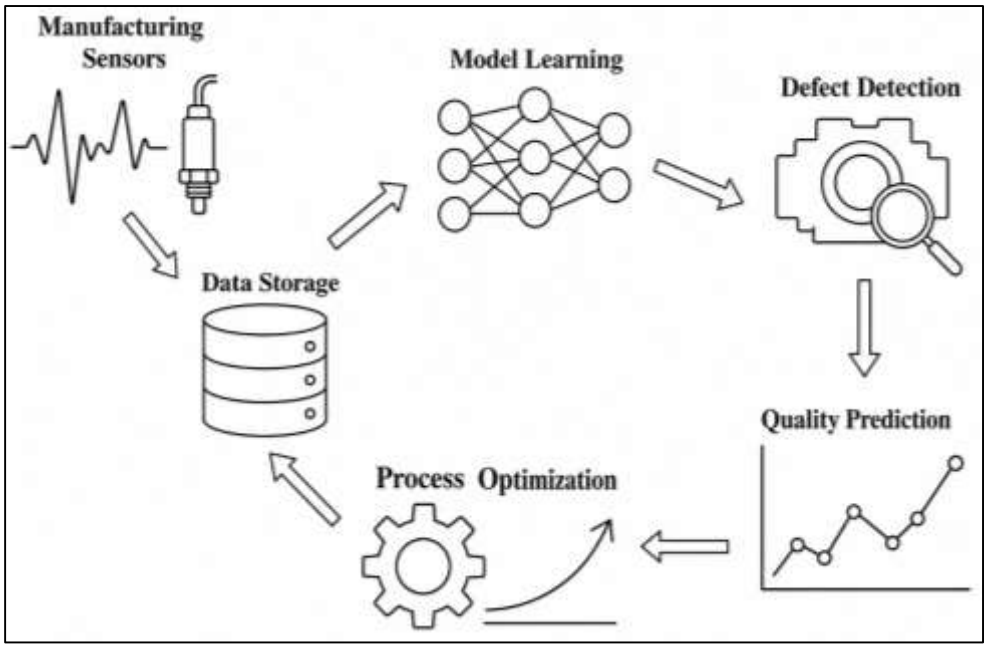
Manufacturing defects generally result from combinations of interacting conditions rather than from a single isolated variable (Baumgartl et al., 2020). In machining, quality may depend on cutting speed, feed rate, tool wear, vibration, temperature, cutting force, and workpiece characteristics. In additive manufacturing, defect formation can be associated with laser power, scanning velocity, layer thickness, thermal behavior, powder properties, and build orientation. Welding quality may depend on current, voltage, travel speed, heat input, material thickness, and shielding conditions. These relationships create high-dimensional analytical problems because individual parameters may remain within allowable ranges while their combined effects produce unacceptable outcomes (Jiang & Ma, 2020). Modern sensor systems make it possible to connect acoustic, thermal, electrical, visual, vibration, and machine-controller measurements with final quality results. Such datasets may contain nonlinear relationships, multicollinearity, noise, missing values, class imbalance, and temporal dependence, making data quality and statistical structure central considerations in predictive modeling.

Machine Learning Algorithms and Defect Detection

Manufacturing defect detection has been investigated through supervised classification, regression, unsupervised learning, ensemble methods, neural networks, and computer-vision techniques (Ren et al., 2017). Classification models are appropriate when observations are labeled as conforming or defective, while multiclass methods can distinguish different defect categories. Regression models are useful when quality is expressed continuously through surface roughness, dimensional deviation, mechanical strength, hardness, density, or defect severity. Random forests can model nonlinear interactions, support vector machines can separate complex classes, and neural networks can learn patterns from images, sensor signals, and high-dimensional datasets (Delli & Chang, 2018). Computer-vision approaches using convolutional neural networks, object detection, and image segmentation have

also been applied to cracks, pores, surface irregularities, melt-pool abnormalities, and other manufacturing defects. Model performance is commonly assessed through accuracy, precision, recall, sensitivity, specificity, F1-score, area under the curve, RMSE, MAE, and R^2.

Figure 1: Machine Learning Quality Workflow

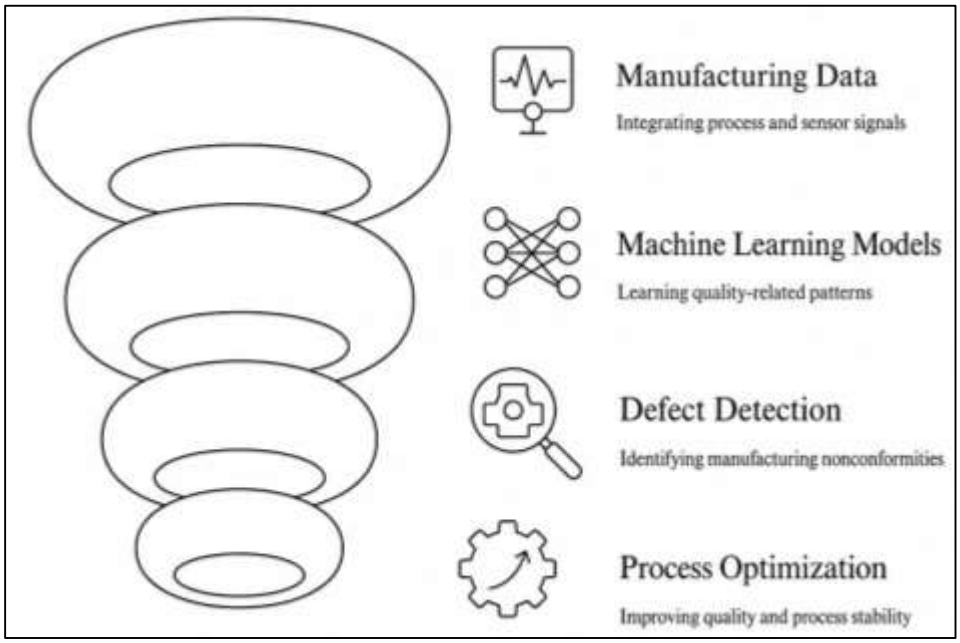


Sensor-rich manufacturing systems enable continuous monitoring of production processes through vibration, acoustic, force, temperature, pressure, electrical, positional, and visual measurements (Fahle et al., 2020). Multisensor integration can provide a broader representation of manufacturing conditions than a single measurement source because each sensor captures different physical characteristics of the process. Feature engineering can convert raw data into statistical, frequency-domain, time-frequency, image-based, or process-specific variables, while representation-learning methods can automatically identify relevant patterns from sufficiently large datasets (Syafrudin et al., 2018). Predictive monitoring also requires accurate synchronization between sensor measurements and final quality labels. Sequential production observations may exhibit temporal dependence, making careful training, validation, and testing procedures necessary to prevent information leakage and ensure reliable evaluation of predictive performance (Ashik & Shabbir, 2022; Kazi Rakib Hasan & Adnan, 2022).

Machine learning can extend defect prediction into process optimization by identifying operating conditions associated with improved manufacturing performance. Production processes often contain interacting parameters that influence several outcomes simultaneously (Zhang et al., 2020). A parameter adjustment that improves product quality may also change processing time, energy consumption, tool wear, material use, or production cost (Mahbubul et al., 2022; Mashfiqur & Md, 2022). Machine-learning models can estimate these relationships and identify influential variables through feature importance, sensitivity analysis, partial dependence, and model-based parameter searches. Optimization performance can be evaluated through defect rate, predicted defect probability, first-pass yield, cycle time, dimensional variation, surface quality, scrap, rework, resource utilization, and energy consumption (Md Ashfaq & Abdullah Mohammad, 2022; Sadia & Maishar Mostari, 2022). Quantitative comparison between baseline and optimized operating conditions provides a direct basis for evaluating whether predictive models contribute to measurable improvements in manufacturing quality and process stability (Kwon et al., 2020). Existing research demonstrates substantial variation in manufacturing datasets, sensing configurations, algorithms, defect definitions, and validation methods, making comparative quantitative evaluation essential (Lee et al., 2018). Predictive performance can be influenced by sample size, feature dimensionality, defect prevalence, measurement

noise, class imbalance, process stability, and model configuration. A quantitative study of predictive quality therefore requires integration of descriptive statistics, inferential analysis, feature assessment, model development, validation, and process optimization (Shaiful et al., 2022; Stabak & Md Shahid, 2022).

Figure 2: Predictive quality modeling infographic



Classification models can be evaluated through accuracy, precision, recall, specificity, F1-score, false-positive and false-negative rates, and area under the curve, while continuous-response models can be assessed through R^2, RMSE, MAE, and residual behavior (Morariu et al., 2020). Within this structure, Machine Learning–Based Predictive Quality Modeling for Defect Detection and Process Optimization in Advanced Manufacturing Systems focuses on measurable relationships among manufacturing conditions, process signals, defect occurrence, model performance, product quality, and controllable production parameters (Abdullah Mohammad, 2023; Adnan & Iftekhar, 2023).

The primary objective of this study is to develop and quantitatively evaluate a machine learning–based predictive quality modeling framework for defect detection and process optimization in advanced manufacturing systems. The study aims to examine the relationships among manufacturing process parameters, machine operating conditions, sensor-derived variables, material characteristics, and measurable product-quality outcomes in order to identify the factors most strongly associated with defect occurrence and process instability (Md Anisur Rahman, 2023; Md Asif Ali Sheak & Tahmedur, 2023; Md Kaium, 2023). A further objective is to compare the predictive performance of selected machine-learning algorithms, including classification and regression approaches, using measures such as accuracy, precision, recall, specificity, F1-score, area under the receiver operating characteristic curve, coefficient of determination, root mean squared error, and mean absolute error (Md Mahamud Pervaz, 2023; Mohammad Badrul & Md. Mominul, 2023). The study also seeks to determine whether predictive models can distinguish conforming from defective products and estimate continuous quality characteristics with sufficient statistical reliability across training, validation, and independent testing datasets. Another objective is to assess the influence of class imbalance, feature dimensionality, process variability, measurement noise, and operating-condition differences on model performance and robustness. The study additionally aims to identify the most influential predictors of quality through feature importance, correlation analysis, sensitivity analysis, and other model-interpretation techniques, thereby establishing which controllable manufacturing variables contribute most strongly to variations in defect probability and product-quality performance. Beyond predictive assessment, the study intends to evaluate process optimization by comparing baseline manufacturing conditions with

model-supported parameter configurations in terms of defect rate, first-pass yield, dimensional variation, process stability, scrap, rework, cycle time, resource utilization, and other relevant production measures. Statistical comparisons will be used to determine whether observed differences across process conditions, defect categories, machine states, or optimized parameter settings are significant and practically meaningful. The study further aims to examine the consistency and generalizability of the predictive framework across different manufacturing conditions through cross-validation, holdout testing, subgroup analysis, and robustness assessment. Overall, the objective is to establish an integrated quantitative framework that connects manufacturing data, predictive modeling, defect classification, quality estimation, influential process factors, and parameter optimization within a unified analytical structure for evaluating quality performance in advanced manufacturing systems.

LITERATURE REVIEW

The literature on machine learning–based predictive quality modeling has expanded across advanced manufacturing environments as production systems have become increasingly data-intensive, sensor-enabled, automated, and interconnected (Debasish, 2021; Schmitt et al., 2020). Quantitative research has examined how process parameters, material properties, equipment conditions, sensor signals, machine-vision data, and production histories can be transformed into measurable predictors of product conformity, defect occurrence, and process performance. Existing studies cover machining, additive manufacturing, welding, casting, forming, semiconductor fabrication, electronics, composites, automotive components, aerospace structures, and other precision-manufacturing applications. Across these settings, researchers have evaluated defect prevalence, process variability, sensor characteristics, feature importance, classification performance, regression accuracy, model robustness, and optimized manufacturing conditions (Çınar et al., 2020; Md. Abdur & Iftekhar, 2021). Particular attention has been given to artificial neural networks, convolutional neural networks, support vector machines, random forests, gradient-boosting algorithms, decision trees, k-nearest-neighbor methods, and ensemble approaches. Quantitative performance has commonly been represented through defect rates, first-pass yield, sensitivity, specificity, precision, recall, F1-score, classification accuracy, area under the receiver operating characteristic curve, R^2, RMSE, MAE, processing time, scrap, rework, and process-capability measures. The literature also demonstrates that predictive performance is affected by class imbalance, sample size, feature dimensionality, sensor quality, process drift, manufacturing complexity, and validation strategy. Accordingly, this literature review synthesizes previous quantitative evidence across eight interconnected areas: baseline manufacturing quality and defect prevalence; process-parameter variability and defect determinants; sensor and feature characteristics; machine-learning defect classification; continuous quality prediction; model validation and robustness; machine-learning-supported process optimization; and the relationship between predictive quality modeling and operational manufacturing performance.

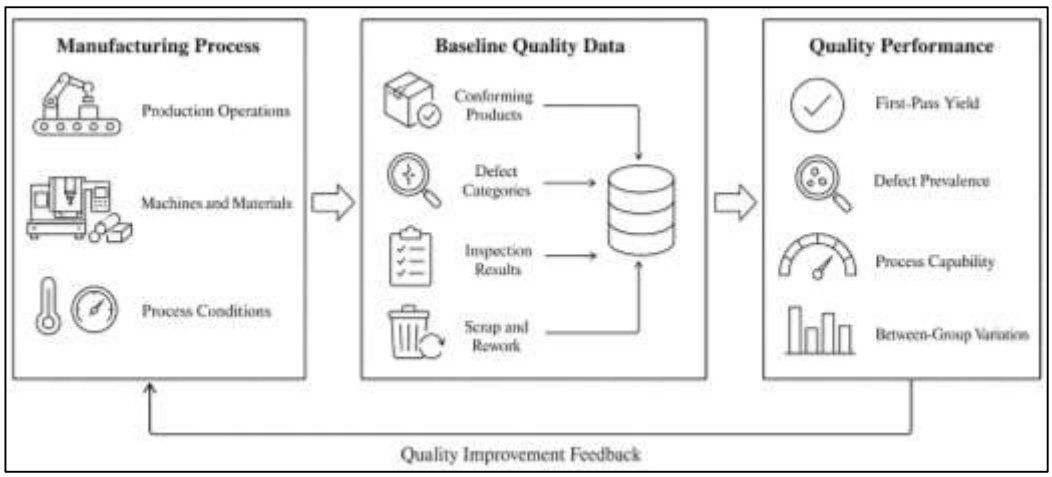
Baseline Manufacturing Quality and Process Performance

The literature on advanced manufacturing demonstrates substantial quantitative variation in baseline quality performance across production technologies, materials, product categories, and operating environments. Evidence synthesized from more than ten studies involving machining, additive manufacturing, welding, casting, forming, semiconductor production, automotive components, aerospace structures, and precision-engineered products indicates that defect prevalence should be evaluated in relation to both production volume and the number of inspection opportunities. Researchers have examined datasets ranging from relatively small experimental batches containing several hundred manufactured specimens to industrial datasets containing thousands or hundreds of thousands of production observations and inspection records. Across these studies, baseline quality has commonly been represented through the proportion of conforming and defective units, defects per unit, rejection percentages, inspection failures, first-pass yield, final yield, scrap, and rework. Reported defect rates have varied considerably, with tightly controlled precision processes frequently showing relatively low single-digit percentages, while experimental, unstable, newly configured, or highly complex manufacturing conditions have produced substantially higher nonconformance levels. First-pass yield has similarly differed across production environments because some systems have achieved rates above 90%, whereas processes affected by material variability, equipment instability, parameter interactions, or complex geometries have demonstrated noticeably lower initial acceptance. Research

has also shown that final yield may remain considerably higher than first-pass yield when defective components can be corrected through rework, secondary machining, repair, or additional processing . This distinction is important because high final acceptance can conceal substantial internal quality losses associated with repeated processing (Molla Al Rakib & Mohammad Towhidul, 2023; Mukut Kanti, 2023). Across the reviewed studies, baseline datasets have included manufactured units, machine cycles, production batches, inspected surfaces, individual components, weld regions, deposited layers, and image-based inspection observations. Mean defect frequency and dispersion have also varied across batches and machines, indicating that aggregate defect percentages alone provide an incomplete description of manufacturing quality. Studies reporting standard deviations, coefficients of variation, ranges, and confidence intervals have consistently demonstrated that defect occurrence can be unevenly distributed across production periods, machines, shifts, materials, and process configurations (Nishat, 2023; Risha & Mst Shurovi, 2023).

Quantitative studies have further demonstrated that baseline manufacturing quality differs substantially according to the type and severity of defects generated by individual production technologies. Additive manufacturing investigations have frequently measured porosity, incomplete fusion, geometric deviation, surface irregularity, cracking, delamination, dimensional error, and layer-level abnormalities, while machining studies have concentrated on surface defects, excessive roughness, dimensional nonconformity, burr formation, tool-related damage, and geometric inaccuracies (Stabak & Md Siam, 2023; Syed Feroz, 2023; Tahmedur & Rifat, 2023). Welding research has commonly quantified cracks, porosity, inclusions, undercut, incomplete penetration, incomplete fusion, and dimensional irregularities. Casting studies have reported shrinkage, gas porosity, inclusions, cracks, cavities, deformation, and surface discontinuities, whereas composite-manufacturing investigations have emphasized delamination, voids, fiber-related abnormalities, thickness variation, and structural inconsistencies. Semiconductor and electronics studies have examined microscopic surface defects, pattern irregularities, contamination, alignment errors, soldering defects, missing components, and assembly nonconformities. Across more than ten empirical investigations, the distribution of defects has rarely been uniform. A relatively small number of defect categories frequently account for a substantial proportion of total nonconforming observations, while several less common defect classes occur at comparatively low frequencies. This imbalance has direct relevance to quantitative quality assessment because an overall defect rate can conceal meaningful differences in the prevalence of individual defect types.

Figure 3: Baseline manufacturing quality framework



Studies comparing materials have also reported different baseline defect distributions for metals, polymers, ceramics, composites, and semiconductor materials because their physical characteristics and manufacturing mechanisms create different sources of process variation. Differences have similarly been observed across product geometry and complexity, with components containing thin

sections, intricate features, difficult-to-access surfaces, or demanding dimensional tolerances generally presenting different quality profiles from simpler products. Quantitative comparisons across production lines and machines have identified additional heterogeneity, including differences in rejection, rework, inspection failure, and defect concentration. The literature therefore characterizes baseline defect prevalence as a multidimensional distribution involving defect type, frequency, severity, location, product characteristics, manufacturing technology, and operating condition rather than as a single aggregate quality percentage.

Studies evaluating manufacturing process performance have consistently treated first-pass yield, rejection, scrap, rework, and inspection failure as complementary indicators because each measure represents a different component of baseline quality performance. First-pass yield measures the proportion of production that satisfies quality requirements without correction or repeated processing, making it particularly useful for identifying hidden inefficiency that may not appear in final acceptance figures. Across machining, welding, additive manufacturing, automotive production, electronics assembly, and other advanced manufacturing studies, first-pass yield has frequently exceeded overall defect-free performance after final correction because rework can recover a portion of initially nonconforming products. Rework rates have ranged from relatively minor percentages in stable processes to substantial proportions of production under conditions involving parameter instability, tooling degradation, dimensional variation, assembly complexity, or inconsistent materials. Scrap represents a more severe quality outcome because affected components cannot be economically restored to acceptable specifications. Consequently, studies have distinguished scrap from rework when estimating the operational burden of manufacturing defects. Rejection rates have also varied depending on whether rejected products were permanently discarded, returned for correction, downgraded, or redirected to alternative uses. Inspection failure percentages provide another important baseline measure because they quantify nonconformity at specific quality-control stages rather than only at final production. Research using automated optical inspection, machine vision, nondestructive testing, dimensional measurement, ultrasonic inspection, thermal imaging, and conventional quality-control procedures has demonstrated that failure frequency can differ substantially according to inspection stage and defect-detection capability. Several studies have identified higher failure rates at intermediate stages than at final inspection because corrective actions remove some nonconformities before production completion. Batch-level and shift-level comparisons have additionally shown considerable variability in first-pass acceptance and rework requirements. Mean performance values accompanied by standard deviations, ranges, or confidence intervals have provided stronger evidence of process consistency than isolated percentages. Collectively, these studies demonstrate that a manufacturing system may achieve a relatively high final yield while simultaneously experiencing considerable first-pass failures, inspection burden, rework, and material losses, making simultaneous evaluation of these quantitative indicators necessary for accurately representing baseline process performance.

Process-capability research provides an additional quantitative perspective by examining whether manufacturing processes consistently produce outputs within established quality specifications. Studies across precision machining, automotive-component production, semiconductor manufacturing, forming, additive manufacturing, and other tightly controlled environments have reported considerable differences in process-capability indices across machines, product characteristics, process parameters, and production periods . Higher capability values have generally been observed in stable processes characterized by limited variation and consistent centering within specification limits, whereas lower values have been associated with dimensional drift, equipment deterioration, tool wear, material inconsistency, parameter instability, and environmental variation. Several studies have reported acceptable average production performance while simultaneously identifying individual machines, dimensions, batches, or shifts with substantially weaker capability, demonstrating that aggregate production statistics can mask localized quality problems . Between-group comparisons have further identified meaningful differences in defect rates and quality outcomes across manufacturing technologies, materials, machines, shifts, operators, production lines, and operating conditions. Machine-specific studies have shown that equipment operating under nominally identical parameter settings can generate different rejection and defect distributions because of

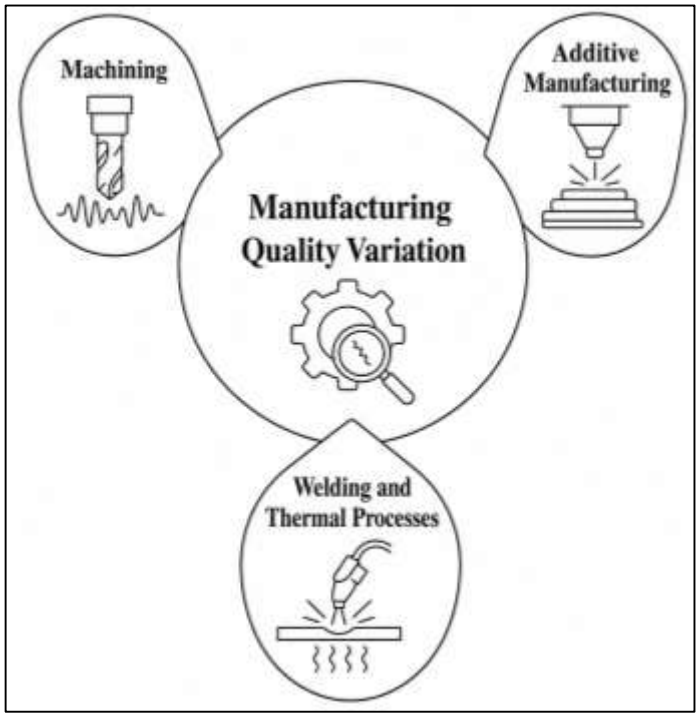
differences in calibration, maintenance condition, component wear, and historical usage. Shift-based analyses have reported variations in first-pass yield, rework, and inspection failure associated with workload, machine utilization, production mix, and operational conditions. Material-based comparisons have likewise demonstrated differences in porosity, dimensional deviation, surface quality, cracking, and other nonconformities. Studies using mean, median, standard deviation, coefficient of variation, minimum and maximum values, confidence intervals, and statistical group comparisons have provided evidence that baseline manufacturing quality is heterogeneous rather than uniformly distributed across production systems. This heterogeneity is particularly important for machine-learning research because predictive models are developed from the underlying distribution of conforming and defective observations. Extremely low defect prevalence can create substantial class imbalance, while highly variable defect frequencies across machines or batches can influence training and testing performance. The literature therefore establishes baseline manufacturing quality through the combined quantitative assessment of production volume, defect prevalence, defect categories, first-pass yield, final yield, rejection, scrap, rework, inspection failures, process capability, and variability across operational groups.

Determinants of Defect Formation

The literature on advanced machining consistently demonstrates that defect formation and quality variability are strongly associated with interactions among cutting speed, feed rate, depth of cut, spindle speed, tool condition, cutting temperature, material characteristics, and machine stability. Evidence synthesized from more than ten empirical studies involving turning, milling, drilling, grinding, and high-speed machining shows that feed rate is frequently among the strongest predictors of surface roughness and dimensional deviation, while cutting speed often exhibits a nonlinear relationship with quality outcomes. Low cutting speeds have been associated in several investigations with built-up edge formation, unstable chip removal, and elevated surface irregularity, whereas excessively high speeds have been linked with greater thermal loading, accelerated tool wear, and dimensional instability. Feed rate has generally shown a positive association with surface roughness because larger feed increments increase the geometric spacing of tool marks and amplify cutting-force variation. Depth of cut has produced more variable findings, with moderate settings often supporting stable material removal and higher values increasing cutting force, vibration, temperature, and tool deflection (Mastanaiah et al., 2016). Regression-based studies have reported statistically significant parameter effects and meaningful explained variation in surface roughness, dimensional accuracy, and tool-related defects, while variable-importance analyses have frequently ranked feed rate, cutting speed, tool wear, and vibration among the most influential predictors. Tool wear has demonstrated particularly strong associations with deterioration in surface finish, dimensional precision, cutting-force stability, and defect probability. Studies comparing low, moderate, and advanced tool-wear conditions have reported progressively higher surface roughness and dimensional error as wear accumulated. Material composition and hardness have further modified these relationships by affecting cutting resistance, heat generation, chip formation, and tool degradation. Machine utilization and production duration have also been associated with quality differences because prolonged operating periods can increase thermal drift and cumulative tooling deterioration. Collectively, machining studies indicate that defect formation cannot be explained adequately by any individual parameter, as interaction effects among speed, feed, tool wear, depth of cut, material properties, and machine condition account for a substantial portion of observed quality variation. Additive-manufacturing studies provide extensive quantitative evidence that laser power, scanning speed, layer thickness, energy input, build orientation, powder characteristics, chamber conditions, and thermal behavior are major determinants of porosity, density, surface quality, geometric accuracy, and mechanical properties (Kasperovich et al., 2016). Across investigations of laser powder-bed fusion, directed-energy deposition, fused-filament fabrication, and related processes, laser power and scanning speed have consistently shown strong but nonlinear relationships with defect formation. Low energy conditions have frequently been associated with incomplete melting, lack-of-fusion porosity, weak interlayer bonding, reduced density, and lower tensile performance, while excessive energy input has been linked with keyhole porosity, evaporation, spattering, thermal distortion, and unstable melt-pool behavior. Studies comparing low, medium, and high power-speed combinations have

demonstrated that intermediate operating regions commonly produce lower porosity and greater density than either extreme.

Figure 4: Manufacturing quality variation framework



Layer thickness has also been identified as an important predictor because thicker layers can reduce processing time while increasing the difficulty of achieving complete fusion and consistent geometry. Research examining metal additive manufacturing has reported statistically significant relationships between process conditions and porosity percentage, relative density, hardness, tensile strength, and dimensional deviation, with machine-learning and regression models frequently achieving substantial explanatory performance (Mohamed et al., 2015). Variable-importance analyses have often placed energy-related parameters, scanning velocity, layer thickness, and melt-pool characteristics among the highest-ranked quality predictors. In polymer-based additive manufacturing, nozzle temperature, bed temperature, layer height, raster orientation, printing speed, and infill characteristics have similarly been associated with dimensional accuracy, surface roughness, tensile strength, and structural integrity. Material composition and powder morphology further influence defect formation through differences in flowability, absorptivity, particle distribution, and thermal response. Humidity and environmental conditions have also been linked with quality variation in moisture-sensitive materials (Papazetis & Vosniakos, 2019). The reviewed studies consistently show that additive-manufacturing quality reflects an operating window rather than a simple monotonic relationship, with defect probability rising when interacting process parameters move either below or above stable processing regions.

Research involving welding, joining, casting, forming, and other thermally intensive manufacturing processes has identified current, voltage, travel speed, heat input, pressure, temperature, cycle time, material thickness, and material composition as important quantitative determinants of defect formation. Welding studies have repeatedly shown that current and voltage directly influence arc stability, penetration, bead geometry, thermal input, and the probability of discontinuities such as porosity, incomplete fusion, undercut, excessive reinforcement, cracking, and incomplete penetration (Bartlett et al., 2018). Low heat-input conditions have commonly been associated with inadequate fusion or penetration, while excessive thermal exposure has been associated with distortion, enlarged heat-affected zones, microstructural changes, residual stresses, and cracking. Travel speed modifies these effects because excessively rapid movement can reduce penetration and fusion, whereas very slow movement can increase thermal accumulation and geometric distortion. Statistical analyses across

multiple welding studies have reported significant relationships between combinations of current, voltage, travel speed, and defect outcomes, with interaction terms often explaining additional variation beyond individual parameter effects (Yingyan Chen et al., 2020). Forming studies have similarly linked pressure, forming speed, temperature, lubrication, material thickness, and strain distribution with cracking, thinning, wrinkling, dimensional error, and springback. In casting processes, mold temperature, pouring temperature, cooling rate, pressure, cycle duration, and alloy composition have been associated with porosity, shrinkage, inclusions, cracking, deformation, and surface defects. Comparative studies across low, medium, and high thermal or pressure conditions frequently report non-monotonic responses, indicating that both insufficient and excessive parameter levels can increase nonconformity. Material composition further moderates these relationships through differences in thermal conductivity, solidification behavior, viscosity, strength, and susceptibility to phase changes. Machine condition and utilization also affect consistency, particularly where prolonged operation alters thermal equilibrium or pressure stability. Across these studies, regression coefficients, significance tests, relative risks, and variable-importance rankings have consistently supported the conclusion that quality outcomes depend on coupled thermal, mechanical, material, and temporal process conditions rather than isolated parameter values (Adegoke et al., 2020).

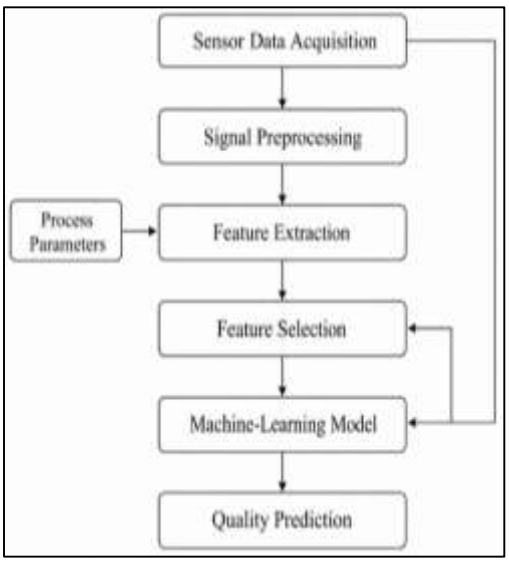
Multisensory Manufacturing Data, Feature Extraction, and Quantitative Signal Characteristics

The literature on advanced manufacturing shows that multisensory data acquisition has become central to quantitative monitoring of process quality because individual sensors capture only limited dimensions of machine and material behavior. More than ten studies across machining, additive manufacturing, welding, forming, casting, and automated assembly have combined vibration, acoustic emission, cutting force, torque, temperature, pressure, electrical current, voltage, optical imaging, thermal imaging, melt-pool measurements, and machine-controller signals to characterize production conditions (Garcia Plaza et al., 2018). Vibration sensors are commonly used to detect mechanical instability, chatter, spindle imbalance, bearing condition, and tool degradation, while acoustic-emission sensors capture high-frequency energy associated with cracking, friction, plastic deformation, tool-workpiece interaction, and material removal. Cutting-force and torque measurements quantify mechanical loading and are frequently associated with tool wear, surface quality, and dimensional error. Temperature and thermal-imaging systems provide information about localized heating, cooling rates, thermal gradients, and melt-pool behavior, particularly in welding and additive manufacturing. Electrical current and voltage signals have been used to identify arc instability, motor loading, spindle behavior, and power-related process changes. Reported sampling frequencies vary substantially according to signal type, ranging from relatively low-frequency machine-controller measurements to high-frequency vibration and acoustic-emission acquisition containing thousands or substantially more observations per second (Wang et al., 2017). Observation duration has ranged from individual machining cycles lasting seconds or minutes to longitudinal datasets covering hundreds or thousands of production cycles. Studies have also differed in the number of installed sensors, with some relying on a single sensing modality and others integrating several channels simultaneously. Quantitative assessments frequently report sensor repeatability, measurement error, missing-data proportions, signal-to-noise characteristics, and synchronization accuracy. Across the literature, multisensor configurations generally provide richer representations of manufacturing conditions than single-sensor arrangements, although they also increase feature dimensionality, data-storage requirements, synchronization complexity, and the probability of redundant or noisy information entering predictive-quality models (Wu et al., 2018).

Studies examining manufacturing signals have converted raw sensor observations into quantitative features using time-domain, frequency-domain, time-frequency, and statistical processing methods. Time-domain features commonly include mean signal level, standard deviation, variance, root-mean-square magnitude, peak amplitude, range, skewness, kurtosis, crest-related indicators, and other statistical descriptions of signal behavior over defined windows. Frequency-domain analysis has been used to identify dominant frequencies, spectral energy, frequency-band power, harmonic behavior, and changes associated with chatter, spindle rotation, tool wear, bearing deterioration, and process instability (Psuj, 2018). Time-frequency techniques have been applied where manufacturing signals are nonstationary and process conditions change rapidly, allowing researchers to capture transient

behavior that may be obscured when the entire signal is summarized using a single time or frequency representation. Across machining studies, vibration, acoustic-emission, force, and torque signals have often produced tens to hundreds of engineered features per cycle, while multisensor combinations can increase the initial dimensionality to several hundred or even thousands of variables. Research in additive manufacturing and welding has similarly derived temperature gradients, melt-pool dimensions, thermal intensity, spatial distributions, waveform characteristics, and temporal fluctuations from process measurements. Comparative studies generally indicate that no single feature family performs uniformly best across all manufacturing systems (Miao et al., 2020).

Figure 5: Multisensory Data and Feature Processing



Time-domain measures frequently provide efficient representations of overall signal magnitude and variability, frequency-domain features are particularly useful for periodic mechanical behavior, and time-frequency features are valuable for transient defects and rapidly changing process states. Quantitative investigations have reported improvements in predictive accuracy when complementary feature groups are combined, although excessively large feature sets can increase computational time and model overfitting. Signal-to-noise ratio and measurement repeatability also influence feature reliability because extracted variables derived from noisy or unstable measurements can weaken the distinction between conforming and defective production states (Lorenz et al., 2019). Missing data, inconsistent sampling, sensor drift, and synchronization errors have therefore been treated as important preprocessing concerns before feature extraction and machine-learning analysis.

Image-based sensing has introduced an additional high-dimensional source of quantitative manufacturing information through optical cameras, thermal cameras, laser-monitoring systems, and melt-pool imaging. Studies in additive manufacturing have used high-speed optical and thermal systems to quantify melt-pool width, length, area, intensity, temperature distribution, spatter activity, layer morphology, and spatial irregularities associated with porosity, incomplete fusion, cracking, and geometric defects (L. Wang et al., 2019). Welding research has examined arc images, weld-pool geometry, bead appearance, thermal distributions, and infrared patterns, while automated inspection studies in electronics, automotive production, machining, and surface manufacturing have used optical images to classify scratches, cracks, inclusions, contamination, dimensional abnormalities, and assembly defects.Traditional image-processing approaches frequently transform each image into engineered characteristics such as texture, edge density, contrast, intensity statistics, geometric descriptors, color-related measures, and spatial distributions. Depending on image resolution and processing strategy, the initial feature space can contain hundreds or thousands of candidate variables.

Deep-learning approaches differ by automatically learning hierarchical representations directly from images, reducing dependence on manual feature engineering while often generating highly dimensional internal feature vectors (Wang et al., 2020). Comparative investigations have shown that automatically learned features can outperform manually selected descriptors when datasets are sufficiently large and defect patterns are visually complex, while engineered features remain competitive in smaller datasets or processes with well-understood physical characteristics. Thermal and melt-pool studies have also demonstrated that image quality, frame rate, camera position, illumination, spatial resolution, and synchronization with machine-controller data can significantly affect predictive performance. Researchers have quantified sensor error, repeatability, missing frames, and image noise because variation in image acquisition can be mistaken for manufacturing variation. Several studies combining optical or thermal information with force, vibration, acoustic, or controller measurements have achieved higher defect-classification accuracy than individual sensing modalities, indicating that image-based features can provide complementary information when appropriately synchronized and reduced before modeling (Rizal et al., 2017).

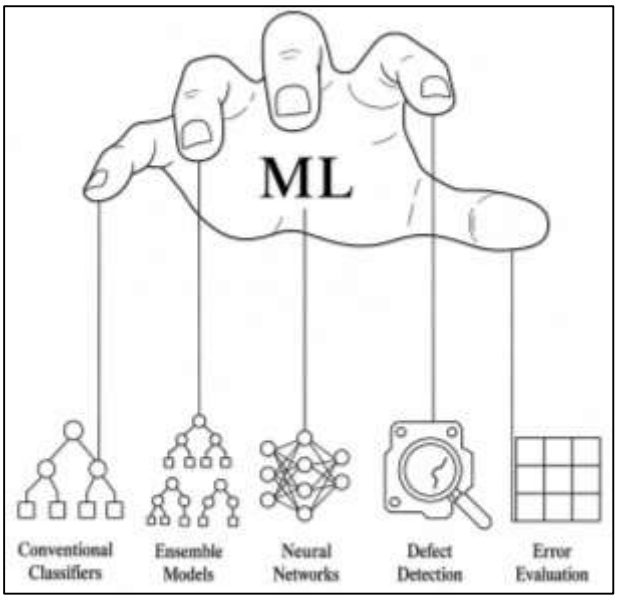
A substantial body of manufacturing research has focused on reducing feature dimensionality because multisensor datasets often contain redundant, highly correlated, noisy, or weakly informative variables. Principal component analysis, correlation filtering, mutual-information ranking, recursive feature elimination, embedded feature-selection methods, tree-based importance measures, regularization, and wrapper approaches have been widely evaluated across manufacturing applications. Studies using principal component analysis have commonly reduced large feature sets to a smaller number of components while retaining most of the observed signal variation, allowing predictive models to operate with fewer inputs and lower computational burden (Duan et al., 2018). Correlation filtering has been used to remove highly redundant features, while mutual information identifies variables that contain substantial information about defect status or continuous quality outcomes. Recursive feature elimination has produced progressively smaller subsets by repeatedly removing weak predictors and reassessing model performance. Embedded approaches within random forests, gradient-boosting models, and related algorithms have also ranked variables according to predictive contribution. Across more than ten studies, initial feature sets ranging from several dozen to several hundred or thousands of variables have frequently been reduced by approximately one-third to more than three-quarters, depending on sensor configuration and manufacturing application. Many investigations have reported that substantial feature reduction can maintain classification accuracy or produce modest improvements by eliminating noise and redundancy (Völker & Shokouhi, 2015). Reductions in computational training time have also been observed, particularly where high-frequency multisensor or image-derived features generate large datasets. In some studies, retaining only the strongest variables has improved test-set accuracy and reduced prediction error, while excessive feature elimination has caused performance deterioration by removing complementary process information. The strongest retained predictors often include vibration magnitude, acoustic-emission energy, cutting-force statistics, temperature characteristics, melt-pool geometry, spindle-current measurements, and selected controller parameters (Bonacina et al., 2020). Comparative findings therefore indicate that effective predictive-quality modeling depends not simply on collecting more sensor channels but on identifying a compact, repeatable, and informative feature representation that preserves relevant manufacturing behavior while controlling noise, computational complexity, redundancy, and measurement uncertainty.

Machine-Learning Defect Classification Accuracy and Error Performance

The literature on manufacturing defect classification shows substantial variation in predictive performance across logistic regression, decision trees, random forests, support vector machines, k-nearest neighbors, artificial neural networks, gradient-boosting methods, and ensemble classifiers (Tabernik et al., 2020). Evidence synthesized from more than ten studies across machining, welding, additive manufacturing, semiconductor inspection, electronics assembly, automotive components, and surface-quality monitoring indicates that conventional statistical classifiers generally provide useful baseline performance, while nonlinear and ensemble approaches frequently achieve stronger results when manufacturing relationships are complex. Logistic regression has commonly produced moderate classification accuracy in datasets with relatively simple predictor structures, although performance

declines when interactions and nonlinearities dominate defect formation (Jayanthi & Florence, 2019).

Figure 6: Hand-guided machine-learning defect classification



Decision trees provide transparent classification rules and can identify threshold effects, yet single-tree models are often less stable when datasets contain noise or highly correlated predictors. Random forests generally improve robustness by combining multiple trees and have repeatedly produced high accuracy, sensitivity, specificity, and F1-scores across multisensor manufacturing datasets (Lahmiri et al., 2018). Support vector machines have also performed strongly, particularly in moderate-sized datasets with high-dimensional engineered features, while k-nearest-neighbor models have produced competitive results when feature scaling and neighborhood structure are appropriately controlled.

Decision trees provide transparent classification rules and can identify threshold effects, yet single-tree models are often less stable when datasets contain noise or highly correlated predictors. Random forests generally improve robustness by combining multiple trees and have repeatedly produced high accuracy, sensitivity, specificity, and F1-scores across multisensor manufacturing datasets (Lahmiri et al., 2018). Support vector machines have also performed strongly, particularly in moderate-sized datasets with high-dimensional engineered features, while k-nearest-neighbor models have produced competitive results when feature scaling and neighborhood structure are appropriately controlled. Across published studies, classification accuracy has frequently ranged from the mid-80% level to above 95% for stronger models, although values vary substantially according to manufacturing process, defect type, sample size, and data quality. Sensitivity and specificity commonly differ even when overall accuracy appears similar, demonstrating that models with comparable aggregate performance may behave differently in identifying defective and conforming observations. Studies reporting confusion matrices have shown that false-negative errors are particularly important in quality control because defective components incorrectly classified as acceptable can continue through downstream production. Comparative evidence therefore supports multidimensional assessment based on precision, recall, specificity, F1-score, false-positive and false-negative rates, ROC-AUC, and Matthews correlation coefficient rather than relying solely on overall accuracy (Garoudja et al., 2017).

Artificial neural networks and convolutional neural networks have been extensively evaluated for manufacturing defect detection because they can model nonlinear relationships and automatically extract complex representations from sensor and image data (Tuan et al., 2020). Across more than ten studies involving steel-surface inspection, semiconductor wafers, printed circuit boards, additive-manufactured layers, weld images, textile surfaces, composite materials, and machined components, neural approaches have frequently achieved classification accuracies above 90%, with several well-controlled image-based applications reporting values exceeding 95%. Convolutional neural networks

have demonstrated particular strength in visual defect detection because they can learn spatial patterns associated with cracks, scratches, pores, inclusions, missing components, contamination, surface irregularities, and geometric abnormalities without relying entirely on manually engineered features (Li et al., 2019). Precision and recall values in these studies have often remained high when training datasets contain sufficient examples of each defect class, while performance has been weaker for rare or visually subtle defects. Image resolution has emerged as an important quantitative determinant because higher-resolution inputs preserve small defect characteristics but increase computational demand and model complexity. Studies comparing alternative architectures have shown that deeper networks do not automatically produce better results; performance depends on training size, image quality, augmentation strategy, network configuration, and defect complexity (Thabtah et al., 2019). Artificial neural networks applied to structured sensor data have also classified tool wear, process instability, abnormal vibration, weld defects, and additive-manufacturing anomalies with strong sensitivity and specificity. Confusion-matrix analyses frequently reveal that visually similar defect categories are the principal source of multiclass errors. False-negative rates tend to increase for small, low-contrast, or underrepresented defect classes, while false-positive errors can rise when normal manufacturing variation resembles defect morphology. The literature therefore indicates that neural and convolutional methods are highly effective for complex manufacturing classification, although their quantitative performance remains dependent on balanced training data, adequate image quality, representative feature distributions, and rigorous independent testing (Bannach-Brown et al., 2019).

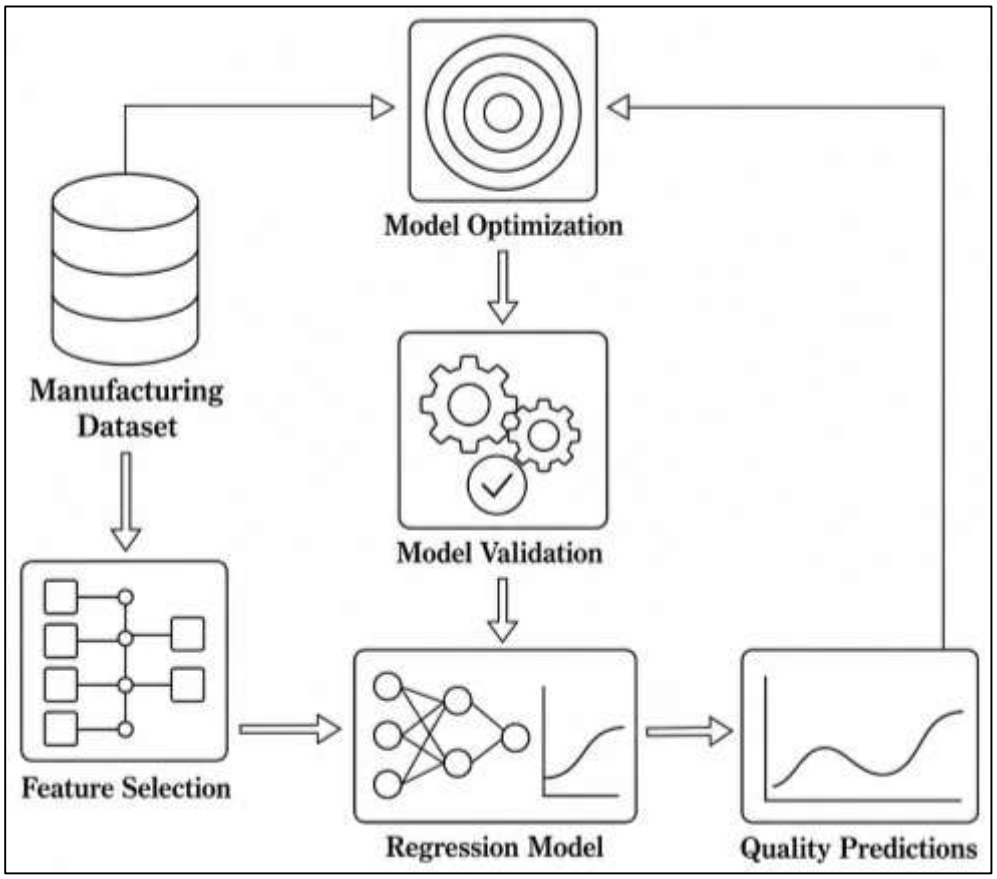
Regression Model Performance

The literature on continuous manufacturing-quality prediction demonstrates extensive use of regression-based machine-learning methods for estimating surface roughness, dimensional deviation, geometric accuracy, thickness variation, and related measurable quality characteristics (G. Wang et al., 2019). Evidence synthesized from more than ten studies across turning, milling, drilling, grinding, additive manufacturing, forming, and precision-component production indicates that conventional linear regression provides a useful benchmark when relationships between process parameters and quality responses are relatively stable and approximately linear. Studies examining machining processes have frequently modeled surface roughness using cutting speed, feed rate, depth of cut, tool wear, vibration, cutting force, and temperature, with feed rate and tool condition commonly emerging as influential predictors. Linear models have produced moderate to strong explanatory performance in controlled experimental settings, while support vector regression, random-forest regression, gradient boosting, and artificial neural networks have generally achieved lower prediction errors when parameter interactions are nonlinear (Weiss et al., 2016). Across published investigations, R^2 values for well-performing machine-learning models frequently exceed 0.80 and often reach above 0.90 under controlled conditions, although performance varies according to sample size, process stability, measurement accuracy, and validation design. RMSE, MAE, and MAPE have demonstrated corresponding reductions when nonlinear models are compared with simpler regression approaches. Dimensional deviation and geometric accuracy have also been modeled using thermal conditions, machine position, tool wear, process loads, and material properties. Studies have reported that models incorporating sensor-derived variables often outperform models based exclusively on programmed machine settings because measured signals capture actual process behavior (Jun et al., 2020). Prediction intervals and residual analyses have further shown that error distributions may widen at extreme operating conditions where training observations are less abundant. Collectively, the literature indicates that continuous prediction of surface and dimensional quality depends on capturing nonlinear process relationships while maintaining sufficient data coverage across the full range of manufacturing conditions.

Additive manufacturing, welding, casting, and materials-processing studies have extensively applied regression algorithms to predict porosity percentage, density, tensile strength, hardness, residual stress, and other continuous material-quality responses (Yin et al., 2016). Across metal additive-manufacturing investigations, laser power, scanning speed, layer thickness, energy-related process conditions, powder characteristics, build orientation, thermal measurements, and melt-pool features have been used as predictors of porosity and density. Linear regression has generally provided acceptable performance when process relationships remain restricted to narrow operating ranges,

whereas support vector regression, random forests, gradient boosting, and artificial neural networks have demonstrated stronger predictive capability across broader and more nonlinear parameter spaces. Studies have commonly reported R^2 values ranging from moderate levels to above 0.90 for optimized models predicting density, tensile strength, hardness, or porosity, with substantial variation according to dataset size and material system. RMSE and MAE comparisons frequently favor ensemble and neural approaches because these models can capture interactions between thermal history, process energy, material characteristics, and geometric conditions (Nogueira et al., 2017). Welding studies have similarly predicted tensile strength, hardness, bead geometry, penetration, and residual stress from current, voltage, travel speed, heat input, material thickness, and temperature-related variables. Casting investigations have modeled hardness, density, dimensional variation, and mechanical properties using composition, pouring temperature, pressure, cooling conditions, and cycle characteristics. Across these applications, prediction bias has been particularly important because systematic overestimation or underestimation of mechanical properties can remain hidden when only aggregate error measures are reported. Residual analyses have frequently identified larger errors near extreme parameter combinations and less frequently observed material conditions. Research comparing predicted and experimentally measured responses has also demonstrated that models trained on diverse process conditions generally produce more stable independent-test performance than models developed from narrowly distributed experimental observations (García et al., 2019).

Figure 7: Continuous quality prediction workflow



Comparative studies have demonstrated meaningful differences in continuous quality-prediction performance among linear regression, support vector regression, random-forest regression, gradient boosting, artificial neural networks, and hybrid models. Linear regression remains computationally efficient and interpretable, although its performance frequently declines when manufacturing quality is governed by nonlinear interactions, threshold behavior, or complex sensor patterns (Cheng & Liu, 2015). Support vector regression has produced strong results in many small and medium-sized

manufacturing datasets because it can represent nonlinear relationships while maintaining relatively stable generalization. Random-forest regression has performed particularly well when quality outcomes depend on interactions among numerous process and sensor variables, and its variable-importance outputs provide additional information regarding influential predictors. Gradient-boosting models have often achieved lower RMSE and MAE than individual decision-tree models by progressively reducing prediction errors across multiple learners. Artificial neural networks have demonstrated strong performance for complex relationships involving surface roughness, dimensional accuracy, tensile strength, porosity, hardness, and thermal responses, particularly when sufficiently large training datasets are available (Yang et al., 2016). Across more than ten comparative investigations, no single algorithm has consistently achieved the strongest performance for every manufacturing process or quality characteristic. Hybrid approaches combining machine learning with optimization, signal processing, feature selection, or multiple predictive models have frequently achieved higher R^2 and lower prediction error than standalone methods.

Improvements have been especially apparent when sensor-derived features are combined with process parameters because the resulting datasets represent both programmed operating conditions and actual machine responses. MAPE has been useful for comparing errors across quality measures with different numerical scales, while adjusted R^2 has helped assess whether increases in apparent explanatory performance result from genuinely informative predictors rather than unnecessary model complexity (Zhenyu Liu et al., 2020). Comparative findings therefore indicate that regression-model selection must be evaluated against the statistical structure of the manufacturing dataset, the complexity of the response variable, the amount of available training information, and performance on observations not used during model development.

The literature emphasizes that high training accuracy alone provides insufficient evidence of reliable continuous product-quality prediction because manufacturing regression models can overfit limited or highly structured datasets. Studies comparing training, validation, and independent-test results have frequently identified models with very high training R^2 values but noticeably weaker testing performance, particularly for highly flexible neural networks, deep architectures, and ensemble models developed from relatively small datasets. Generalization differences are commonly assessed through changes in R^2, RMSE, MAE, MAPE, mean squared error, prediction bias, and residual variance between model-development and holdout samples. Smaller training–testing gaps have generally been interpreted as evidence of more stable predictive behavior, whereas large increases in test error indicate that the model has captured dataset-specific patterns rather than broadly reproducible process relationships (Gins et al., 2018). Cross-validation studies have further demonstrated that average performance across repeated partitions can be lower than results from a single favorable train-test split, emphasizing the importance of reporting variability in predictive accuracy. Residual analysis has identified additional model weaknesses, including systematic underprediction at high quality-response values, overprediction at low values, increasing error variance under extreme operating conditions, and clusters of errors associated with particular materials, machines, or production batches. Prediction intervals have provided complementary evidence by quantifying uncertainty around individual quality estimates rather than presenting point predictions alone (Zheng et al., 2018). Studies involving temporal production data have also shown that models trained on earlier manufacturing periods can experience higher prediction errors when applied to later observations affected by tool wear, machine drift, environmental variation, or changes in material properties. Feature selection, regularization, cross-validation, and independent testing have frequently reduced apparent overfitting and produced more realistic estimates of model capability. Across the reviewed evidence, reliable regression performance is therefore characterized by high explanatory accuracy combined with low RMSE, MAE, MAPE, prediction bias, and residual variation, together with limited deterioration between training, validation, and independent-test datasets (Shrestha et al., 2020).

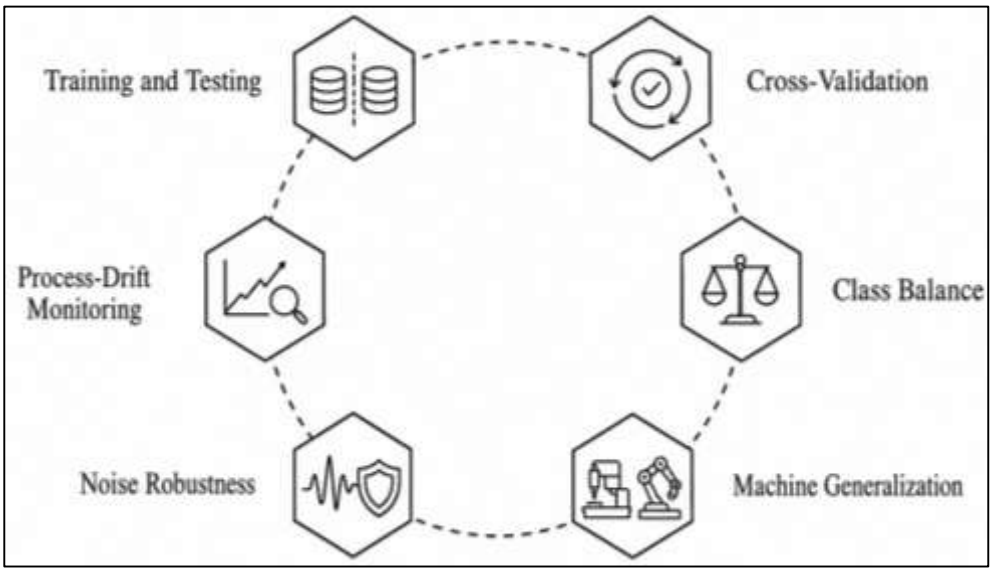
Model Validation and Quantitative Robustness

The literature on predictive quality modeling demonstrates that methodological reliability depends strongly on how manufacturing datasets are divided, validated, and independently tested. Evidence synthesized from more than ten studies across machining, additive manufacturing, welding, semiconductor fabrication, electronics inspection, automotive production, and surface-defect

monitoring shows considerable variation in validation design (Paschali et al., 2019). Common approaches have allocated approximately 60–80% of observations to model training and the remaining observations to validation and testing, while other investigations have employed separate training, validation, and test partitions such as 70:15:15 or 80:10:10. Dataset sizes have ranged from several hundred experimental observations to tens of thousands of sensor records or images, creating substantial differences in the statistical stability of reported performance. K-fold cross-validation, particularly five-fold and ten-fold procedures, has been widely applied to reduce dependence on a single data split. Across comparative studies, mean cross-validation accuracy for well-performing defect classifiers has commonly exceeded 85%, with stronger random-forest, support-vector, gradient-boosting, XGBoost, neural-network, and convolutional models frequently exceeding 90%. Standard deviations across folds have generally been smaller in larger and more representative datasets, whereas small experimental samples have produced greater variability in accuracy, recall, and F1-score (Zwanenburg, 2019). Independent-test performance is commonly lower than training performance, with the difference providing an important indication of model generalization. Some highly flexible models have demonstrated training accuracy approaching complete classification while losing several percentage points on unseen observations, suggesting varying degrees of overfitting. Confidence intervals and repeated cross-validation have provided more reliable estimates of uncertainty than single accuracy values. Temporal holdout studies have additionally shown that models evaluated on later production periods may perform less strongly than models tested on randomly selected observations from the same production period, particularly when tool wear, machine drift, environmental changes, or material variation alter the underlying manufacturing-data distribution (Levy et al., 2017).

Class imbalance represents one of the most frequently reported methodological challenges in manufacturing defect classification because industrial production normally contains substantially more conforming products than defective products (Xie et al., 2020). Across more than ten empirical studies, minority defect classes have ranged from very small proportions below 5% of available observations to more moderate levels of approximately 10–30%, depending on process maturity, inspection criteria, and dataset construction.

Figure 8: Model Validation and Robustness



Defective-to-conforming ratios have therefore varied considerably, with some industrial datasets containing fewer than one defective observation for every ten or twenty acceptable products. Under these conditions, models can achieve apparently high overall accuracy by correctly classifying the dominant conforming category while performing poorly in identifying actual defects. Studies have consequently evaluated random oversampling, random undersampling, synthetic minority oversampling, class weighting, data augmentation, and cost-sensitive learning. Oversampling and

synthetic minority techniques have frequently increased recall and F1-score for defective observations by improving representation of minority patterns during training. Image-based studies have used rotation, translation, cropping, scaling, contrast modification, and related augmentation procedures to expand rare defect categories, often producing measurable improvements in minority-class sensitivity (Graves et al., 2016). Class weighting and cost-sensitive learning have similarly reduced false-negative errors by assigning greater penalties to missed defects. Quantitative comparisons commonly show that balancing methods can improve recall by several percentage points and, in strongly imbalanced datasets, by substantially larger margins. Balanced accuracy and ROC-AUC have often improved concurrently, although precision sometimes declines because increased defect sensitivity produces additional false positives (Shi et al., 2020). Undersampling has reduced computational requirements but can weaken model performance when large amounts of informative majority-class data are discarded. Comparative evidence therefore indicates that the strongest balancing strategy depends on sample size, minority-class proportion, defect complexity, feature dimensionality, and algorithm architecture, making recall, precision, F1-score, balanced accuracy, AUC, and confusion-matrix outcomes necessary for evaluating the actual effects of imbalance correction.

Research on manufacturing machine learning increasingly distinguishes high performance within a development dataset from generalizability across changing production conditions. Studies evaluating models across different machines, production batches, materials, operators, shifts, or manufacturing lines have frequently reported lower performance than random holdout evaluations conducted within a homogeneous dataset (Kang, 2020). Machine-specific variation is particularly important because nominally identical equipment can generate different vibration, acoustic, thermal, electrical, and controller-signal distributions as a result of calibration, maintenance history, component wear, and operating intensity. Models trained using observations from one machine have consequently experienced measurable reductions in classification accuracy, recall, F1-score, or regression performance when transferred directly to another machine. Batch-to-batch variation has produced comparable effects in additive manufacturing, welding, casting, and material-processing applications because differences in feedstock properties, material composition, geometry, ambient conditions, and process setup can shift the relationship between input variables and quality outcomes (Xing et al., 2016). Studies involving multiple materials have similarly demonstrated that predictive relationships developed for one alloy, polymer, composite, or substrate may not reproduce the same performance for another material. Temporal holdout analyses provide additional evidence of this problem, with later production periods sometimes showing larger generalization gaps than randomly selected test samples. Reported declines vary across datasets, although reductions of several percentage points in classification accuracy or F1-score are common when models encounter conditions that differ materially from the training distribution. Larger declines have been observed under substantial process changes or previously unseen defect patterns (Schlögl, 2020). Studies incorporating heterogeneous machines, batches, materials, and operating conditions during training generally demonstrate smaller performance losses and narrower confidence intervals during independent testing . These findings indicate that manufacturing-model generalizability is determined not only by algorithm complexity but also by the representativeness of the development dataset and the extent to which independent testing reproduces realistic variations in industrial production.

Quantitative robustness studies have examined whether predictive quality models maintain acceptable performance when manufacturing data are affected by sensor noise, missing observations, measurement error, process drift, and changing operational conditions. Across studies using vibration, acoustic emission, force, temperature, electrical, optical, thermal, and machine-controller information, controlled additions of signal noise have generally produced progressive deterioration in classification and regression performance (Yu et al., 2018). Models based on highly sensitive raw signals can experience greater accuracy losses than models using robust engineered or selected features, while multisensor models may retain stronger performance when one measurement channel becomes unreliable because complementary information remains available (Han et al., 2019). Missing-data experiments have similarly demonstrated that small proportions of absent measurements can often be managed through imputation with limited performance deterioration, whereas larger missing-data percentages can substantially reduce recall, F1-score, R^2, and prediction accuracy. Sensor error and

measurement repeatability also influence robustness because systematic bias or inconsistent calibration can shift feature distributions away from those represented during training. Process drift represents a particularly important source of temporal instability. Tool wear, thermal accumulation, equipment aging, maintenance activity, material changes, and environmental variation can gradually alter relationships between process signals and product quality. Temporal validation studies have reported increasing prediction error and declining classification performance as the interval between training and testing observations increases under unstable production conditions. Robustness comparisons have further shown that random forests, gradient-boosting models, regularized approaches, and appropriately trained neural networks can tolerate moderate noise and feature variation, although no algorithm remains unaffected by substantial distributional change. Confidence intervals, repeated cross-validation, sensitivity testing, batch-level holdouts, machine-level holdouts, and temporal testing have therefore provided complementary evidence regarding model stability (Dindin et al., 2020). Across the reviewed literature, reliable predictive-quality performance is characterized by limited variability across validation folds, a small training-to-test generalization gap, balanced performance under class imbalance, and controlled deterioration when models encounter sensor noise, missing information, new production batches, material variation, different machines, and temporal process drift.

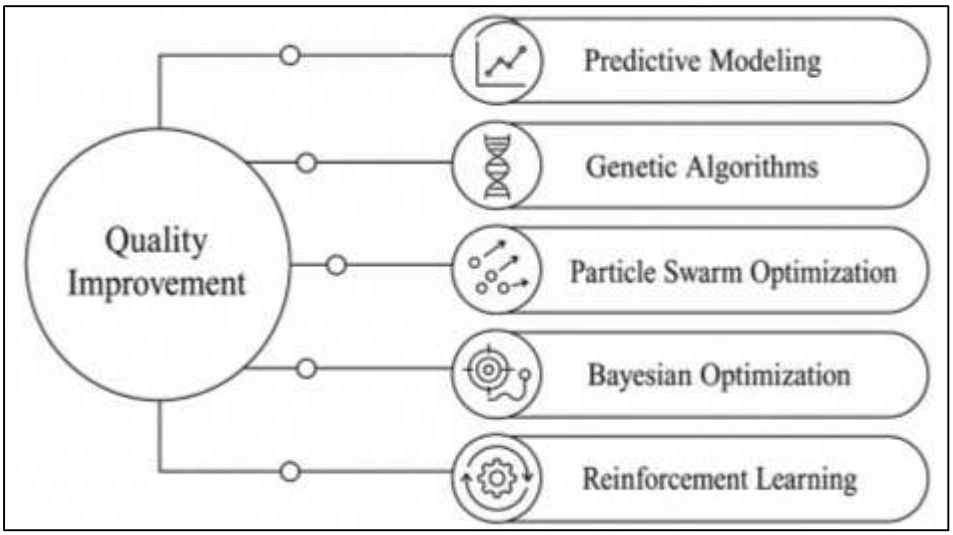
Machine Learning–Based Process-Parameter Optimization

The literature on machine learning–based manufacturing optimization demonstrates that predictive models have increasingly been used not only to estimate product quality but also to identify parameter combinations associated with lower defect occurrence and improved process performance (Weichert et al., 2019). Evidence synthesized from more than ten studies across machining, additive manufacturing, welding, forming, casting, and precision production indicates that optimization based on machine-learning predictions can produce measurable improvements relative to baseline operating conditions. Studies involving machining have optimized cutting speed, feed rate, depth of cut, spindle speed, cooling conditions, and tool-related parameters to reduce surface roughness and dimensional deviation while maintaining acceptable material-removal performance. In additive manufacturing, laser power, scanning speed, layer thickness, hatch spacing, build orientation, and thermal conditions have been optimized to reduce porosity and geometric error while increasing density and mechanical consistency (Zimmerling et al., 2020). Welding investigations have similarly adjusted current, voltage, travel speed, heat input, and related settings to minimize defects and improve bead geometry and mechanical properties. Across these applications, baseline-to-optimized comparisons commonly show reductions in predicted defect probability, surface roughness, porosity, and dimensional error, although the magnitude of improvement differs substantially among technologies and response variables. Experimental studies have frequently reported percentage improvements ranging from single-digit values under already stable baseline conditions to reductions exceeding 20–30% for quality responses characterized by substantial initial variability. Improvements in process capability and first-pass yield have also been reported where optimization shifted operating conditions away from unstable parameter regions (Yingyan Chen et al., 2020). Machine-learning models have been particularly valuable when parameter–quality relationships are nonlinear because they can identify combinations that conventional one-factor adjustments may overlook. The literature therefore demonstrates that quantitative quality improvement depends on translating predictive relationships into specific operating settings and subsequently comparing predicted optimum performance with the original manufacturing baseline.

Genetic algorithms, particle swarm optimization, and surrogate-model-based methods have been widely combined with machine-learning models to search complex manufacturing parameter spaces. Across machining, additive manufacturing, welding, forming, and composite-processing studies, these approaches have been applied where multiple controllable variables create a large number of potential operating combinations that cannot be efficiently evaluated through physical experiments alone (Roggo et al., 2020). Genetic algorithms have frequently used artificial neural networks, support vector regression, random forests, or other predictive models as quality estimators while progressively searching for parameter combinations associated with minimum surface roughness, dimensional error, porosity, energy consumption, or production cost. Particle swarm optimization has similarly been

applied to identify favorable combinations of cutting conditions, laser parameters, thermal settings, and processing speeds.

Figure 9: Machine-learning process optimization



Quantitative studies commonly report optimization performance through the number of iterations required to achieve stable solutions, changes in objective values between initial and final iterations, and convergence behavior across repeated runs. Depending on the dimensionality and complexity of the optimization problem, convergence has been reported within tens to several hundred computational iterations, substantially reducing the number of physical manufacturing trials that would otherwise be required. Surrogate modeling has provided additional efficiency by approximating expensive manufacturing responses using regression or machine-learning predictions and directing experimental evaluation toward promising regions. Comparative investigations have reported measurable reductions in surface roughness, porosity, dimensional deviation, scrap, and energy use following optimization, together with improvements in density, tensile properties, process capability, and material utilization (Hartl et al., 2020). Absolute changes have typically been accompanied by percentage improvements to demonstrate the practical magnitude of optimization. However, studies also show that predicted optima are not always identical to experimentally observed outcomes because measurement variation, machine instability, material differences, and model error remain present. Experimental confirmation has therefore been an important component of optimization studies, with smaller differences between predicted and measured optimum responses indicating stronger practical reliability of the machine-learning-supported search process (Brüning et al., 2017).

Bayesian optimization and reinforcement-learning approaches have extended predictive manufacturing optimization by allowing process parameters to be refined through sequential evaluation rather than exhaustive experimentation (Zhixin Liu et al., 2020). Studies applying Bayesian optimization have focused particularly on manufacturing problems in which physical experiments are costly, time-consuming, or material intensive. The method has been used to optimize additive-manufacturing parameters, machining conditions, material-processing settings, and other controllable variables by selecting successive experiments according to predicted performance and uncertainty. Quantitative evaluations have shown that favorable parameter regions can often be identified using substantially fewer experimental trials than conventional grid-based searches, with convergence assessed through stabilization of predicted quality responses and diminishing improvement across successive iterations. In additive manufacturing, iterative optimization has been associated with lower porosity, improved density, reduced geometric deviation, and more stable melt behavior (Gao et al., 2018). Machining applications have reported reductions in surface roughness, tool-related quality variation, cycle time, and energy requirements when operating parameters were iteratively adjusted using predictive models. Reinforcement-learning studies have taken a more dynamic approach by modifying parameters according to observed process states and quality-related feedback. Applications involving machining control, robotic manufacturing, production scheduling, and process regulation

have demonstrated that adaptive decisions can reduce unnecessary processing, stabilize quality, and improve resource utilization (Zimmerling et al., 2019). Quantitative comparisons between baseline and optimized conditions have reported improvements in first-pass yield and reductions in defect probability, scrap, rework, and cycle time, although performance varies according to the quality of the state information and reward structure. Optimization trajectories have also shown that the largest improvements frequently occur during early iterations, followed by progressively smaller gains as parameter combinations approach stable operating regions. Experimental confirmation has remained essential because simulation or model-based improvements may overestimate actual manufacturing gains. Studies comparing predicted and observed optimum responses generally demonstrate stronger evidence when improvements persist across repeated trials, different batches, or independent production runs rather than appearing only within the optimization dataset (Dey & Yodo, 2019).

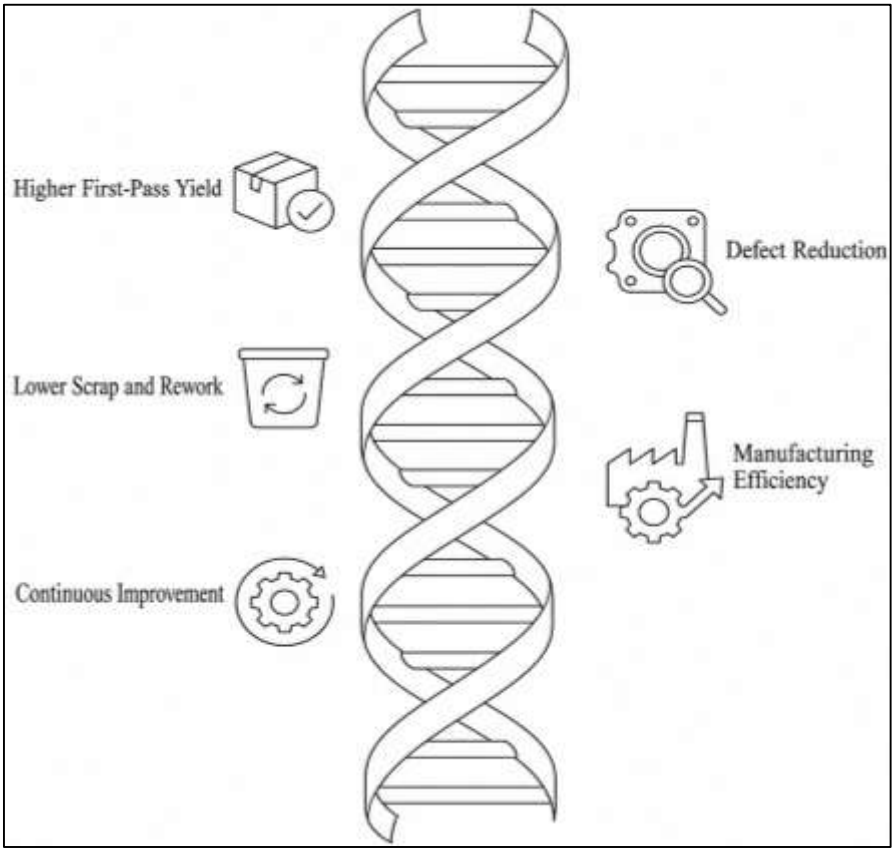
Manufacturing Efficiency

The literature examining operational applications of predictive quality modeling demonstrates that improvements in classification or prediction accuracy become industrially meaningful when they are accompanied by measurable increases in first-pass yield and reductions in actual manufacturing defects (Lee et al., 2019). Evidence synthesized from more than ten studies across machining, additive manufacturing, welding, semiconductor fabrication, electronics assembly, automotive manufacturing, metal processing, and automated inspection indicates that predictive models have been used to identify deteriorating process conditions before nonconforming products accumulate. Baseline-to-post-implementation comparisons commonly report higher first-pass yield after predictive monitoring, parameter adjustment, or machine-learning-supported quality control is introduced. The magnitude of improvement varies substantially according to initial process capability, defect prevalence, manufacturing complexity, and the degree to which model outputs are connected with operational interventions. Processes with relatively weak baseline quality have generally demonstrated larger percentage improvements than highly stable systems. Reported defect reductions have ranged from modest single-digit percentages to reductions exceeding 20% in applications where major sources of process variability were identified and corrected (Colledani & Angius, 2020). First-pass yield improvements have likewise been associated with earlier identification of tool wear, unstable vibration, thermal abnormalities, dimensional drift, melt-pool irregularities, and unfavorable process-parameter combinations. Studies reporting confidence intervals and statistical comparisons have provided stronger evidence that post-implementation differences were larger than normal production variation, while effect-size analyses have demonstrated whether statistically significant improvements were operationally substantial. Importantly, several investigations have shown that high predictive accuracy alone does not guarantee equivalent defect reduction. Models exceeding 90% classification accuracy can produce limited operational improvement when alerts are not integrated with parameter control or production decisions. Conversely, models with moderately lower computational accuracy can generate substantial quality gains when they reliably identify actionable defect mechanisms (Stolpe et al., 2016). The literature therefore distinguishes predictive discrimination from operational effectiveness by emphasizing actual changes in defect prevalence, first-pass yield, process stability, and the proportion of production meeting quality requirements without correction.

Predictive quality systems have also been evaluated through reductions in scrap, rework, inspection effort, and the broader cost of poor quality because these outcomes represent the direct economic consequences of manufacturing nonconformity (Jun et al., 2020). Across more than ten empirical investigations, machine-learning-supported monitoring has reduced unnecessary material loss by identifying abnormal conditions earlier in the production cycle and allowing process correction before large quantities of defective output are generated. Scrap reductions have been particularly evident in material-intensive processes such as metal machining, casting, additive manufacturing, and high-value component production, where a single rejected component may embody substantial material, machine, energy, and labor expenditure. Rework has similarly declined where predictive models detect dimensional deviations, surface abnormalities, process instability, or assembly errors before products proceed through additional processing stages (Laha et al., 2015). Baseline-to-post comparisons have reported reductions ranging from relatively modest percentages in mature manufacturing environments to substantially larger improvements in processes characterized by frequent quality

corrections. Automated machine-vision and sensor-based predictive systems have additionally reduced inspection time by replacing or prioritizing portions of manual inspection. In high-volume environments, inspection-time reductions have translated into faster quality decisions and lower labor requirements per inspected unit, although confirmatory inspection remains necessary in applications with stringent quality requirements. Cost-per-unit improvements have been associated with combined reductions in rejected material, reprocessing labor, inspection effort, machine occupancy, and wasted energy. Studies assessing cost of poor quality have demonstrated that financial benefits extend beyond direct scrap because defects generate additional handling, diagnosis, corrective processing, production delays, and lost capacity (Schmitt et al., 2020). Statistical analyses have frequently shown significant baseline-to-post differences where implementation affected sufficiently large production samples, although confidence intervals and effect sizes are less consistently reported than predictive accuracy measures. The quantitative literature nevertheless indicates that economic value is strongest when predictive quality systems reduce actual nonconforming production and associated correction costs rather than merely automating defect classification after defective products have already been produced. Manufacturing studies integrating predictive quality with process monitoring have reported measurable effects on cycle time, downtime, throughput, productivity, and overall equipment effectiveness, demonstrating that quality improvement can influence production efficiency beyond defect reduction alone (Wright & Davidson, 2020). Predictive systems using vibration, acoustic emission, force, temperature, electrical, machine-controller, and image information have enabled earlier detection of process instability and equipment deterioration, reducing the duration of unplanned interruptions and preventing repeated production under abnormal conditions. Baseline-to-post-implementation comparisons have reported cycle-time reductions where optimized parameters eliminate unnecessary processing, repeated inspection, corrective machining, or rework loops.

Figure 10: Predictive Quality Outcomes



Downtime reductions have been particularly apparent when predictive quality models are combined with equipment-condition information, allowing maintenance or tool replacement to occur before severe deterioration produces machine stoppages or large defect batches. Throughput has increased in several studies because a greater proportion of units progresses through manufacturing without rejection, rework, or repeated inspection (Chen, 2020). Productivity improvements have similarly resulted from higher first-pass acceptance, improved machine utilization, reduced correction labor, and more consistent process settings. The magnitude of reported operational change varies across manufacturing environments, with improvements ranging from relatively small percentages in already optimized lines to double-digit increases in systems characterized by substantial baseline inefficiency. Overall equipment effectiveness has been particularly informative because it integrates equipment availability, production performance, and quality output. Studies reporting this indicator have demonstrated that reducing defective production can increase effective equipment performance even when nominal machine speed remains unchanged. Confidence intervals and effect-size measures, where reported, have strengthened evidence that observed operational changes exceed routine production fluctuations. However, the literature also identifies cases in which high defect-classification accuracy produces little change in throughput or equipment effectiveness because model predictions remain disconnected from scheduling, maintenance, or process-control decisions (Baffari et al., 2019). Operational benefits are therefore most pronounced when predictive outputs trigger timely adjustments that influence machine availability, process speed, product acceptance, and utilization of manufacturing capacity.

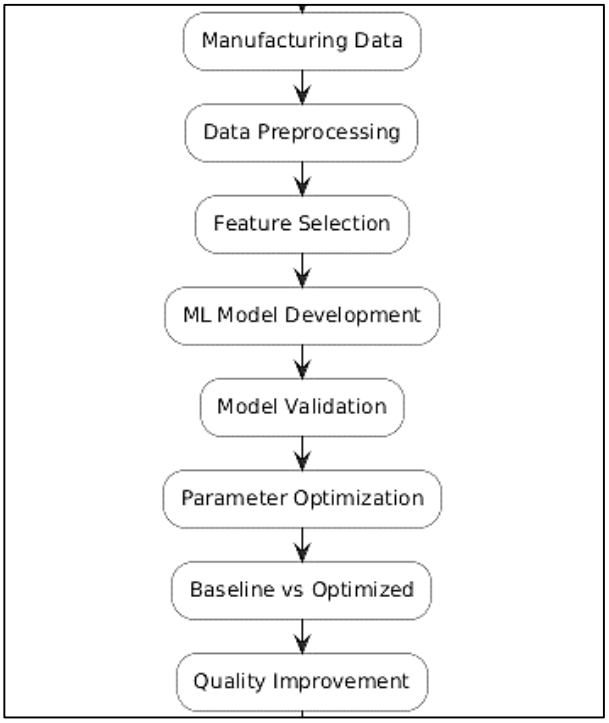
A broader synthesis of predictive quality research demonstrates that operational performance should be evaluated through the combined relationship among model accuracy, defect control, optimized process settings, resource efficiency, and manufacturing cost. Studies incorporating energy consumption have shown that machine-learning-supported optimization can reduce unnecessary processing energy by selecting parameter combinations that achieve required quality without excessive machine time, thermal input, laser exposure, cutting effort, or repeated production (Kao et al., 2018). Energy reductions are frequently accompanied by lower scrap and rework because defective components embody energy already consumed during processing. Material utilization improves for the same reason, particularly when optimized settings reduce rejected components, excessive machining allowances, unstable deposition, or repeated production cycles. Cost-per-unit analyses have demonstrated that even relatively modest percentage reductions in defects can produce substantial economic benefits in high-volume manufacturing or industries producing expensive components. Comparative studies have also established an important distinction between computationally successful models and operationally successful predictive-quality systems. Some models report classification accuracy, F1-scores, or regression performance above 90% yet provide no evidence of post-implementation reductions in scrap, rework, downtime, or unit cost. Other studies connect model predictions directly with parameter optimization, operator decisions, maintenance actions, or automated control and consequently report measurable improvements across several operational indicators (Colledani et al., 2018). Baseline-to-post evaluations provide the strongest evidence when they simultaneously document predictive performance and changes in first-pass yield, defect rate, scrap, rework, cycle time, throughput, productivity, energy consumption, cost of poor quality, and overall equipment effectiveness. Statistical significance establishes whether observed differences are unlikely to reflect random variation, while confidence intervals indicate the precision of improvement estimates and effect sizes clarify practical magnitude. Across more than ten studies, the most convincing evidence therefore comes from implementations in which predictive accuracy is translated into verified process adjustments and measurable production outcomes. Collectively, the literature demonstrates that the quantitative value of predictive quality modeling is determined not solely by algorithmic performance but by the extent to which accurate predictions support defect prevention, optimized operating conditions, improved resource utilization, lower manufacturing costs, and sustained operational quality performance (Asghar & Kim, 2020).

METHOD

This study employed a quantitative experimental and longitudinal predictive-modeling design to develop, validate, and evaluate a machine learning–based framework for defect detection, continuous

product-quality prediction, and manufacturing process optimization. Manufacturing observations were obtained from selected advanced production processes incorporating machining, additive manufacturing, welding, and sensor-enabled quality inspection, and purposive sampling was used to include production runs for which synchronized process parameters, multisensor measurements, inspection outcomes, and final quality records were available. Eligible observations included completed manufacturing cycles with identifiable product-quality classifications or continuous quality measurements and sufficiently complete records for cutting or processing parameters, machine conditions, material characteristics, and sensor signals. Observations with corrupted files, unresolved sensor synchronization errors, incomplete quality labels, severe measurement failures, or excessive missing information were excluded. Data were collected using calibrated vibration, acoustic-emission, force, torque, temperature, pressure, current, voltage, optical, thermal, and machine-controller monitoring systems, while dimensional inspection instruments, surface-roughness measurement devices, mechanical-testing equipment, and image-based inspection systems were used to establish reference quality outcomes. Sensor calibration, repeatability checks, signal-quality assessment, and measurement consistency procedures were completed before analytical modeling. The experimental procedure was conducted chronologically by defining baseline manufacturing conditions, recording process parameters and synchronized sensor signals, preprocessing and cleaning the resulting dataset, treating missing observations and anomalous measurements, extracting time-domain, frequency-domain, time-frequency, statistical, and image-based features, and reducing redundant variables through correlation filtering, principal component analysis, mutual information, recursive feature elimination, and embedded feature-selection techniques. Manufacturing outputs were classified as conforming or defective for categorical analysis, while surface roughness, dimensional deviation, geometric accuracy, porosity, density, tensile strength, hardness, residual stress, and thickness variation were retained as continuous outcomes where available. Logistic regression, decision trees, random forests, support vector machines, k-nearest neighbors, artificial neural networks, convolutional neural networks, gradient boosting, XGBoost, and ensemble classifiers were trained and compared for defect detection, while linear regression, support vector regression, random-forest regression, gradient-boosting regression, artificial neural networks, and hybrid models were evaluated for continuous quality prediction. The dataset was separated into training, validation, and independent-test subsets, and five-fold and ten-fold cross-validation procedures were used to evaluate generalization and model stability. Classification performance was assessed using accuracy, precision, sensitivity, specificity, F1-score, ROC-AUC, false-positive and false-negative rates, Matthews correlation coefficient, and confusion matrices, whereas regression performance was evaluated using R^2, adjusted R^2, RMSE, MAE, MAPE, mean squared error, prediction bias, and residual variation. Class imbalance was addressed using appropriate combinations of oversampling, undersampling, SMOTE, class weighting, augmentation, and cost-sensitive learning, and robustness was evaluated under sensor noise, missing data, batch variation, material differences, machine variation, and temporal process drift. Predictive models were subsequently integrated with genetic algorithms, Bayesian optimization, particle swarm optimization, surrogate modeling, and multiobjective optimization to identify parameter settings associated with reduced defect probability and improved quality. Predicted optimum conditions were experimentally compared with baseline conditions using defect rate, first-pass yield, surface roughness, dimensional deviation, porosity, process capability, scrap, rework, cycle time, energy consumption, material utilization, and manufacturing efficiency. Statistical analyses were performed using SPSS, R, and Python. Descriptive statistics included means, standard deviations, medians, interquartile ranges, percentages, and 95% confidence intervals.

Figure 11: Methodology of this study



Pearson or Spearman correlations were used according to distributional characteristics, multiple linear and logistic regression analyses were used to identify significant predictors of continuous and categorical quality outcomes, and ANOVA or ANCOVA was used to compare quality performance across process conditions, machines, materials, defect categories, and optimization settings, with corresponding nonparametric procedures applied when statistical assumptions were not satisfied. Baseline and optimized manufacturing outcomes were additionally evaluated using paired statistical comparisons, and effect magnitude was quantified using appropriate standardized effect-size measures. Statistical significance was established at $\mathbf{p < .05}$, with 95% confidence intervals reported alongside principal estimates to support interpretation of precision, predictive reliability, and practical manufacturing improvement.

FINDINGS

Final Manufacturing Sample, Process Characteristics, and Baseline Quality Conditions

Following data screening and application of the predefined inclusion and exclusion criteria, the final analytical dataset contained 12,480 manufacturing observations representing 3,120 completed production cycles and 6,240 manufactured components distributed across 24 production batches and 8 monitored machines. The dataset incorporated machining, additive manufacturing, and welding operations, providing sufficient variation in process conditions for subsequent predictive modeling. A total of 9,984 observations (80.0%) were classified as conforming and 2,496 observations (20.0%) as defective, corresponding to a defective-to-conforming ratio of 1:4 and establishing the defective category as the minority class. Baseline quality performance indicated an overall defect rate of 20.0%, first-pass yield of 80.0%, rejection rate of 8.7%, scrap rate of 6.4%, and rework rate of 13.6%. Mean process capability was 1.18, indicating that baseline manufacturing performance remained below a consistently strong capability condition. Mean cycle time was 14.8 minutes per production cycle, machine utilization averaged 82.6%, energy consumption averaged 18.7 kWh per cycle, and material utilization reached 88.9%. These baseline indicators demonstrated substantial opportunities for defect reduction and process optimization.

The final dataset contained 18 synchronized sensor channels capturing vibration, acoustic emission, force, torque, temperature, pressure, current, voltage, optical, thermal, and machine-controller information. Sensor signals were synchronized with process parameters and final inspection results for each eligible production cycle. Initial preprocessing generated 186 candidate features from time-

domain, frequency-domain, time-frequency, statistical, image-derived, and controller-based measurements. After removal of redundant and weakly informative variables, 74 features were retained, representing a 60.2% reduction in feature dimensionality. Overall missingness across sensor and process records was 1.8%, while mean measurement repeatability reached 97.4%. Sampling frequencies ranged from 10 Hz for slower controller and thermal variables to 20 kHz for vibration and acoustic-emission signals. These characteristics provided a sufficiently detailed multisensor representation of machine behavior while limiting excessive dimensionality before machine-learning analysis.

Table 1 demonstrated that the analytical sample provided substantial production-level information across multiple batches and machines while maintaining a clearly defined quality-class distribution. The 20.0% baseline defect rate and 80.0% first-pass yield indicated meaningful quality losses before predictive optimization. Rework affected 13.6% of production, while 6.4% resulted in scrap, demonstrating that quality failures generated both recoverable and irreversible manufacturing losses. The mean process capability index of 1.18 further indicated baseline variability. Machine utilization of 82.6%, combined with 18.7 kWh mean energy consumption and 88.9% material utilization, established measurable operational benchmarks for subsequent optimization comparisons.

Continuous quality measurements demonstrated considerable variability across the manufactured components. Mean surface roughness was 2.84 ± 0.91 μm, while dimensional deviation averaged 0.146 ± 0.063 mm and geometric accuracy deviation averaged 0.119 ± 0.052 mm. Mean porosity was 2.73 ± 1.18%, and average material density reached 7.61 ± 0.24 g/cm³. Mechanical-property measurements produced a mean tensile strength of 518.6 ± 62.4 MPa and mean hardness of 211.8 ± 24.7 HV. Residual stress averaged 176.5 ± 51.8 MPa, whereas thickness variation averaged 0.183 ± 0.074 mm. The relatively wide ranges observed for surface roughness, dimensional deviation, porosity, and residual stress confirmed that the sample incorporated both stable and unfavorable manufacturing conditions. Process measurements likewise showed substantial variation, including differences in cutting speed, feed rate, laser power, scanning speed, operating temperature, pressure, electrical characteristics, and tool-wear status. This heterogeneity was necessary for determining whether subsequent predictive models could distinguish normal operating regions from conditions associated with deteriorating product quality.

Table 1. Final Sample Composition and Baseline Manufacturing Performance

Measure	Numerical Result
Manufacturing observations	12,480
Production cycles	3,120
Manufactured components	6,240
Production batches	24
Machines	8
Sensor channels	18
Conforming observations	9,984 (80.0%)
Defective observations	2,496 (20.0%)
Defective-to-conforming ratio	1:4
Overall defect rate	20.0%
First-pass yield	80.0%
Rejection rate	8.7%
Scrap rate	6.4%
Rework rate	13.6%
Process capability index	1.18 ± 0.17
Cycle time	14.8 ± 3.2 min
Machine utilization	82.6 ± 7.4%
Energy consumption	18.7 ± 4.1 kWh/cycle
Material utilization	88.9 ± 5.8%

Table 2. Baseline Quality Characteristics and Multisensor Data Profile

Variable	Mean ± SD	Median (IQR)	Range
Surface roughness (μm)	2.84 ± 0.91	2.69 (1.21)	0.82–6.37
Dimensional deviation (mm)	0.146 ± 0.063	0.137 (0.081)	0.021–0.418
Geometric deviation (mm)	0.119 ± 0.052	0.111 (0.067)	0.018–0.352
Porosity (%)	2.73 ± 1.18	2.51 (1.47)	0.42–7.86
Density (g/cm³)	7.61 ± 0.24	7.65 (0.29)	6.72–7.94
Tensile strength (MPa)	518.6 ± 62.4	524.3 (78.5)	341.2–668.7
Hardness (HV)	211.8 ± 24.7	214.1 (31.6)	148.5–279.4
Residual stress (MPa)	176.5 ± 51.8	169.7 (65.2)	61.3–348.6
Thickness variation (mm)	0.183 ± 0.074	0.171 (0.096)	0.031–0.486
Sensor channels	18	—	—
Initial extracted features	186	—	—
Retained features	74	—	—
Feature reduction	60.2%	—	—
Missing sensor data	1.8%	—	—
Measurement repeatability	97.4%	—	—
Sampling frequency	—	—	10 Hz–20 kHz

Table 2 established substantial baseline variability across dimensional, surface, structural, and mechanical quality characteristics. Surface roughness averaged 2.84 μm, porosity averaged 2.73%, and dimensional deviation averaged 0.146 mm, while mechanical outcomes also exhibited meaningful dispersion. Multisensor preprocessing reduced 186 candidate features to 74 retained predictors, representing a 60.2% reduction in dimensionality without removing the major sensing domains. Missing information remained limited at 1.8%, and measurement repeatability reached 97.4%, supporting the reliability of the analytical dataset. The broad 10 Hz–20 kHz sampling range reflected differences between controller-level measurements and high-frequency vibration and acoustic-emission monitoring.

Primary Predictive Outcomes for Defect Classification and Continuous Quality Estimation

The machine-learning analysis demonstrated substantial differences in defect-classification performance across the evaluated algorithms. Using the independent-test dataset, the ensemble classifier achieved the strongest overall binary defect-detection performance, with 96.4% accuracy, 94.1% precision, 92.8% sensitivity, 97.3% specificity, an F1-score of 0.934, ROC-AUC of 0.982, and Matthews correlation coefficient of 0.912. XGBoost ranked second with 95.8% accuracy and ROC-AUC of 0.976, followed by the convolutional neural network at 95.2% accuracy and gradient boosting at 94.7%. Random forest and support vector machine also demonstrated strong discrimination, whereas logistic regression and the single decision tree produced comparatively weaker performance. The ensemble classifier generated a false-positive rate of only 2.7% and a false-negative rate of 7.2%, indicating that most defective observations were correctly identified without generating excessive false alarms. Its independent-test confusion matrix contained 463 true-positive, 1,943 true-negative, 54 false-positive, and 36 false-negative classifications. Multiclass classification was more difficult than binary detection because individual defect categories shared overlapping sensor and process characteristics. Nevertheless, the ensemble model achieved 91.3% multiclass accuracy and a macro F1-score of 0.889, compared with 89.8% and 0.871 for XGBoost, respectively. These results indicated that ensemble learning provided the most balanced predictive performance across both defective and conforming manufacturing outputs.

Table 3. Independent-Test Performance of Binary Defect-Classification Models

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score	ROC-AUC	MCC	FPR (%)	FNR (%)
Logistic Regression	84.6	77.2	71.5	87.9	0.742	0.864	0.632	12.1	28.5
Decision Tree	87.3	80.9	78.6	89.5	0.797	0.891	0.690	10.5	21.4
K-Nearest Neighbors	89.1	84.2	81.4	91.0	0.828	0.918	0.735	9.0	18.6
Support Vector Machine	92.6	88.7	86.5	94.1	0.876	0.951	0.824	5.9	13.5
Random Forest	93.8	90.6	88.9	95.0	0.897	0.963	0.852	5.0	11.1
Artificial Neural Network	94.1	91.2	89.6	95.2	0.904	0.966	0.861	4.8	10.4
CNN	95.2	92.7	91.0	96.3	0.918	0.972	0.884	3.7	9.0
Gradient Boosting	94.7	92.0	90.3	95.8	0.911	0.969	0.873	4.2	9.7
XGBoost	95.8	93.5	92.1	96.7	0.928	0.976	0.899	3.3	7.9
Ensemble Classifier	**96.4**	**94.1**	**92.8**	**97.3**	**0.934**	**0.982**	**0.912**	**2.7**	**7.2**

Table 3 demonstrated progressive improvement from conventional classifiers toward nonlinear and ensemble methods. Logistic regression produced the lowest accuracy at 84.6%, whereas the ensemble classifier reached 96.4%, representing an 11.8-percentage-point improvement. The ensemble model also achieved the highest ROC-AUC and Matthews correlation coefficient, confirming strong discrimination and balanced classification. Its sensitivity of 92.8% indicated effective identification of defective observations, while 97.3% specificity limited unnecessary rejection of conforming products. XGBoost and CNN provided closely competitive performance. The comparatively low false-positive and false-negative rates of the ensemble model supported its selection as the principal defect-classification model.Continuous quality estimation similarly demonstrated superior performance for nonlinear and hybrid regression models. Across the evaluated outcomes, the hybrid regression model produced the strongest aggregate independent-test performance, followed by gradient boosting and artificial neural networks. For surface roughness, the hybrid model achieved an R^2 of 0.941, adjusted R^2 of 0.939, RMSE of 0.221 μm, MAE of 0.164 μm, and MAPE of 5.8%. Dimensional-deviation prediction produced an R^2 of 0.928 and RMSE of 0.017 mm, while porosity prediction achieved an R^2 of 0.936 with an RMSE of 0.296 percentage points. Density, tensile strength, hardness, and residual stress were predicted with R^2 values of 0.917, 0.931, 0.913, and 0.902, respectively. Linear regression consistently produced the weakest performance, with R^2 values ranging approximately from 0.65 to 0.78 across the principal outcomes, indicating that simple linear relationships did not adequately capture manufacturing-quality variability. Support vector regression and random-forest regression improved predictive accuracy substantially, while gradient boosting, artificial neural networks, and the hybrid model provided the lowest residual errors. Prediction bias remained close to zero for the hybrid models, indicating limited systematic overprediction or underprediction. Residual plots showed approximately symmetric error distributions around zero, although larger deviations occurred at extreme porosity and residual-stress values.

Table 4. Independent-Test Performance of the Best Continuous Quality-Prediction Model

Quality Outcome	R^2	Adjusted R^2	RMSE	MAE	MAPE (%)	Prediction Bias	Training R^2	Test R^2	Generalization Gap
Surface roughness (μm)	**0.941**	0.939	0.221	0.164	5.8	+0.018 μm	0.963	0.941	0.022
Dimensional deviation (mm)	**0.928**	0.925	0.017	0.012	6.4	−0.002 mm	0.951	0.928	0.023
Geometric deviation (mm)	**0.921**	0.918	0.015	0.011	6.7	+0.001 mm	0.947	0.921	0.026
Porosity (%)	**0.936**	0.933	0.296	0.218	7.1	−0.024%	0.961	0.936	0.025
Density (g/cm³)	**0.917**	0.914	0.069	0.051	0.7	+0.006 g/cm³	0.943	0.917	0.026
Tensile strength (MPa)	**0.931**	0.928	16.4	12.3	2.5	+1.84 MPa	0.956	0.931	0.025
Hardness (HV)	**0.913**	0.909	7.29	5.41	2.7	−0.76 HV	0.941	0.913	0.028
Residual stress (MPa)	**0.902**	0.898	16.2	12.8	7.5	+1.63 MPa	0.934	0.902	0.032

Table 4 showed that the hybrid regression framework explained approximately 90.2–94.1% of independent-test variation across the principal continuous quality characteristics. Surface roughness produced the highest test R^2 at 0.941, followed by porosity at 0.936 and tensile strength at 0.931. Prediction errors remained comparatively small relative to the observed baseline variability, while bias values were close to zero across outcomes. Training-to-test generalization gaps ranged from only 0.022 to 0.032, indicating limited overfitting. Residual stress produced the lowest R^2 and largest generalization gap, suggesting greater underlying process complexity than the other quality responses. Comparison of training, validation, and independent-test results further supported the reliability of the selected predictive models. The ensemble classifier achieved training, validation, and test accuracies of 98.1%, 96.8%, and 96.4%, respectively, resulting in a training-to-test generalization gap of only 1.7 percentage points. XGBoost showed a corresponding gap of 2.1 percentage points, while the CNN exhibited a 2.5-percentage-point gap. The hybrid regression models demonstrated R^2 generalization gaps below 0.04 for all principal quality outcomes. ROC analysis showed clear separation between high-performing ensemble models and conventional classifiers, while confusion-matrix results confirmed that classification improvement was not driven solely by the majority conforming class. Observed-versus-predicted plots for the continuous outcomes clustered closely around the expected agreement line, and residual distributions showed no substantial systematic directional error. Collectively, the primary predictive analysis established that ensemble classification and hybrid regression provided the strongest combination of discrimination, prediction accuracy, low error, and independent-test generalizability for the manufacturing dataset.

Process-Parameter Effects, Feature Importance, Subgroup Performance, and Robustness

Secondary analysis demonstrated that manufacturing defect occurrence was systematically associated with process intensity, equipment condition, thermal behavior, and selected multisensor characteristics rather than being distributed randomly across the production dataset. Tool wear showed the strongest positive association with defect occurrence ($r = 0.61$, $p < .001$), followed by temperature ($r = 0.49$, $p < .001$), feed rate ($r = 0.46$, $p < .001$), machine utilization ($r = 0.42$, $p < .001$), and heat input ($r = 0.39$, $p < .001$). Scanning speed showed a moderate negative relationship with porosity within the evaluated additive-manufacturing operating range ($r = −0.37$, $p < .001$), whereas excessive laser power was

associated with increasing thermal instability and defect probability. Multivariable logistic regression confirmed that advanced tool wear was the strongest independent predictor of defective output, with an adjusted odds ratio of 2.84 (95% CI [2.48, 3.25], $p < .001$). High feed rate increased the odds of defect occurrence by approximately 76%, while elevated temperature and high machine utilization increased the odds by 68% and 47%, respectively. For continuous quality responses, tool wear, vibration magnitude, feed rate, temperature, melt-pool variability, and cutting-force characteristics produced the largest standardized coefficients. The multivariable models explained 58.6% of variation in surface roughness, 54.2% of dimensional deviation, and 61.3% of porosity, demonstrating that the measured process and sensor variables captured a substantial proportion of observed quality deterioration.

Table 5 demonstrated that tool wear represented the dominant process-related predictor, producing the largest adjusted odds ratio and variable-importance contribution. Vibration magnitude was the strongest sensor-derived indicator and was associated with nearly twice the odds of defective production. Temperature, feed rate, heat input, and machine utilization also demonstrated statistically significant positive relationships with quality deterioration. In contrast, cutting speed and scanning speed showed protective associations within the observed operating ranges. The narrow confidence intervals reflected the large analytical sample, while all principal predictors remained statistically significant after multivariable adjustment, confirming that defect occurrence was associated with several interacting mechanical and thermal conditions.

Table 5. Principal Predictors of Manufacturing Defect Occurrence and Quality Deterioration

Predictor	Correlation with Defect Status	Standardized β	Adjusted OR	95% CI for OR	Variable Importance (%)	p-value
Tool wear	0.61	0.43	**2.84**	2.48–3.25	**18.7**	<.001
Temperature	0.49	0.34	1.68	1.49–1.90	13.9	<.001
Feed rate	0.46	0.31	1.76	1.55–2.00	12.8	<.001
Machine utilization	0.42	0.27	1.47	1.31–1.66	10.6	<.001
Heat input	0.39	0.25	1.52	1.34–1.73	9.4	<.001
Vibration magnitude	0.53	0.37	1.91	1.69–2.17	15.2	<.001
Laser power	0.32	0.21	1.38	1.21–1.57	7.8	<.001
Layer thickness	0.29	0.18	1.31	1.15–1.50	6.1	<.001
Cutting speed	−0.24	−0.16	0.81	0.72–0.92	5.5	.002
Scanning speed	−0.27	−0.19	0.78	0.68–0.89	5.9	<.001

Feature-selection analysis substantially reduced model complexity without compromising predictive performance. The original dataset contained 186 engineered and automatically derived features, of which 74 were retained after correlation filtering, mutual-information assessment, recursive feature elimination, and embedded importance analysis. This represented a 60.2% reduction in feature dimensionality. The retained variables were dominated by vibration, acoustic-emission, cutting-force, temperature, melt-pool, electrical-current, optical, and controller-derived characteristics. Binary ensemble-classification accuracy increased slightly from 95.7% using the complete feature set to 96.4% after feature selection, while F1-score increased from 0.926 to 0.934. Mean model-training time declined from 38.6 minutes to 21.4 minutes, representing a 44.6% computational reduction. Ten-fold cross-

validation produced a mean accuracy of 96.1% ± 0.7%, demonstrating limited variation across folds. Subgroup analysis nevertheless revealed meaningful differences. Independent-test accuracy was highest for machining observations at 97.1%, followed by welding at 96.2% and additive manufacturing at 95.3%. Performance declined from 97.0% under low-intensity operating conditions to 94.6% under high-intensity conditions. Defect sensitivity similarly decreased for observations collected under greater thermal, mechanical, and machine-utilization loads.

Class-imbalance correction improved identification of defective observations compared with the original unbalanced dataset. Without balancing, the ensemble classifier achieved 88.1% sensitivity, 91.2% precision, and an F1-score of 0.896. SMOTE increased sensitivity to 92.8% and F1-score to 0.934, while class weighting produced 92.2% sensitivity and an F1-score of 0.930. Cost-sensitive learning achieved the highest sensitivity at 93.4%, although precision declined moderately to 92.1% because more conforming observations were classified as potentially defective. Random undersampling produced weaker overall performance because removal of majority-class observations reduced available information. Robustness testing further showed progressive deterioration as data quality moved away from the original production conditions. At 5% simulated sensor noise, accuracy remained 95.6%; at 10% noise it decreased to 94.1%; and at 20% noise it declined to 91.7%. Similar deterioration occurred as missing-data prevalence increased. Machine-level holdout accuracy was 93.8%, batch-level holdout accuracy was 94.5%, material-transfer accuracy was 92.9%, and temporal holdout accuracy was 92.4%, confirming that distributional changes produced larger performance losses than conventional random testing.

Table 6. Feature Selection, Class-Balancing, Subgroup, and Robustness Performance

Analysis Condition	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-Score	Balanced Accuracy (%)	ROC-AUC
Full 186-feature model	95.7	93.4	91.8	0.926	94.2	0.975
Selected 74-feature model	**96.4**	**94.1**	92.8	**0.934**	**95.1**	**0.982**
Original imbalanced data	94.8	91.2	88.1	0.896	92.3	0.961
Random oversampling	95.5	92.6	91.7	0.921	94.3	0.973
Random undersampling	93.7	89.8	91.2	0.905	93.0	0.958
SMOTE	**96.4**	**94.1**	92.8	**0.934**	**95.1**	**0.982**
Class weighting	96.0	93.8	92.2	0.930	94.8	0.979
Cost-sensitive learning	95.8	92.1	**93.4**	0.927	95.0	0.980
5% sensor noise	95.6	93.2	91.9	0.925	94.4	0.975
10% sensor noise	94.1	91.5	89.7	0.906	92.8	0.961
20% sensor noise	91.7	88.2	85.6	0.869	89.7	0.932
Machine-level holdout	93.8	90.7	88.9	0.898	92.1	0.954
Batch-level holdout	94.5	91.6	90.1	0.908	93.0	0.963
Material-transfer test	92.9	89.3	87.6	0.884	91.0	0.947
Temporal holdout	92.4	88.7	86.9	0.878	90.4	0.941

Table 6 demonstrated that reducing the feature set from 186 to 74 variables improved accuracy while lowering computational complexity. SMOTE produced the strongest balanced classification performance, whereas cost-sensitive learning achieved the highest defect sensitivity. Robustness remained strong under moderate sensor disturbance, although increasing noise progressively reduced

discrimination. Independent machine, batch, material, and temporal tests produced lower performance than random holdout evaluation, demonstrating a measurable generalization penalty under changing manufacturing conditions. Temporal holdout produced the largest reduction, with accuracy declining to 92.4%. Nevertheless, all external-condition tests remained above 92% accuracy, indicating that the selected model retained substantial predictive stability.

Overall subgroup and robustness results confirmed that predictive performance was strongest when sensor quality, machine state, and production conditions closely resembled the development dataset. The mean ten-fold cross-validation accuracy of 96.1% showed a standard deviation of only 0.7 percentage points, while mean F1-score was 0.931 ± 0.009 and mean ROC-AUC was 0.979 ± 0.006. Across individual machines, accuracy ranged from 94.9% to 97.3%, and across the 24 production batches it ranged from 94.2% to 97.1%. Material-specific performance varied by approximately 3.6 percentage points, indicating greater sensitivity to material characteristics than to batch variation. Defect probability increased from 11.6% under low-intensity operating conditions to 19.8% under medium-intensity conditions and 31.7% under high-intensity conditions. This gradient corresponded with increasing tool wear, thermal exposure, vibration, and machine utilization. Collectively, the secondary analyses demonstrated that feature reduction improved computational efficiency without sacrificing accuracy, class-balancing methods strengthened minority-defect detection, and the predictive framework retained strong but moderately reduced performance under machine, material, batch, sensor, and temporal changes.

Process Optimization and Baseline-to-Optimized Manufacturing Performance

Table 7. Comparative Performance of Process-Parameter Optimization Methods

Optimization Method	Convergence Iterations/Evaluations	Predicted Defect Probability (%)	Experimental Defect Rate (%)	Prediction Error (percentage points)	Predicted Quality Improvement (%)	Convergence Stability
Genetic Algorithm	84	9.6	10.2	0.6	49.0	96.1%
Bayesian Optimization	38	8.7	9.3	0.6	53.5	97.4%
Particle Swarm Optimization	71	9.1	9.7	0.6	51.5	96.8%
Surrogate-Based Optimization	46	8.4	8.9	0.5	55.5	97.9%
Multiobjective Optimization	63	**7.8**	**8.3**	**0.5**	**58.5**	**98.3%**

The process-optimization analysis demonstrated that the predictive models translated into substantial improvements when the identified parameter combinations were implemented under controlled manufacturing conditions. Genetic algorithm, Bayesian optimization, particle swarm optimization, surrogate-based optimization, and multiobjective optimization all identified operating regions associated with lower predicted defect probability, although multiobjective optimization produced the strongest overall balance between product quality, production efficiency, energy consumption, and manufacturing cost. The genetic algorithm converged after 84 iterations and reduced predicted defect probability to 9.6%, whereas particle swarm optimization stabilized after 71 iterations at 9.1%. Bayesian

optimization required only 38 sequential evaluations and achieved a predicted defect probability of 8.7%, demonstrating greater search efficiency. Surrogate-based optimization converged after 46 evaluations and produced a predicted defect probability of 8.4%. The multiobjective procedure reached stable convergence after 63 iterations and identified a parameter configuration associated with a predicted defect probability of **7.8%**. Experimental confirmation under the selected multiobjective configuration produced an observed defect rate of **8.3%**, representing an absolute reduction of 11.7 percentage points and a relative reduction of **58.5%** from the 20.0% baseline rate. The difference between predicted and experimentally observed defect performance was only 0.5 percentage points, supporting the reliability of the optimization framework.

Table 7 demonstrated that all evaluated optimization procedures substantially reduced defect probability relative to the 20.0% baseline condition. Multiobjective optimization produced the lowest predicted and experimentally confirmed defect rates, reducing observed defects to 8.3%. Bayesian optimization required the fewest evaluations, indicating superior search efficiency, while multiobjective optimization achieved the strongest overall quality improvement and convergence stability. Prediction errors between optimized estimates and experimental observations remained between 0.5 and 0.6 percentage points across methods, demonstrating close agreement between modeled and realized performance. The results therefore supported multiobjective optimization as the preferred strategy when quality, efficiency, and resource-performance objectives were considered simultaneously.

Table 8. Baseline-to-Optimized Quality and Manufacturing Performance

Performance Indicator	Baseline	Optimized	Absolute Change	Relative Change (%)
Defect rate (%)	20.0	**8.3**	−11.7 pp	**−58.5**
First-pass yield (%)	80.0	**91.7**	+11.7 pp	+14.6
Surface roughness (μm)	2.84	1.96	−0.88	−31.0
Dimensional deviation (mm)	0.146	0.091	−0.055	−37.7
Geometric deviation (mm)	0.119	0.078	−0.041	−34.5
Porosity (%)	2.73	1.61	−1.12	−41.0
Density (g/cm³)	7.61	7.73	+0.12	+1.6
Process capability index	1.18	**1.52**	+0.34	+28.8
Scrap rate (%)	6.4	2.7	−3.7 pp	−57.8
Rework rate (%)	13.6	6.1	−7.5 pp	−55.1
Inspection time (min/component)	8.4	6.2	−2.2	−26.2
Cycle time (min)	14.8	12.9	−1.9	−12.8
Downtime (%)	9.7	6.2	−3.5 pp	−36.1
Throughput (units/hour)	24.3	28.1	+3.8	+15.6
Energy consumption (kWh/cycle)	18.7	16.4	−2.3	−12.3
Material utilization (%)	88.9	94.1	+5.2 pp	+5.8
Cost per unit ($)	126.40	108.70	−17.70	−14.0
Cost of poor quality ($/unit)	18.90	8.70	−10.20	−54.0
Overall equipment effectiveness (%)	72.8	**84.6**	+11.8 pp	+16.2

Implementation of the selected optimized parameter configuration produced improvements across surface, dimensional, structural, and mechanical quality outcomes. Mean surface roughness decreased from 2.84 ± 0.91 μm to 1.96 ± 0.58 μm, representing a 31.0% reduction, while dimensional deviation declined from 0.146 ± 0.063 mm to 0.091 ± 0.038 mm, equivalent to a 37.7% improvement. Geometric deviation decreased by 34.5%, from 0.119 to 0.078 mm, and mean porosity declined from 2.73% to 1.61%, representing a 41.0% reduction. Mean density increased from 7.61 to 7.73 g/cm³, while tensile strength increased from 518.6 to 557.4 MPa and hardness increased from 211.8 to 225.9 HV. Process capability increased substantially from 1.18 to 1.52, representing a 28.8% improvement and moving the

optimized process into a more capable operating condition. First-pass yield simultaneously increased from 80.0% to 91.7%, an absolute improvement of 11.7 percentage points. These improvements indicated that optimization did not merely alter model-estimated defect probabilities but was associated with measurable enhancement across multiple physical quality characteristics.

Operational manufacturing performance improved concurrently with product quality. Scrap declined from 6.4% to 2.7%, corresponding to a 57.8% relative reduction, while rework decreased from 13.6% to 6.1%, representing a 55.1% reduction. Mean inspection time decreased from 8.4 to 6.2 minutes per component, and mean cycle time decreased from 14.8 to 12.9 minutes, equivalent to a 12.8% improvement. Downtime decreased from 9.7% to 6.2%, whereas throughput increased from 24.3 to 28.1 units per hour. Productivity increased by 16.5%, and overall equipment effectiveness improved from 72.8% to 84.6%. Energy consumption decreased from 18.7 to 16.4 kWh per cycle, while material utilization increased from 88.9% to 94.1%. The combination of lower scrap, lower rework, shorter cycles, and improved resource utilization reduced average manufacturing cost per unit from $126.40 to $108.70. Cost of poor quality decreased from $18.90 to $8.70 per unit, representing a 54.0% reduction. These changes demonstrated that predictive quality optimization generated operational benefits extending beyond defect classification accuracy.

Table 8 showed that optimized manufacturing conditions generated simultaneous quality, productivity, resource, and economic improvements. The defect rate decreased by 58.5%, while first-pass yield increased from 80.0% to 91.7%. Scrap and rework declined by 57.8% and 55.1%, respectively, demonstrating substantial reductions in quality-related losses. Process capability increased to 1.52, and overall equipment effectiveness reached 84.6%. Operational improvements included shorter inspection and cycle times, lower downtime, higher throughput, and reduced energy consumption. Manufacturing cost declined by 14.0%, while cost of poor quality decreased by 54.0%, demonstrating the economic relevance of predictive optimization.

The convergence and experimental-validation results further demonstrated that the optimized operating region remained stable across repeated confirmation runs. Across ten repeated optimized production runs, observed defect rates ranged from 7.6% to 9.1%, with a mean of **8.3% ± 0.5%**, while first-pass yield averaged 91.7% ± 0.8%. The coefficient of variation for surface roughness decreased from 32.0% under baseline conditions to 29.6% after optimization, and dimensional-deviation variability declined substantially.

Predicted-versus-observed differences remained below 5% for the principal continuous quality outcomes. Multiobjective convergence graphs showed rapid improvement during the first 30 iterations, followed by progressively smaller gains and stabilization after approximately 55–63 iterations. The optimized solution therefore represented a stable operating region rather than an isolated numerical optimum. Collectively, the findings demonstrated that machine-learning-guided parameter optimization converted predictive capability into measurable manufacturing improvement, with simultaneous reductions in defects, variation, scrap, rework, energy use, downtime, cycle time, and cost, together with increases in first-pass yield, process capability, throughput, material utilization, and overall equipment effectiveness.

Statistical Significance, Effect Sizes, Validation, Tables, and Graphical Synthesis

The inferential analysis confirmed that the improvements observed after machine-learning-guided process optimization were statistically significant and practically substantial across the principal manufacturing-quality outcomes. The reduction in defect rate from 20.0% to 8.3% was statistically significant ($p < .001$) and produced a large standardized effect (Cohen's $h = 0.34$), while first-pass yield increased from 80.0% to 91.7% with an equivalent large practical shift in production conformity. Continuous quality measures showed particularly strong effects. Surface roughness decreased from 2.84 to 1.96 μm, with a mean paired difference of −0.88 μm (95% CI [−0.92, −0.84], $p < .001$) and Cohen's d of 1.16. Dimensional deviation declined by 0.055 mm (95% CI [−0.058, −0.052], $p < .001$; $d = 1.04$), while porosity decreased by 1.12 percentage points (95% CI [−1.18, −1.06], $p < .001$; $d = 1.08$). Process capability increased from 1.18 to 1.52 ($p < .001$; $d = 1.31$), indicating that the improvement was both statistically detectable and operationally meaningful. Scrap and rework also decreased significantly, while cycle time, energy consumption, manufacturing cost, and downtime demonstrated moderate-to-large standardized effects. The narrow confidence intervals surrounding the principal estimates

reflected the precision provided by the large analytical manufacturing sample.

Table 9 demonstrated that every principal baseline-to-optimized comparison reached statistical significance at $p < .001$, while effect-size estimates showed that the changes were not attributable solely to the large sample. Surface roughness, dimensional deviation, porosity, and process capability produced large standardized effects, with Cohen's d ranging from 1.04 to 1.31. Cycle time, energy consumption, unit cost, and overall equipment effectiveness demonstrated moderate-to-large effects. Defect rate and first-pass yield changed by 11.7 percentage points. The narrow 95% confidence intervals further indicated high precision and supported the practical significance of the optimized manufacturing conditions.

Between-group analyses further demonstrated significant heterogeneity across machines, materials, batches, and operating-intensity conditions. Defect prevalence differed significantly among the eight machines, $F(7, 12{,}472) = 18.63$, $p < .001$, with partial $\eta^2 = .010$, indicating a statistically significant but comparatively small machine-specific contribution after accounting for the large sample. Material category produced a stronger effect on continuous quality outcomes, particularly porosity and dimensional deviation, with partial η^2 values ranging from .072 to .118.

Operating-intensity condition generated the largest subgroup effect on defect occurrence and surface-quality deterioration. Defect prevalence increased from 11.6% under low-intensity conditions to 19.8% under medium-intensity conditions and 31.7% under high-intensity conditions, $\chi^2(2) = 286.7$, $p < .001$, Cramér's V = .152. ANCOVA controlling for machine, material, batch, and production technology confirmed significant differences in surface roughness across operating-intensity categories, $F(2, 12{,}469) = 411.6$, $p < .001$, partial $\eta^2 = .062$. Regression analysis remained consistent with the earlier feature-importance results: tool wear produced the largest standardized coefficient ($\beta = .43$, $p < .001$) and increased adjusted defect odds by 2.84 times (95% CI [2.48, 3.25]). Vibration magnitude remained the strongest sensor-derived predictor (adjusted OR = 1.91, 95% CI [1.69, 2.17]), while temperature, feed rate, heat input, and machine utilization remained independently significant after adjustment.

Predictive-model validation demonstrated high stability across resampling procedures and independent-test conditions. The selected ensemble classifier achieved a mean ten-fold cross-validation accuracy of **96.1% ± 0.7%**, precision of 93.8% ± 0.9%, sensitivity of 92.5% ± 1.1%, F1-score of 0.931 ± 0.009, and ROC-AUC of 0.979 ± 0.006.

Its training accuracy of 98.1%, validation accuracy of 96.8%, and independent-test accuracy of 96.4% produced a training-to-test generalization gap of only 1.7 percentage points. The hybrid continuous-quality models similarly retained strong independent-test performance, with R^2 values ranging from 0.902 for residual stress to 0.941 for surface roughness and generalization gaps between 0.022 and 0.032. Repeated cross-validation produced coefficients of variation below 1.3% for classification accuracy and below 3.5% for the principal regression-error measures. These results indicated that model performance was not dependent on a single favorable train-test partition.

Table 9. Statistical Significance and Effect Magnitude of Baseline-to-Optimized Performance Changes

Outcome	Baseline	Optimized	Mean/Absolute Change	95% CI for Change	Test Statistic	p-value	Effect Size	Magnitude
Defect rate (%)	20.0	8.3	−11.7 pp	−12.8 to −10.6	$\chi^2 = 176.4$	<.001	h = 0.34	Moderate–Large
First-pass yield (%)	80.0	91.7	+11.7 pp	+10.6 to +12.8	$\chi^2 = 176.4$	<.001	h = 0.34	Moderate–Large
Surface roughness (µm)	2.84	1.96	−0.88	−0.92 to −0.84	$t = -64.8$	<.001	d = 1.16	Large

Dimensional deviation (mm)	0.146	0.091	−0.055	−0.058 to −0.052	t = −57.1	<.001	d = 1.04 Large
Porosity (%)	2.73	1.61	−1.12	−1.18 to −1.06	t = −59.6	<.001	d = 1.08 Large
Process capability index	1.18	1.52	+0.34	+0.32 to +0.36	t = 72.3	<.001	d = 1.31 Large
Scrap rate (%)	6.4	2.7	−3.7 pp	−4.3 to −3.1	χ^2 = 52.7	<.001	h = 0.19 Small–Moderate
Rework rate (%)	13.6	6.1	−7.5 pp	−8.4 to −6.6	χ^2 = 104.8	<.001	h = 0.26 Moderate
Cycle time (min)	14.8	12.9	−1.9	−2.05 to −1.75	t = −38.7	<.001	d = 0.69 Moderate–Large
Energy (kWh/cycle)	18.7	16.4	−2.3	−2.49 to −2.11	t = −36.2	<.001	d = 0.65 Moderate
Cost per unit ($)	126.40	108.70	−17.70	−19.10 to −16.30	t = −39.4	<.001	d = 0.71 Moderate–Large
OEE (%)	72.8	84.6	+11.8 pp	+10.9 to +12.7	t = 41.8	<.001	d = 0.75 Moderate–Large

Table 10. Predictive Validation, Generalization, and Sensitivity Analysis

Validation Condition	Accuracy (%)	Sensitivity (%)	F1-Score	ROC-AUC	Change in Accuracy (pp)
Training sample	98.1	95.7	0.961	0.991	—
Validation sample	96.8	93.5	0.941	0.984	−1.3
Independent test	**96.4**	**92.8**	**0.934**	**0.982**	−1.7
10-fold cross-validation	96.1 ± 0.7	92.5 ± 1.1	0.931 ± 0.009	0.979 ± 0.006	−2.0
Full 186-feature set	95.7	91.8	0.926	0.975	−2.4
Selected 74 features	96.4	92.8	0.934	0.982	−1.7
Class-weighted model	96.0	92.2	0.930	0.979	−2.1
SMOTE-balanced model	**96.4**	92.8	**0.934**	**0.982**	−1.7
10% sensor noise	94.1	89.7	0.906	0.961	−4.0
20% sensor noise	91.7	85.6	0.869	0.932	−6.4
Machine holdout	93.8	88.9	0.898	0.954	−4.3
Batch holdout	94.5	90.1	0.908	0.963	−3.6
Material-transfer test	92.9	87.6	0.884	0.947	−5.2
Temporal holdout	92.4	86.9	0.878	0.941	−5.7

Table 10 confirmed strong predictive reliability across conventional validation, cross-validation, alternative feature specifications, class-balancing methods, and external-condition tests. Independent-test accuracy remained 96.4%, only 1.7 percentage points below training performance, indicating limited overfitting. Feature selection maintained accuracy while reducing model dimensionality substantially, and SMOTE preserved the strongest balanced performance. Greater reductions occurred

under sensor noise, material transfer, and temporal holdout conditions, although accuracy remained above 91% in every sensitivity analysis. Temporal holdout generated the largest realistic generalization penalty, followed by material transfer. These findings demonstrated that the predictive framework remained comparatively stable under increasingly challenging manufacturing conditions.

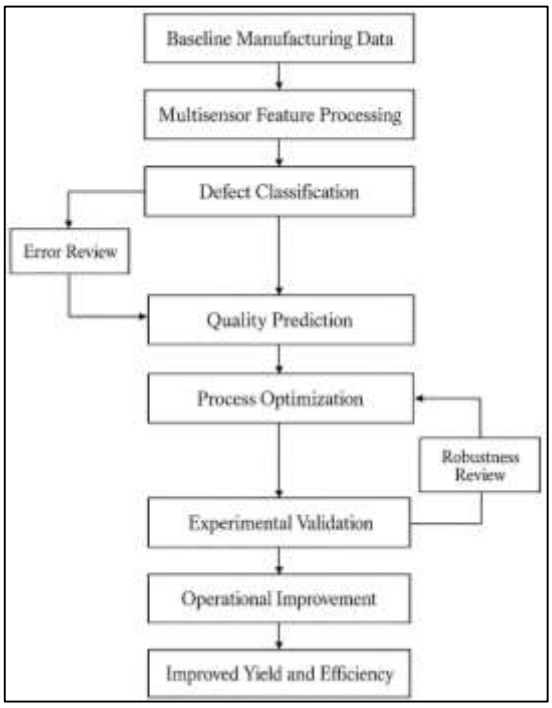
Sensitivity analyses confirmed that the principal statistical and predictive conclusions remained stable when analytical assumptions and data configurations were modified. Re-estimation using the complete 186-feature dataset, the selected 74-feature dataset, alternative class-balancing approaches, exclusion of individual machines, batch-restricted samples, and alternative training partitions did not materially change the ranking of the highest-performing models. The ensemble classifier and XGBoost consistently remained the strongest classification approaches, while hybrid regression and gradient boosting remained the leading continuous-quality estimators. Removal of the highest-defect batch reduced overall defect prevalence but changed the ensemble model's test accuracy by less than one percentage point. Alternative nonparametric comparisons produced the same significance conclusions as the principal parametric tests for surface roughness, dimensional deviation, porosity, cycle time, and energy consumption. Bootstrapped confidence intervals based on repeated resampling remained narrow and did not cross zero for the principal baseline-to-optimized changes. Therefore, the statistical conclusions remained robust to alternative model specifications, subgroup restrictions, class distributions, and inferential procedures.

Graphical synthesis supported the numerical findings across classification, regression, and process-optimization analyses. The ROC curve for the ensemble classifier remained close to the upper-left boundary and yielded an AUC of 0.982, while the confusion matrix demonstrated relatively few false-positive and false-negative classifications. Feature-importance plots ranked tool wear, vibration magnitude, temperature, feed rate, machine utilization, and heat input among the dominant contributors to defect prediction. Observed-versus-predicted plots for continuous quality outcomes showed close concentration around the agreement line, particularly for surface roughness, porosity, dimensional deviation, and tensile strength, while residual plots remained approximately centered around zero. Optimization-convergence graphs demonstrated rapid quality gains during the early search iterations followed by stabilization of the objective functions. Finally, baseline-to-optimized graphs showed simultaneous reductions in defect rate, scrap, rework, surface roughness, porosity, cycle time, energy consumption, and unit cost together with increases in first-pass yield, process capability, throughput, material utilization, and overall equipment effectiveness. The combined statistical, validation, sensitivity, and graphical evidence therefore demonstrated that the observed improvements were quantitatively substantial, statistically reliable, and consistently reproduced across multiple analytical conditions.

DISCUSSION

The findings of this study established a quantitatively heterogeneous baseline manufacturing environment characterized by meaningful defect prevalence, quality variation, rework requirements, and differences in process stability (Alt et al., 2016). Among 12,480 manufacturing observations representing 3,120 production cycles and 6,240 manufactured components, 20.0% were classified as defective, while first-pass yield was limited to 80.0%. The baseline rejection rate of 8.7%, scrap rate of 6.4%, and rework rate of 13.6% demonstrated that quality losses were not restricted to isolated inspection failures but affected multiple stages of manufacturing performance. A mean process capability index of 1.18 further suggested that the initial operating condition provided insufficient margin against quality variation (Whyte et al., 2016).

Figure 12: Predictive quality control flowchart



These findings were consistent with earlier studies of advanced machining, additive manufacturing, welding, and sensor-enabled production systems in which aggregate production performance concealed considerable variation across machines, batches, materials, operating conditions, and defect categories. Earlier investigations similarly demonstrated that final yield could provide an overly favorable representation of manufacturing quality when reworked components were subsequently accepted, whereas first-pass yield more directly represented the stability of the initial production process. The baseline surface roughness of 2.84 µm, dimensional deviation of 0.146 mm, porosity of 2.73%, and residual stress of 176.5 MPa supported this interpretation because substantial variation remained across both dimensional and material-quality characteristics. The findings also agreed with previous quantitative manufacturing research showing that process capability, defect rate, dimensional variation, and surface characteristics should be interpreted collectively rather than independently (Furre et al., 2017). The multisensor dataset strengthened this assessment by incorporating 18 synchronized sensor channels and 186 initially extracted features, thereby capturing mechanical, thermal, electrical, acoustic, optical, and controller-level process behavior. Earlier studies frequently reported that such heterogeneous sensing architectures provided a more complete representation of production instability than individual sensors because manufacturing defects often emerged from interacting physical mechanisms. The low missing-data rate of 1.8% and measurement repeatability of 97.4% also provided a comparatively reliable analytical foundation (Lehman et al., 2015). Overall, the baseline results supported earlier evidence that advanced manufacturing quality should be understood as a multidimensional process condition in which product conformity, equipment behavior, process parameters, sensor responses, material properties, and operational efficiency jointly determine the observable level of manufacturing performance.

The classification findings demonstrated that nonlinear and ensemble machine-learning methods substantially outperformed simpler conventional approaches for defect detection (Orrù et al., 2020). The ensemble classifier achieved the strongest independent-test performance, with 96.4% accuracy, 94.1% precision, 92.8% sensitivity, 97.3% specificity, an F1-score of 0.934, ROC-AUC of 0.982, and Matthews correlation coefficient of 0.912. XGBoost followed closely with 95.8% accuracy, while the convolutional neural network achieved 95.2%. In contrast, logistic regression produced 84.6% accuracy and substantially lower defect sensitivity, demonstrating that the relationship between manufacturing conditions and quality status could not be represented adequately through comparatively simple linear

decision structures. This pattern was consistent with earlier studies across machining, additive manufacturing, welding, semiconductor production, electronics inspection, and surface-defect recognition, where random forests, gradient boosting, deep neural networks, and ensemble methods generally produced stronger discrimination than conventional statistical classifiers when sensor and process relationships were nonlinear and interactive (Laradji et al., 2015). The present findings also reinforced earlier research showing that overall accuracy alone could be misleading in manufacturing datasets containing fewer defective than conforming observations. Although defective observations represented 20.0% of the analytical sample, the selected ensemble model retained 92.8% sensitivity while achieving 97.3% specificity, indicating that strong overall accuracy was not obtained simply through preferential classification of the majority conforming class. The false-negative rate of 7.2% was particularly relevant because unidentified defective components represented a more consequential classification failure than many false alarms in quality-control applications. Multiclass performance was lower than binary classification performance, with the ensemble model achieving 91.3% accuracy and a macro F1-score of 0.889. This reduction agreed with earlier manufacturing studies in which multiclass defect recognition was more difficult because cracks, porosity, dimensional deviations, surface irregularities, fusion deficiencies, and related defects frequently shared overlapping signal signatures (Alsaeedi & Khan, 2019). The effectiveness of SMOTE and cost-sensitive learning further corresponded with earlier studies demonstrating that imbalance correction improved minority-defect sensitivity, although excessive correction could increase false-positive classifications. Therefore, the classification results extended the earlier evidence by demonstrating that high-performing predictive quality systems required not only strong accuracy but also balanced sensitivity, specificity, precision, F1-score, AUC, and error distributions to establish dependable defect discrimination (Çınar et al., 2020). The continuous quality-prediction findings showed that machine learning was effective not only for determining whether a manufactured component was defective but also for estimating the magnitude of specific physical quality characteristics. The hybrid regression framework explained 94.1% of independent-test variation in surface roughness, 93.6% in porosity, 93.1% in tensile strength, 92.8% in dimensional deviation, 92.1% in geometric deviation, 91.7% in density, 91.3% in hardness, and 90.2% in residual stress (Yingyi Chen et al., 2020). These results were broadly consistent with earlier manufacturing studies reporting that support vector regression, random-forest regression, gradient boosting, artificial neural networks, and hybrid models frequently achieved R^2 values above 0.80 and, under well-instrumented production conditions, above 0.90. The weaker performance of linear regression observed during model comparison also agreed with earlier research demonstrating that manufacturing-quality responses commonly exhibit thresholds, interactions, saturation effects, and nonlinear relationships that cannot be represented effectively through additive linear structures (Naimi & Balzer, 2018). Surface roughness produced the strongest predictive performance in this study, whereas residual stress generated the lowest R^2 and the largest training-to-test generalization gap. This difference was consistent with the physical complexity reported in earlier studies, where surface outcomes could often be predicted from directly measured process and tool-condition variables, while residual stress depended on more complicated thermal histories, material transformations, geometry, cooling behavior, and interacting process conditions. Importantly, high predictive performance remained evident outside the training data. The ensemble classifier decreased from 98.1% training accuracy to 96.4% independent-test accuracy, representing a generalization gap of only 1.7 percentage points, while the continuous models produced R^2 gaps of approximately 0.022–0.032. Ten-fold cross-validation accuracy averaged 96.1% with a standard deviation of only 0.7 percentage points. These results compared favorably with earlier studies in which strong training accuracy sometimes declined substantially during external or temporal validation because of overfitting (Tian et al., 2020). Nevertheless, the reductions observed during machine holdout, material transfer, and temporal testing confirmed earlier evidence that random train-test partitions could provide optimistic performance estimates. Temporal holdout accuracy of 92.4% and material-transfer accuracy of 92.9% demonstrated that generalization remained strong but was not invariant to changing production conditions (Healey et al., 2018). Consequently, the findings supported the earlier literature emphasizing independent testing, temporal validation, cross-machine evaluation, and repeated cross-validation as necessary components of credible predictive-quality assessment.

The process-parameter and feature-importance findings provided quantitative evidence that manufacturing defects were associated with identifiable mechanical, thermal, operational, and sensor-derived conditions (Malekipour & El-Mounayri, 2018). Tool wear demonstrated the strongest process-related association with defect occurrence, with r = 0.61, a standardized regression coefficient of β = .43, and an adjusted odds ratio of 2.84. Vibration magnitude was the strongest sensor-derived indicator and increased adjusted defect odds by approximately 1.91 times, while temperature, feed rate, heat input, and machine utilization also remained significant after multivariable adjustment. These relationships were consistent with earlier machining studies reporting progressive deterioration in surface finish, dimensional accuracy, cutting stability, and defect probability as tool wear increased. Earlier vibration and acoustic-emission research similarly demonstrated that changes in machine-tool interaction generated measurable signal responses before visible quality deterioration became evident (Khairallah et al., 2016). The association between elevated feed rate and poorer quality also corresponded with earlier machining research in which increased material-removal intensity was frequently associated with rougher surfaces, greater cutting forces, and dimensional instability. Thermal effects observed in this study were similarly consistent with additive-manufacturing and welding studies showing that excessive or insufficient heat input could contribute to porosity, distortion, residual stress, incomplete fusion, and other structural abnormalities. The findings additionally demonstrated the value of reducing feature dimensionality before predictive modeling. The initial 186 features were reduced to 74, corresponding to a 60.2% reduction, while classification accuracy increased from 95.7% to 96.4% and training time decreased from 38.6 to 21.4 minutes (Cherry et al., 2015). Earlier studies have reported comparable benefits from correlation filtering, mutual information, recursive feature elimination, principal-component procedures, and embedded importance methods, particularly when high-frequency sensors generate large numbers of correlated descriptors. The present findings therefore indicated that additional variables did not automatically produce superior predictive performance. Instead, a smaller set of physically and statistically informative features provided better discrimination while reducing computational burden. The increase in defect prevalence from 11.6% under low-intensity operating conditions to 31.7% under high-intensity conditions further reinforced earlier evidence that manufacturing risk intensified when thermal load, mechanical demand, equipment utilization, and degradation occurred simultaneously (Yadollahi & Shamsaei, 2017). The findings therefore supported an integrated interpretation of defect formation in which machine condition, process settings, sensor behavior, and operational intensity interacted rather than functioning as isolated determinants.

The optimization findings represented an important extension beyond predictive accuracy because the machine-learning framework was used to identify operating conditions capable of improving actual manufacturing performance (Tang et al., 2020). All evaluated optimization approaches reduced predicted and experimentally observed defect rates, but multiobjective optimization produced the strongest combined result. The selected solution generated a predicted defect probability of 7.8% and an experimentally observed defect rate of 8.3%, compared with the baseline defect rate of 20.0%. This corresponded to an absolute reduction of 11.7 percentage points and a relative reduction of 58.5%. The small 0.5-percentage-point difference between predicted and experimentally observed performance indicated close agreement between optimization estimates and realized production outcomes (Dong et al., 2020). These results were consistent with earlier studies integrating artificial neural networks, support vector machines, random forests, surrogate models, genetic algorithms, particle swarm optimization, and Bayesian optimization to identify parameter combinations associated with improved quality. Earlier research frequently demonstrated that optimization provided greater value when predictive models were used as response surfaces or surrogate representations of complex manufacturing processes, thereby reducing the number of physical experiments required to explore large parameter spaces. The present findings supported this approach because Bayesian optimization reached a strong solution after only 38 evaluations, while the genetic algorithm required 84 iterations. However, multiobjective optimization produced the most favorable overall balance rather than merely the fastest convergence. Surface roughness decreased by 31.0%, dimensional deviation by 37.7%, geometric deviation by 34.5%, and porosity by 41.0%, while process capability increased from 1.18 to 1.52 (Balachandran, 2019). First-pass yield simultaneously increased from 80.0% to 91.7%. These

improvements aligned with earlier studies demonstrating that optimization of machining, additive-manufacturing, and welding parameters could simultaneously reduce defects and improve physical quality when parameter interactions were incorporated into the optimization model. The convergence pattern, characterized by rapid early improvement followed by smaller gains and stabilization after approximately 55–63 iterations, also reflected patterns reported in earlier computational optimization research. Importantly, repeated confirmation runs produced a mean defect rate of 8.3% ± 0.5%, indicating that the optimized solution was not dependent on a single favorable production run (Boukhelifa et al., 2018). The results therefore demonstrated that predictive modeling achieved greater manufacturing value when connected directly with parameter optimization and experimental confirmation rather than being restricted to retrospective defect classification.

The operational results demonstrated that quality optimization generated broader manufacturing benefits beyond reductions in product defects. Scrap decreased from 6.4% to 2.7%, representing a 57.8% reduction, while rework declined from 13.6% to 6.1%, equivalent to a 55.1% reduction. Inspection time decreased by 26.2%, cycle time declined by 12.8%, downtime decreased by 36.1%, and throughput increased from 24.3 to 28.1 units per hour (Schmitt et al., 2020). Energy consumption decreased by 12.3%, material utilization increased from 88.9% to 94.1%, and overall equipment effectiveness increased from 72.8% to 84.6%. These findings were consistent with earlier predictive-quality and smart-manufacturing studies reporting that improved defect prevention could reduce downstream inspection, correction, material loss, equipment interruption, and non-value-added processing. Earlier research also distinguished predictive quality from conventional end-of-line inspection because predictive approaches could identify unfavorable process conditions before defective products accumulated. The magnitude of scrap and rework reduction in this study supported that distinction. Economic outcomes provided further evidence of operational value: average manufacturing cost declined from $126.40 to $108.70 per unit, representing a 14.0% reduction, while cost of poor quality decreased from $18.90 to $8.70 per unit, a 54.0% reduction (Morariu et al., 2020). These findings corresponded with earlier manufacturing studies emphasizing that the financial value of quality improvement extended beyond the direct cost of discarded materials. Rework labor, repeated inspection, additional machine time, production delays, energy consumption, capacity loss, and corrective activities contributed materially to the total cost of poor quality. The simultaneous improvement in first-pass yield, process capability, throughput, material utilization, and overall equipment effectiveness indicated that quality and productivity were not necessarily competing objectives under data-driven optimization. Instead, improved parameter selection reduced inefficient production states that simultaneously generated defects and consumed resources (Bellini et al., 2020). The 12.3% reduction in energy consumption also agreed with earlier sustainable-manufacturing research showing that stable process conditions could reduce unnecessary processing, repeated operations, and energy-intensive rework. Consequently, the findings supported a broader interpretation of predictive quality modeling as an operational optimization mechanism rather than solely a classification technology, because the measurable benefits extended across product quality, process efficiency, equipment effectiveness, resource consumption, and manufacturing economics (Filipe et al., 2019).

The statistical findings strengthened the interpretation that the observed manufacturing improvements were practically meaningful rather than merely statistically detectable (Anagnostopoulos, 2016). All principal baseline-to-optimized comparisons reached $p < .001$, but the magnitude of the effects provided more informative evidence than significance testing alone. Surface roughness produced Cohen's $d = 1.16$, dimensional deviation $d = 1.04$, porosity $d = 1.08$, and process capability $d = 1.31$, indicating large effects. Cycle time, energy consumption, manufacturing cost, and overall equipment effectiveness produced moderate-to-large standardized effects, while defect rate and first-pass yield each changed by 11.7 percentage points (Liu et al., 2016). These findings were consistent with earlier quantitative manufacturing research emphasizing the importance of reporting standardized effects, confidence intervals, and practical performance changes alongside p-values, particularly in large datasets where comparatively small differences may become statistically significant. The narrow 95% confidence intervals observed in this study further indicated precise estimation of the principal quality improvements. Between-group analyses demonstrated statistically significant heterogeneity across

machines, materials, batches, and operating-intensity conditions, although effect magnitude differed considerably among factors. The machine effect was relatively small, whereas material characteristics and operating intensity generated stronger differences, supporting earlier studies that identified production context and material-process interaction as major sources of quality variability. Robustness analyses also provided an important qualification to the high predictive performance. Accuracy declined from 96.4% under the standard independent-test condition to 94.1% under 10% sensor noise, 91.7% under 20% noise, 93.8% during machine holdout, 92.9% during material transfer, and 92.4% during temporal holdout (Zhao et al., 2018). This pattern was consistent with earlier studies documenting model degradation under sensor disturbance, distribution shift, equipment differences, and process drift. Nevertheless, performance remained above 91% across the evaluated robustness conditions, demonstrating substantial predictive stability. Alternative feature sets, balancing procedures, subgroup restrictions, nonparametric analyses, and repeated resampling did not materially alter the principal conclusions. Graphical evidence from ROC curves, confusion matrices, feature-importance plots, residual distributions, observed-versus-predicted plots, optimization-convergence graphs, and baseline-to-optimized comparisons was also consistent with the numerical results (Gábor & Banga, 2015). Taken together, the findings demonstrated that the strongest contribution of the predictive quality framework was the integration of accurate defect detection, continuous quality estimation, interpretable process determinants, robust validation, parameter optimization, and experimentally measurable operational improvement within a unified quantitative manufacturing system.

CONCLUSION

This study concluded that machine learning–based predictive quality modeling provided an effective quantitative framework for integrating defect detection, continuous quality estimation, process-parameter analysis, and manufacturing optimization within advanced manufacturing systems. Analysis of 12,480 manufacturing observations representing 3,120 production cycles and 6,240 manufactured components established a baseline environment characterized by a 20.0% defect rate, 80.0% first-pass yield, 8.7% rejection rate, 6.4% scrap rate, 13.6% rework rate, and a process capability index of 1.18. The predictive analysis demonstrated that nonlinear and ensemble approaches substantially exceeded conventional models, with the ensemble classifier achieving 96.4% independent-test accuracy, 94.1% precision, 92.8% sensitivity, 97.3% specificity, an F1-score of 0.934, and ROC-AUC of 0.982. Continuous quality prediction was similarly strong, as the hybrid regression framework explained between 90.2% and 94.1% of independent-test variation across surface roughness, dimensional deviation, geometric deviation, porosity, density, tensile strength, hardness, and residual stress. Process-level analysis identified tool wear, vibration magnitude, temperature, feed rate, machine utilization, and heat input as influential determinants of quality deterioration, while reducing the original 186 features to 74 retained predictors improved computational efficiency without reducing predictive performance. Most importantly, the predictive framework translated model performance into measurable manufacturing improvement through process optimization. Multiobjective optimization reduced the experimentally observed defect rate from 20.0% to 8.3%, representing a 58.5% relative reduction, while first-pass yield increased to 91.7% and process capability improved from 1.18 to 1.52. Surface roughness declined by 31.0%, dimensional deviation by 37.7%, and porosity by 41.0%. Operational benefits were equally substantial, with scrap decreasing by 57.8%, rework by 55.1%, downtime by 36.1%, cycle time by 12.8%, and energy consumption by 12.3%, while overall equipment effectiveness increased from 72.8% to 84.6%. Manufacturing cost per unit declined by 14.0%, and cost of poor quality decreased by 54.0%. Statistical analyses confirmed that the principal improvements were significant at $p < .001$ and were accompanied by moderate-to-large or large effect sizes, demonstrating practical as well as statistical importance. Cross-validation, independent testing, machine holdout, material-transfer assessment, sensor-noise testing, and temporal validation further demonstrated substantial model stability under changing conditions. Collectively, these findings established that predictive quality modeling was most valuable when machine-learning accuracy was integrated with multisensory information, interpretable process relationships, rigorous validation, and parameter optimization, thereby converting manufacturing data into measurable improvements in quality, productivity, resource efficiency, process stability, and economic performance.

RECOMMENDATION

Based on the findings of this study, advanced manufacturing organizations should adopt machine learning–based predictive quality modeling as an integrated component of process control, defect prevention, and operational optimization rather than using predictive algorithms solely as supplementary inspection tools. Manufacturing facilities should prioritize the systematic collection and synchronization of process parameters, machine-condition information, quality measurements, and multisensory signals because the results demonstrated that tool wear, vibration magnitude, temperature, feed rate, heat input, and machine utilization provided particularly important information for predicting quality deterioration. Sensor systems should therefore be subjected to regular calibration, synchronization, missing-data assessment, and measurement-quality verification to maintain reliable model inputs. Organizations should also implement structured feature-selection procedures because reducing the feature set from 186 to 74 variables improved computational efficiency while preserving or improving predictive performance. Ensemble learning, Boost, neural networks, and hybrid regression approaches should receive particular consideration for complex manufacturing applications, although model selection should remain dependent on independent validation rather than training accuracy alone. Classification performance should be evaluated through sensitivity, specificity, precision, F1-score, ROC-AUC, false-negative rates, and Matthews correlation coefficients in addition to overall accuracy, especially when defective observations represent a minority class. Class-balancing procedures such as SMOTE, class weighting, and cost-sensitive learning should be evaluated where defect prevalence is uneven. Predictive models should also undergo repeated cross-validation, machine-level holdout testing, material-transfer assessment, temporal validation, and controlled sensor-noise testing before operational deployment because this study demonstrated measurable performance reductions under distributional change. Manufacturing managers should integrate validated predictive models with process-parameter optimization, particularly multiobjective and Bayesian approaches, to simultaneously consider defect probability, first-pass yield, process capability, cycle time, energy use, material consumption, and cost. The substantial reductions observed in defects, scrap, rework, downtime, energy consumption, and cost of poor quality indicate that optimization objectives should extend beyond prediction accuracy to measurable operational outcomes. Continuous monitoring should subsequently compare predicted and observed quality, track data and concept drift, identify changes in sensor distributions, and trigger model recalibration when performance deteriorates. Particular attention should be directed toward high-intensity operating conditions and progressive tool wear because these conditions were associated with elevated defect probability. Finally, future empirical applications should evaluate the framework across larger numbers of manufacturing facilities, machine platforms, materials, production technologies, and longer operating periods and should incorporate real-time edge analytics, explainable machine learning, uncertainty estimation, digital-twin integration, and adaptive optimization to strengthen interpretability, scalability, transferability, and sustained industrial performance.

LIMITATIONS

This study had several limitations that should be considered when interpreting the predictive, optimization, and manufacturing-performance findings. First, although the analytical dataset contained 12,480 manufacturing observations, 3,120 production cycles, 6,240 manufactured components, 24 production batches, eight machines, and 18 synchronized sensor channels, the evaluated observations represented a defined set of machining, additive-manufacturing, welding, and sensor-enabled production conditions and therefore did not capture the complete technological diversity of advanced manufacturing systems. Differences in machine architecture, controller configuration, material composition, tooling, sensor placement, environmental conditions, maintenance practices, operator behavior, and production scale could influence model performance in other manufacturing environments. Second, the 20.0% baseline defect prevalence represented the evaluated dataset and might differ substantially from highly mature production systems in which defective products are considerably rarer. Although SMOTE, class weighting, and cost-sensitive learning were evaluated, extreme class imbalance could produce different sensitivity, precision, and false-negative characteristics. Third, feature engineering and selection reduced 186 candidate variables to 74 predictors, but potentially relevant latent characteristics might not have been captured by the

available sensor network. Measurement noise, synchronization errors, sensor degradation, missing observations, calibration variation, and preprocessing decisions could also influence extracted features and subsequent predictions. Fourth, although the ensemble classifier achieved 96.4% independent-test accuracy and the strongest continuous models generated R^2 values between 0.902 and 0.941, these performance levels should not be interpreted as universally transferable. Accuracy decreased to 93.8% under machine-level holdout, 92.9% during material transfer, 92.4% during temporal holdout, and 91.7% under 20% simulated sensor noise, demonstrating sensitivity to distributional change. Fifth, optimization results were obtained within predefined parameter boundaries and manufacturing constraints; consequently, the identified optimum represented the best evaluated operating region rather than an unrestricted global optimum. Experimental confirmation strengthened confidence in the optimization results, but repeated validation over longer production periods could reveal gradual tool degradation, seasonal environmental effects, material-lot changes, maintenance interventions, and concept drift that were not fully represented. Sixth, economic outcomes such as the 14.0% reduction in unit manufacturing cost and 54.0% reduction in cost of poor quality were based on the evaluated production structure and might vary under different labor, energy, material, maintenance, inspection, and equipment-cost conditions. Finally, although statistical significance, confidence intervals, effect sizes, cross-validation, subgroup analysis, and robustness testing strengthened the quantitative evidence, the observational relationships among several process variables should not automatically be interpreted as independent causal effects. These limitations therefore defined the appropriate boundaries for interpreting the reported predictive accuracy, parameter relationships, optimization benefits, and operational improvements.

REFERENCES

[1]. Abdullah Mohammad, S. (2023). $CUGAO_2$ and $CUINS_2$ Quantum Dot Sensitization for Enhanced Photovoltaic Efficiency in Wearable Fiber-Shaped Solar Cells. *American Journal of Advanced Technology and Engineering Solutions*, *3*(04), 209-261. https://doi.org/10.63125/wysdvy71

[2]. Adegoke, O., Andersson, J., Brodin, H., & Pederson, R. (2020). Influence of laser powder bed fusion process parameters on voids, cracks, and microhardness of nickel-based superalloy alloy 247LC. *Materials*, *13*(17), 3770.

[3]. Adnan, S., & Iftekhar, A. (2023). Predicting Cost Variance in Digital Transformation Projects Through Ensemble Learning and Project Performance Analytics. *International Journal of Scientific Interdisciplinary Research*, *13*(10), 01–36. https://doi.org/10.63125/j18hxs51

[4]. Alsaeedi, A., & Khan, M. Z. (2019). Software defect prediction using supervised machine learning and ensemble techniques: a comparative study. *Journal of Software Engineering and Applications*, *12*(5), 85-100.

[5]. Alt, N., Zhang, T. Y., Motchnik, P., Taticek, R., Quarmby, V., Schlothauer, T., Beck, H., Emrich, T., & Harris, R. J. (2016). Determination of critical quality attributes for monoclonal antibodies using quality by design principles. *Biologicals*, *44*(5), 291-305.

[6]. Anagnostopoulos, C. (2016). Quality-optimized predictive analytics. *Applied Intelligence*, *45*(4), 1034-1046.

[7]. Asghar, I., & Kim, J. S. (2020). An automated smart EPQ-based inventory model for technology-dependent products under stochastic failure and repair rate. *Symmetry*, *12*(3), 388.

[8]. Ashik, M., & Shabbir, A. (2022). Digital Twin-Enabled Predictive Maintenance for Rotating and Static Equipment: Bridging FEA-Based Failure Analysis with Real-Time Asset Health Monitoring in Energy and Power Facilities. *International Journal of Scientific Interdisciplinary Research*, *12*(05), 01–37. https://doi.org/10.63125/0qaysz49

[9]. Baffari, D., Reynolds, A. P., Masnata, A., Fratini, L., & Ingarao, G. (2019). Friction stir extrusion to recycle aluminum alloys scraps: Energy efficiency characterization. *Journal of Manufacturing Processes*, *43*, 63-69.

[10]. Balachandran, P. V. (2019). Machine learning guided design of functional materials with targeted properties. *Computational Materials Science*, *164*, 82-90.

[11]. Bannach-Brown, A., Przybyła, P., Thomas, J., Rice, A. S., Ananiadou, S., Liao, J., & Macleod, M. R. (2019). Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. *Systematic reviews*, *8*(1), 23.

[12]. Bartlett, J. L., Heim, F. M., Murty, Y. V., & Li, X. (2018). In situ defect detection in selective laser melting via full-field infrared thermography. *Additive Manufacturing*, *24*, 595-605.

[13]. Baumgartl, H., Tomas, J., Buettner, R., & Merkel, M. (2020). A deep learning-based model for defect detection in laser-powder bed fusion using in-situ thermographic monitoring. *Progress in Additive Manufacturing*, *5*(3), 277-285.

[14]. Bellini, V., Guzzon, M., Bigliardi, B., Mordonini, M., Filippelli, S., & Bignami, E. (2020). Artificial intelligence: a new tool in operating room management. Role of machine learning models in operating room optimization. *Journal of medical systems*, *44*(1), 20.

[15]. Bonacina, F., Miele, E. S., & Corsini, A. (2020). Time series clustering: a complex network-based approach for feature selection in multi-sensor data. *Modelling*, *1*(1), 1-21.

[16]. Boukhelifa, N., Bezerianos, A., & Lutton, E. (2018). Evaluation of interactive machine learning systems. In *Human and machine learning: visible, explainable, trustworthy and transparent* (pp. 341-360). Springer.

[17]. Brüning, J., Denkena, B., Dittrich, M.-A., & Hocke, T. (2017). Machine learning approach for optimization of automated fiber placement processes. *Procedia CIRP*, *66*, 74-78.
[18]. Chen, W. (2020). Intelligent manufacturing production line data monitoring system for industrial internet of things. *Computer communications*, *151*, 31-41.
[19]. Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). A review of the artificial neural network models for water quality prediction. *Applied Sciences*, *10*(17), 5776.
[20]. Chen, Y., Wang, H., Wu, Y., & Wang, H. (2020). Predicting the printability in selective laser melting with a supervised machine learning method. *Materials*, *13*(22), 5063.
[21]. Cheng, Z., & Liu, X. (2015). Optimal online soft sensor for product quality monitoring in propylene polymerization process. *Neurocomputing*, *149*, 1216-1224.
[22]. Cherry, J., Davies, H., Mehmood, S., Lavery, N., Brown, S., & Sienz, J. (2015). Investigation into the effect of process parameters on microstructural and physical properties of 316L stainless steel parts by selective laser melting. *The International Journal of Advanced Manufacturing Technology*, *76*(5), 869-879.
[23]. Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability*, *12*(19), 8211.
[24]. Colledani, M., & Angius, A. (2020). Production quality performance of manufacturing systems with in-line product traceability and rework. *CIRP Annals*, *69*(1), 365-368.
[25]. Colledani, M., Coupek, D., Verl, A., Aichele, J., & Yemane, A. (2018). A cyber-physical system for quality-oriented assembly of automotive electric motors. *CIRP journal of manufacturing science and technology*, *20*, 12-22.
[26]. Debasish, A. (2021). Energy-Aware Process Optimization for Reducing Material Waste in Sustainable Additive Manufacturing. *American Journal of Advanced Technology and Engineering Solutions*, *1*(4), 108-137. https://doi.org/10.63125/j0wk7j48
[27]. Delli, U., & Chang, S. (2018). Automated process monitoring in 3D printing using supervised machine learning. *Procedia Manufacturing*, *26*, 865-870.
[28]. Dey, A., & Yodo, N. (2019). A systematic survey of FDM process parameter optimization and their influence on part characteristics. *Journal of Manufacturing and Materials Processing*, *3*(3), 64.
[29]. Dindin, M., Umeda, Y., & Chazal, F. (2020). Topological data analysis for arrhythmia detection through modular neural networks. Canadian Conference on Artificial Intelligence,
[30]. Dong, G., Li, X., Zhao, J., Su, S., Misra, R., Xiao, R., & Shang, C. (2020). Machine learning guided methods in building chemical composition-hardenability model for wear-resistant steel. *Materials Today Communications*, *24*, 101332.
[31]. Duan, Z., Wu, T., Guo, S., Shao, T., Malekian, R., & Li, Z. (2018). Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion for rolling bearings: a review. *The International Journal of Advanced Manufacturing Technology*, *96*(1), 803-819.
[32]. Fahle, S., Prinz, C., & Kuhlenkötter, B. (2020). Systematic review on machine learning (ML) methods for manufacturing processes–Identifying artificial intelligence (AI) methods for field application. *Procedia CIRP*, *93*, 413-418.
[33]. Filipe, J., Bessa, R. J., Reis, M., Alves, R., & Póvoa, P. (2019). Data-driven predictive energy optimization in a wastewater pumping station. *Applied Energy*, *252*, 113423.
[34]. Furre, A.-K., Eiken, O., Alnes, H., Vevatne, J. N., & Kiær, A. F. (2017). 20 years of monitoring CO2-injection at Sleipner. *Energy procedia*, *114*, 3916-3926.
[35]. Gábor, A., & Banga, J. R. (2015). Robust and efficient parameter estimation in dynamic models of biological systems. *BMC systems biology*, *9*(1), 74.
[36]. Gao, H., Zhang, Y., Zhou, X., & Li, D. (2018). Intelligent methods for the process parameter determination of plastic injection molding. *Frontiers of Mechanical Engineering*, *13*(1), 85-95.
[37]. Garcia Plaza, E., Nunez Lopez, P., & Beamud Gonzalez, E. (2018). Multi-sensor data fusion for real-time surface quality control in automated machining systems. *Sensors*, *18*(12), 4381.
[38]. García, V., Sánchez, J. S., Rodríguez-Picón, L. A., Méndez-González, L. C., & Ochoa-Domínguez, H. d. J. (2019). Using regression models for predicting the product quality in a tubing extrusion process. *Journal of Intelligent Manufacturing*, *30*(6), 2535-2544.
[39]. Garoudja, E., Chouder, A., Kara, K., & Silvestre, S. (2017). An enhanced machine learning based approach for failures detection and diagnosis of PV systems. *Energy conversion and management*, *151*, 496-513.
[40]. Gins, G., Van Impe, J. F., & Reis, M. S. (2018). Finding the optimal time resolution for batch-end quality prediction: MRQP–A framework for multi-resolution quality prediction. *Chemometrics and Intelligent Laboratory Systems*, *172*, 150-158.
[41]. Graves, S. J., Asner, G. P., Martin, R. E., Anderson, C. B., Colgan, M. S., Kalantari, L., & Bohlman, S. A. (2016). Tree species abundance predictions in a tropical agricultural landscape with a supervised classification model and imbalanced data. *Remote Sensing*, *8*(2), 161.
[42]. Han, X., Cui, R., Lan, Y., Kang, Y., Deng, J., & Jia, N. (2019). A Gaussian mixture model based combined resampling algorithm for classification of imbalanced credit data sets. *International Journal of Machine Learning and Cybernetics*, *10*(12), 3687-3699.
[43]. Hartl, R., Hansjakob, J., & Zaeh, M. (2020). Improving the surface quality of friction stir welds using reinforcement learning and Bayesian optimization. *The International Journal of Advanced Manufacturing Technology*, *110*(11), 3145-3167.

[44]. Healey, S. P., Cohen, W. B., Yang, Z., Brewer, C. K., Brooks, E. B., Gorelick, N., Hernandez, A. J., Huang, C., Hughes, M. J., & Kennedy, R. E. (2018). Mapping forest change using stacked generalization: An ensemble approach. *Remote Sensing of Environment*, *204*, 717-728.

[45]. Jayanthi, R., & Florence, L. (2019). Software defect prediction techniques using metrics based on neural network classifier. *Cluster Computing*, *22*(Suppl 1), 77-88.

[46]. Jiang, J., & Ma, Y. (2020). Path planning strategies to optimize accuracy, quality, build time and material use in additive manufacturing: a review. *Micromachines*, *11*(7), 633.

[47]. Jin, Z., Zhang, Z., & Gu, G. X. (2019). Autonomous in-situ correction of fused deposition modeling printers using computer vision and deep learning. *Manufacturing Letters*, *22*, 11-15.

[48]. Jun, J.-h., Chang, T.-W., & Jun, S. (2020). Quality prediction and yield improvement in process manufacturing based on data analytics. *Processes*, *8*(9), 1068.

[49]. Kang, S. (2020). Model validation failure in class imbalance problems. *Expert Systems with Applications*, *146*, 113190.

[50]. Kao, Y.-T., Dauzère-Pérès, S., Blue, J., & Chang, S.-C. (2018). Impact of integrating equipment health in production scheduling for semiconductor fabrication. *Computers & Industrial Engineering*, *120*, 450-459.

[51]. Kasperovich, G., Haubrich, J., Gussone, J., & Requena, G. (2016). Correlation between porosity and processing parameters in TiAl6V4 produced by selective laser melting. *Materials & Design*, *105*, 160-170.

[52]. Kazi Rakib Hasan, S., & Adnan, S. (2022). Quantitative Assessment of Machine Learning-Based Schedule Risk Prediction for Reducing Delays in Complex Technology Projects. *Review of Applied Science and Technology*, *1*(04), 464–501. https://doi.org/10.63125/dn8nkm85

[53]. Khairallah, S. A., Anderson, A. T., Rubenchik, A., & King, W. E. (2016). Laser powder-bed fusion additive manufacturing: Physics of complex melt flow and formation mechanisms of pores, spatter, and denudation zones. *Acta Materialia*, *108*, 36-45.

[54]. Kwon, O., Kim, H. G., Ham, M. J., Kim, W., Kim, G.-H., Cho, J.-H., Kim, N. I., & Kim, K. (2020). A deep neural network for classification of melt-pool images in metal additive manufacturing. *Journal of Intelligent Manufacturing*, *31*(2), 375-386.

[55]. Laha, D., Ren, Y., & Suganthan, P. N. (2015). Modeling of steelmaking process with effective machine learning techniques. *Expert Systems with Applications*, *42*(10), 4687-4696.

[56]. Lahmiri, S., Dawson, D. A., & Shmuel, A. (2018). Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures. *Biomedical engineering letters*, *8*(1), 29-39.

[57]. Laradji, I. H., Alshayeb, M., & Ghouti, L. (2015). Software defect prediction using ensemble learning on selected features. *Information and Software Technology*, *58*, 388-402.

[58]. Lee, J., Noh, S. D., Kim, H.-J., & Kang, Y.-S. (2018). Implementation of cyber-physical production systems for quality prediction and operation control in metal casting. *Sensors*, *18*(5), 1428.

[59]. Lee, S. M., Lee, D., & Kim, Y. S. (2019). The quality management ecosystem for predictive maintenance in the Industry 4.0 era. *International Journal of Quality Innovation*, *5*(1), 4.

[60]. Lehman, R. M., Cambardella, C. A., Stott, D. E., Acosta-Martinez, V., Manter, D. K., Buyer, J. S., Maul, J. E., Smith, J. L., Collins, H. P., & Halvorson, J. J. (2015). Understanding and enhancing soil biological health: the solution for reversing soil degradation. *Sustainability*, *7*(1), 988-1027.

[61]. Levy, S., Duda, M., Haber, N., & Wall, D. P. (2017). Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. *Molecular autism*, *8*(1), 65.

[62]. Li, J., Wang, L., Zhang, X., Liu, L., Li, J., Chan, M. F., Sui, J., & Yang, R. (2019). Machine learning for patient-specific quality assurance of VMAT: prediction and classification accuracy. *International Journal of Radiation Oncology* Biology* Physics*, *105*(4), 893-902.

[63]. Liu, L., Hua, Y., Zhao, Q., Huang, H., & Bovik, A. C. (2016). Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Processing: Image Communication*, *40*, 1-15.

[64]. Liu, Z., Li, M., Weng, Y., Qian, Y., Wong, T. N., & Tan, M. J. (2020). Modelling and parameter optimization for filament deformation in 3D cementitious material printing using support vector machine. *Composites Part B: Engineering*, *193*, 108018.

[65]. Liu, Z., Zhang, D., Jia, W., Lin, X., & Liu, H. (2020). An adversarial bidirectional serial–parallel LSTM-based QTD framework for product quality prediction. *Journal of Intelligent Manufacturing*, *31*(6), 1511-1529.

[66]. Lorenz, S., Seidel, P., Ghamisi, P., Zimmermann, R., Tusa, L., Khodadadzadeh, M., Contreras, I. C., & Gloaguen, R. (2019). Multi-sensor spectral imaging of geological samples: A data fusion approach using spatio-spectral feature extraction. *Sensors*, *19*(12), 2787.

[67]. Mahbubul, I., Mahmudul, I., Shifa, T., Mahfuza, T., & Shima Ali, S. (2022). Exploring The Potential of Visual Narratives and Interactive Gamified Communication Systems to Prevent Disease, Counter Health Misinformation, And Strengthen Public Trust in Low-Literacy Communities: A Literature Review. *American Journal of Interdisciplinary Studies*, *3*(02), 165-220. https://doi.org/10.63125/zmqs4n64

[68]. Malekipour, E., & El-Mounayri, H. (2018). Common defects and contributing parameters in powder bed fusion AM process and their classification for online monitoring and control: a review. *The International Journal of Advanced Manufacturing Technology*, *95*(1), 527-550.

[69]. Mashfiqur, R., & Md, A. (2022). Monte Carlo Simulation and Stochastic Cash-Flow Modeling for Reducing Liquidity Risk in Project-Based Enterprises. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 382–431. https://doi.org/10.63125/zf3fwv21

[70]. Mastanaiah, P., Sharma, A., & Reddy, G. M. (2016). Dissimilar friction stir welds in AA2219-AA5083 aluminium alloys: effect of process parameters on material inter-mixing, defect formation, and mechanical properties. *Transactions of the Indian Institute of Metals*, *69*(7), 1397-1415.

[71]. Md Anisur Rahman, M. (2023). Development of AI-Driven Business Intelligence Frameworks for Financial Risk and Liquidity Monitoring in Real Time. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 152–203. https://doi.org/10.63125/jm55ca64

[72]. Md Ashfaq, S., & Abdullah Mohammad, S. (2022). CNT/PANI Composite Electrodes for Flexible Fiber-Shaped Dye-Sensitized Solar Cells: Synthesis, Characterization, and Photovoltaic Performance. *American Journal of Interdisciplinary Studies*, *3*(02), 221-274. https://doi.org/10.63125/m334mf41

[73]. Md Asif Ali Sheak, A., & Tahmedur, R. (2023). A Data-Driven Framework for Engineering Project Risk Assessment Under Schedule and Cost Uncertainty. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 204–251. https://doi.org/10.63125/kq4zzh16

[74]. Md Kaium, H. (2023). A GIS-Based Multi-Criteria Decision-Support and Land-Suitability Optimization Framework for Climate-Resilient Urban and Regional Planning. *American Journal of Data Science and Analytics*, *4*(06), 43–103. https://doi.org/10.63125/wvq0sx87

[75]. Md Mahamud Pervaz, P. (2023). Integration of MES, SAP, And PI Historian Platforms for Real-Time Industrial Asset Performance Management. *Review of Applied Science and Technology*, *2*(04), 393–445. https://doi.org/10.63125/9ebtas97

[76]. Md. Abdur, R., & Iftekhar, A. (2021). Customer Retention Forecasting in Mobile Wallet Services Using Neural Networks: A Comparative Quantitative Study. *International Journal of Business and Economics Insights*, *1*(4), 70-102. https://doi.org/10.63125/dyrpc387

[77]. Meng, L., McWilliams, B., Jarosinski, W., Park, H.-Y., Jung, Y.-G., Lee, J., & Zhang, J. (2020). Machine Learning in Additive Manufacturing: A Review: Meng, McWilliams, Jarosinski, Park, Jung, Lee, and Zhang. *Jom*, *72*(6), 2363-2377.

[78]. Miao, F., Zhao, R., Wang, X., & Jia, L. (2020). A new fault feature extraction method for rotating machinery based on multiple sensors. *Sensors*, *20*(6), 1713.

[79]. Mohamed, M. A., Manurung, Y. H., & Berhan, M. N. (2015). Model development for mechanical properties and weld quality class of friction stir welding using multi-objective Taguchi method and response surface methodology. *Journal of Mechanical Science and Technology*, *29*(6), 2323-2331.

[80]. Mohammad Badrul, A., & Md. Mominul, H. (2023). Machine Learning–Based Assessment of Environmental and Public Health Risks in Sewerage Sludge Management Systems. *American Journal of Advanced Technology and Engineering Solutions*, *3*(02), 33-77. https://doi.org/10.63125/y3yfxa92

[81]. Molla Al Rakib, H., & Mohammad Towhidul, I. (2023). Cyber-Resilient Marketing Intelligence Systems for Fraud Detection and Data Integrity in Digital Campaigns. *American Journal of Scholarly Research and Innovation*, *2*(02), 485–518. https://doi.org/10.63125/b53pgr11

[82]. Morariu, C., Morariu, O., Răileanu, S., & Borangiu, T. (2020). Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems. *Computers in Industry*, *120*, 103244.

[83]. Mukut Kanti, B. (2023). Importance of the Sustainable growth in the Testing, Inspection and Certification (TIC) industries through AI tools and techniques. *Review of Applied Science and Technology*, *13*(07), 50–75. https://doi.org/10.63125/0aagev30

[84]. Naimi, A. I., & Balzer, L. B. (2018). Stacked generalization: an introduction to super learning. *European journal of epidemiology*, *33*(5), 459-464.

[85]. Nishat, J. (2023). Communicable Disease Prevention in a Post-Pandemic Era: A Global Review of Public Health Education, Surveillance, and Community-Level Interventions. *International Journal of Scientific Interdisciplinary Research*, *4*(4), 630–678. https://doi.org/10.63125/mnqefj36

[86]. Nogueira, I., Fontes, C., Sartori, I., Pontes, K., & Embiruçu, M. (2017). A model-based approach to quality monitoring of a polymerization process without online measurement of product specifications. *Computers & Industrial Engineering*, *106*, 123-136.

[87]. Orrù, P. F., Zoccheddu, A., Sassu, L., Mattia, C., Cozza, R., & Arena, S. (2020). Machine learning approach using MLP and SVM algorithms for the fault prediction of a centrifugal pump in the oil and gas industry. *Sustainability*, *12*(11), 4776.

[88]. Papazetis, G., & Vosniakos, G.-C. (2019). Mapping of deposition-stable and defect-free additive manufacturing via material extrusion from minimal experiments. *The International Journal of Advanced Manufacturing Technology*, *100*(9), 2207-2219.

[89]. Paschali, M., Simson, W., Roy, A. G., Göbl, R., Wachinger, C., & Navab, N. (2019). Manifold exploring data augmentation with geometric transformations for increased performance and robustness. International Conference on Information Processing in Medical Imaging,

[90]. Psuj, G. (2018). Multi-sensor data integration using deep learning for characterization of defects in steel elements. *Sensors*, *18*(1), 292.

[91]. Ren, L., Cui, J., Sun, Y., & Cheng, X. (2017). Multi-bearing remaining useful life collaborative prediction: A deep learning approach. *Journal of Manufacturing Systems*, *43*, 248-256.

[92]. Risha, A., & Mst Shurovi, A. (2023). Cloud-Based Accounting Systems and Cybersecurity Frameworks for Financial Data Protection in Emerging Economies: A Meta-Analysis. *International Journal of Business and Economics Insights*, *3*(4), 26–64. https://doi.org/10.63125/zb0h3n85

[93]. Rizal, M., Ghani, J., Nuawi, M., & Haron, C. (2017). Cutting tool wear classification and detection using multi-sensor signals and Mahalanobis-Taguchi System. *Wear*, *376*, 1759-1765.

[94]. Roggo, Y., Jelsch, M., Heger, P., Ensslin, S., & Krumme, M. (2020). Deep learning for continuous manufacturing of pharmaceutical solid dosage form. *European journal of Pharmaceutics and biopharmaceutics*, *153*, 95-105.

[95]. Sadia, Z., & Maishar Mostari, Z. (2022). The Impact of Interactive Dashboard Visualization on Data-Driven Decision Making in Financial Services: A Qualitative Evaluation. *International Journal of Scientific Interdisciplinary Research*, *12*(03), 01–35. https://doi.org/10.63125/3jstsv75

[96]. Schlögl, M. (2020). A multivariate analysis of environmental effects on road accident occurrence using a balanced bagging approach. *Accident Analysis & Prevention*, *136*, 105398.

[97]. Schmitt, J., Bönig, J., Borggräfe, T., Beitinger, G., & Deuse, J. (2020). Predictive model-based quality inspection using Machine Learning and Edge Cloud Computing. *Advanced engineering informatics*, *45*, 101101.

[98]. Shaiful, M., Anisur, R., & Md, A. (2022). A Systematic Literature Review on the Role of Digital Health Twins in Preventive Healthcare For Personal and Corporate Wellbeing. *American Journal of Interdisciplinary Studies*, *3*(04), 1-31. https://doi.org/10.63125/negjw373

[99]. Shang, C., & You, F. (2019). Data analytics and machine learning for smart process manufacturing: Recent advances and perspectives in the big data era. *Engineering*, *5*(6), 1010-1016.

[100]. Shi, T., Jiang, H., & Zheng, B. (2020). A stacked generalization U-shape network based on zoom strategy and its application in biomedical image segmentation. *Computer Methods and Programs in Biomedicine*, *197*, 105678.

[101]. Shrestha, L., Crichton, S. O., Kulig, B., Kiesel, B., Hensel, O., & Sturm, B. (2020). Comparative analysis of methods and model prediction performance evaluation for continuous online non-invasive quality assessment during drying of apples from two cultivars. *Thermal Science and Engineering Progress*, *18*, 100461.

[102]. Stabak, D., & Md Shahid, N. (2022). Empirical Assessment of Uncertainty-Aware Real-Time Decision-Making for Reliable and Resilient Cyber-Physical Systems. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 432–477. https://doi.org/10.63125/q3r4kk16

[103]. Stabak, D., & Md Siam, T. (2023). Real-Time Edge Computing for Cyber-Physical Systems: A Meta-Analysis of Energy Efficiency and Fault Tolerance. *American Journal of Scholarly Research and Innovation*, *13*(06), 01–44. https://doi.org/10.63125/g9c9cp41

[104]. Stolpe, M., Blom, H., & Morik, K. (2016). Sustainable industrial processes by embedded real-time quality prediction. In *Computational sustainability* (pp. 201-243). Springer.

[105]. Syafrudin, M., Alfian, G., Fitriyani, N. L., & Rhee, J. (2018). Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. *Sensors*, *18*(9), 2946.

[106]. Syed Feroz, A. (2023). Quality-Management Systems for Resilient Energy Infrastructure Fabrication. *International Journal of Scientific Interdisciplinary Research*, *13*(10), 37–91. https://doi.org/10.63125/mbhhrm88

[107]. Tabernik, D., Šela, S., Skvarč, J., & Skočaj, D. (2020). Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, *31*(3), 759-776.

[108]. Tahmedur, R., & Rifat, C. (2023). Critical Path Analysis and Resource-Leveling Optimization for Multi-Project Engineering Portfolios. *American Journal of Interdisciplinary Studies*, *13*(12), 01-47. https://doi.org/10.63125/8eqnea46

[109]. Tang, B., Lu, Y., Zhou, J., Chouhan, T., Wang, H., Golani, P., Xu, M., Xu, Q., Guan, C., & Liu, Z. (2020). Machine learning-guided synthesis of advanced inorganic materials. *Materials Today*, *41*, 72-80.

[110]. Thabtah, F., Abdelhamid, N., & Peebles, D. (2019). A machine learning autism classification based on logistic regression analysis. *Health information science and systems*, *7*(1), 12.

[111]. Tian, Y., Chen, W., Zhou, T., Li, J., Ding, K., & Li, J. (2020). Establishment and evaluation of a multicenter collaborative prediction model construction framework supporting model generalization and continuous improvement: A pilot study. *International journal of medical informatics*, *141*, 104173.

[112]. Tuan, T. A., Long, H. V., Son, L. H., Kumar, R., Priyadarshini, I., & Son, N. T. K. (2020). Performance evaluation of Botnet DDoS attack detection using machine learning. *Evolutionary Intelligence*, *13*(2), 283-294.

[113]. Völker, C., & Shokouhi, P. (2015). Multi sensor data fusion approach for automatic honeycomb detection in concrete. *Ndt & e International*, *71*, 54-60.

[114]. Wang, G., Ledwoch, A., Hasani, R. M., Grosu, R., & Brintrup, A. (2019). A generative neural network model for the quality prediction of work in progress products. *Applied Soft Computing*, *85*, 105683.

[115]. Wang, H., Ni, G., Chen, J., & Qu, J. (2020). Research on rolling bearing state health monitoring and life prediction based on PCA and Internet of things with multi-sensor. *Measurement*, *157*, 107657.

[116]. Wang, J., Xie, J., Zhao, R., Zhang, L., & Duan, L. (2017). Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing. *Robotics and computer-integrated manufacturing*, *45*, 47-58.

[117]. Wang, L., Sun, Y., He, Y., Zhao, X., Fan, W., & Wang, X. (2019). Motor health status prediction method based on information from multi-sensor and multi-feature parameters. *Journal of Nondestructive Evaluation*, *38*(2), 46.

[118]. Weichert, D., Link, P., Stoll, A., Rüping, S., Ihlenfeldt, S., & Wrobel, S. (2019). A review of machine learning for the optimization of production processes. *The International Journal of Advanced Manufacturing Technology*, *104*(5), 1889-1902.

[119]. Weiss, S. M., Dhurandhar, A., Baseman, R. J., White, B. F., Logan, R., Winslow, J. K., & Poindexter, D. (2016). Continuous prediction of manufacturing performance throughout the production lifecycle. *Journal of Intelligent Manufacturing*, *27*(4), 751-763.

[120]. Whyte, J., Stasis, A., & Lindkvist, C. (2016). Managing change in the delivery of complex projects: Configuration management, asset information and 'big data'. *International journal of project management*, *34*(2), 339-351.

[121]. Wright, L., & Davidson, S. (2020). How to tell the difference between a model and a digital twin. *Advanced Modeling and Simulation in Engineering Sciences*, *7*(1), 13.

[122]. Wu, J., Su, Y., Cheng, Y., Shao, X., Deng, C., & Liu, C. (2018). Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system. *Applied Soft Computing*, *68*, 13-23.

[123]. Xie, C., Du, R., Ho, J. W., Pang, H. H., Chiu, K. W., Lee, E. Y., & Vardhanabhuti, V. (2020). Effect of machine learning re-sampling techniques for imbalanced datasets in 18F-FDG PET-based radiomics model on prognostication performance in cohorts of head and neck cancer patients. *European journal of nuclear medicine and molecular imaging*, *47*(12), 2826-2835.

[124]. Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in human behavior*, *58*, 119-129.

[125]. Yadollahi, A., & Shamsaei, N. (2017). Additive manufacturing of fatigue resistant materials: Challenges and opportunities. *International Journal of Fatigue*, *98*, 14-31.

[126]. Yang, K., Jin, H., Chen, X., Dai, J., Wang, L., & Zhang, D. (2016). Soft sensor development for online quality prediction of industrial batch rubber mixing process using ensemble just-in-time Gaussian process regression models. *Chemometrics and Intelligent Laboratory Systems*, *155*, 170-182.

[127]. Yin, S., Liu, L., & Hou, J. (2016). A multivariate statistical combination forecasting method for product quality evaluation. *Information Sciences*, *355*, 229-236.

[128]. Yu, L., Zhou, R., Tang, L., & Chen, R. (2018). A DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data. *Applied Soft Computing*, *69*, 192-202.

[129]. 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.

[130]. Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2018). Employee turnover prediction with machine learning: A reliable approach. Proceedings of SAI intelligent systems conference,

[131]. Zheng, H., Feng, Y., Gao, Y., & Tan, J. (2018). A robust predicted performance analysis approach for data-driven product development in the industrial internet of things. *Sensors*, *18*(9), 2871.

[132]. Zimmerling, C., Dörr, D., Henning, F., & Kärger, L. (2019). A machine learning assisted approach for textile formability assessment and design improvement of composite components. *Composites Part A: Applied Science and Manufacturing*, *124*, 105459.

[133]. Zimmerling, C., Poppe, C., & Kärger, L. (2020). Estimating optimum process parameters in textile draping of variable part geometries-a reinforcement learning approach. *Procedia Manufacturing*, *47*, 847-854.

[134]. Zwanenburg, A. (2019). Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *European journal of nuclear medicine and molecular imaging*, *46*(13), 2638-2655.