Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Risk-Aware Rework Prevention in Personalized Hearing Aid Manufacturing

Sudhakavya Bodapati Venkata
Starkey Laboratories Inc, Eden Prairie, MN, 55344, USA
Corresponding Author: Sudhakavya Bodapati Venkata, **E-mail**: bvsudhakavya@gmail.com

| ABSTRACT

The manufacturing of custom hearing aids is done under a degree of variability that is hard to manage using standard production logic since each order is an expression of an indi- vidual anatomy, fitting needs, and design-specific decisions about the product. This means that it is frequently found out too late, after line time, operator effort and material have been used, resulting in avoidable delay and unnecessary burden on manufacturing resources. This paper builds a risk aware execution model to detect probable rework-prone orders earlier in the production pro- cess and to take action on the risk before downstream cost is incurred. The framework integrates order-context capture, risk profiling, rework-risk prediction, intervention selection, execution rout- ing and feedback-based updating in a closed operational loop. Actions that are triggered by the system may include parameter checking, expert checking, alternate routing, or controlled release, depending on the estimated risk level before the order is sent to more costly downstream stages. The main point is that prediction can only be valuable when it alters the way the order is man- aged. In that perspective, rework-risk estimation is included in manufacturing control, as opposed to an independent analytical output. The anticipated advantage is reduced rework load, enhanced turnaround performance, increased attention to specialists, minimized material waste, and more stable production flow in the personalized production of healthcare devices.

| KEYWORDS

Hearing aid manufacturing · Rework prevention · Manufacturing execution · Predictive analytics · Personalized healthcare devices · Risk-aware production

| ARTICLE INFORMATION

ACCEPTED: 01 September 2022 **PUBLISHED:** 25 September 2022 **DOI:** 10.32996/jcsts.2022.4.2.25

1. Introduction

The production of custom hearing aids is not governed by the logic of standard electronics production where the units go through a relatively homogeneous process with few cases to case variation. Each order is shipped with its ear geometry, acoustic goals, shell design options, and fit-related limitations, and these variations directly influence the job flow through the modeling, shell generation, curing, finishing, and subsequent handling processes. This is not the peripheral variability of the process. It alters the produc- tion load in a very tangible manner. Two orders issued on the same day may act quite differently on the line although they may fall into the same broad product category. Previous research on hearing aid shell manufacturing has realized this long before AI became a household discussion in the manufactur- ing industry. Digital scanning, CAD/CAM processes and additive techniques enhanced consistency and throughput, but failed to reduce the operational variability of patient-specific geometry and extremely sensitive downstream processing [3,10,12].

That is important since the narrative surrounding digital manufacturing is usually purer than what tran- spires on the shop floor. As soon as the shell design transitioned to digital workflows and 3D processing was more developed, it was not difficult to believe that the issue of customization was finally resolved operationally. As a matter of fact, that did not happen. The generation of digital shells is definitely quicker and more predictable than the older manual methods, yet

orders do not follow the desired trajec- tory. Others require additional review, others must be diverted, and others attract additional operator work once material and line time have been allocated. That is where the actual cost is realized in a personalized device environment, not just in scrap or remake rates, but also in unstable queues, delays in handoffs, and the use of skilled labor on the same case. More current research on automated virtual shell modeling indicates the same practical conclusion: enhanced design automation does not necessarily eliminate the execution risk that manifests itself later in the manufacturing [11].

The same trend is observed in the field of personalized medical manufacturing in general. Individualized devices can now be produced at a viable scale using additive methods, particularly when the geometry is complicated and the traditional tooling is either too stiff or too slow. Scalable personalization, however, does not eliminate tight control of processes. Patient-specific manufacturing continues to require disci- plined processes, intermediate verification, and staged validation, especially when variability introduced at an early stage may cause significantly greater issues later on [2,9]. The production of hearing aids is in the same category of production with one additional complication: it must be able to deal with the variation imposed by anatomy without sacrificing the speed and responsiveness that is required in a commercial manufacturing setting.

What is usually lacking in such a discussion is a perspective of rework as a problem of execution and not a post-production statistic. In hearing aid production, rework is normally recorded once the interruption has already occurred, instead of being handled as an issue that can still be resolved and the order can be saved. That difference is important on the line. When a difficult order is realized, it is often too late, as it has already gone through several phases, attracted the attention of operators several times, and already disrupted the production schedule. A prediction model alone is not a solution to that. The more practical question is whether the system is capable of identifying a probable rework-prone order early enough to alter its treatment before more cost is hardened. It is not defect inspection and it is not retrospective quality reporting. It is more in line with manufacturing execution control.

Studies in manufacturing have already demonstrated that production information may be useful in making predictions and risk evaluation when it is maintained in the context of its process rather than being diminished to disconnected measurements [6]. The reviews of additive manufacturing and multi- stage production also find a similar conclusion: the signal that is relevant is often the interaction of the characteristics of the order, the conditions of the processes, and the interdependence of the stages and not a single variable considered in isolation [8,4,5]. However, in an actual factory, a risk score is of little use unless it alters the course of action. Assuming that routing, review intensity and release decisions remain constant, the model output would be reduced to a mere additional figure on a dashboard.

A better approach to managing rework risk is to use it as an indicator of action instead of a post-factum score. Orders that are judged to be of low risk can remain in the normal route, and medium-risk orders may receive a mid-manufacturability inspection or parameter inspection prior to being placed in more costly downstream operations. Orders that are considered high risk might require tighter release control, specialist review or even a different routing path. In this context, the issue is brought much nearer to manufacturing execution than to independent predictive analytics. It is also aligned with the larger trend of industrial control systems where model output is supposed to act directly rather than passively reported [1,7].

This paper is based on that operational view. Rework is addressed as a production-control problem in the manufacture of personalized hearing aids and not as a label of quality that is historically assigned. The actual problem is how to convert early order and process signals into handling decisions before moredownstream effort is squandered. That requires an architecture that links order context, risk profiling, prediction, intervention logic, execution routing, and feedback of actual outcomes. It is not just to demonstrate that rework is predictable. This is to provide the production line with a means of responding earlier, with fewer surprises at the end of the day and less firefighting.

The rest of the paper elaborates on that argument operationally. It initially poses rework as an execution- control issue, and then outlines a risk-conscious architecture that bridges prediction with intervention and routing, and lastly explores the predictive layer on a public manufacturing benchmark and then goes back to what this would imply in a real personalized hearing aid production environment.

2. Literature Review

The literature most closely related to this study comes from three overlapping areas: digital hearing aid shell manufacturing, patient-specific additive manufacturing, and predictive analytics for industrial production systems. Earlier work on hearing aid shell fabrication described the shift from manual shell processing to digitally supported workflows built around impression scanning, CAD/CAM modeling, and automated shell generation. Those advances

improved repeatability and removed some of the limitations of conventional shell production, but they did not eliminate the downstream execution burden created by patient-specific variability [3,10,11,12].

The second source of literature is personalized and additive medical manufacturing in general. Studies on this field render it evident that design flexibility alone does not support patient-specific production. It also relies on strict control of the process, intermediate controls, and progressive validation. Research on additive manufacturing of medical devices and patient-specific implants consistently indicates that structured control gates are significant since variation at an early stage of the process can be significantly more expensive at a later stage of the process [2,9]. Combined, that literature justifies considering the personalized device production as a managed workflow, rather than as a fabrication problem.

The third body of literature focuses on machine learning in manufacturing. Surveys of production-line learning and analytics in additive manufacturing indicate that manufacturing information can be useful in prediction when process context and cross-stage dependencies are not flattened away but instead stored in memory as features of the production line [6,8,4,5,7]. Nevertheless, much of that work is still devoted to prediction itself. Reporting of performance metrics is done, yet there is less talk about how those model outputs are to be recapitulated into the real production control. It is that gap that is particularly significant in high-mix, low-volume settings, where the cost of identifying trouble late can be significantly greater than the cost of dragging a challenging case into earlier consideration.

This paper is at the crossroad of those three threads of work. It uses predictive manufacturing analytics as the foundation of early risk estimation, hearing aid manufacturing as the application environment, and intervention and routing as components of the production issue itself and not an addition. In that regard, the paper takes the discussion further than standalone classification by presenting rework prevention as an execution issue in personalized healthcare manufacturing.

3. Problem Statement

There is hardly a single point of failure in rework in the production of personalized hearing aids. It is more frequently expressed as a sequence of corrective measures that can be avoided and which appear only after the order has consumed line time, operator labor, and material. The operational issue is not that rework happens, but that it is often realized too late and the production system has already incurred the cost.

That recognition lag is not a small process problem, it is a structural flaw in the way the manufacturing flow takes variability. Custom-made hearing aid manufacturing is based on the anatomy of an individual and this difference alters the workload per stage. Orders are not fungible. The shape of the canal, style of shell, configuration of the vents, quality of the scan and other case specific considerations may affect manufacturability in a way that cannot be easily observed at intake [10,12]. Online processes made them more consistent and minimized some of the variability of manual processing, but did not change the fact that customized jobs behave differently once they are in production [3,11]. There are orders that go through the line with little friction and those that receive extra verification, re-handling, re-routing or remake effort.

Such disruptions in most production settings are still dealt with in a reactive manner. An order starts losing its course and the system only reacts when the problem becomes evident. At that point, the price is no longer restricted to a local repair. Rework begins to impact queue stability, specialist availability, reliability of due-dates and throughput. One difficult order may use up limited professional resources as well as hold up unrelated tasks behind it. In a high-mix, low-volume environment, such instability is more important than just a number of defects.

The gap in practice does not lie in the fact that manufacturing systems do not have data. The majority of the lines already collect order metadata, process parameters, queue conditions, operator assignments and outcome history. The problem is that this information is hardly ever structured into an execution system that can identify an order that is prone to rework early enough and change the way it is handled before the normal course of action turns out to be expensive. Studies on production analytics have shown that manufacturing information can be useful in prediction when stage context and process dependencies are maintained [6,4]. Additive manufacturing reviews are consistent in telling the same general lesson: predictive models are much more practical when tied to decisions and not isolated analytical results. This is a very significant point in this regard.

The issue may be formulated in a more direct way: The manufacturing of personalized hearing aids does not have a

closed-loop execution mechanism that may detect high-risk orders prior to being sent to more expensive downstream steps and transform that risk into practical measures such as controlled routing, parameter checking, expert inspection, or conditional release. In the absence of such a mechanism, rework is still a late response rather than an early control object. The disruption has already propagated by the time the line reacts.

This problem has two sides. One is predictive: with the available order-level and process-level informa- tion prior to completion, can the system predict the possibility of a job needing rework? The other is functional: when we have such an estimate, how are the line to react in such a manner that it will be of practical value in production, and not merely of analytical interest? A risk score is of little use when it fails to modify routing, review policy or release behavior. Whether a classifier can be trained or not is not the real question, but whether the risk that is predicted can be transformed into execution logic that minimizes downstream disruption.

The concept of rework is used in this research to refer to any corrective production process that causes an order to be taken out of track once it has been released. That may consist of repeat processing, additional hand correction, re-route to a previous step, a controlled hold to review manufacturability, or handling related to remake. The specific term applied in a production system can vary across organizations, but the functional meaning is the same: the order did not flow smoothly through its intended path and needed some extra effort.

The model adopted in the present paper views early rework prevention as an execution-control problem, which is played out in a series of manufacturing decisions. Let x_i denote the feature vector for order i, formed from intake characteristics, customization-related attributes, process context, and production- state indicators available before the order reaches critical downstream stages. Let $y_i \in \{0, 1\}$ denote the observed outcome, where $y_i = 1$ indicates that the order required rework and $y_i = 0$ indicates that it completed along its intended path without corrective manufacturing effort. The first task is to estimate a risk function

$$r_i = f(x_i),$$

where $r_i \in [0, 1]$ represents the predicted likelihood that order i will require rework.

Prediction on its own is not enough. The manufacturing system also has to translate the estimated risk into an intervention that can actually be executed on the floor. Let $a_i \in A$ denote the action assigned to order i, where the action set may include standard release, intermediate verification, expert review, alternate routing, or controlled release. The second task is therefore to define a policy

$$a_i = \pi(r_i, x_i),$$

that converts forecasted risk together with existing production environment into a concrete production reaction. Operationally, it is not about maximizing the statistical classification performance, but about minimizing the expected burden of rework, delay and unnecessary intervention.

When viewed from that angle, the basic research problem is not limited to the prediction of rework. The design of a risk-conscious execution architecture to avoid the need for downstream correction is the use of early manufacturing signals. The argument that is presented in this paper is still based on the hearing aid production, but the problem is deeper than that. Specialized manufacturing requirements require logic that is capable of accommodating variation without having the variation disrupt performance. It is at that point that predictive analytics is no longer a reporting layer, but starts to become a useful manufacturing tool.

4. Execution-Level Architecture

The execution logic shown in Fig. 1 is based on a practical assumption, which is very basic: the prevention of rework is only important when the risk estimate alters the manner in which an order is processed prior to the further downstream commitment of effort. That is why the architecture is not structured as a traditional analytics pipeline. It is configured as a closed manufacturing control loop where order information is decoded, risk is estimated, intervention is selected, execution is modified and the result observed is fed back to refine it later.

The first stage, *order intake and production context*, serves as the operational entry point of the frame- work. At this stage, the system gathers signals which are already present to make a personalized hearing aid order, such as order-specific attributes, geometry-related indicators, current process state, queue con- ditions, and other manufacturing constraints. These signals are not likely to work independently in a custom production environment. A shell setup that seems to be self-manageable might prove challenging when paired with a low scan quality, queue overflow or a time-

dependent release criterion. That is why intake is considered more than a record-keeping step. It is where the manufacturing context of the order starts to become operational.

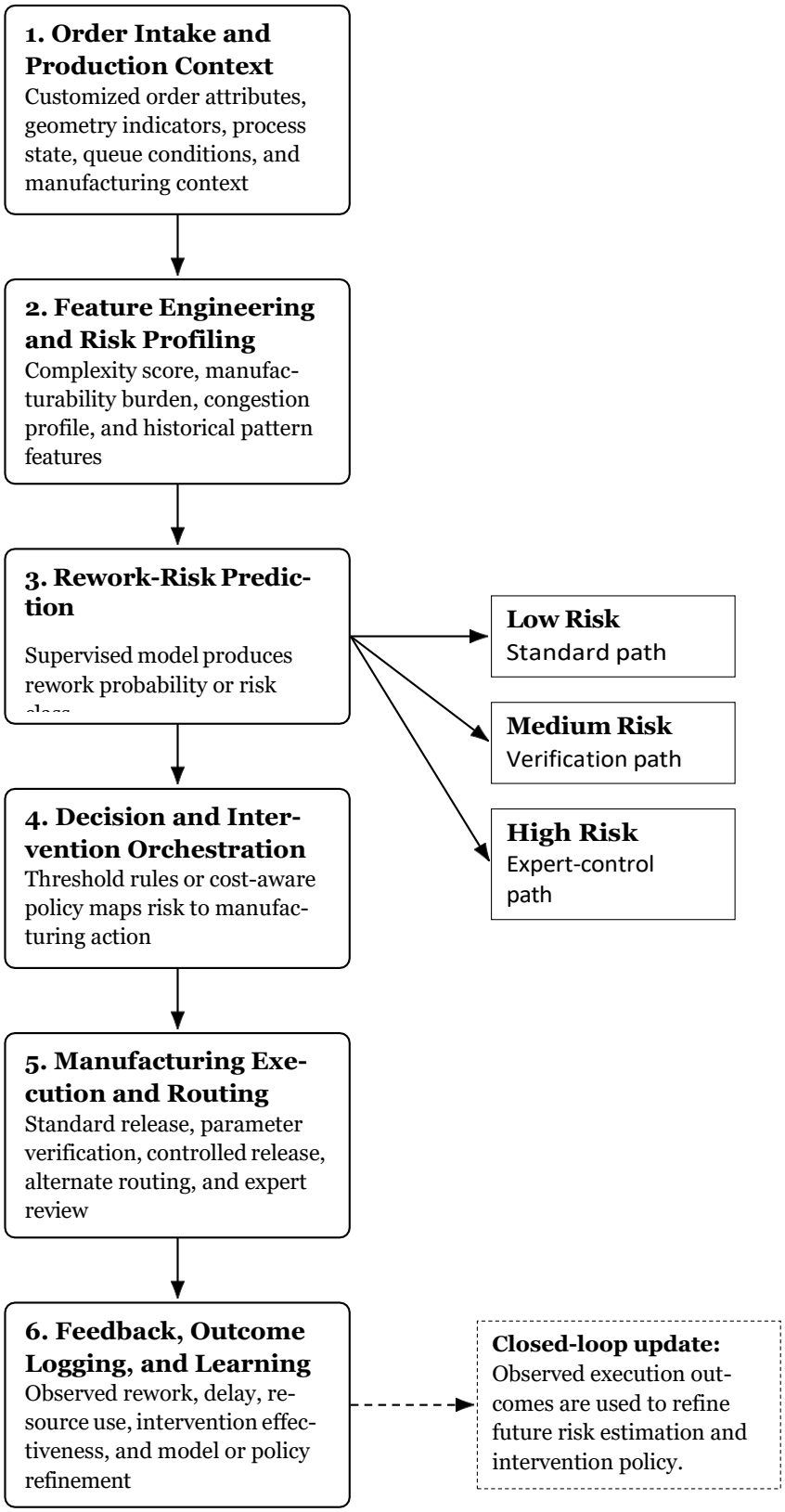

Fig. 1. Execution-level architecture for early rework prevention in personalized hearing aid manufacturing. Risk estimation is treated as an operational signal that governs intervention and routing decisions rather than as a standalone analytical output.

The second stage, *feature engineering and risk profiling*, converts raw records of operations into variables that are more indicative of how an order is expected to perform in production. This involves the derived measures of complexity score, manufacturability burden, queue congestion and historical pattern indica- tors. The idea here is not to simply make the data usable to a classifier. It is to construct a risk profile that is interpretable and still represents conditions that are important in manufacturing.

The third stage, *rework-risk prediction*, is an estimation of whether an order will leave the planned flow and will need the corrective effort in the future of the production process. The prediction element can produce a probability score or a risk category. In Fig. 1, the output is represented in the form of low-, medium- and high-risk, and that decision is not accidental. A risk estimate can only be useful in an execution environment when it is understandable to make a handling decision. Due to that fact, prediction is considered here as a middle control signal and not the ultimate goal of the system.

The right-hand branch of Fig. 1 makes the control logic explicit. Orders with low risk are left on the standard route. Orders of medium risk are diverted to a verification route, where further verifications or controlled processing may be implemented before the order proceeds. Orders with a high risk are shifted to an expert-control route, where review and routing are done more stringently. This is what makes the framework have an execution-level nature. The model is not applied to only find probable trouble. It is applied to modify the way the order is processed before such a bother turns operationally costly.

That policy is formalized in the fourth stage, *decision and intervention orchestration*. This is where the forecasted risk is translated into a production activity. The most straightforward implementation is based on threshold-based regulations, but a cost-conscious policy can be employed as well when the trade-off between the effort of intervention and downstream disruption must be addressed more explicitly. The principle remains the same regardless of the type of the rule: a predicted risk has to result in a tangible intervention decision. In the absence of that, the result will be just an output of analysis and not something that the production system can do something about.

The fifth stage, *manufacturing execution and routing*, translates that decision to the real production route. At this point, an order can be left on standard release, transferred to parameter verification, transferred to controlled release, transferred to an alternate routing path, or transferred to expert review. Here the structure is materially different to a traditional predictive model. The architecture is not halted after estimation of rework likelihood. It changes execution. Such difference is important since the manufacturing load imposed by rework is not only determined by the presence of a problematic order, but also by the wait time of the line before it addresses the order.

The final stage, *feedback, outcome logging, and learning*, completes the loop. Once an order has gone through its designated route, the system logs the measured rework result, delay, resource usage, inter- vention efficacy, and any other data required to determine whether the decision was helpful or not. The observations are then used to optimize the prediction model as well as the intervention policy. The ab- sence of this step would make the architecture a one-way control chain. It makes the framework adaptive, so that past experience in execution can be used to make decisions in the future.

Collectively, these six steps constitute a closed operation architecture of early rework prevention in the production of personalized hearing aids. It begins with the context of order, passes through risk profiling and prediction, and splits into intervention based on the estimated risk, which is then fed back via the observed outcomes. The design choice of the framework is that loop. Rework is not treated as a post factum label. It is addressed as an execution event which in most instances can be predicted and in most instances mitigated by previous and more focused control.

5.Methodology

This paper aims to be a factual analysis of the early rework prevention in the custom-made hearing aid production. The methodological process is based on the execution logic presented in the previous section: order and process recordings are arranged into a structured production dataset, manufacturing-specific characteristics are extracted out of the records, a rework-risk model is trained, and the risk signal is converted to intervention decisions that influence the routing or reviewing of an order on the line. The methodology is built to answer two related questions, first, can rework prone orders be detected before more downstream work is committed and second, can that risk signal be converted into an execution

policy that is actually useful in production and not just informative in retrospective analysis?

5.1 Data Sources and Order Representation

The data is built up based on the historical production data related to the custom hearing aid orders. The records are linked to manufacturing orders that went through the standard production system and the final outcome of which is known. The image-centric representation here is replaced by the order-centric representation, and this is not accidental. This is not intended to be component level defect inspection, but rather to identify orders that are likely to need corrective manufacturing activity once released.

The order level representation is a combination of intake information, customization attributes, pro- cess context and downstream outcome labels. Depending on what is available in the production system, candidate fields can include product family, shell style, ear side, vent configuration, scan or impression quality indicators, geometry-related descriptors, material type, build batch, queue state, due-date prior- ity, workstation assignment, and stage-level handling markers. Additional production-state variables can also be included, if they capture such conditions as congestion, timing pressure, or local manufacturing burden. These variables are assembled into a feature vector x_i for each order i, with the intention of preserving the production conditions that were present before major downstream cost had already been committed.

The target variable is a binary rework label $y_i \in \{0, 1\}$. In this formulation, $y_i = 1$ indicates that the order required corrective manufacturing activity outside its intended path, while $y_i = 0$ indicates completion without that kind of intervention. Corrective activity could be repeat handling, additional manual correction, return to an earlier stage, controlled hold for manufacturability review, or remake related processing. The binary setup is used here to keep the execution policy stable and easy to interpret. A more detailed multi-class scheme could be developed later, but for initial deployment it is more practical to pose the decision as rework versus no rework.

5.2 Data Preparation and Feature Engineering

Raw manufacturing records do not often come in a form that can be used for modeling without further preparation. Before training, the dataset is cleaned to deal with missing values, duplicate entries, incon- sistent categorical labels, and malformed operational fields. Numeric variables are standardized where appropriate and categorical variables are encoded in a manner that is suitable for the learning model being used for training. Orders with no reliable final outcome are removed from the supervised training set as ambiguous labels would undermine the downstream execution logic.

Feature engineering is manufacturing relevant rather than generic feature expansion. Several sets of derived variables are particularly useful in this setting. The first group captures *customization complexity*, including compounded indicators based on style of shell, geometry burden, vent configuration, and fit- related attributes. The second group is reflective of *process pressure* such as queue congestion, due-date urgency and stage load at the time of release. The third group captures *historical similarity*, for example, the observed rework tendency among previously completed orders with similar characteristics. The fourth group reflects *manufacturability burden*, that is, whether an order is in production under conditions that have been historically associated with higher corrective effort.

This stage is important because rework in personalized manufacturing is seldom caused by a single variable. More frequently, it arises out of the combination of customization burden and local production conditions. For that reason, features are designed to maintain operational meaning and not just increase model complexity. The goal is to reflect the types of relationships that production engineers would consider to be of practical importance on the line.

5.3 Rework-Risk Modeling

The predictive component estimates the rework likelihood r_i for each order i as

$$r_i = f(x_i),$$

where x_i is the engineered feature vector and $r_i \in [0, 1]$ represents the predicted probability that the order will require rework. Because the data are tabular and operationally structured, the modeling approach favors methods that work well on heterogeneous production records, while at the same time being interpretable enough to support execution decisions.

A baseline model is first determined using logistic regression. This gives a simple and transparent reference point, and it helps to determine if the rework signal is linearly separable to a meaningful degree. A stronger non-linear model is then trained using either random forests or gradient boosted decision trees. These methods are more suited to interactions between customization, process pressure, and contextual variables, which are unlikely to be linearly related in a personalized manufacturing setting.

The dataset is divided into training and testing subsets with stratified split to maintain the distribution of class of rework events. When the number of rework cases is relatively small, class imbalance is addressed using weighted loss, resampling, or some other balancing technique that is selected to prevent artificially inflating performance. Hyperparameter tuning is only performed on the training part, usually using cross validation, so that the test results are a valid estimate of generalization performance. Model quality is measured in terms of precision, recall, F 1-score, and area under the ROC curve. Accuracy alone is not enough here as a model that is good at predicting the majority class may not necessarily identify the orders that are important operationally.

5.4 Intervention Policy Design

Prediction is not considered as the endpoint of the system. Once a rework score has been estimated, the framework translates the score into a manufacturing action using a policy function

$$a_i = \pi(r_i, x_i),$$

where $a_i \in A$ denotes the selected intervention for order i, and A is the available action set. In the current formulation, the action space consists of standard release, parameter verification, expert review, alternate routing, and controlled release. The exact set can be adjusted to reflect the manufacturing environment under study, but the central principle remains the same: the output of the model must change order handling before avoidable downstream effort is expended.

A threshold based policy is used as the first mechanism of execution. Orders with risk below a lower threshold are put into the standard path. Orders that have risk between the lower and upper thresholds are placed on a verification path, where additional checks for review or manufacturability are applied before proceeding. Orders above the upper threshold are placed on an expert-control path, which can involve conditional release, alternative routing or specialist oversight. This design ensures that the inter- vention layer remains interpretable and operationally plausible. It is also a reflection of how many of the factories actually do absorb predictive information, first through controlled rules and only later through more advanced optimization.

5.5 Execution Feedback Loop

Each order passes through its designated handling path and produces an observed execution outcome. The framework captures whether rework actually occurred, what intervention was used, whether the action caused delay or avoided subsequent disruption and the amount of specialist capacity consumed. These observations are not considered as passive logs. They are part of the learning loop that helps refine the prediction model as well as the intervention policy later.

The feedback design is especially important because the quality of execution cannot be determined by prediction

metrics alone. A model may rank risk well but may still yield poor operational results if the policy thresholds are badly selected or if the intervention burden is misallocated. For that reason, the methodology evaluates the framework on two levels. The first level is predictive performance, which is measured according to classification metrics on held-out data. The second level is execution usefulness, which is measured using such indicators as high-risk capture rate, intervention precision, reduction in exposure to late stage rework, and distribution of review effort across risk groups. This dual evaluation keeps the study on track with its true objective, which is not only to estimate the likelihood of rework, but make it actionable in personalized hearing aid manufacturing.

6. Experimental Design

The experimental design was designed to test the predictive and execution-oriented elements of the framework separately, while maintaining the link between them. The predictive layer was validated using the publicly available SECOM manufacturing dataset, which contains high-dimensional production data with labels of pass-fail outcomes. Although SECOM does not represent hearing aid manufacturing specifically, it provides a useful industrial benchmark to test whether the proposed risk-aware modeling approach is able to identify problematic production cases from structured manufacturing signals before final outcome is known. This makes it suitable for validating the learning and classification parts of the architecture, but not the order-routing and intervention records that would be present in a deployment- specific hearing aid production environment.

The SECOM dataset has 1,567 manufacturing instances and 590 process related features. Among these instances, 104 of them belong to the positive outcome class, corresponding to a failure rate of approxi- mately 6.64%. The strong class imbalance makes the dataset realistic for early-risk prediction but also makes it impossible to use accuracy as the main evaluation criterion. In this context, the more important question is whether the model can detect a meaningful fraction of problematic cases without flooding the system with false alarms.

Each instance was considered as a production unit which is represented as a high dimensional feature vector. Missing values were addressed during the preprocessing stage by median imputation. Features with no measurable variance after imputation were removed and model-specific preprocessing was then applied. For the linear baseline, the features were standardised before classification. For the ensemble model, tree-based learning was used directly on the processed tabular representation.

The dataset was divided with the help of stratified split so that the minority class distribution is main- tained in the training, validation, and testing subsets. A baseline logistic regression model was chosen as the first classifier as it offers a transparent and interpretable reference point for imbalanced manu- facturing data. A random forest model was then introduced as a more powerful non-linear alternative, as interaction effects between production variables are unlikely to be adequately captured by a linear boundary alone. This comparison was designed to test whether the rework-risk formulation benefits from more flexible decision structure under industrial class imbalance.

Threshold tuning was done on the validation subset, rather than the default decision threshold of 0.5. This step was needed because the data set is so imbalanced and because the execution settings are less focused on default class assignment and more focused on capturing useful risk. The tuned threshold was found by maximizing the validation-set F1-score and then the final model was trained on the entire training data and tested on the held-out test data. This strategy decoupled the process of threshold selection from the final test assessment and still produced a decision boundary that more accurately reflects operational needs.

Model performance was assessed in terms of precision, recall, F1-score, ROC-AUC, and PR-AUC. Pre- cision tells us how many flagged units were actually problematic, recall tells us how many problematic units were successfully flagged and *F* 1-score is a balance between the two. ROC-AUC was used to eval- uate the quality of ranking across thresholds, and PR-AUC was included as it is more informative than ROC-AUC in highly imbalanced settings. Confusion matrices were also analyzed to better understand the practical tradeoff between missed failures and false alarms.

The SECOM experiment was used to validate the predictive layer of the framework under realistic indus- trial class imbalance. The benchmark does not replicate the full hearing aid manufacturing environment, but does provide a credible test of whether structured manufacturing data can be used to support early- risk classification before final outcome is known.

7. Results and Discussion

The experimental results demonstrate the viability of the prediction layer of the framework, but they also make it clear why execution policy is important. The SECOM benchmark is very unbalanced, with only 104 positive instances among 1567 total records, which corresponds to a positive-class rate of 6.64%. In a dataset like this, a model may be able to achieve superficially strong accuracy, but it may still miss out on a large proportion of the cases that are operationally important. That is precisely why the evaluation here focuses on recall, *F* 1-score, ROC-AUC and PR-AUC and not just on accuracy.

Table 2 summarizes the test-set performance of two models that were used in the study. Logistic regression achieved an accuracy of 0.8852, and its recall was low at 0.2308 and its F1-score was low at 0.2034. Random forest gave less accuracy of 0.8571 but improved recall significantly to 0.4615 and improved *F* 1- score to 0.2824. The same pattern is observed in the ranking metrics. Random forest yielded a ROC-AUC value of 0.7854 and a PR-AUC value of 0.1980, whereas logistic regression yielded a ROC-AUC value of 0.6929 and a PR-AUC value of 0.1415. In practical terms, the ensemble model was able to capture a much greater proportion of problematic manufacturing instances, even at the expense of some false positives.

Table 1. Summary of the SECOM benchmark dataset used for predictive-layer validation.

A. Dataset Property	*Value*
Total instances	1567
Total features	590
Positive / fail instances	104
Negative / pass instances	1463
Positive class rate	6.64% Learning setting
	Binary classification
Application role	Public manufacturing benchmark

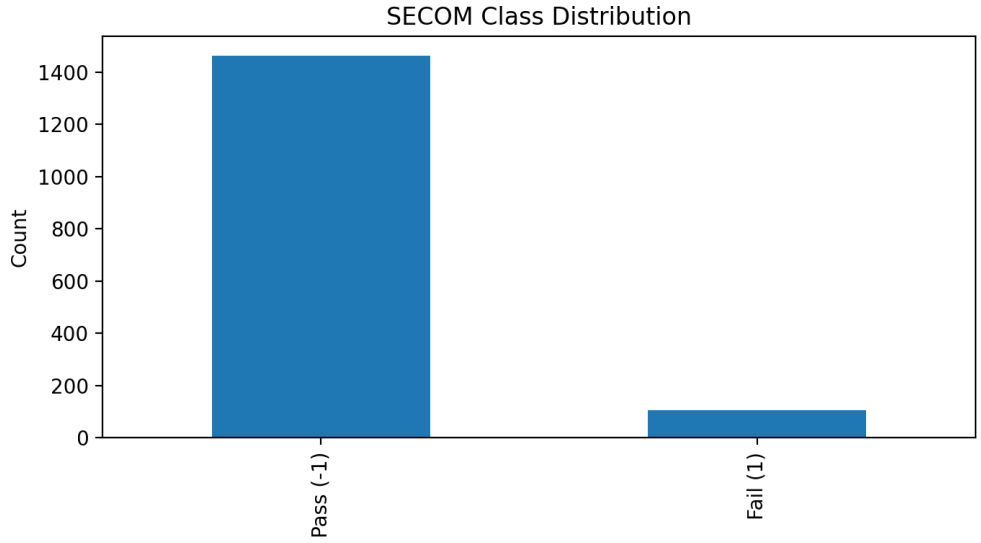

Fig. 2. Class distribution in the SECOM benchmark dataset. The minority positive class represents only 6.64% of all instances, which makes the prediction task strongly imbalanced.

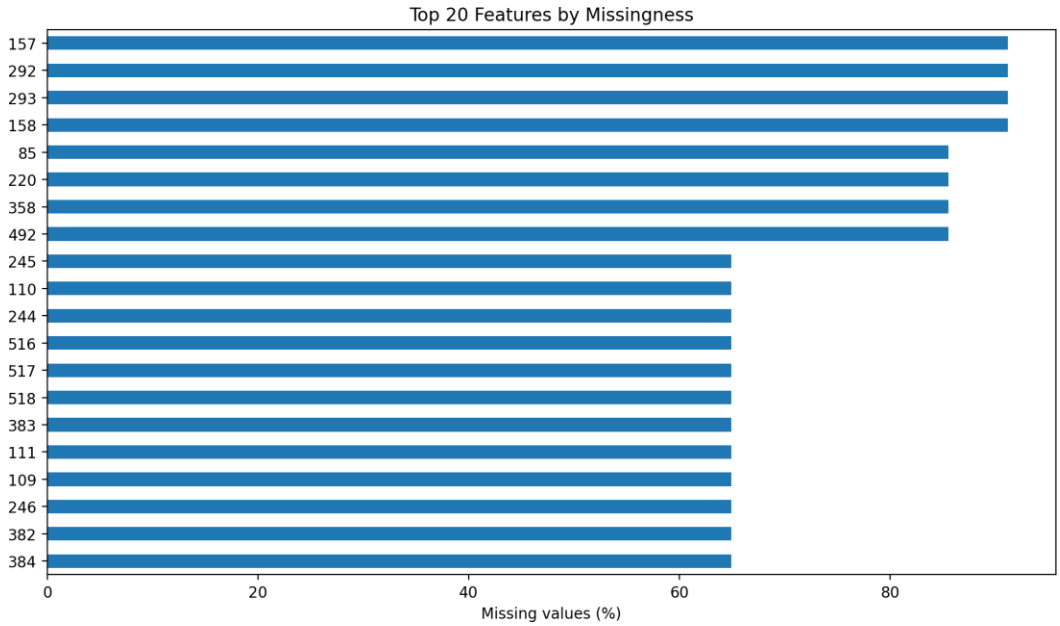

Fig. 3. Top 20 SECOM features by percentage of missing values. The dataset contains substantial incompleteness, which motivated the use of median imputation during preprocessing.

Table 2. Predictive performance on the SECOM benchmark dataset.

Metric	Logistic Regression	Random Forest
Accuracy	0.8852	0.8571
Precision	0.1818	0.2034
Recall	0.2308	0.4615
F_1-score	0.2034	0.2824
ROC-AUC	0.6929	0.7854
PR-AUC	0.1415	0.1980

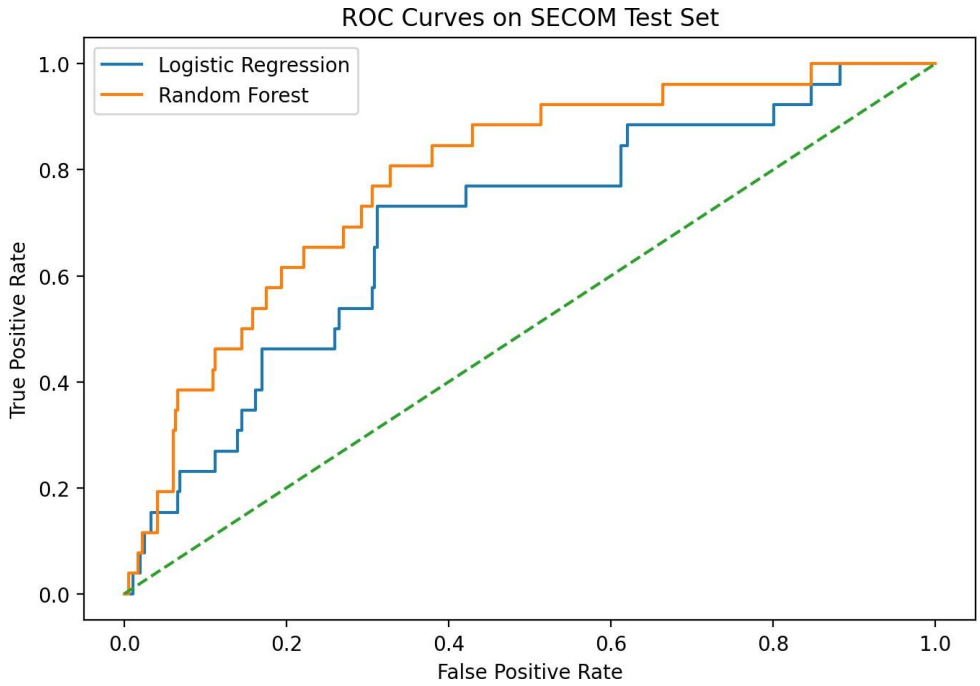

Fig. 4. ROC curves for logistic regression and random forest on the SECOM test set. Random forest achieved the stronger ranking performance, with a ROC-AUC of 0.7854 compared with 0.6929 for logistic regression.

From a traditional classification point of view, one might be tempted to prefer the model with the cleaner accuracy figure. In this setting that would be the wrong reading. The architecture described above is designed to support early intervention, not to maximize the number of passive "correct" predictions in an imbalanced situation. In an execution environment, the cost of missing a high-risk order is usually more expensive than reviewing a moderate number of additional orders. For that reason, the recall gain offered by random forest is more meaningful than the slight decrease in overall accuracy. The model was able to retrieve almost twice as many positive cases as logistic regression, which is more consistent with verification paths and expert-control routing.

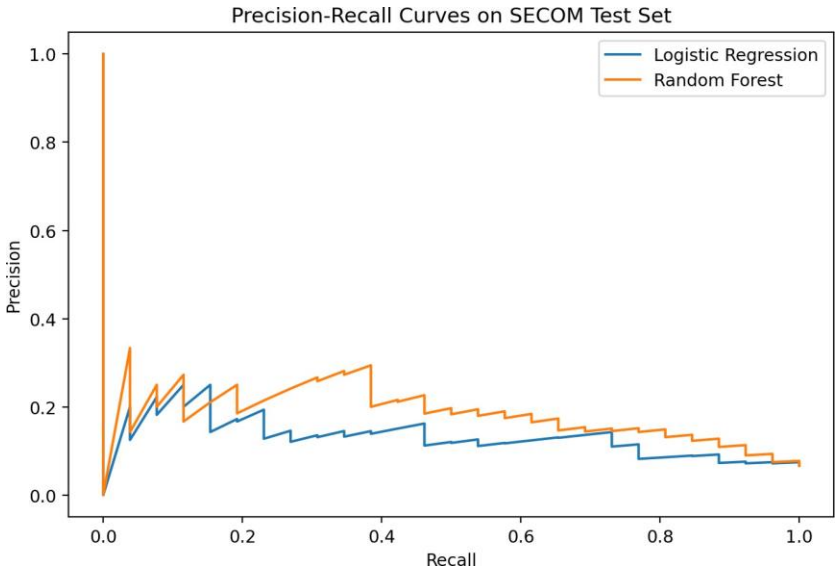

Fig. 5. Precision-recall curves on the SECOM test set. Because the positive class is rare, PR behavior provides a stricter view of model usefulness than accuracy alone.

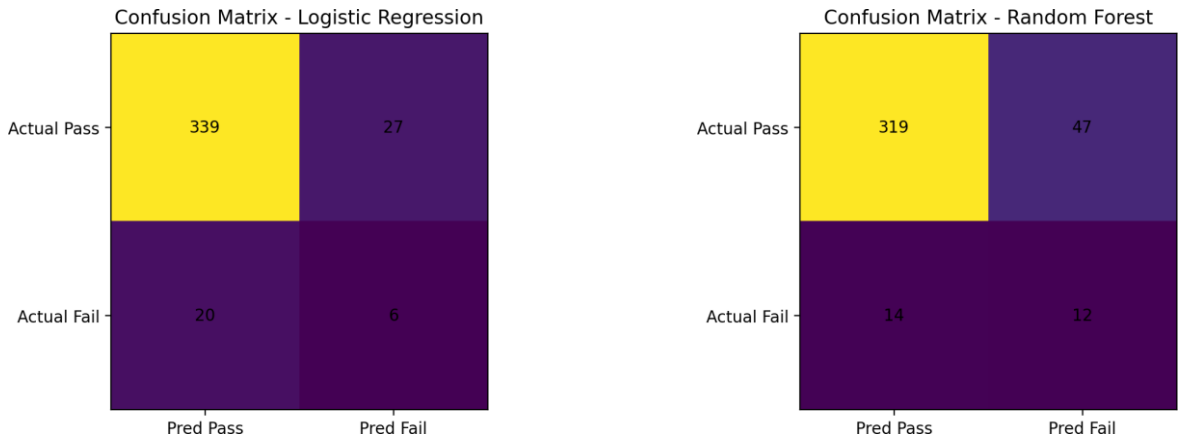

Fig. 6. Confusion matrices for the two evaluated models on the SECOM test set. The random forest model captured a larger number of positive cases than logistic regression, which is more consistent with the intervention- oriented role of the prediction layer.

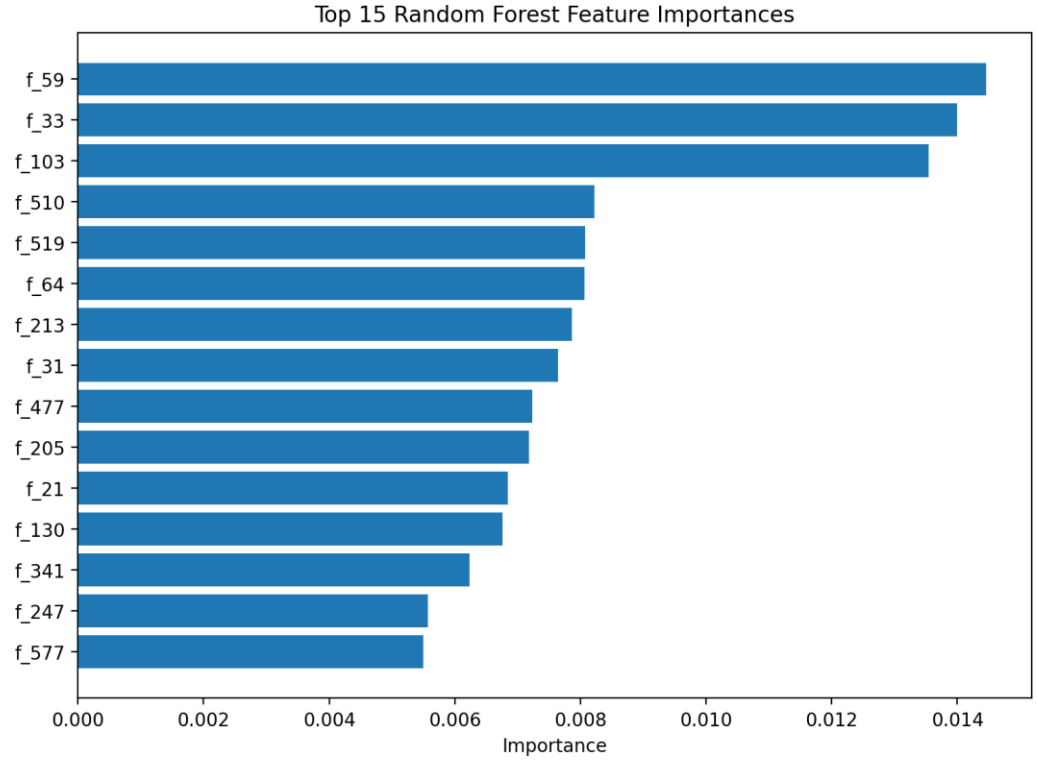

Fig. 7. Top feature importances from the random forest model. The importance distribution indicates that the predictive signal is spread across multiple process variables rather than concentrated in a single dominant feature.

The precision values are still modest for both models, which is not surprising considering the imbalance and complexity of the manufacturing signal. A precision of 0.2034 for random forest means that many of the flagged cases would not result in actual failure or rework. In the case of a pure accept-reject pipeline that is automated, that would be a serious limitation. In the current framework, however, the intervention policy is intentionally less firm than a final classification decision. Medium- and high-risk assignments are not meant to throw the order out the window, but to make the reviewing more intense. Under that interpretation, false positives are operationally tolerable up to a point, especially if the alternative is to let really difficult cases pass unexamined into costly downstream stages.

The difference between ROC-AUC and PR-AUC is also worth a mention. The ROC-AUC values indicate reasonably good ranking performance, in particular random forest. The PR-AUC values, however, are still low in absolute terms as the positive class is rare and difficult to isolate cleanly. That is a more honest reflection on the challenge. In the case of heavily imbalanced manufacturing data, PR-AUC is a more stringent measure of model usefulness because it measures how well the system concentrates true positives among flagged cases. The observed values indicate that the dataset is indeed containing some meaningful predictive signal, but not enough to be overconfident about automating it. That finding supports the architecture developed in this paper. A guarded, execution level intervention policy is more realistic than a brittle all or nothing classification rule.

The results also strengthen the importance of threshold tuning. Using a default threshold in a rare event manufacturing setting would have biased both models to conservative majority class behavior. Tuning on the validation subset resulted in an improvement of the balance between precision and recall and a more appropriate form of resulting decision rules for operational use. That is important because the execution framework is based on a stable mapping between risk estimate and handling policy. If the model is too conservative then high-risk cases get through the normal path. If it is too aggressive, review resources are consumed indiscriminately. Threshold selection therefore becomes part of execution design, not just a modeling detail.

From the point of view of the architecture proposed, the random forest model is the better candidate for the risk estimation layer. It is not a solution to the problem in itself, and does not eliminate the need for human judgment. What it does give is a more useful ordering of manufacturing risk, one that can support differentiated handling. Low-risk cases can go down the normal path with little friction. Medium-risk cases can be referred to parameter verification or controlled handling. High-risk cases can be escalated to expert review or alternate routing. The results of the model make more sense to justify that kind of selective intervention than to justify any kind of autonomous pass-fail decision.

Overall, the findings offer support for three practical conclusions. First, manufacturing risk prediction is possible on structured, imbalanced production data, which validates the analysis core of the architecture. Second, non-linear models are more suitable than the linear baseline to account for the interactions that motivate problematic cases in complex production settings. Third, the true value of prediction only comes when it is integrated into a control policy that converts risk into differentiated handling. That last point brings us back to the main point of the paper: Prediction is not sufficient. What is important is whether the factory will be able to use predicted risk early enough to minimise downstream correction, preserve scarce expert time and prevent difficult orders destabilising the line.

8. Threats to Validity

Several limitations should be kept explicit in the interpretation of the present results. The first is the validity of the dataset. The experimental evaluation was carried out according to the SECOM man- ufacturing benchmark and not on the production data from the manufacture of personalized hearing aids. SECOM is useful in that it contains high-dimensional measurements of industrial processes and a highly imbalanced distribution of outcomes making it suitable for validating the prediction layer in a realistic manufacturing environment. However, it lacks hearing aid specific order attributes, anatomical complexity indicators, intervention traces, and routing outcomes. As a result, the benchmark is helping to support the technical credibility of the risk estimation component, but it is not entirely replicating the operational environment in which the proposed execution framework would ultimately be deployed.

The second limitation is in terms of *construct validity*. In the paper, rework is defined as corrective manufacturing activity that causes an order to deviate from its intended path. That interpretation is central to the execution argument, but the SECOM dataset has pass-fail labels instead of explicit rework events. The experimental study therefore uses manufacturing failure as a proxy for problematic production outcomes which would warrant early intervention. This is a defensible approximation for validating the predictive architecture, but it is still a proxy. A deployment grade evaluation would require a more direct label based on actual hearing aid production records such as repeat handling, return to an earlier stage, controlled hold, remake, or other forms of corrective effort.

A third limitation is that of *feature correspondence*. The proposed framework is based on a personalized hearing aid context where order behavior may be dependent on shell style, vent configuration, scan quality, geometry burden and other domain specific variables. None of these are made explicit in SECOM. Instead, the benchmark offers generic manufacturing features whose semantic interpretation is related to another industrial process. That means that the experiment is validating the feasibility of early manufacturing risk inference but not the exact domain structure of the final application. A future study based on real hearing aid production records would mean that the feature engineering stage could be more in line with real manufacturing practice.

There is also a limitation of *policy validity*. The intervention logic described in the architecture is oper- ationally

plausible, but has not yet been tried on a live manufacturing line. The present study demon- strates the feasibility of developing risk scores and that they separate problematic cases better under the stronger model than under the linear baseline. What it does not yet show is the full downstream effect of those scores when they are connected to review capacity, release control and alternative routing in a real production environment. In other words, the predictive layer is supported by empirical data while the execution layer still needs to be validated at the field level.

Another problem to be worried about is *class imbalance sensitivity*. The positive class in SECOM is rare, which makes the experiment realistic but also makes the metrics obtained sensitive to threshold selection and data partitioning. Threshold tuning has been done carefully using validation split, but the final operating point still represents a certain trade-off between recall and false positive. In a deployment setting, that trade-off would have to be re-evaluated in relation to actual intervention cost, expert ca- pacity, and the relative burden of missed high-risk orders. The chosen threshold is therefore appropriate for the validation of the method, but not necessarily final for operational use.

Finally, there is a *generalization limitation*. Personalized hearing aid manufacturing is a high-mix, low- volume environment with a high degree of dependency on case-specific variability and expert judgment. Even with a better domain-specific dataset, the resulting model would be likely to be sensitive to factory practices, labeling conventions and local process design. The framework is designed to be transferable on the architectural level, but not to be assumed to be generalisable perfectly across organisations without recalibration. That is typical in manufacturing analytics and should not be considered as a flaw that is unique to this study. It just means that actual deployment would require local validation, feature adaptation and policy tuning before the framework could be trusted as an execution tool.

These limitations do not affect the main contribution of the paper, but they do clarify the current scope of the paper. The current research is a validation of the predictive logic of the proposed framework on a public industrial dataset and an argument for its application inside an execution-oriented control archi- tecture. The next step is a deployment-specific evaluation based on historical hearing aid manufacturing records where both the rework label and the intervention pathway can be measured directly.

9. Conclusion

Rework in personalized hearing aid manufacturing is more of an execution problem rather than a retro- spective quality outcome. The main difficulty is not that some orders require extra handling, but that they are often recognised too late, after the line has already absorbed avoidable effort in the form of delay, repeat handling and specialist intervention.

The framework presented here addresses that gap by integrating early risk estimation within a closed manufacturing loop of order-context acquisition, risk profiling, intervention selection, execution routing and feedback-driven refinement. Under this view, prediction is useful only when it makes a difference in the way the order is handled before disruption downstream becomes expensive.

The results of the SECOM benchmark study support the feasibility of the predictive layer under real- istic manufacturing imbalance. Using 1,567 industrial instances with 590 process-related features and a positive outcome rate of 6.64%, the study proved that structured manufacturing data can be used for early-risk classification. Logistic regression gave a transparent baseline and random forest gave a better overall performance with recall of 0.4615, F1 score of 0.2824, ROC-AUC of 0.7854 and PR-AUC of 0.1980. These results do not support autonomous decision-making, but they do support the use of differentiated handling policies such as verification paths, expert-control routes and controlled release.

The contributions of the paper are on two levels. Conceptually, it is an early rework prevention in personalized hearing aid manufacturing as a risk aware execution problem. Empirically it shows that the predictive essence of that framework is technically plausible on a public industrial benchmark. The next step is simple, to test the whole architecture on real hearing aid production data so that the intervention pathway, not just the prediction layer, can be measured directly.

Statements and Declarations

Funding. This research was not funded by any external source.

Conflicts of Interest. The author declares no conflict of interest.

Acknowledgments. No other acknowledgments are reported.

Data Availability Statement. The public benchmark data set used in the experimental study is the SECOM manufacturing data set. The target deployment context described in the paper is that of personalized hearing aid manufacturing, but no proprietary hearing aid production dataset was used in the present benchmark evaluation.

References

[1]. Attaran, M., Celik, B., Bhardwaj, S.: Digital twin: Benefits, use cases, challenges, and opportunities. Decision Analytics Journal **6**, 100165 (2023). https://doi.org/10.1016/j.dajour.2023.100165, https:
[2]. //www.sciencedirect.com/science/article/pii/S277266222300005X
[3]. Auriemma, G., Tommasino, C., Falcone, G., Esposito, T., Sardo, C., Aquino, R.P.: Additive manufacturing strategies for personalized drug delivery systems and medical devices: Fused filament fabrication and semi solid extrusion. Molecules **27**(9), 2784 (2022). https://doi.org/10.3390/molecules27092784
[4]. Cortez, R., Dinulescu, N., Skafte, K., Olson, B., Keenan, D., Kuk, F.: Changing with the times: Applying digital technology to hearing aid shell manufacturing. The Hearing Re- view (Mar 2004), https://hearingreview.com/hearing-products/accessories/components/ changing-with-the-times-applying-digital-technology-to-hearing-aid-shell-manufacturing, accessed 2026-03-14
[5]. Gonzalez, L.F.A., et al.: Predictive modeling for quality prediction in multi-stage manufacturing systems: A review. Smart Manufacturing **3**, 100097 (2025). https://doi.org/10.1016/j.smman.2025.100097, https:
[6]. //www.sciencedirect.com/science/article/pii/S2666827025001379
[7]. Inayathullah, S., et al.: Review of machine learning applications in additive manufacturing. Re- sults in Engineering **25**, 103756 (2025). https://doi.org/10.1016/j.rineng.2024.103756, https://www. sciencedirect.com/science/article/pii/S2590123024019194
[8]. Kang, Z., Catal, C., Tekinerdogan, B.: Machine learning applications in production lines: A systematic literature review. Computers & Industrial Engineering **149**, 106773 (2020). https://doi.org/10.1016/j. cie.2020.106773
[9]. Kausik, A.K., Rashid, A.B., Baki, R.F., Maktum, M.M.J.: Machine learning algorithms for man- ufacturing quality assurance: A systematic review of performance metrics and applications. Array **26**, 100393 (2025). https://doi.org/10.1016/j.array.2025.100393, https://www.sciencedirect.com/ science/article/pii/S2590005625000207
[10]. Kumar, S., et al.: Machine learning techniques in additive manufacturing. Journal of Intelligent Manufacturing **34**(6), 2161–2189 (2023). https://doi.org/10.1007/s10845-022-02029-5, https://link.springer.com/ article/10.1007/s10845-022-02029-5
[11]. Martinez-Marquez, D., Jokymaityte, M., Mirnajafizadeh, A., Carty, C.P., Lloyd, D., Stewart, R.A.: Develop- ment of 18 quality control gates for additive manufacturing of error free patient-specific implants. Materials **12**(19), 3110 (2019). https://doi.org/10.3390/ma12193110, https://www.mdpi.com/1996-1944/12/19/
[12]. 3110
[13]. Pirzanski, C.: Earmolds and hearing aid shells: A tutorial. The Hearing Re- view (2006), https://hearingreview.com/hearing-products/accessories/earmolds/
[14]. earmolds-and-hearing-aid-shells-a-tutorial, accessed 2026-03-14
[15]. Pirzanski, C.: A new solution for automated 3d virtual modeling of custom hearing aid shells. The Hearing Review (Feb 2020), https://hearingreview.com/hearing-products/accessories/earmolds/ new-solution-for-automated-3d-modeling-of-custom-hearing-aid-shells, accessed 2026-03-14
[16]. Tognola, G., Parazzini, M., Svelto, C., Galli, M., Ravazzani, P.: Design of hearing aid shells by three- dimensional laser scanning and mesh reconstruction. Journal of Biomedical Optics **9**(4), 835–843 (2004). https://doi.org/10.1117/1.1756595