A Fuzzy Linguistic Approach to Functional Adaptability Evaluation in Industrial Forecasting Systems

Francisco P. Romero[1,2][0000−0002−6993−2434], Jesus R. Oviedo[1,2][0000−0001−7962−1042], Jared Guerrero-Sosa[1][0000−0001−7999−9870], Moises Rodriguez[1,2][0000−0003−2155−7409], Andres Montoro-Montarroso[1,2][0000−0003−1893−3346], Javier Verdugo[1,2][0000−0002−2526−2918], Jesus Serrano-Guerrero[1,2][0000−0002−6177−8188], Mario Piattini[1,2][0000−0002−7212−8279], and Jose A. Olivas[1,2][0000−0003−1893−3346]

[1] Department of Information Technologies and Systems, University of Castilla La Mancha, Ciudad Real, Spain
franciscop.romero@uclm.es
[2] Intelligent Information Systems Certification, Ciudad Real, Spain
jverdugo@i2sc.es

Abstract. Artificial Intelligence systems deployed in industrial forecasting environments are exposed to temporal drift, seasonality and evolving operational conditions. ISO/IEC 25059 defines functional adaptability as an AI quality subcharacteristic, but it does not specify how to assess whether a deployed system effectively learns from new information and maintains or improves predictive behaviour over time. This proposal addresses that gap through an operational, temporal-error-based assessment, using indicators such as temporal-error trends, degradation ratios, recovery gains and stability measures. Since the interpretation of these indicators is uncertain, context-dependent and often expressed by experts through linguistic judgements such as low degradation, acceptable delay or high stability, the proposal incorporates a fuzzy linguistic model. It defines fuzzy variables, expert-based membership functions and ISO/IEC 25059-aligned aggregation rules. The model provides an interpretable fuzzy linguistic assessment, complemented by a defuzzified score for operational monitoring and comparison.

Keywords: Artificial Intelligence Quality · ISO/IEC 25059 · Functional Adaptability · Fuzzy Logic · Industrial Forecasting

1 Introduction

Artificial Intelligence (AI) systems are increasingly embedded in industrial decision making processes. In these settings, model quality cannot be assessed only through static predictive metrics computed at development time. Once deployed, these systems are exposed to changing data distributions, seasonality, shifts in

user behaviour, and evolving operational constraints. Therefore, their ability to maintain or improve performance over time is a key quality requirement.

ISO/IEC 25059 [1] extends the SQuaRE quality model [2] to AI systems and introduces AI-specific quality subcharacteristics. Among them, functional adaptability refers to the extent to which an AI system can acquire information from data or past actions and use it to inform future predictions or behaviours. This property is especially relevant for forecasting systems that are periodically retrained with updated operational data. However, ISO/IEC 25059 provides a conceptual definition and does not prescribe a concrete measurement procedure for real industrial deployments. Recent work has shown that ISO/IEC 25059 can be operationalised in real AI quality evaluations, highlighting the importance of acceptable specification limits, representative test cases, and context-dependent assessment criteria, especially in safety-critical applications [3].

A previous operational approach evaluates functional adaptability by analysing the temporal evolution of forecasting error across successive operational executions. Under this view, adaptation is evidenced when retraining with new operational information preserves or improves predictive performance [4, 6], consistent with time-adaptive forecasting and concept-drift adaptation methods. Error degradation, recovery gain, and post-adaptation stability are useful indicators, although their interpretation remains context-dependent because acceptable degradation and detection delay vary with the operational environment and forecast horizon [7].

Existing work relevant to this proposal can be organised into three complementary topics: 1) Concept-drift and time-adaptive forecasting studies focused on detecting distributional changes and maintaining predictive performance [8]; 2) Production-oriented machine learning frameworks that emphasise data validation, monitoring, retraining, infrastructure reliability, and technical readiness [9] and; 3) fuzzy linguistic models provide mechanisms for representing gradual and expert dependent judgements [5]. However, these strands do not jointly provide an ISO/IEC 25059-aligned procedure that transforms temporal adaptation evidence into an auditable linguistic quality assessment. Accordingly, this paper reformulates functional adaptability as a fuzzy linguistic assessment problem, by integrating temporal adaptability indicators, expert knowledge, and interpretable linguistic outputs within an ISO/IEC 25059-aligned framework.

The remainder of the paper is organised as follows. Section 2 introduces functional adaptability in ISO/IEC 25059 and its temporal, uncertainty-aware assessment. Section 3 presents the proposed fuzzy linguistic model, including indicators, membership functions, rule-based aggregation, and linguistic output. Section 4 illustrates its application through an industrial predictive maintenance scenario. Section 5 discusses the contributions, limitations, and validation requirements of the proposal. Finally, Section 6 presents the conclusions and future work.

2 Functional Adaptability in Industrial Forecasting Systems

Industrial forecasting systems commonly operate in batch or periodically retrained settings. Unlike online learning systems, these solutions do not update their internal parameters continuously after each new observation. Instead, they are retrained at predefined intervals or when sufficient new operational data becomes available. Each retraining process produces a new operational state of the system, whose predictive behaviour can be evaluated in subsequent forecasting windows. This operational view is consistent with the broader literature on production machine learning, where system quality depends on model accuracy, data validation, monitoring, retraining practices, infrastructure reliability and technical-debt control [9, 10]. In this sense, functional adaptability should be assessed through observable behaviour during operation, rather than inferred only from the existence of a retraining pipeline.

In industrial environments, this issue is particularly relevant because forecasting systems are rarely exposed to stationary data-generating processes. Demand patterns, production regimes, customer behaviour, machine conditions and external factors may evolve over time, leading to changes in the relationship between input variables and target values. This phenomenon is closely related to concept drift, which has been extensively studied in machine learning as a central challenge for models deployed under evolving distributions [8]. In time-dependent prediction problems, concept drift is especially critical because the temporal order of observations determines both the training context and the future operational conditions under which the model is evaluated.

Let $t = 1, \ldots, T$ denote ordered operational executions or evaluation windows. For each execution, the forecasting system generates predictions $\hat{y}_{i,t}$ for observed values $y_{i,t}$. Forecasting error may be measured using application dependent metrics such as RMSE, sMAPE or WAPE. In industrial settings with heterogeneous demand volumes, WAPE is particularly useful because it expresses the aggregate absolute deviation relative to the observed magnitude:

$$WAPE_t = \frac{\sum_i |y_{i,t} - \hat{y}_{i,t}|}{\sum_i |y_{i,t}|}. \tag{1}$$

The ordered sequence $E = \{E_1, E_2, \ldots, E_T\}$, where E_t denotes the error in execution t, provides the empirical basis for analysing functional adaptability. In the original operational proposal, the temporal behaviour of this sequence is examined through trend and stability analysis. A simple linear trend can be estimated as:

$$E_t = \beta_0 + \beta_1 t + \epsilon_t, \tag{2}$$

where $\beta_1 < 0$ suggests improvement, $\beta_1 \approx 0$ suggests stability, and $\beta_1 > 0$ suggests degradation. This analysis can be complemented with hypothesis testing, non-parametric trend tests, bootstrap confidence intervals or segmented analysis by forecasting horizon, region, product family or business context. Such

temporal analysis aligns with the idea that functional adaptability, as introduced in ISO/IEC 25059, should reflect an AI system's ability to acquire information from data or previous actions and use it to inform future predictions.

However, functional adaptability cannot be reduced to a regression coefficient. A deployed AI system may temporarily degrade due to drift, recover after retraining and then remain stable. Therefore, we consider four complementary dimensions: performance degradation, drift detection delay, recovery gain and post-adaptation stability. These dimensions also connect functional adaptability with production-readiness concerns: the system must detect that predictive behaviour has changed and support recovery actions and demonstrate stable post-adaptation behaviour [10].

Let t_c denote the change point at which the operational environment changes, t_d the window in which drift is detected, and t_a the window following an adaptation action, such as retraining. Let E_{base}, E_{drift} and E_{post} represent the mean errors in the baseline, degradation and post-adaptation periods, respectively. The following indicators are then defined:

$$D = \frac{E_{drift} - E_{base}}{E_{base}}, \quad \Delta_{det} = t_d - t_c, \quad G = \frac{E_{drift} - E_{post}}{E_{drift} - E_{base}}. \tag{3}$$

Post-adaptation stability can be estimated as the variability of the error after adaptation:

$$S = \sqrt{\frac{1}{K-1} \sum_{j=0}^{K-1} \left(E_{t_a+j} - \overline{E}_{post}\right)^2}. \tag{4}$$

where $\overline{E}_{post}$ denotes the mean prediction error over the K post-adaptation windows:

$$\overline{E}_{post} = \frac{1}{K} \sum_{j=0}^{K-1} E_{t_a+j}. \tag{5}$$

Lower degradation, shorter detection delay, higher recovery gain, and lower post-adaptation instability indicate better functional adaptability. However, indicators like these cannot be interpreted using universal crisp thresholds, since their acceptability depends on the business context, forecasting horizon, error costs and the volatility of the operational environment. For instance, a small degradation may be acceptable in aggregate demand forecasting but critical in predictive maintenance or safety-sensitive monitoring. Therefore, the proposed fuzzy model preserves the original temporal-error-based assessment and extends it with a linguistic interpretation layer. Degradation, detection delay, recovery gain and post-adaptation stability are still computed from the temporal evolution of prediction error, but they are subsequently interpreted and aggregated under uncertainty, expert judgement and domain-dependent criteria using labels such as *low degradation*, *acceptable delay*, *high recovery* and *stable post-adaptation behaviour*.

3 Fuzzy Linguistic Functional Adaptability Model

The proposed model interprets functional adaptability as a fuzzy quality property derived from observable temporal behaviour. Instead of imposing crisp thresholds, the model represents each adaptability dimension through linguistic variables and expert-defined membership functions. This design reflects the fact that AI quality evaluation often involves gradual transitions between acceptable and unacceptable behaviour.

3.1 Input and Output Linguistic Variables

The fuzzy assessment model uses four input variables: degradation D, detection delay Δ_{det}, recovery gain G, and post-adaptation instability S. The output variable represents functional adaptability, denoted as FA, through the linguistic labels *Very Low*, *Low*, *Moderate*, *High*, and *Very High*. A scalar value, denoted as FA_{score}, is obtained only after defuzzification as a complementary representation for operational monitoring and comparison. Table 1 summarises the proposed linguistic structure.

Table 1. Linguistic variables for fuzzy functional adaptability assessment.

Variable	**Type**	**Linguistic labels**
Performance degradation D	Input	Low, Medium, High
Detection delay Δ_{det}	Input	Fast, Acceptable, Slow
Recovery gain G	Input	Poor, Partial, Good
Post-adaptation instability S	Input	Stable, Moderate, Unstable
Functional adaptability FA	Output	Very Low, Low, Moderate, High, Very High

The proposed labels are intentionally simple in order to provide an more interpretable description. Additional labels may be introduced in domains requiring finer granularity.

3.2 Expert-defined Membership Functions

Each linguistic label is represented through a membership function. Triangular and trapezoidal functions are suitable for an initial industrial deployment because they are interpretable, easy to calibrate and compatible with expert elicitation. For a generic variable x, a triangular membership function can be defined as:

Table 2. Illustrative membership function parameters for normalised indicators.

Variable	Labels	Illustrative membership functions
D	Low; Medium; High	Low: $(0, 0, 0.25)$; Medium: $(0.15, 0.40, 0.65)$; High: $(0.50, 1, 1)$
Δ_{det}	Fast; Acceptable; Slow	Fast: $(0, 0, 0.30)$; Acceptable: $(0.20, 0.45, 0.70)$; Slow: $(0.60, 1, 1)$
G	Poor; Partial; Good	Poor: $(0, 0, 0.35)$; Partial: $(0.25, 0.50, 0.75)$; Good: $(0.65, 1, 1)$
S	Stable; Moderate; Unstable	Stable: $(0, 0, 0.30)$; Moderate: $(0.20, 0.45, 0.70)$; Unstable: $(0.60, 1, 1)$
FA	Very Low; Low; Moderate; High; Very High	Five overlapping fuzzy sets uniformly distributed in $[0, 1]$

$$\mu_A(x; a, b, c) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x \leq b, \\ \frac{c-x}{c-b}, & b < x < c, \\ 0, & x \geq c. \end{cases} \tag{6}$$

For example, the linguistic variable degradation may be calibrated using domain-specific parameters. In a relatively stable forecasting domain, low degradation may correspond to values close to 0, medium degradation to moderate increases in error, and high degradation to values exceeding an agreed-upon operational tolerance. The precise parameters should be elicited from domain experts, historical performance distributions or service-level requirements.

A possible normalised configuration is shown in Table 2. The values are illustrative and should be adapted to each domain. In this configuration, all input indicators are normalised to $[0, 1]$, except recovery gain, where higher values are better.

This fuzzy formalisation aligns with ISO/IEC 25059 because it assesses the system effectively uses new information to inform future predictions, while avoiding the assumption that adaptability can be judged through universal crisp thresholds.

3.3 Fuzzy Rule Base

The rule base encodes expert knowledge about how the different indicators interact. A Mamdani-type fuzzy inference system [11] is appropriate because it provides transparent linguistic rules and facilitates communication with technical and non-technical stakeholders. The general model can be expressed as:

$$FA = \mathcal{F}(D, \Delta_{det}, G, S), \tag{7}$$

where $\mathcal{F}$ denotes the fuzzy inference process. Table 3 presents examples of rules.

The rule base can be extended or weighted based on the application's criticality. For instance, in short-term operational forecasting, detection delay may be more relevant than long-term stability. In strategic forecasting, post-adaptation stability may receive greater importance.

Table 3. Example fuzzy rules for functional adaptability evaluation.

Rule	**Antecedent**	**Conseq.**
R1	If degradation is Low and recovery is Good and instability is Stable	Very High
R2	If degradation is Low and recovery is Partial and delay is Acceptable	High
R3	If degradation is Medium and recovery is Good and delay is Fast	High
R4	If degradation is Medium and instability is Moderate and recovery is Partial	Moderate
R5	If degradation is High and recovery is Good and instability is Stable	Moderate
R6	If degradation is High and recovery is Poor	Low
R7	If delay is Slow and recovery is Poor	Very Low
R8	If instability is Unstable and degradation is Medium or High	Low

3.4 Aggregation and Defuzzification

The fuzzy inference process produces an aggregated fuzzy output over the linguistic terms defined for functional adaptability, such as *Low*, *Medium*, and *High*. Therefore, the primary result of the model is a linguistic assessment represented by the activation degrees of the output labels.

Let $\mathcal{L}_{FA}$ denote the set of output linguistic terms. The linguistic assessment is represented as

$$\mathcal{A}_{FA} = \{(L_k, \gamma_k) \mid L_k \in \mathcal{L}_{FA}\}, \tag{8}$$

where $\gamma_k \in [0, 1]$ is the activation degree of linguistic label L_k. The dominant assessment is given by

$$L^* = \arg \max_{L_k \in \mathcal{L}_{FA}} \gamma_k. \tag{9}$$

Since several labels may be partially activated simultaneously, the complete linguistic profile is retained. These activation degrees are not probabilities and are therefore not required to sum to one.

For operational monitoring and comparison, the aggregated fuzzy output is also transformed into a scalar value using centroid defuzzification:

$$FA_{\text{score}} = \frac{\int_0^1 z\, \mu_{\widetilde{FA}}(z)\, dz}{\int_0^1 \mu_{\widetilde{FA}}(z)\, dz}. \tag{10}$$

The model reports both a linguistic assessment and a complementary numerical score. For example, a result may be expressed as *Medium functional adaptability, with partial activation of High*, together with its corresponding defuzzified value. The complete linguistic profile constitutes the primary interpretable output of the model, whereas the scalar score facilitates longitudinal monitoring, comparison across models and forecasting horizons, and integration into industrial dashboards. The inference process operates on fuzzy sets and preserves their activation profile, while defuzzification provides only an auxiliary operational representation. Accordingly, the scalar value does not replace the underlying fuzzy linguistic assessment.

Interpretability is operationalised by retaining an assessment trace composed of the normalised indicator values, their membership degrees, the activated rules and their corresponding firing strengths, the aggregated linguistic output, and the complementary defuzzified score. Thus, an evaluator can trace a final assessment back to the observed degradation, detection delay, recovery, and post-adaptation stability values, and identify which expert rules determined the result. The explanation therefore concerns the final linguistic label and the complete path from temporal operational evidence to the resulting quality assessment.

4 Illustrative Industrial Predictive Maintenance Scenario

To illustrate how the proposed framework can be instantiated, we consider a predictive maintenance scenario based on the AI4I 2020 Predictive Maintenance Dataset [12]. The purpose of this section is to define the operational windows, adaptation events, temporal indicators, and fuzzy assessment process that would be used in an empirical application. AI4I is a synthetic benchmark designed to reproduce relevant characteristics of an industrial manufacturing process. It includes air temperature, process temperature, rotational speed, torque, tool wear, machine-failure labels, and specific failure modes, thereby providing a controlled setting in which different operational changes and adaptation situations can be defined.

In this scenario, a maintenance system is periodically retrained to estimate the probability of machine failure in future production windows. Let $W_1, \ldots, W_T$ denote successive operational evaluation windows. For each window W_t, the system receives new sensor observations and produces a failure-risk prediction for each machine instance. Once the actual maintenance outcome is observed, the predictive error is computed using a window-level metric such as the false negative rate, the balanced error rate, or a cost-sensitive failure prediction error. This generates a temporal error sequence:

$$E = \{E_1, E_2, \ldots, E_T\},$$

where E_t represents the predictive error observed in window W_t. In this setting, functional adaptability is the observable capacity of the predictive maintenance system to maintain or recover its predictive behaviour after changes in the operational regime.

The AI4I variables allow the simulation of several industrial adaptation situations. For example, a gradual increase in tool wear may represent progressive machine ageing; changes in torque and rotational speed may indicate modified production regimes; and variations in process temperature may represent changes in operating conditions. These changes may alter the distribution of the input variables and lead to temporal degradation of the predictive maintenance model. Therefore, the system should be evaluated in terms of static classification performance and its temporal behaviour before, during and after adaptation.

Following the original temporal-error-based assessment, four quantitative indicators are computed. First, degradation D measures the relative increase in prediction error after the operational change:

$$D = \frac{E_{drift} - E_{base}}{E_{base}},$$

where E_{base} is the mean error before the change and E_{drift} is the mean error during the degradation phase. Second, detection delay Δ_{det} measures the number of evaluation windows required to detect a relevant change in predictive behaviour:

$$\Delta_{det} = t_d - t_c,$$

where t_c is the estimated change point and t_d is the drift detection point. Third, recovery gain G measures the extent to which the system recovers its previous predictive performance after retraining:

$$G = \frac{E_{drift} - E_{post}}{E_{drift} - E_{base}},$$

where E_{post} is the mean error after the adaptation action. Finally, post-adaptation instability S measures the variability of the error after adaptation:

$$S = \sqrt{\frac{1}{K-1} \sum_{j=0}^{K-1} (E_{t_a+j} - \overline{E}_{post})^2},$$

where t_a is the adaptation point and K is the number of post-adaptation windows considered.

Interpreting these indicators through crisp thresholds is problematic. For instance, an 8% degradation may be considered acceptable in a low-risk production context but critical in a safety-sensitive maintenance scenario. Similarly, a detection delay of three windows may be acceptable when failures evolve slowly, but inadequate when failures are abrupt and costly.

Therefore, the proposed model maps these numerical indicators into fuzzy linguistic variables. The degradation indicator is represented by the linguistic labels *Low*, *Medium*, and *High*. Detection delay is represented as *Fast*, *Acceptable*, and *Slow*; recovery gain as *Poor*, *Partial*, and *Good*; and post-adaptation instability as *Stable*, *Moderate*, and *Unstable*. These labels are defined through expert-based triangular or trapezoidal membership functions, allowing gradual transitions between quality levels.

The fuzzy inference system aggregates these four dimensions into a fuzzy linguistic assessment of functional adaptability, complemented by a defuzzified operational score. The output variable is defined using five linguistic labels: *very low*, *low*, *moderate*, *high* and *very high*. A simplified fragment of the rule base is shown below:

R_1 : IF degradation is low AND recovery is good
THEN functional adaptability is very high.

R_2 : IF degradation is high AND recovery is poor
THEN functional adaptability is low.

R_3 : IF degradation is medium AND detection delay is acceptable
AND recovery is partial THEN functional adaptability is moderate.

R_4 : IF detection delay is slow AND post-adaptation instability is unstable
THEN functional adaptability is very low.

This formulation allows the predictive maintenance system to be assessed in a way that is both quantitative and interpretable. The temporal indicators preserve the original operational assessment of functional adaptability, while the fuzzy layer reflects the uncertainty and context-dependence involved in judging whether the system has adapted adequately. As a result, the model provides an ISO/IEC 25059-aligned mechanism for evaluating whether the predictive maintenance system effectively incorporates new operational information and maintains reliable behaviour under changing industrial conditions.

In this scenario, functional adaptability requires limited degradation under changes in operating conditions, timely detection, substantial recovery after retraining, and stable post-adaptation performance. By contrast, a persistent increase in error, delayed detection, poor recovery, or unstable performance indicates insufficient adaptability. The model reports the dominant linguistic assessment, the activation degrees of the remaining categories, and a complementary defuzzified score. The linguistic profile supports expert interpretation, while the numerical score enables comparisons across models, forecasting horizons, and evaluation periods.

5 Discussion and Conclusions

This paper proposes an ISO/IEC 25059-aligned fuzzy linguistic framework for evaluating functional adaptability in industrial forecasting systems. The ap-

proach preserves the original operational view that adaptability should be assessed through observable temporal predictive behaviour, using degradation, detection delay, recovery gain, and post-adaptation stability. Its main contribution is a fuzzy inference architecture that integrates several temporal indicators into an interpretable assessment framework that combines quantitative evidence with expert knowledge.

The fuzzy model extends crisp threshold-based evaluation by representing the gradual and context-dependent nature of adaptability criteria. Membership functions and linguistic rules make explicit how domain experts interpret acceptable degradation, timely detection, effective recovery, and stable post-adaptation behaviour. The primary output is therefore a fuzzy linguistic assessment, while the defuzzified score is retained only as a complementary representation for longitudinal monitoring, comparison across models and forecasting horizons, and integration into industrial dashboards.

The expected benefit of the proposal is representational and decision-oriented due to it avoids abrupt changes around fixed thresholds, preserves partial satisfaction of quality criteria, and provides a traceable link between temporal indicators, activated rules, and the final assessment. Nevertheless, membership functions may introduce subjectivity, the rule base may contain inconsistencies if it is not carefully validated, and the reliability of the assessment depends on the quality of the underlying drift-detection and temporal-error estimation procedures.

Future work will empirically compare the fuzzy model with a crisp threshold-based baseline using the same temporal indicators. The evaluation will consider agreement with expert assessments, stability near decision boundaries, sensitivity to membership-function parameters, and explanation traceability. Further studies will also examine the transferability of the calibrated model across forecasting horizons and industrial domains, as well as its possible extension to other AI quality subcharacteristics defined in ISO/IEC 25059.

References

1. ISO/IEC: ISO/IEC 25059:2023. Software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality model for AI systems. International Organization for Standardization, Geneva (2023)
2. ISO/IEC: ISO/IEC 25010:2023. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Product quality model. International Organization for Standardization, Geneva (2023)
3. Albertini, W., Oviedo, J., Motta, E., Araujo, J., Araujo, M., Rodriguez, M., Lopes, H., Kalinowski, M., Piattini, M.: Practical insights into AI system product quality evaluation. IT Professional **28**(1), 81–87 (2026). https://doi.org/10.1109/MITP.2025.3646118
4. Lang, M.N., Lerch, S., Mayr, G.J., Simon, T., Stauffer, R., Zeileis, A.: Remember the past: A comparison of time-adaptive training schemes for non-homogeneous regression. Nonlinear Processes in Geophysics **27**(1), 23–34 (2020). https://doi.org/10.5194/npg-27-23-2020

5. Herrera, F., Herrera-Viedma, E.: Linguistic decision analysis: Steps for solving decision problems under linguistic information. Fuzzy Sets and Systems **115**(1), 67–82 (2000). https://doi.org/10.1016/S0165-0114(99)00024-X
6. Oliveira, G.H.F.M., Cavalcante, R.C., Cabral, G.G., Minku, L.L., Oliveira, A.L.I.: Time series forecasting in the presence of concept drift: A PSO-based approach. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 239–246. IEEE (2017). https://doi.org/10.1109/ICTAI.2017.00042
7. Zhang, Y., Wen, Q., Chen, W., Sun, L., Zhang, Z., Wang, L., Wang, X., Jin, R., Tan, T.P.: OneNet: Enhancing time series forecasting models under concept drift by online ensembling. In: Advances in Neural Information Processing Systems, vol. 36, pp. 69949–69980 (2023)
8. Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM Computing Surveys **46**(4), 1–37 (2014). https://doi.org/10.1145/2523813
9. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.F., Dennison, D.: Hidden technical debt in machine learning systems. In: NeuroIPS 2015, pp. 2503–2511 (2015)
10. Breck, E., Cai, S., Nielsen, E., Salib, M., Sculley, D.: The ML test score: A rubric for ML production readiness and technical debt reduction. In: IEEE International Conference on Big Data, pp. 1123–1132. IEEE (2017). https://doi.org/10.1109/BigData.2017.8258038
11. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies **7**(1), 1–13 (1975). https://doi.org/10.1016/S0020-7373(75)80002-2
12. Matzka, S.: Explainable artificial intelligence for predictive maintenance applications. In: 2020 Third International Conference on Artificial Intelligence for Industries (AI4I), pp. 69–74. IEEE (2020). https://doi.org/10.1109/AI4I49448.2020.00023