International Journal of Information Management Data Insights 6 (2026) 100417

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

FashionDSS: A Human-in-the-Loop Decision Support System for forecasting production efficiency and predicting item defects in the luxury fashion industry

Giovanna Migliorelli [a], Rocco Pietrini [b], Alessandro Galdelli [c], Emanuele Frontoni [d], Marina Paolanti [d,*]

[a] Department of Law, University of Macerata, Piaggia dell'Università 2, 62100, Macerata, Italy
[b] Department of Engineering and Science, Universitas Mercatorum, Piazza Mattei 10, Rome, 00186, Italy
[c] Department of Information Engineering (DII), Università Politecnica delle Marche, Via Brecce Bianche 12, 60131, Ancona, Italy
[d] Department of Political Sciences, Communication and International Relations (SPOCRI), University of Macerata, Via Don Minzoni 22A, 62100, Macerata, Italy

ARTICLE INFO

Keywords:
Decision support system
Defects prediction
Luxury fashion industry
Artificial Intelligence
Regression
Classification

ABSTRACT

In the luxury fashion industry, where quality directly impacts brand reputation and customer satisfaction, detecting product defects is crucial. This paper introduces FashionDSS: a Human-in-the-Loop Decision Support System designed to facilitate proactive quality management during the design phase. The system integrates three structured data sources, Product Registry, Bill of Materials (BOM) and Claim Records, into a unified predictive framework. Using machine learning models that are tailored to moderate-sized, information-rich datasets and validated through expert interaction, FashionDSS can perform two tasks: forecasting prototype production time (Task T1) and predicting potential defects in newly designed products (Task T2). Experimental results on real-world luxury fashion data demonstrate the feasibility of extracting meaningful predictive signals from structured complaint and production information despite data constraints. Although validation is currently limited to a single company's dataset, the findings emphasise the potential benefits of combining predictive analytics with human expertise for design-oriented quality management.

1. Introduction

The fashion industry, which includes clothing, footwear, makeup and accessories, is estimated to be worth over 3 trillion US dollars.[1] The digitalisation of the fashion retail supply chain and evolving consumer behaviour have significantly accelerated the growth of fashion e-commerce (Gazzola et al., 2024; Jing et al., 2026). The e-commerce fashion sector is expected to grow at a compound annual growth rate (CAGR) of 14.2% from 2017 to 2025, reaching a valuation of $1 trillion by 2024 (Pavan & Samant, 2024). In the United States alone, online fashion sales have reached $1204.9 billion and continue to grow.[2] The luxury fashion industry occupies a distinctive position within this broader ecosystem. It sets trends, drives creative innovation and combines tradition with modernity to create highly appealing products (Pereira et al., 2022; Tam & Lung, 2025). IIn this context, quality is a strategic imperative that directly influences brand perception, customer experience, and long-term competitiveness (Chen et al., 2022). Despite rigorous craftsmanship and quality control practices (Chakraborty et al., 2022), product defects may still occur, leading to complaints and customer dissatisfaction (Lysenko-Ryba & Zimon, 2021). A defect in a fashion product is any irregularity that falls short of the expected quality, design or performance standards (Eldessouki, 2018). Such defects may include stitching errors, faulty closures, inconsistent colouring, incorrect sizing, or discrepancies in the material used. These defects can be detected before retail, during internal quality control or after purchase by the customer. While traditional quality management relies on inspection and post-production control, the increasing availability of data and predictive analytics enables us to anticipate defects earlier in the product lifecycle (Alkahtani et al., 2019; Zajec et al., 2024). Although predictive approaches have been

* Correspondence to: Department of Political Sciences, Communication and International Relations (SPOCRI), University of Macerata, Via Don Minzoni 22A, 62100, Macerata, Italy
E-mail addresses: g.migliorelli3@unimc.it (G. Migliorelli), rocco.pietrini@unimercatorum.it (R. Pietrini), a.galdelli@staff.univpm.it (A. Galdelli), emanuele.frontoni@unimc.it (E. Frontoni), marina.paolanti@unimc.it (M. Paolanti).

[1] https://fashionunited.com/global-fashion-industry-statistics/
[2] https://www.shopify.com/enterprise/ecommerce-fashion-industry

https://doi.org/10.1016/j.jjimei.2026.100417

Available online 25 May 2026

widely adopted in manufacturing contexts, a significant gap remains in the luxury fashion domain. Existing systems largely operate in a reactive mode, addressing issues after they occur rather than supporting proactive decision-making during design and production planning (Pereira et al., 2022). Decision Support Systems (DSS) enriched by Artificial Intelligence (AI) and customer modelling techniques offer promising opportunities in this direction (Barrera et al., 2024; Caro & de Tejada Cuenca, 2023; Oskouei et al., 2023; Wang et al., 2024). When combined with human expertise, such systems enable context-aware interpretation, iterative refinement and more reliable decision-making. Human-in-the-loop approaches allow designers, quality managers and production planners to validate predictions, adjust inputs and improve model performance while preserving domain knowledge and practical relevance (Loureiro et al., 2018). A data-driven approach applied during the early stages of the product lifecycle, particularly in the design phase, can significantly enhance quality management (Filz et al., 2024; Wang & Liu, 2021). By leveraging historical information from similar products, predictive systems can estimate defect probabilities before production begins (Nikseresht et al., 2024; Silva et al., 2024). Similar approaches have been successfully adopted in other industries, such as automotive manufacturing (Babakmehr et al., 2024; Rasheed et al., 2024), yet their application to luxury fashion remains limited, especially in integrating production efficiency forecasting with defect prediction within a unified framework. To address this gap, this paper proposes FashionDSS, a human-in-the-loop decision support system for the luxury fashion industry. FashionDSS integrates AI-based predictive modelling with expert validation to support upstream decision-making. Its objective is twofold: to forecast the production efficiency of new prototypes based on material composition, operational requirements and production complexity (Task T1); to predict the most probable defects of newly designed products before manufacturing begins (Task T2). The system leverages a unique dataset combining three complementary information sources: the Product Registry, the Bill of Materials (BOM) and the Claim Record. This dataset captures detailed product characteristics, material quantities and functions, manufacturing operations and times, as well as structured defect reports submitted by retailers or customers. Acquiring such granular and sensitive information in the luxury sector is particularly challenging due to confidentiality constraints and limited production volumes. Although the dataset includes 600 products and 1800 defect claims, its richness and multidimensional structure provide a rare and valuable foundation for predictive modelling. By transforming historical complaint data into actionable predictive knowledge, FashionDSS shifts quality management from a reactive to a proactive paradigm. It enables designers and production planners to adjust materials, processes and resource allocation before production starts, thereby reducing defect risks, limiting rework and improving operational efficiency. Crucially, the human-in-the-loop design ensures transparency, iterative refinement and contextual validation of predictions. While developed for the luxury fashion sector, the proposed methodology has broader applicability across retail domains characterised by seasonal product turnover, absence of historical time series for new items, and the availability of structured product registries. By integrating heterogeneous data sources into a unified predictive framework, this work contributes to the literature on information management and data-driven decision support. The empirical investigation is guided by the following research questions:

- **RQ1**: *To what extent can the production time of newly designed luxury fashion products be forecast using structured product and bill-of-material data prior to manufacturing?*
- **RQ2**: *To what extent can historical complaint and production data be used to predict the most probable defect characteristics of newly designed products during the design phase?*

This paper makes several key contributions. It introduces an innovative approach that considers customer complaint data to be a strategic, predictive resource. It also develops and validates an integrated analytical framework that combines product, material and defect information. Furthermore, it implements AI-based models to forecast production efficiency (Task T1) and predict defects (Task T2) before manufacturing begins. Additionally, it introduces a unique, high-value dataset specific to the luxury fashion domain. Finally, it demonstrates the system's potential applicability beyond fashion retail.

The paper is structured as follows: Section 2 examines existing literature and research on the use of data analytics in the fashion industry, with a particular focus on predictive modelling for quality assurance and operational efficiency. Section 3 provides a comprehensive overview of the research design, data processing and analysis techniques that are the basis of our study, ensuring replicability and transparency. In Section 4, we present the results of our research and demonstrate the predictive accuracy of our models in terms of process time estimation and defect prediction. Finally, Section 5 concludes by summarising the key contributions and findings, reiterating the importance of our innovative approach to using customer complaint data for predictive analytics. This section also outlines directions for future research, suggesting how the methods and findings of this study can be further refined and applied to other areas of the retail industry. Potential areas for further development, including the application of more advanced analytical techniques and the exploration of additional data sources, are identified to encourage continued innovation and research in this area.

2. State of the art

This section reviews existing research on predictive quality management and defect prevention. The discussion is structured around three main areas of literature: predictive analytics in manufacturing, Zero-Defect Manufacturing (ZDM) frameworks and human-centred design in fashion production. This approach clarifies remaining limitations and motivates the development of FashionDSS.

Traditional quality control systems are predominantly reactive, addressing defects only after they have been detected post-production or through customer feedback. While these approaches are still necessary, they cannot prevent defective products from reaching consumers, which can result in returns, dissatisfaction, and reputational damage. In response to these limitations, predictive analytics has been widely adopted in manufacturing-intensive industries such as electronics and automotive (Bertolini et al., 2021; Theissler et al., 2021). These systems rely on statistical learning and machine learning models trained on historical operational data to anticipate equipment failures or product defects.

In the automotive domain, for example, the analysis of large-scale claim datasets has demonstrated strong predictive potential. In Lee et al. (2021), 305.965 claims collected over 6 to 14 years were used to train deep learning architectures, including 1D convolutional neural networks, recurrent neural networks (RNNs) and sequence-to-sequence (Seq2Seq) models, to predict component failures and reliability trends. Such approaches illustrate the power of predictive modelling in high-volume industrial contexts. However, these systems typically rely on extensive longitudinal datasets and standardised production environments, conditions that are not easily replicated in luxury fashion, where production volumes are limited and variability is high.

A second line of research focuses on ZDM, a paradigm that aims to eliminate defects through preventive, predictive and corrective strategies. Chen (2016). Originally introduced in 1965 within the US Army's Perishing Missile System (Fouch, 1965), ZDM has evolved into a comprehensive quality framework. As described in Psarommatis et al. (2020), ZDM can be implemented from both a product-oriented and a process-oriented perspective, yet both approaches converge into a unified operational structure. The framework encompasses four core strategies—Detect, Predict, Prevent and Repair (May & Kiritsis, 2019; Psarommatis et al., 2020). While ZDM introduces predictive elements, its implementation is primarily oriented towards equipment monitoring

and process anomaly detection in structured manufacturing systems. In low-volume, highly customised production environments, predictive modelling becomes more complex due to limited historical data and greater contextual variability. For instance, Verna et al. (2023) propose a diagnostic tool linking defect rates to workstation complexity, yet this approach remains focused on process decomposition rather than design-stage intervention. Furthermore, the human-centric dimension of ZDM has often been underexplored, with roles and responsibilities of managers, engineers and operators remaining insufficiently defined in predictive workflows (Wan & Leirmo, 2023).

A third area of literature emphasises the important role of designers in human-centred production within the fashion industry (Thiel & Postlethwaite, 2023). Designers must balance creativity, functionality, and user experience (Hassenzahl et al., 2021; Hisrich & Soltanifar, 2021), influencing material selection, assembly choices and usability outcomes. Their early-stage decisions significantly affect product reliability and defect likelihood (Murzyn-Kupisz & Hołuj, 2021; Sun & Zhao, 2018). However, despite their strategic importance, designers are rarely supported by structured predictive tools capable of quantifying defect risks before production begins. Taken together, the existing literature reveals three main limitations. First, predictive quality systems are predominantly developed for high-volume industries with abundant sensor or operational data. Second, ZDM frameworks focus more on process monitoring than on design-stage risk estimation. Third, the integration of structured customer complaint data into upstream predictive decision-making remains largely unexplored, particularly in luxury fashion.

In this context, this paper makes a significant contribution to the field by proposing a comprehensive predictive framework that can estimate the location and severity of potential defects in luxury fashion products before manufacturing begins. Unlike traditional approaches, which operate during or after production, FashionDSS uses structured complaint records, product registry information, and BOM data to generate probabilistic defect risk assessments during the design phase. Defects may manifest immediately at the point of sale or emerge after customer use, potentially leading to claims. Our models quantify these risks by assigning probability scores that reflect latent vulnerabilities in materials or processes. By incorporating this predictive capability into a human-in-the-loop decision support system, FashionDSS allows designers and production planners to proactively modify material selections and manufacturing processes before production begins. In doing so, it extends predictive quality management from process-level monitoring to design-oriented risk anticipation, addressing a gap in the existing literature that has not been sufficiently resolved. The integration of complaint-driven analytics, low-volume production modelling, and human-centred decision support constitutes the core conceptual advancement of our work.

3. Materials and methods

In this section, we outline the basic elements and analytical techniques that constitute our FashionDSS. Central to our investigation are two tasks, an extensive data set, and the predictive models we employed to forecast manufacturing times and anticipate product defects in the luxury fashion sector. Each component is critical to the understanding the novel approach our study introduces with the use of customer feedback and manufacturing data for accomplishing quality assurance and improve operational efficiency in fashion retailing. FashionDSS is being deployed in an Italian fashion company of a prestigious luxury fashion brand, embodying a state-of-the-art DSS designed to significantly improve the brand's product development and quality assurance processes. FashionDSS addresses two tasks and we refer to the first task as **T1** and the second task as **T2**. Here is a detailed description for each of them:

- **T1**: For a prototype, i.e. a product that has never been manufactured before, the framework is asked to predict the process time. The process time, a sub-interval of the manufacturing time, is the time required to manufacture the final product from the raw materials.
- **T2**: At some point during its life, any product may prove to be or become defective. This event may occur earlier, after the product has been distributed but before it has been retailed, or later, after the product has been retailed, in which case it is usually the customer who has purchased the product who makes a claim. Here, the framework is asked to predict the most likely defect for a given new product.

From a methodological perspective, the study employs an empirical, case-based research design, which is grounded in real industrial data and structured around supervised learning techniques. Each task is formulated as a predictive modelling problem that aligns with the relevant research question. Task T1 is formulated as a regression problem aimed at forecasting production time based on structured product and bill-of-material features. In contrast, Task T2 is framed as a multi-class classification problem focused on predicting defect-related characteristics from historical complaint and product information. The selection of analytical techniques reflects the tabular nature and moderate size of the dataset, prioritising models that allow for controlled complexity, interpretability and robustness under data constraints.

The importance of **T1** and **T2** cannot be overstated, as an estimate of the process time can help the company to allocate resources carefully, or even to decide whether a product should be adapted to reduce the process time. On the other hand, any foreknowledge of possible defects can help to take prompt corrective actions to improve customer satisfaction.

The details of FashionDSS components are explained in the following subsections to ensure a thorough understanding of our methodology and its application. Fig. 1 schematically depicts the workflow of our FashionDSS.

3.1. Dataset

As shown in Fig. 1(a), the dataset encompasses three main sources of information and it is based on original industrial data that were collected from the internal information systems of a Made in Italy luxury fashion brand. The Product registry, which contains the general characteristics of the product, such as colour, product type, product line; the BOM, which contains a detailed list of all the materials used to manufacture the final product and, for each material, the exact quantity required, the function of the material (inner, outer, reinforcement), and also the list of all the operations performed during the manufacturing process along with the time required to complete each step; the Record of Claims provides detailed information about any product claim, including details of the customer who filed the claim if the defect was discovered post-purchase, or the retailer (or their representative) if the defect was identified pre-retail. Furthermore, it also includes comprehensive information about the type of defect discovered. The number of products in the dataset is 600, while the number of claims is 1800. For training, **T1** relies mainly on information from the BOM available for each of the 600 products, and **T2** relies mainly on information from the Record of claims for each of the 1800 claims. Although the dataset comprises 600 products and 1800 defect claims, its size is inherently limited by the unique and highly specialised content of the information it contains. Table 1 summarises the numerical and categorical features and targets for tasks T1 and T2. Fig. 2 presents the distribution of categories and their corresponding sample counts for each defect-related attribute, clearly illustrating the substantial class imbalance present in the data.

In the luxury fashion industry, access to such detailed and sensitive data is extremely rare. Luxury brands work in competitive environments where manufacturing processes and defect data are closely protected due to concerns over intellectual property and brand reputation.

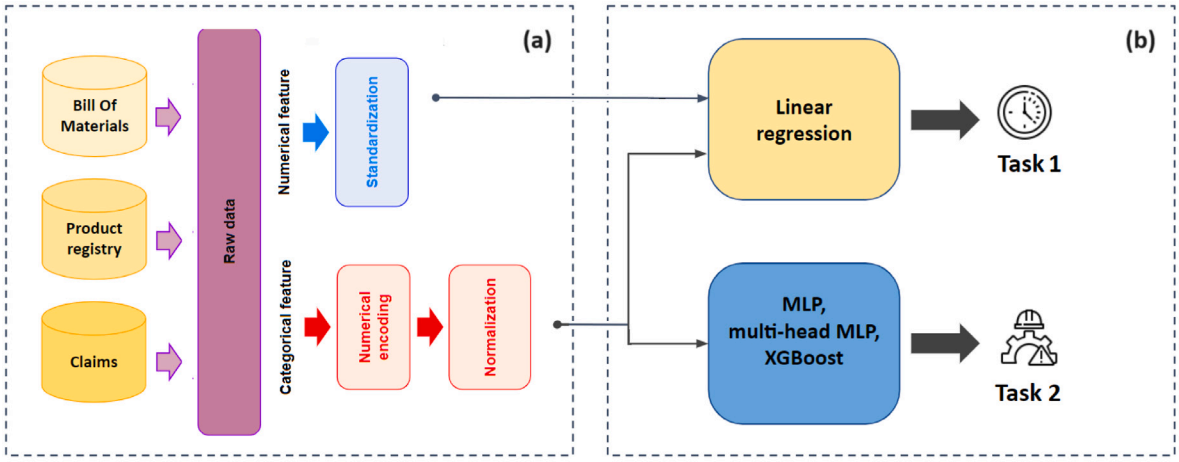
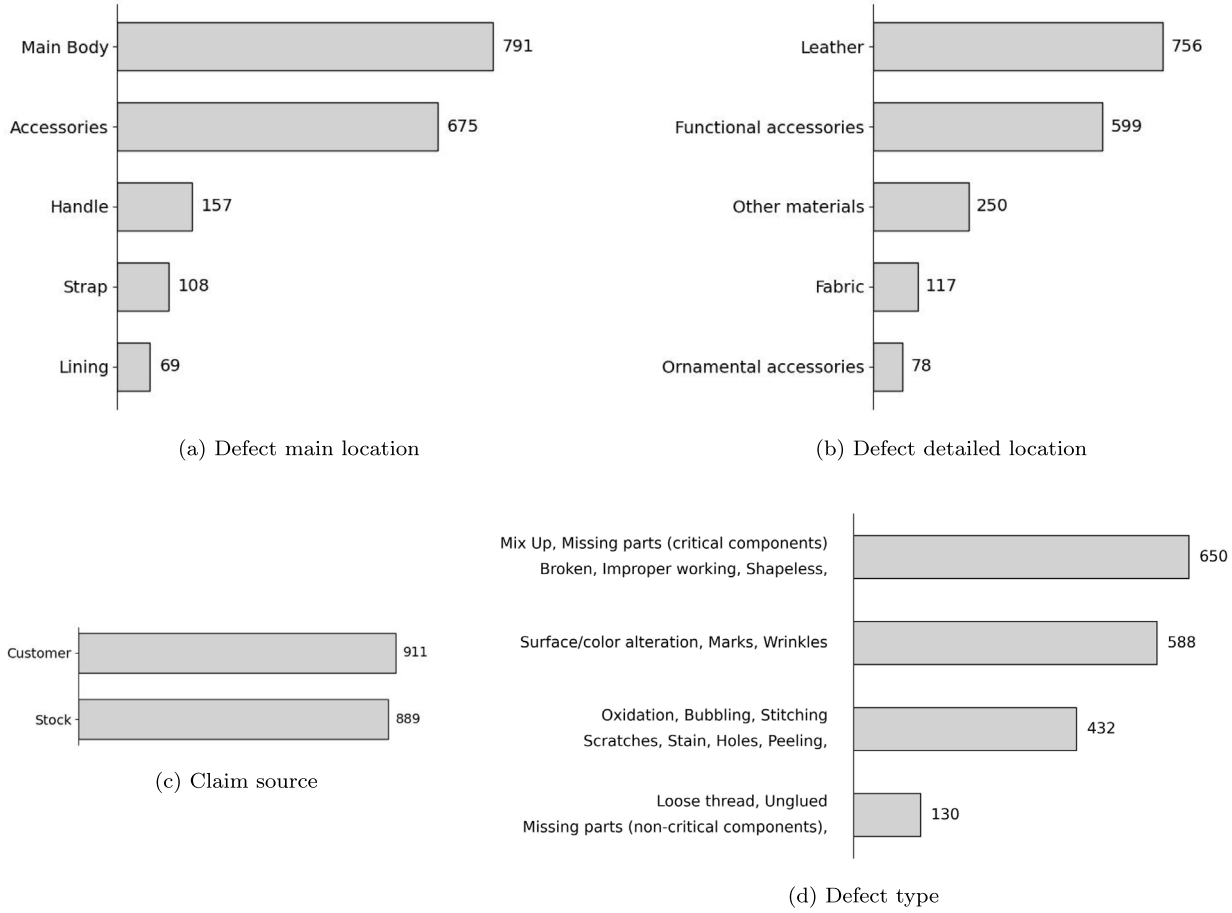


Fig. 1. FashionDSS workflow. (a) Data pre-processing and feature extraction (b) Proposed models to solve task **T1** and task **T2**. All models are trained by exploiting data collected for products already manufactured and commercialised, for which both the complete BOM and the claims record are available.



(a) Defect main location

(b) Defect detailed location

(c) Claim source

(d) Defect type

Fig. 2. Distribution of samples across categories for each categorical target considered in Task **T2**: (a) defect main location, (b) defect detailed location, (c) claim source, and (d) defect type. For each category, the corresponding number of observations in the dataset is reported, highlighting class imbalance across targets.

Gathering this level of granular data — including material specifications, manufacturing workflows and detailed defect reports — is both costly and demanding logistically. Luxury fashion production typically involves limited product quantities, reflecting a business model based on exclusivity and craftsmanship rather than mass production. As a result, production and defect data in this sector is naturally smaller in scale, but much richer in detail. In addition, the complexity of standardising and systematising defect reporting across tailored products further limits the scalability of such datasets. Each product may have unique designs, materials and craftsmanship, making defect classification and reporting inherently complex. However, the depth of information compensates for the limited size of the dataset, providing

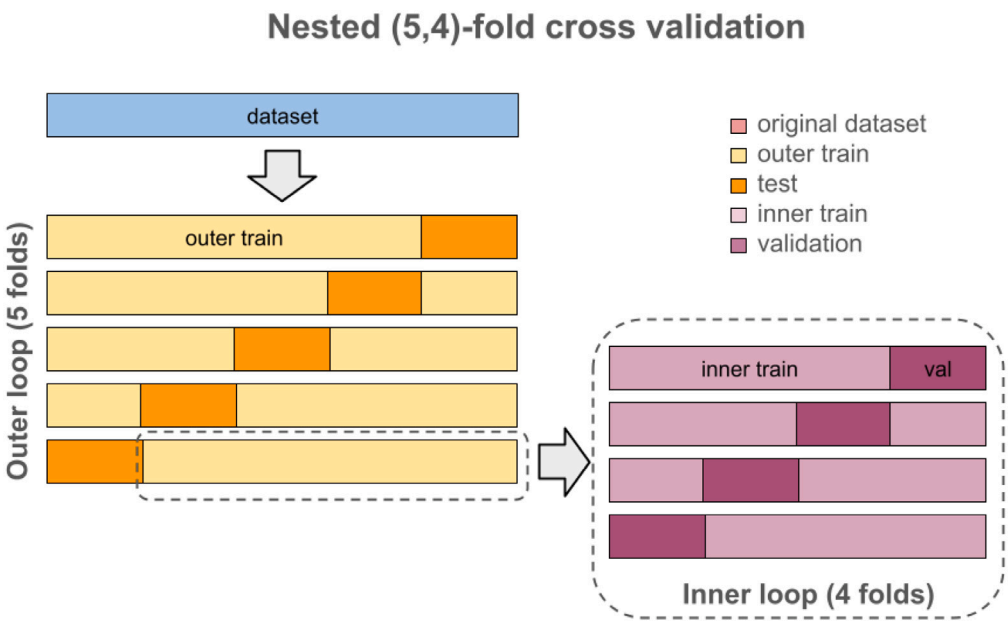


Fig. 3. Nested cross validation is applied with 5 outer folders and 4 inner folders respectively. The inner loop is responsible for hyper-parameter tuning whereas the outer loop is used to evaluate each model mitigating the risk of overconfidence.

Table 1
Overview of input features and target variables used in the two predictive tasks addressed by FashionDSS. For each task, the table reports the feature names and their type — categorical (c) or numerical (n) — as well as the corresponding target variables to be predicted.

Task	Features		Targets	
	Name	Type	Name	Type
T1	*collection*	(c)	*process time*	(n)
	product category	(c)		
	$material_1$	(n)		
	$material_2$	(n)		
	$material_7$	(n)		
	combinations	(n)		
	difficulty	(n)		
T2	*collection*	(c)	*defect main location*	(c)
	product category	(c)	*defect detailed location*	(c)
	product line	(c)	*claim origin*	(c)
	lifespan	(c)	*defect type*	(c)
	launch season	(c)		
	$material_1$	(n)		
	$material_2$	(n)		
	$material_7$	(n)		

a high-quality, multi-dimensional resource that enables sophisticated analysis of production processes and product performance.

3.2. Dataset split during training and evaluation

Dataset was split in training, validation and test sets in the ratio 80:10:10, whereas for **T2** a more complex approach is applied, as shown in Fig. 3. In this case, nested cross-validation is used, which has a twofold advantage: on the one hand, the final evaluation by the 5-fold outer splits helps to reduce the risk of overestimating performance, and on the other hand, the 4-fold inner loops allow a grid search through the hyperparameter space. The split approach is stratified in that for each class it retains the same proportion of samples found in the original dataset. For the sake of reproducibility and fair comparison between different models, both outer and inner splits are computed and stored beforehand.

It is important to consider the size of the dataset in relation to the structural characteristics of the luxury fashion industry. Unlike high-volume manufacturing sectors, the luxury fashion industry is characterised by limited product quantities, strong seasonality, high design variability, and strict confidentiality constraints. Therefore, large-scale historical datasets comparable to those in the automotive or electronics industries are unavailable. From a methodological perspective, this constraint directly informed our modelling strategy. Rather than adopting the deep learning architectures typically required for large amounts of data, we deliberately selected models well-suited to structured, medium-sized tabular datasets that incorporate implicit or explicit regularisation mechanisms. In addition, strict data separation protocols were enforced to mitigate the risk of overfitting. For Task T1, a train–validation–test split of 80:10:10 ensures evaluation on unseen data. For Task T2, nested cross-validation decouples hyperparameter tuning from performance estimation to reduce optimistic bias. Feature dimensionality was carefully controlled through frequency encoding and domain-driven feature selection, and low-level operational variables were excluded to avoid sparsity issues. These methodological choices reflect an explicit bias–variance trade-off tailored to a limited yet information-rich dataset, prioritising robustness, interpretability, and generalisability under realistic industry constraints.

3.3. Data pre-processing, feature selection and target definition

With reference to Fig. 1, data is parsed from the product registry, the claims record, and the BOM. Zero-variance data is filtered out, as well as any data that lacks essential information (e.g. the name of the customer who made the claim, the progress of the claim). Numerical features are standardised, while categorical features are first numerically encoded and then normalised in the range $[0, 1]$. Numerical encoding was preferred: one-hot encoding, usually considered the method of choice for neural networks, could not be applied here for features with more than 10 categories, and numerical encoding came to the rescue as an acceptable option. The encoding method used is frequency encoding, also referred to as count encoding, which encodes categorical features by assigning numerical values based on how frequently each category appears in the dataset. After pre-processing, a minimal subset of relevant features is defined based on experience and domain knowledge and collected in a round of discussions with our technology partners.

In order to train **T1**, *collection* and *product category*, two general features from the Product registry, are used in combination with more specific features from the BOM: $matx, y$ represents the total number of different materials used in the manufacturing process, where $x, y \in \{1, 2, 7\}$ are indices identifying the function of the material (1=*inner*, 2=*outer*, 7=*reinforcement*). Optionally, it is possible to use the total amount of materials used for each macro category instead of the number of different materials from it. *Combination* is a technical feature that takes into account combinations of materials and shapes. Finally, *Difficulty* is a feature whose value has been subjectively estimated for each product by experts in the manufacturing process. It varies in the range $[3, 9]$ and is very informative as it strongly correlates with the complexity of a given product and thus with the process time. Although available, low-level details about each operation have been deliberately omitted to avoid the risk of sparse data due to the limited number of products/samples in the dataset. For **T1** the numerical target is time.

In order to train **T2**, features relating to the characteristics of the product that turned out to be defective are taken from the Record of claims. **T2** performs the prediction for four different targets: *stock origin*, which returns the identity of the subject who filed the claim, either the retailer or the customer, and helps to identify the point in time when a defect tends to occur; *defect main location* (e.g. *main body*, *handle*, *lining*) reports the exact location of the defect in the product; *defect detailed location* (e.g. *fabric*, *leather*) refines the information and reports the defective material in the product; *defect type* (e.g. *broken*, *stained*, *stitched*, *unstitched*) provides an accurate description of the type of defect detected. For each target, those categories referring to similar defects or defect location are aggregated to support the decision systems with additional information. A few different types of aggregation are then experimented.

3.4. FashionDSS predictive models

To tackle **T1**, a linear regression approach is implemented in the form of a shallow neural network (NN), with no hidden layer, to allow for easy experimentation with different configurations, whereas a larger variety of models are experimented for **T2**, including an ensemble of Multi Layer Perceptron (MLP) models, working independently, one for each target; a unique multi head MLP (mhMLP) model, simultaneously making predictions over all targets; a set of ensembled binary XGBoost models, each ensemble separately trained for a specific target according to One vs All (OvA) strategy, finally providing a different binary classifier for each category of the given target. During hyperparameter grid search, the shallow architecture at the core of the MLP/mhMLP is occasionally transformed into a deeper architecture to increase complexity, by adding either one or three hidden layers.

It is worth clarifying that the selection of predictive models was guided by the characteristics of the two tasks and by the size and structure of the dataset. Task **T1** is formulated as a regression problem with a relatively limited number of samples (600 products) and a small set of structured, mostly low-dimensional features derived from the BOM and product registry. For this reason, model complexity was deliberately controlled to reduce overfitting risk and to preserve interpretability. Task **T2**, instead, is a multi-target, multi-class classification problem based on 1800 claims, characterised by heterogeneous categorical features and class imbalance. Consequently, multiple model families were explored to assess different modelling hypotheses, including independent models per target, shared multi-head representations, and tree-based ensemble methods suited for structured tabular data.

In the following section, details are provided for each of the model.

3.4.1. Linear regression

This model, in charge for the forecasting of process times, defined as task **T1**, implements linear regression, relying on a shallow NN. More formally, we want to learn a linear regression function f defined as follows:

$$f : \mathbb{R} \cup \{0, 1\}^n \rightarrow \mathbb{R} \tag{1}$$

where the target represents the predicted process time and features can be either numerical or categorical. Mean Squared Error (MSE) is used for the loss:

$$MSE = \sum_{i=1}^{N} (x_i - y_i)^2 \tag{2}$$

where N gives the number of samples/products in the dataset, x_i represents the predicted process time for sample i and y_i the actual process time recorded during the manufacturing of sample i.

In most experiments, a shallow/linear NN is used, in which case a linear activation function is added at the top of the network. In some experiments, during an initial exploratory phase, when non-linearity was of interest, hidden layers were added to make the shallow NN deeper, and an activation function of the following form was added between each hidden layer and the one immediately preceding it in the network:

$$Relu(z) = max(0, z) \tag{3}$$

The choice of a linear regression model implemented through a shallow neural network reflects a deliberate trade-off between bias and variance. Given the moderate size of the dataset and the interpretable nature of the selected features (e.g. material counts, combinations and expert-assessed difficulty), a linear formulation provides sufficient expressive power while limiting variance. More complex architectures were only explored during a controlled hyperparameter search to assess potential nonlinear effects. However, they were not adopted as the primary configuration in order to maintain robustness under limited data conditions.

3.4.2. Supervised multiclass classification via multiple single head MLP

In this first approach, task **T2** is solved relying on multiple single-head MLP models, whose architecture allows for a number of hidden layers which can be equal to 0, 1 or 3. In **T2**, four different targets undergo prediction, as described in Section 3.3, and for each of them a different training of the MLP is performed. The loss consists in Cross Entropy (CE), weighted according to the class imbalance present in the training set and SoftMax is used as the final activation function.

3.4.3. Supervised multiclass classification via a single multi head MLP

In this second approach, task **T2** is solved relying on a unique MLP model having four head/outputs. As a consequence, the four different targets undergo prediction simultaneously and only one training process is performed. Also in this case, the number of hidden layers can be equal to 0, 1 or 3, the loss consists in weighted Cross Entropy (CE) and SoftMax is used as the final activation function.

3.4.4. Supervised multiclass classification via XGBoost

In this third approach, task **T2** is solved via XGBoost (Chen & Guestrin, 2016). For each target, the classification problem is binarized and one binary model is trained for each single class/category of the target. Training data are rearranged according to OvR strategy.

XGBoost was included due to its strong performance on structured tabular datasets and its inherent regularisation mechanisms, which are particularly advantageous for moderate-sized datasets. Unlike high-capacity deep neural networks, gradient-boosted trees are less sensitive to feature scaling and can effectively handle heterogeneous encoded variables. These characteristics make them well-suited to Task T2, where categorical features are frequency-encoded and class imbalance is present.

Table 2
Median with IQR in brackets for all metrics and targets, inclusive of alternative aggregation modes. With reference to each column, the highest median and the lowest IQR (best results) are highlighted in red whereas the lowest median and highest IQR (worst results) are highlighted in blue.

		stock	main loc	detail loc (A)	detail loc (B)	type (A)	type (B)
f1-score	MLP	69.51 (3.67)	67.12 (2.06)	65.89 (3.63)	58.48 (4.9)	46.35 (3.57)	47.82 (2.96)
	mhMLP	66.6 (4.17)	63.5 (4.95)	63.88 (5.15)	47.57 (10.03)	49.71 (8.69)	41.55 (2.98)
	XGBoost	71.04 (5.15)	67.78 (2.33)	67.99 (3.25)	77.47 (2.9)	74.39 (1.4)	79.81 (1.05)
b-acc	MLP	69.5 (3.65)	68.53 (2.96)	67.83 (3.7)	60.35 (4.78)	47.49 (3.72)	52.35 (4.62)
	mhMLP	66.66 (3.93)	64.48 (4.51)	64.85 (4.95)	51.61 (6.24)	49.49 (8.77)	47.35 4.31
	XGBoost	71.07 (5.48)	69.4 (1.84)	69.93 (2.82)	71.45 (4.55)	71.87 (1.49)	73.23 (0.38)
prec	MLP	69.51 (3.67)	67.12 (2.06)	65.89 (3.63)	58.48 (4.9)	46.35 (3.57)	47.82 (2.96)
	mhMLP	66.6 (4.17)	63.5 (4.95)	63.88 (5.15)	47.57 (10.03)	49.71 (8.69)	41.55 (2.98)
	XGBoost	71.15 (5.91)	68.43 (2.35)	68.75 (2.55)	81.51 (2.74)	81.04 (1.51)	87.97 (0.88)
recall	MLP	69.51 (3.67)	67.12 (2.06)	65.89 (3.63)	58.48 (4.9)	46.35 (3.57)	47.82 (2.96)
	mhMLP	66.6 (4.17)	63.5 (4.95)	63.88 (5.15)	47.57 (10.03)	49.71 (8.69)	41.55 (2.98)
	XGBoost	71.46 (4.76)	68.9 (3.1)	69.64 (3.21)	74.51 (4.04)	69.55 (2.36)	73.53 (1.72)

3.4.5. Training phase

With reference to task **T2**, for all models, the training procedure is repeated for each inner split returned by nested cross validation and the balanced accuracy of the splits is then averaged to identify the best performing hyper parameters. Hyperparameters for the MLP/mhMLP include number of *epochs* (200 or 300), initial *learning rate* for Adam optimizer (0.001 or 0.0001) and number of *hidden layers* (no hidden layers, 1 hidden layer with 16 units or 3 hidden layers with 16, 32 and 64 units respectively). In all experiments, the batch size was 32. Hyperparameters for the XGBoost include *learning rate* (0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.300000012, 0.310000012, 0.320000012), *gamma* (0, 0.05, 0.1, 0.15, 0.2), *max depth* (3, 5, 6, 8, 10) and *min child weight* (1, 3, 5).

With reference to task **T1**, our experiments use a batch size of 32 and the model is trained over 1000 epochs. Adam optimiser is used with a learning rate of 0.1. It is important to note that process times of less than 20 min were excluded from our analysis because likely erroneous.

To ensure reproducibility and prevent overfitting, hyperparameter optimisation was strictly confined to the validation data within the inner cross-validation loops. No information from the outer test folds was used during model selection. All dataset splits were precomputed and stored to guarantee consistency across model comparisons. Separating model tuning from final evaluation reduces optimistic bias and supports transparent assessment of generalisation performance.

3.4.6. Evaluation phase

For task **T2**, after the training phase, for each model category, the model whose hyperparameters performed best through inner cross validation is retrained. Then, for each outer fold, the retained model is retrained through the complete outer train set and evaluated through the (outer) test set. Results for all the 5 outer fold are finally presented via box plots.

For task **T1**, evaluation is performed on the test set after standard training.

3.4.7. Evaluation metrics

For regression task **T1**, performance is evaluated by means of Eq. (4), as follows:

$$RMSE = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2} \tag{4}$$

where the argument of the square is the MSE defined in (2)

For the multi-class classification task **T2**, the performance of different models is evaluated by means of balanced accuracy (b-acc), f1-score, precision (prec) and recall, whose general formulas follow:

$$prec = \frac{TP}{TP + FP} \tag{5}$$

$$recall = \frac{TP}{TP + FN} \tag{6}$$

$$f1\text{-}score = 2 \cdot \frac{prec \cdot recall}{prec + recall} \tag{7}$$

$$b\text{-}acc(y, \hat{y}, w) = \frac{1}{\sum \hat{w}_i} \cdot \sum_i \mathbb{1}(y_i = \hat{y}_i) \cdot \hat{w}_i \tag{8}$$

where $\hat{w}_i = \frac{w_i}{\sum_j \mathbb{1}(y_j = y_i) \cdot \hat{w}_j}$

In particular, balanced accuracy is used during hyperparameter search to pick the best configuration. Although the splits of the dataset are stratified, considering the imbalance all targets (apart from *stock origin*) are affected by, with some categories much larger than other, balanced accuracy seems a reasonable metric to study performance. In addition to the original data imbalance, when multi-class problems are binarized in the case of XGBoost, further imbalance can be introduced as an artifact of binarization, supporting the use of balanced accuracy. F1-score, precision and recall are to be intended as *micro*, that is computed globally through all available classes. In any case, the behaviour of each model is finally discussed in view of all proposed metrics. For each of them, metrics are computed for all the 5 outer folds of nested cross validation and reported via box-plots in Fig. 4. Table 1 numerically sums up the same results via the Median and the Interquartile Range (IQR).

4. Results and discussions

In this section, we present the results of our study. On the base of them, we define our proposed workflow FashionDSS and discuss its applicability to real world scenarios.

As far as results relating task **T2** are concerned, with reference to Fig. 4, all metrics support the superiority of XGBoost. Table 2 further confirms this, showing that XGBoost consistently outperforms both MLP methods in terms of the median and almost always exhibits the narrowest IQR, indicating greater stability in results across different outer folds. Precision and recall are also usually above 70% for XGBoost. From Table 2, a change in the aggregations of classes for target *detailed location* and target *type* clearly benefits XGBoost suggesting that a search in the domain of different aggregations may lead to much better results for this method. With reference to Fig. 5, we tried two different aggregations for target *defect detailed location* and target *type*. The former, target *defect detailed location*, includes five categories/classes initially arranged into two groups as follows: <*Fabric*, *Leather*, *Other Materials*> and <*Functional Accessories*, *Ornamental Accessories*>. As an alternative aggregation, we tried to fatherly split the first group into <*Fabric*, *Leather*> and <*Other Materials*>. Similarly we reasoned for

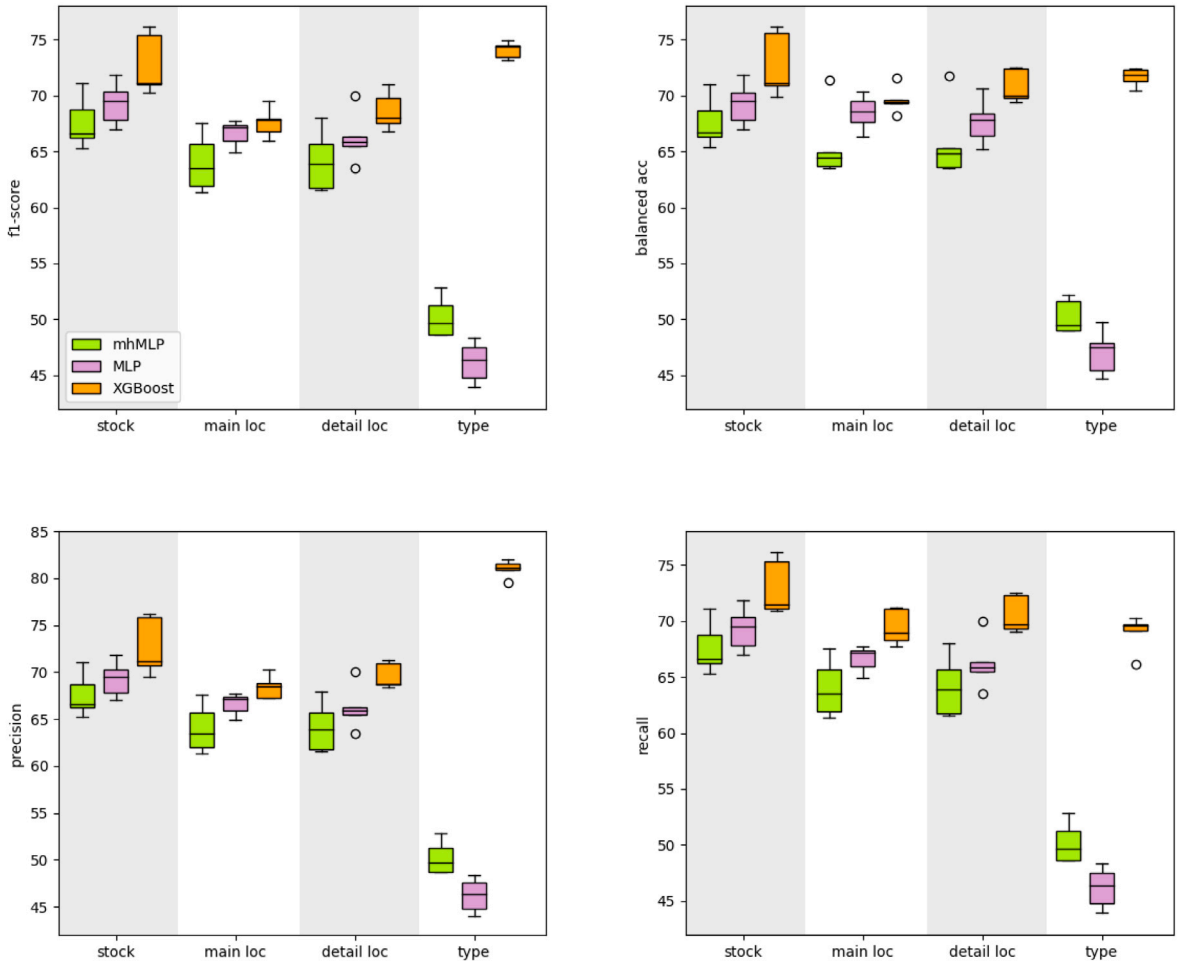


Fig. 4. Box-plots for Balanced accuracy, F1-score. Precision and Recall obtained when **mhMLP** (green), **MLP** (purple) or **XGBoost** (orange) is applied.

Table 3
Results for the forecast of process time in **T1**.

Target range in minutes	RMSE on test set
[20, 250]	27.72
[20, 270]	27.83
[20, 280]	38.90
[20, 300]	47.60
[20, 325]	41.98
[20, 350]	51.38
[20, 375]	45.62
[20, 400]	42.81

target *type*, investigating a different and more informative aggregation, which proved to be advantageous.

Compared to XGBoost, MLP-based models lag behind for all targets, with a particularly dramatic drop for target *type*, the one with more categories. A comparison of the multiple single-head MLP models *versus* the multi-head single MLP model shows that simultaneous prediction reduces performance.

Our attention back to **T1**, with reference to Table 3, we see that the prediction of the process time is characterised by an $RMSE$ that depends on the range considered. On the one hand, up to 250 min, the error seems to be reasonable, although we could aim for better results. On the other hand, if we consider longer processing times, the performance deteriorates. This could be a consequence of the imbalance in the dataset, where the longer process times are less frequent than the shorter ones. We could not do much about this kind of imbalance, as the number of samples with longer process times was very small.

Considering the results we have collected, we can define our proposed workflow, FashionDSS, as a combination of a linear regression component, implemented in the form of a shallow NN, to handle task **T1**, and an ensemble of carefully tuned XGBoost models to handle task **T2**. The results of our experiments demonstrate how FashionDSS has the potential to significantly impact decision-making processes in luxury fashion manufacturing and quality control. Finally, given that a minimal set of features has been exploited, we expect that relevant improvements can be achieved by introducing new informative variables.

While the quantitative improvements observed across the models are statistically significant, their practical importance becomes clearer when considered in the context of the operational processes involved in producing luxury fashion items. In Task T1, for example, even modest reductions in forecasting error can lead to more accurate resource planning, improved production scheduling and better coordination between material procurement and workshop allocation. In low-volume, high-margin production environments, improvements of a few minutes per unit in process time estimation can result in significant efficiency gains over the course of a collection cycle. Similarly, improvements in balanced accuracy and defect prediction reliability in Task T2 have direct managerial implications. Identifying high-risk material combinations or defect-prone product components during the design phase enables designers and production managers to weigh up the trade-offs between aesthetic choices, manufacturing complexity, and quality risk. Even incremental improvements in defect prediction can justify early design adjustments, thereby preventing rework, warranty costs and reputational damage. In this context, predictive performance should be

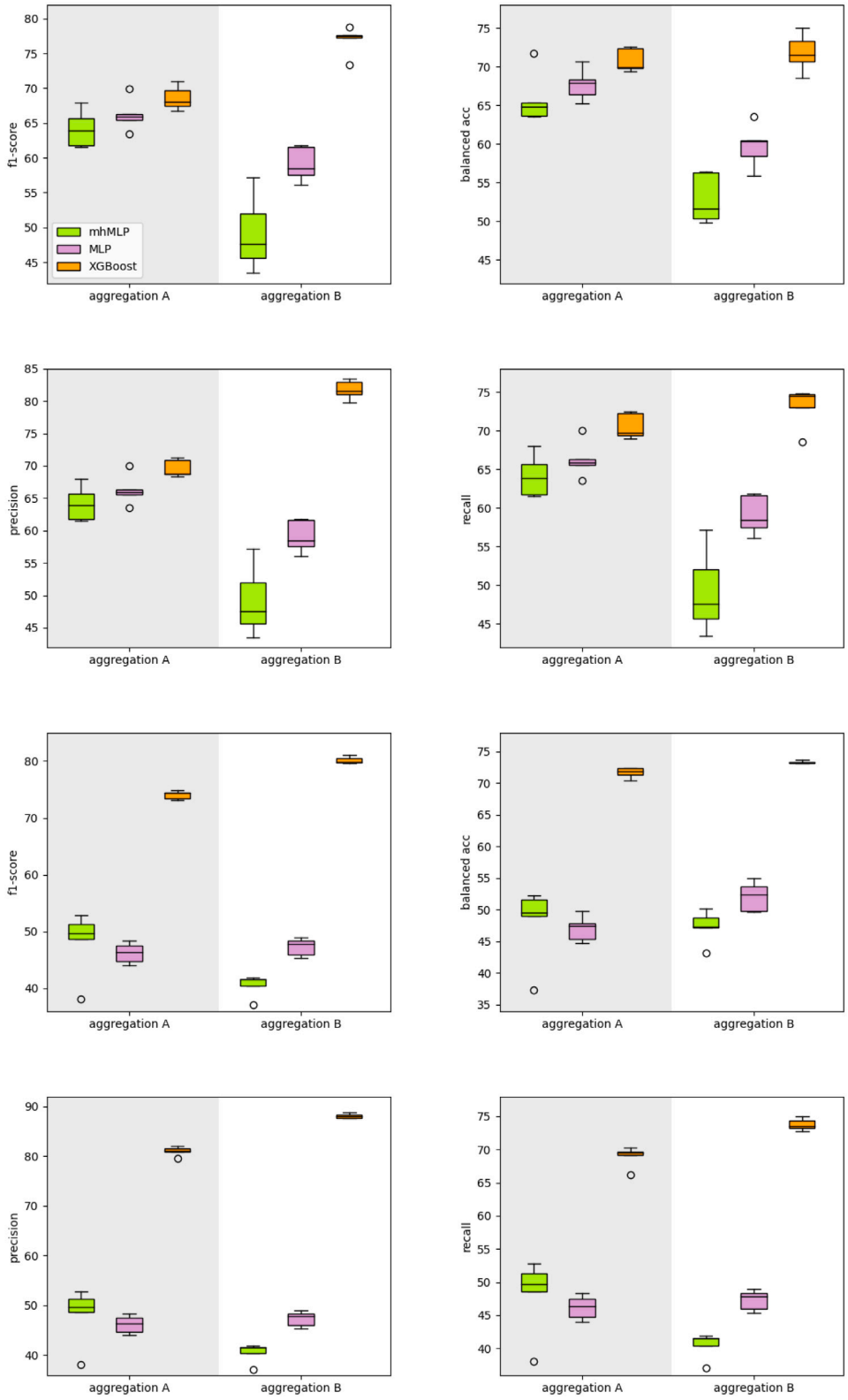


Fig. 5. Box-plots for two different aggregations A and B of target *defect detailed location> and target defect type.*

viewed not only as a classification metric, but also as a risk reduction tool that supports informed design and planning decisions in situations of uncertainty.

Beyond the numerical performance indicators, the results provide valuable insights into the purpose of this research study. The forecasting accuracy achieved in Task T1 demonstrates that the structured design and BOM features contain a sufficiently strong predictive signal to support early-stage production planning. This extends prior research, which has predominantly focused on post-production optimisation rather than upstream anticipation. Similarly, the defect prediction outcomes in Task T2 confirm that historical complaint data can function as a forward-looking informational asset, rather than merely

as a reactive quality indicator. This finding directly addresses the gap in the literature concerning the limited integration of structured customer feedback into design-stage decision support systems. From a practical standpoint, the predictive signals generated by FashionDSS provide designers and production planners with actionable managerial insights, enabling them to evaluate material choices, complexity levels, and risk exposure before manufacturing begins. In this sense, the empirical findings not only validate the feasibility of the proposed framework, but also advance a more proactive, data-informed approach to quality management in low-volume, high-variability production environments.

5. Conclusions and future works

This paper presents FashionDSS: a DSS designed to explore the feasibility of anticipating potential defects and forecasting production efficiency in the luxury fashion industry. It does this by integrating product registry data, BOM information and structured claim records. The results show that predictive modelling can support the identification of defect-prone components and provide reasonable production time estimates for newly designed products within the scope of the available dataset. The study shows that, when systematically structured and combined with production-related features, customer complaint data can be used as informative input for upstream decision support. In particular, the predictive models developed for Tasks T1 and T2 illustrate that process time estimation and defect risk assessment can both be addressed within a unified analytical framework. Although the predictive performance is necessarily limited by the size of the dataset and the specificity of the domain, the findings suggest that meaningful risk signals can be extracted and used to inform decisions regarding early-stage design and production planning. A key feature of FashionDSS is its human-in-the-loop configuration, which allows domain experts to interpret, validate and refine model outputs iteratively. Rather than replacing expert judgement, the system is intended to complement it by providing probabilistic insights that support more informed trade-offs between design complexity, material selection, and quality risk. Within the context of the present study, this approach demonstrates the practical feasibility of integrating predictive analytics into design-oriented quality management. This research provides managers and practitioners with actionable insights by demonstrating how structured production and complaint data can inform design-stage decisions. This enables more informed trade-offs to be made regarding material selection, production planning and risk mitigation before manufacturing begins.

Future research should focus on extensions that strengthen external validity. In particular, validating the model on larger datasets and across multiple companies would enable a more robust assessment of its stability and transferability. Cross-company validation would also clarify whether defect patterns and production dynamics are company-specific or shared across the sector. Additionally, longitudinal data covering multiple seasonal cycles would enable the evaluation of temporal robustness. Further methodological refinements could involve exploring alternative modelling strategies under small-data constraints and investigating semi-supervised or transfer learning approaches. Finally, as predictive systems increasingly rely on customer-related information, future work should also consider aspects of governance, transparency, and data protection to ensure responsible deployment. This study provides an initial step towards integrating structured complaint data and production features into proactive, design-stage decision support in the luxury fashion industry. While further validation is required, the proposed framework demonstrates how information management principles can be implemented in environments with limited data and high variability.

CRediT authorship contribution statement

Giovanna Migliorelli: Software, Methodology, Formal analysis, Data curation, Conceptualization. **Rocco Pietrini:** Visualization, Resources, Investigation, Methodology, Formal analysis, Data curation, Conceptualization. **Alessandro Galdelli:** Visualization, Validation, Resources, Investigation.. **Emanuele Frontoni:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition. **Marina Paolanti:** Writing – review & editing, Writing – original draft, Methodology, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Marina Paolanti reports was provided by University of Macerata. Marina Paolanti reports a relationship with University of Macerata that includes:. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alkahtani, Mohammed, Choudhary, Alok, De, Arijit, & Harding, Jennifer Anne (2019). A decision support system based on ontology and data mining to improve design using warranty data. *Computers & Industrial Engineering*, *128*, 1027–1039.

Babakmehr, Mohammad, Baumanns, Sascha, Chehade, Abdallah, Hochkirchen, Thomas, Kalantari, Mahdokht, Krivtsov, Vasiliy, & Schindler, David (2024). Data-driven framework for warranty claims forecasting with an application for automotive components. *Engineering Reports*, *6*(5), Article e12764.

Barrera, Felipe, Segura, Marina, & Maroto, Concepción (2024). Multiple criteria decision support system for customer segmentation using a sorting outranking method. *Expert Systems with Applications*, *238*, Article 122310.

Bertolini, Massimo, Mezzogori, Davide, Neroni, Mattia, & Zammori, Francesco (2021). Machine learning for industrial applications: A comprehensive literature review. *Expert Systems with Applications*, *175*, Article 114820.

Caro, Felipe, & de Tejada Cuenca, Anna Sáez (2023). Believing in analytics: Managers' adherence to price recommendations from a DSS. *Manufacturing & Service Operations Management*, *25*(2), 524–542.

Chakraborty, Samit, Moore, Marguerite, & Parrillo-Chapman, Lisa (2022). Automatic defect detection for fabric printing using a deep convolutional neural network. *International Journal of Fashion Design, Technology and Education*, *15*(2), 142–157.

Chen, Hsi-Tien (2016). Quality function deployment in failure recovery and prevention. *The Service Industries Journal*, *36*(13–14), 615–637.

Chen, Tianqi, & Guestrin, Carlos (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Chen, Lihong, Halepoto, Habiba, Liu, Chunhong, Yan, Xinfeng, & Qiu, Lijun (2022). Research on influencing mechanism of fashion brand image value creation based on consumer value co-creation and experiential value perception theory. *Sustainability*, *14*(13), 7524.

Eldessouki, Mohamed (2018). Computer vision and its application in detecting fabric defects. In *Applications of computer vision in fashion and textiles* (pp. 61–101). Elsevier.

Filz, Marc-André, Bosse, Jan Philipp, & Herrmann, Christoph (2024). Digitalization platform for data-driven quality management in multi-stage manufacturing systems. *Journal of Intelligent Manufacturing*, *35*(6), 2699–2718.

Fouch, George (1965). A guide to zero defects. Quality and reliability assurance handbook 4155.12-H.

Gazzola, Patrizia, Grechi, Daniele, Iliashenko, Iuliia, & Pezzetti, Roberta (2024). The evolution of digitainability in the fashion industry: a bibliometric analysis. *Kybernetes*, *53*(13), 101–126.

Hassenzahl, Marc, Burmester, Michael, & Koller, Franz (2021). User experience is all there is: twenty years of designing positive experiences and meaningful technology. *I-Com*, *20*(3), 197–213.

Hisrich, Robert D., & Soltanifar, Mariusz (2021). Unleashing the creativity of entrepreneurs with digital technologies. *Digital Entrepreneurship: Impact on Business and Society*, 23–49.

Jing, Teng, Shi, Guicheng, Liu, Matthew Tingchi, & Zhu, Mingxia (2026). Fast fashion vs. Luxury fashion: who is better fitting with sustainable extension? The comparison of recycling and upcycling. *Journal of Product & Brand Management*, *35*(1), 148–163.

Lee, Jun-Guel, Kim, Taehyeong, Sung, Ki Woo, & Han, Sung Won (2021). Automobile parts reliability prediction based on claim data: The comparison of predictive effects with deep learning. *Engineering Failure Analysis*, *129*, Article 105657.

Loureiro, Ana L. D., Miguéis, Vera L., & Da Silva, Lucas F. M. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, *114*, 81–93.

Lysenko-Ryba, Kateryna, & Zimon, Dominik (2021). Customer behavioral reactions to negative experiences during the product return. *Sustainability*, *13*(2), 448.

May, Gökan, & Kiritsis, Dimitris (2019). Zero defect manufacturing strategies and platform for smart factories of industry 4.0. In *Proceedings of the 4th international conference on the industry 4.0 model for advanced manufacturing: AMP 2019 4* (pp. 142–152). Springer.

Murzyn-Kupisz, Monika, & Hołuj, Dominika (2021). Fashion design education and sustainability: towards an equilibrium between craftsmanship and artistic and business skills? *Education Sciences*, *11*(9), 531.

Nikseresht, Ali, Shokouhyar, Sajjad, Tirkolaee, Erfan Babaee, Nikookar, Ethan, & Shokoohyar, Sina (2024). An intelligent decision support system for warranty claims forecasting: Merits of social media and quality function deployment. *Technological Forecasting and Social Change*, *201*, Article 123268.

Oskouei, Amin Golzari, Balafar, Mohammad Ali, & Motamed, Cina (2023). RDEIC-LFW-DSS: ResNet-based deep embedded image clustering using local feature weighting and dynamic sample selection mechanism. *Information Sciences*, *646*, Article 119374.

Pavan, M., & Samant, Lata (2024). Digitalization and E-commerce trends in the industry. In *Consumption and production in the textile and garment industry: a comparative study among Asian countries* (pp. 253–272). Springer.

Pereira, Artur M., Moura, J. Antao B., Costa, Evandro De B., Vieira, Thales, Landim, Andre R. D. B., Bazaki, Eirini, & Wanick, Vanissa (2022). Customer models for artificial intelligence-based decision support in fashion online retail supply chains. *Decision Support Systems*, *158*, Article 113795.

Psarommatis, Foivos, May, Gökan, Dreyfus, Paul-Arthur, & Kiritsis, Dimitris (2020). Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research. *International Journal of Production Research*, *58*(1), 1–17.

Rasheed, Rohail, Qazi, Farheen, Ahmed, Aarish, Asif, Alyan, Shams, Hussain, et al. (2024). Machine learning approaches for in-vehicle failure prognosis in automobiles: A review. *VFAST Transactions on Software Engineering*, *12*(1), 169–182.

Silva, Anabela Costa, Machado, José, & Sampaio, Paulo (2024). Predictive quality model for customer defects. *The TQM Journal*, *36*(9), 155–174.

Sun, Lushan, & Zhao, Li (2018). Technology disruptions: Exploring the changing roles of designers, makers, and users in the fashion industry. *International Journal of Fashion Design, Technology and Education*, *11*(3), 362–374.

Tam, Fung Yi, & Lung, Jane (2025). Digital marketing strategies for luxury fashion brands: A systematic literature review. *International Journal of Information Management Data Insights*, *5*(1), Article 100309.

Theissler, Andreas, Pérez-Velázquez, Judith, Kettelgerdes, Marcel, & Elger, Gordon (2021). Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. *Reliability Engineering & System Safety*, *215*, Article 107864.

Thiel, Katharina, & Postlethwaite, Susan (2023). Human-centric research of skills and decision-making capacity in fashion garment manufacturing to support robotic design tool development. *Human Aspects of Advanced Manufacturing*, 20–30.

Verna, Elisa, Genta, Gianfranco, Galetto, Maurizio, & Franceschini, Fiorenzo (2023). Zero defect manufacturing: a self-adaptive defect prediction model based on assembly complexity. *International Journal of Computer Integrated Manufacturing*, *36*(1), 155–168.

Wan, Paul Kengfai, & Leirmo, Torbjørn Langedahl (2023). Human-centric zero-defect manufacturing: State-of-the-art review, perspectives, and challenges. *Computers in Industry*, *144*, Article 103792.

Wang, Mengyue, Li, Xin, Liu, Yidi, Chau, Patrick, & Chen, Yubo (2024). A contrast-composition-distraction framework to understand product photo background's impact on consumer interest in E-commerce. *Decision Support Systems*, *178*, Article 114124.

Wang, Lei, & Liu, Zhengchao (2021). Data-driven product design evaluation method based on multi-stage artificial neural network. *Applied Soft Computing*, *103*, Article 107117.

Zajec, Patrik, Rožanec, Jože M., Theodoropoulos, Spyros, Fontul, Mihail, Koehorst, Erik, Fortuna, Blaž, & Mladenić, Dunja (2024). Few-shot learning for defect detection in manufacturing. *International Journal of Production Research*, 1–20.