

AI Go-Around Decisions: Predicting Unstable Approaches

Published May 2, 2026 38 min read



Executive Summary

Unstable approaches during landing remain a perilous hazard in aviation safety. By regulation and industry best practice, approaches **must** be stable by defined “gates” – typically 1000 ft above field elevation in instrument conditions (IFR) or 500 ft in visual conditions (Source: [skybrary.aero](#)). If these stability criteria are not met, standard operating procedures mandate an immediate go-around (Source: [skybrary.aero](#)) (Source: [flightsafety.org](#)). Yet many accidents and incidents persist due to continuation of unstabilized approaches. Recent research shows that **machine learning (ML) and AI techniques applied to flight data (especially FOQA/FDM data)** can flag unstabilized approach profiles well *before* the conventional 1000/500 ft gates. Notably, ML models trained on Flight Operational Quality Assurance (FOQA) data can identify “unstable approach” precursors at higher altitudes (e.g. ~2000 ft above touchdown) with high accuracy (Source: [www.researchgate.net](#)) (Source: [www.researchgate.net](#)). For example, Random Forest classifiers have demonstrated ~80–85% accuracy in distinguishing unstable approaches from normal ones more than 2000 ft prior to landing (Source: [www.researchgate.net](#)) (Source: [www.researchgate.net](#)). Deep neural networks on publicly-available ADS-B data have similarly achieved ~82–85% accuracy in predicting unstable approach patterns (Source: [www.mdpi.com](#)). Furthermore, regression models (e.g. Random Forest regression) can reliably predict critical **landing performance metrics** (e.g. touchdown airspeed, ground speed) at approach decision altitudes, with errors on the order of only 2–3 knots (Source: [www.researchgate.net](#)). These results suggest that AI/ML can provide actionable early warnings – effectively moving the “go-around decision” gate upward by hundreds or even thousands of feet.

This report examines the state-of-the-art in **AI-assisted go-around decision-making**. We first review background on stabilized approach criteria, the FOQA/FDM program (data sources, parameters), and traditional safety analysis approaches. We then survey recent research where ML models are trained on FOQA (and related) data to detect unstable approach precursors. Key studies and case analyses are summarized, including quantitative results and identified features (e.g. high airspeed, glideslope deviations, thrust settings, etc. as critical predictors of instability (Source: [jglobal.jst.go.jp](#)) (Source: [www.researchgate.net](#)). Two comparative tables summarize: (1) major ML case studies and their performance, and (2) flight parameters found most indicative of unstable approaches. We also discuss practical considerations: real-time implementation challenges, **human factors in pilot decision support**, **regulatory implications**, and future work. The evidence indicates that **AI-driven early-warning systems could significantly enhance approach safety** by complementing existing stabilized-approach gates with predictive alerts, thereby giving crews more lead time to initiate go-arounds if required.

Introduction and Background

Stabilized Approach Criteria and Go-Around Decision Gates

Approach and landing phases account for a disproportionate share of aviation accidents and incidents (Source: [flightsafety.org](#)) (Source: [skybrary.aero](#)). To mitigate these risks, the **stabilized approach** concept has become a cornerstone of airline operating procedures. As summarized by the Flight Safety Foundation's Approach-and-Landing Accident Reduction (ALAR) program, "*all flights must be stabilised by 1000 feet above airport elevation in instrument meteorological conditions (IMC) and 500 feet above airport elevation in visual meteorological conditions (VMC)*" (Source: [skybrary.aero](#)). An approach is considered stabilized when the aircraft is on the [correct flight path](#), in the proper landing configuration (flaps, gear, speed), with only small changes in pitch/roll or descent rate needed to maintain the glidepath, and with other criteria (airspeed within $V_{REF} \pm 20$ kt, vertical speed ≤ 1000 ft/min, etc.) (Source: [skybrary.aero](#)) (Source: [skybrary.aero](#)). Critically, **if an approach becomes unstabilized below these gates, crews are trained to execute an immediate go-around** (Source: [skybrary.aero](#)) (Source: [flightsafety.org](#)). The U.S.NTSB has strongly reinforced this rule: "*If an approach becomes unstabilized, an immediate go-around is required... Never attempt to 'save' an unstabilized approach.*" (Source: [flightsafety.org](#)).

Despite these stringent criteria, the stabilised approach policy remains fundamentally **reactive**: lack of stability is confirmed only when one of the predefined criteria is exceeded at or below the gate altitude. In practice this means crews may not know an approach is irretrievably unstable until very late in the descent. Indeed, several high-profile accidents (and serious incidents) have occurred where flight crews continued unstable approaches below 1000 feet, resulting in runway overruns, controlled-flight-into-terrain (CFIT), or loss of control (Source: [flightsafety.org](#)) (Source: [flightsafety.org](#)). For example, the 2017 Teterboro Learjet crash was attributed to the captain "*attempt to salvage an unstabilized visual approach*", which led to a stall near the runway (Source: [flightsafety.org](#)). Conversely, if that crew had more advance warning of the instability, a go-around at 1000–1500 feet could have averted the accident. In fact, industry guidance increasingly encourages earlier "should-gate" calls (some operators designate a "*should go-around*" at about 1000 ft AGL and a "*must go-around*" at 500 ft AGL) (Source: [skybrary.aero](#)). However, even a 1000 ft warning can sometimes be barely adequate to initiate a go-around safely, especially in high-workload or degraded-visual conditions.

FOQA/FDM: Flight Data Monitoring Programs

Modern airliners generate massive amounts of flight data. The Flight Operations Quality Assurance (FOQA) / Flight Data Monitoring (FDM) program captures this data for safety analysis. FOQA is a **voluntary (often now practically mandatory)** safety program in which airlines collect and analyze data recorded by onboard flight recorders (Source: [www.mdpi.com](#)) (Source: [scaledanalytics.com](#)). A commonly cited FOQA definition (Enders, 1993) is "[a] program for obtaining and analyzing data recorded in flight to improve flight crew performance, air carrier training programs and operating procedures... and aircraft operations and design" (Source: [scaledanalytics.com](#)). In practice, a FOQA/FDM system continuously uploads thousands of parameters from each flight's Quick Access Recorder (QAR) to ground analysis tools. Typical FOQA data consist of high-resolution, multivariate time series (speeds, vertical rates, altitudes, attitude, configuration, thrust, control inputs, etc.) sampled at up to 16 Hz (Source: [www.mdpi.com](#)) (Source: [www.mdpi.com](#)). These parameters span system categories (atmospheric, attitude, navigation, engine, control surfaces, etc.) and yield an exhaustive depiction of each approach, landing, and all flight phases.

Because FOQA data are **standardized and comprehensive**, they are ideal for machine learning. Past FOQA analyses were largely based on rule-based exceedance reports: each parameter has pre-defined thresholds (Event Exceedance Analysis) and FOQA reports flag any threshold breaches (e.g. excessive angle of attack, high sink rate, or bank angle) after flight. While valuable, this approach only identifies events *after* they occur, rather than predicting them. In contrast, modern ML techniques can ingest FOQA datasets to discover complex patterns and **precursors** to safety events. Indeed, researchers note that the availability of standardized FOQA data with labeled events makes possible a repeatable ML pipeline for safety analysis (Source: [www.mdpi.com](#)).

By integrating FOQA/FDM with machine learning, airlines and regulators hope to shift from retrospective analysis to **proactive risk management**. In Europe, FOQA/FDM has even become de facto mandatory for major carriers' Safety Management Systems (Source: [scaledanalytics.com](#)). The stage is now set to leverage FOQA with AI: numerous studies have explored using ML algorithms (random forests, neural networks, SVMs, etc.) to identify safety-related events or precursors based on FOQA inputs (Source: [www.mdpi.com](#)) (Source: [www.mdpi.com](#)). The unstable-approach event is clearly a prime target for this approach, since it naturally involves detecting an emerging unsafe condition as the approach progresses.

Defining “Unstable Approach” in Data

In order to train ML models, one must define the “unstable approach” event in data terms. This is not trivial, since “unstabilized” can involve multiple parameters: speed deviations, descent angle, configuration, flight path, etc. Airlines typically have SOPs that specify stabilized approach criteria; correspondingly, some FOQA systems already tag an “unstable approach” event if ANY criterion (e.g. speed, sink rate, glide slope, flaps/gear) is not within limits at the 1000/500 ft gates. However, this single event label may occur very low in the approach – often triggered only when the crew already missed the gate.

In ML studies, researchers often *reformulate* the problem as **precursor identification**: spotting the early warning signs of an approach that eventually will be labeled “unstable.” For example, Ackley *et al.* (2020) collected FOQA data from 4500 flights and identified 1300 flights that contained an unstable-approach event as defined by FOQA thresholds (Source: www.researchgate.net). Using supervised learning, they then examined the sequences of FOQA parameters at various altitudes (2000 ft, 3000 ft, ...) prior to the event to see which combinations of variables best discriminated unstable from normal approaches (Source: www.researchgate.net). This time-series feature engineering – generating point-specific or staggered time-slice feature vectors – is common in aviation ML. Similarly, other authors have defined unstable approach precursors using FOQA or QAR data, often supplementing FOQA with contextual info (e.g. approach type, weather).

Flight Data Sources: Besides FOQA, researchers have used public tracking data (ADS-B) to explore these problems. For example, Chiu and Lai (2023) used ADS-B flight paths (open-source data) to derive energy-based features for Taiwan approaches (Source: www.mdpi.com) (Source: www.mdpi.com). While ADS-B lacks cockpit parameters, it illustrates the potential to apply ML even with limited data. In practice, however, FOQA/QAR remains the richest source, containing direct cockpit and control measurements used by most recent studies.

Machine Learning Goals: The typical ML goals are either (a) classification – distinguishing an unstable vs. stable approach at a given point in the approach, or (b) regression – predicting continuous landing metrics (e.g. touchdown speed, sink rate, landing distance) from approach data. Classification models (e.g. Random Forest, SVM, neural nets) output a probability that the current approach is “in jeopardy” of becoming unstable. Regression models predict landing performance so that deviations from normal can signal a risk (e.g. predicted touchdown speed much higher than reference suggests excess energy). Notably, Puranik *et al.* (2020) developed an **offline-online framework**: a Random Forest regression model predicted landing true airspeed and ground speed using FOQA approach data at specified altitudes. This model was accurate (RMSE ~2.6–3.0 knots) across multiple aircraft types (Source: www.researchgate.net). Crucially, because the model is fast to compute, it could generate these predictions at decision altitudes (~2000 ft or higher) – offering pilots a real-time indicator of whether the approach energy state was within safe limits (Source: www.researchgate.net).

In the sections that follow, we review these ML approaches in detail, summarizing data characteristics, modeling methods, and key outcomes. We examine what flight parameters emerge as the most relevant predictors (the “*precursors*”), how early in the approach detection is feasible, and what accuracy and lead-time the models achieve. We also discuss case studies and example incidents to illustrate how AI-assisted warning might alter crew decisions.

FOQA Data and Machine Learning Approaches

FOQA Data Characteristics

FOQA data archives contain long time-series traces of each flight’s parameters. A typical dataset might include hundreds of variables sampled at 1–16 Hz over each flight (up to 1 billion data points for a year of flights). As an example, Ackley *et al.* (2020) worked with 4500 flights’ FOQA data, originally 623 variables, pared down to ~270 after preprocessing and correlation analysis (Source: www.researchgate.net). The raw data include:

- **Navigation:** position, altitude (barometric and radio), distance-to-runway, lateral deviation, glidepath deviation (ILS dots), localizer deviation, heading error, etc.
- **Performance:** calibrated airspeed, ground speed, Mach number, vertical speed (V/S), angle of attack, accelerating vertical acceleration, etc.
- **Configuration:** flaps position, landing gear (up/down), spoilers, autopilot/flight director modes, autothrottle engaged, etc.
- **Power and Energy:** engine targets (N1, EPR, thrust lever angle), actual engine params (N1, EGT), total energy (combined speed/altitude), etc.
- **Attitude:** pitch angle, roll angle, bank angle, pitch rate, roll rate.
- **Atmospheric:** static pressure, temperature, wind vector if available, etc.
- **State Indicators:** landing light on, seatbelt sign (as proxy for landing phase), weight-on-wheels, etc.

Many of these variables are directly relevant to approach stability. For example, multiple “glideslope” and “localizer” deviation signals indicate how well the aircraft is aligned laterally and vertically with the runway. The “distance from airport” and “air miles to touchdown” features (derived from navigation) give situational context (e.g. how far out the plane is at each point). Speed and descent rates are obvious energy indicators. Everything is timestamped, so FOQA data can be sliced into segments at specific altitudes or time relative to the runway.

Crucially, FOQA datasets are **heterogeneous**: they contain numeric, discrete, boolean, and even some text fields (e.g. recorded METAR conditions) (Source: www.mdpi.com) (Source: www.mdpi.com). Preprocessing is therefore needed: features lacking data are dropped, outliers cleaned, etc. Studies often downsample or resample to uniform rates (commonly 1 Hz) to align time series for all flights (Source: www.researchgate.net) (Source: www.mdpi.com). Highly correlated parameters (e.g. multiple flaps surrogates, dual sensors) are filtered out to reduce redundancy. At the end, the ML pipeline typically works with a manageable set (tens to hundreds) of numerical features per time step, plus possibly some engineered features (e.g. *glideslope deviation envelope*, or *total energy sum*).

ML Modeling Strategies

Most studies use **supervised learning**, since FOQA allows labeling of flights/events. For unstable approach, label assignment is straightforward post-hoc (did the flight have an unstable approach exceedance?). Then one can train a binary classifier to predict that label from earlier data. Ackley *et al.* (2020) followed this route, building classification datasets at multiple “decision altitudes” (e.g., feature vectors at 3000 ft, 2000 ft, etc.) (Source: www.researchgate.net) (Source: www.researchgate.net). Each vector was labeled “event” or “non-event” based on whether the flight eventually triggered an unstable-approach exceedance. Random Forests and K-Nearest Neighbors were among the classifiers tried, often paired with a **feature-selection** step (Sequential Backward Selection) to reduce to a small subset of the most predictive features (Source: www.researchgate.net) (Source: www.researchgate.net).

Other studies have explored many algorithms. For example, a Japanese research summary (J-Global 2024) compared logistic regression, Long Short-Term Memory (LSTM) networks, random forests, and support vector machines for unstable-approach prediction across diverse fleets. They consistently found Random Forests to perform best in high-dimensional, imbalanced FOQA data, identifying **airspeed, glideslope deviation, flap position, and thrust** as the dominant features (Source: jglobal.jst.go.jp). Chiu & Lai (2023) used a deep neural network on ADS-B-derived features (total energy and lateral deviation) to both detect and analyze unstable approaches, achieving ~85% accuracy on their test set (Source: www.mdpi.com).

In parallel, **regression** approaches have been used to predict continuous landing metrics. The Puranik (2020) work described an “offline-online” pipeline: the model is trained offline on historical FOQA to predict, say, touchdown airspeed and ground speed from data above 1000–2000 ft (Source: www.researchgate.net) (Source: www.researchgate.net). Then at run-time the “online” model ingests live approach data to forecast the expected landing state. The idea is that if the predicted values deviate significantly from the planning values (e.g. predicted landing speed is 10 knots high), the crew may be alerted. Puranik *et al.* reported very low prediction errors (RMSE ~2.6 kt for airspeed, ~3.0 kt for ground speed) across multiple aircraft (Source: www.researchgate.net), and pointed out that the model can produce these predictions *before the go-around decision gate*.

Overall, the flow is: **(1)** Compile a FOQA dataset with labeled stable/unstable approaches. **(2)** Preprocess and possibly reduce features. **(3)** Split into training/testing sets. **(4)** Train supervised models at a chosen altitude (or multiple altitudes). **(5)** Evaluate performance (accuracy, precision/recall, RMSE for regression). Many papers report performance improving after dimensionality reduction – e.g. Ackley *et al.* saw similar accuracy with only 5–35 features compared to hundreds (Source: www.researchgate.net) (Source: www.researchgate.net).

Technical challenges include class imbalance (far fewer unstable events than normal flights) and overfitting. Researchers mitigate these by balancing the training data, using cross-validation, and focusing on a reduced feature set. Visualization and statistical tests (e.g. p-values comparing feature distributions (Source: www.researchgate.net) are often used to interpret the model – essentially confirming that the ML-identified precursors align with pilot knowledge.

Key Predictive Features

A major outcome of these ML studies is the identification of which FOQA parameters are most “predictive” of an unstable approach. Across different algorithms and altitudes, **consistently emergent features** include:

- **Airspeed / Excess Energy:** High or rising approach speed (relative to V_{REF}) is a strong precursor. Most unstable approaches have higher mean airspeeds than nominal (e.g. 165 kt vs 159 kt) (Source: www.researchgate.net). ML rankings and statistical tests highlight airspeed as a top discriminator (Source: www.researchgate.net) (Source: jglobal.jst.go.jp). (A fast aircraft has more kinetic energy and is harder to slow down in

time.)

- **Glideslope Deviation:** Both single-dot and double-dot ILS deviations figure prominently. Studies list “Glideslope Deviation (2 dots)” and (1 dot) among top features (Source: www.researchgate.net). Large, persistent deviation from glideslope early on signals an unstable approach path. Similarly, high values of *localizer* deviation or “lateral distance from centerline” are predictors of misalignment (Source: www.researchgate.net).
- **Configuration (Flaps/Thrust):** Inadequate or late configuration is key. A low flap setting (higher “flap position”) correlates with instability, as does low thrust (low N1 target) (Source: jglobal.jst.go.jp) (Source: www.researchgate.net). In practical terms, if an aircraft is still in a high-speed flap setting or idle thrust too close to the runway, an ML model flags it. For example, Ackley et al. found “N1 Target” and “Glideslope Hold Engaged” among the few features needed for accurate prediction at 2000 ft (Source: www.researchgate.net).
- **Descent Rate / Vertical Speed:** High descent rates (exceeding ~1000-1500 fpm) often precede an unstable outcome. Although vertical speed alone may not top the list, it is used in combination: ML models often include it in their minimal feature sets (Source: www.researchgate.net).
- **Kinematic Features:** Other related metrics like energy height (combining airspeed & altitude), and “pitch attitude” or angle-of-attack can be predictive. Some models derive “total energy” measures and show that abnormal energy trends precede instability (Source: www.mdpi.com).
- **Environment/Terrain:** A few models even use airport elevation or runway slope as predictors, since a given performance may be stable at one field but not at another. For instance, “Terrain Database Elevation” was selected among top features in the 2000 ft classifier (Source: www.researchgate.net) – perhaps reflecting that approaches into high-altitude airports have narrower safety margins.

Table 2 below summarizes key features frequently identified by ML as unstable-approach precursors, along with their significance. These are drawn from multiple sources (Source: www.researchgate.net) (Source: jglobal.jst.go.jp) (Source: www.researchgate.net).

FLIGHT PARAMETER	ROLE AS UNSTABLE-APPROACH INDICATOR
Airspeed	High airspeed (above reference $V_{REF} \pm 20$ kn) indicates excess energy. ML studies consistently rank airspeed as a top feature (Source: jglobal.jst.go.jp). Unstable approaches often maintain above-normal speeds late in descent.
Glideslope Deviation	Large vertical guidance deviations (on ILS, >1 dot or 2 dots) signal misalignment with glide path. ML models list <i>GS deviation (1-dot/2-dot)</i> among top predictors (Source: www.researchgate.net). Persistent positive glideslope error (undershooting) was noted as a precursor.
Lateral Deviation	Lateral offset from runway centerline indicates poor alignment. "Distance from centerline" is often used in classification features (Source: www.researchgate.net). Significant offset increases landing risk.
Flap/Gear Configuration	Inadequate/sluggish configuration changes are warning signs. Sustaining a higher flap setting or delay in gear extension at low altitudes was flagged by ML. For example, an ML feature "flap position" was identified as critical in one study (Source: jglobal.jst.go.jp).
Thrust Setting (N1)	Low thrust (idle or below-minimum) in late approach is a precursor. ML models highlight <i>N1</i> or <i>thrust-lever target</i> as important (Source: www.researchgate.net) (Source: jglobal.jst.go.jp). Insufficient thrust causes high descent rates.
Vertical Speed	Excessive sink rates (>1,000 fpm) often presage a failure of stabilization. Though not always a singled-out feature, high descent was statistically correlated with unstable events (Source: flightsafety.org).
Total Energy / Pitch	Unusually high total energy or nose-up pitch inputs close to runway (without increase in power) can predict a stall/too-slow condition. Some DNN models learned to detect these subtle energy states (Source: www.mdpi.com) (Source: jglobal.jst.go.jp).
Distance to Runway	The remaining distance can contextualize other cues: for example, being far from runway with high speed/altitude might be safer, but being <i>close</i> and too fast is critical. "Distance from landing airport" was one of the top features in a Random Forest model (Source: www.researchgate.net).
Terrain Elevation	High airport elevation (thin air) effectively reduces climb margin. ML occasionally uses terrain elevation as a factor (Source: www.researchgate.net). Approaches in mountainous areas have tighter margins.

Each of these parameters can be monitored continuously during an approach. The advantage of ML is that it does **not require a single "hard" threshold** for each; instead, it learns patterns and combinations of these cues that together indicate impending instability. For example, a moderately high airspeed alone might not trigger an alert, but high airspeed *combined with* glideslope undershoot *and* idle thrust would. The ML model weights such combinations based on training data.

AI/ML Models for Early Unstable-Approach Detection

In this section we review specific research efforts that apply ML to predict unstable approaches, focusing on **when** and **how well** they flag problems relative to the traditional gates.

Random Forest Classifiers (Data-Driven Precursors)

Ackley et al. (2020) – In an AIAA conference paper, Ackley and colleagues (affiliated with Delta and Boeing/GA Tech) implemented a sequential machine-learning pipeline on real airline FOQA data (Source: www.researchgate.net) (Source: www.researchgate.net). They first identified all flights labeled as having an unstable approach (by standard FOQA exceedances). From ~4500 flights, 1300 unstable flights were found (Source: www.researchgate.net). They then extracted fixed-point feature vectors at several altitudes above ground (3000 ft, 2000 ft, 1000 ft, 500 ft) to train/test classifiers. Using Random Forest models with sequential backward feature selection, they found that at **2000 ft** AGL an unstable approach could be

predicted with “a reasonable level of accuracy” using as few as 5–35 features (Source: www.researchgate.net). In fact, they report test-set accuracies exceeding 80% with only ~20 parameters (and even >82% with just 12–35 features) (Source: www.researchgate.net) (Source: www.researchgate.net). Performance remained robust even with only 5 selected features, implying a small subset of parameters is highly indicative.

Perhaps most impressively, the study found that *two thousand feet prior* to the eventual unstable event, the model could already distinguish unstable flights with “reasonably high accuracy” (Source: www.researchgate.net). In their words: “*the algorithms can predict unstable approach events at least two thousand feet prior to the event with a reasonably level of accuracy.*” (Source: www.researchgate.net). (By comparison, the current industry gate is 1000 ft for IFR.) This implies that an AI system could flag the danger twice as early as the mandated threshold. Ackley *et al.* also performed statistical analyses on the top features and found their identified precursors aligned with intuitive factors: e.g. unstable flights at 2000 ft had significantly higher average airspeeds and more variable flaps/pitch settings than stable flights (Source: www.researchgate.net) (Source: www.researchgate.net).

Lepez Da Silva Duarte et al. (2024) – A recent (Japanese-translated) study similarly applied ML to FOQA data for unstable approaches. They compared logistic regression, LSTM neural nets, random forests, and SVMs, and again found Random Forests clearly superior for detecting instability in heterogeneous data (Source: jglobal.jst.go.jp). Their feature-importance analysis confirmed that **airspeed, glideslope deviations, flap position, and thrust setting** came out on top as predictors (Source: jglobal.jst.go.jp). While exact accuracy figures were not disclosed in abstracts, the authors emphasize the consistency of this result across multiple aircraft types. Their work supports Ackley’s finding that a small set of energy and configuration parameters suffices for early warning.

Additional Classifier Studies: Various other efforts have reported comparable results with different algorithms. For instance, tree-based ensemble methods (Random Forest, Gradient Boosting) often achieve 75–90% classification accuracy for unstable approaches using FOQA (Source: jglobal.jst.go.jp) (Source: www.mdpi.com). Some research also applies conformal methods or anomaly detection (unsupervised) for flagging outlier approaches, though supervised learning has been most common.

Table 1 (below) summarizes key ML studies on unstable approach/pre-cursor detection, including data sources, methods, and reported performance.

STUDY (YEAR)	DATA SOURCE	ML METHOD	APPROACH & ALTITUDE	PERFORMANCE / NOTES
Ackley <i>et al.</i> (2020) (Source: www.researchgate.net) (Source: www.researchgate.net)	FOQA data (4500 flights, ~1300 unstable)	Random Forest + feature selection	Classification at points (1000–3000 ft). Best results at 2000 ft.	~82–83% accuracy with ~35 features at 2000 ft (Source: www.researchgate.net). Predict unstable events ≥2000 ft before landing (Source: www.researchgate.net).
Lepez Da Silva <i>et al.</i> (2024) (Source: jglobal.jst.go.jp)	FOQA data (several thousand flights)	Random Forest (compared LR, SVM, LSTM)	Classification of unstable approach based on FOQA	RF consistently best. Identified key precursors (airspeed, glideslope, flaps, thrust) (Source: jglobal.jst.go.jp). (Exact accuracy not quoted.)
Chiu & Lai (2023) (Source: www.mdpi.com)	ADS-B (public) data (Taipei approaches)	Deep Neural Network (FDNN)	Detection of unstable approaches (energy-based)	Accuracy ≈85.15% (energy), 82.11% (trajectory dev) (Source: www.mdpi.com). Robust across weather conditions.
Kumar <i>et al.</i> (2021) (Source: www.mdpi.com)	ADS-B (9M landings, GOAR dataset)	Machine Learning (various)	Go-around classification (not specifically FDM)	Introduced large GA dataset; used GLMs and ML to model runway GA rates. Relevant for GA risk but not direct FOQA.
Puranik <i>et al.</i> (2020) (Source: www.researchgate.net)	FOQA (multi-airframe)	Random Forest regression	Regression to predict landing metrics (IAS, GS)	RMSE 2.62 kn (TAS), 2.98 kn (GS) (Source: www.researchgate.net). Model fast, can provide predictions at decision altitudes for go-around aid.
Monstein <i>et al.</i> (2022) (Source: www.mdpi.com)	ADS-B (OpenSky dataset)	Statistical & ML analysis	GA rate modeling (macroscopic)	Offered a large GA dataset (9M flights) for GA studies. Not directly on unstable approach but highlights GA risk factors.
Lorente <i>et al.</i> (2024) (Source: www.sciencedirect.com)	Proprietary FOQA-like (China)	Deep MIL + Knowledge Distillation	Real-time precursor detection	Reported high accuracy (≈95%) for general safety event precursors, with ~0.951 final accuracy (Source: www.sciencedirect.com). Demonstrated precursors can provide ~8 seconds warning on average.

Notes: The performance figures above are indicative; exact metrics depend on data selection and definitions. Nonetheless, the studies agree that ML identification of unstable approaches is viable at altitudes well above the 1000-ft gate.

Regression-Based Landing Performance Models

While classification predicts a binary “unstable” label, some ML work instead forecasts numeric landing outcomes, which can implicitly signal instability. Puranik *et al.* (2020) exemplified this by training Random Forest regressors on FOQA data to predict **landing airspeed** and **ground speed**. They used data from six airline fleets at over 70 airports. The model achieved very low errors (RMS 2.62 and 2.98 knots) (Source: www.researchgate.net). Importantly, because prediction is extremely fast (~milliseconds on modern hardware), they proposed using this offline-trained model in an onboard setting: input the approach data (e.g. up to 3000 ft) and obtain a prediction of touchdown speed. If the predicted final speed is much higher than planned V_{REF} , the flight crew could be alerted to initiate a go-around. The authors specifically note that the framework “**can provide prediction at an altitude where a go-around decision may be made**”, making it “particularly suitable for online application” (Source: www.researchgate.net).

A similar concept is to predict *landing distance* or *vertical speed at touchdown*. While direct ML studies on those are fewer, the approach is analogous. Because FOQA contains all needed predictors, and results show high cross-type robustness (Source: www.researchgate.net), this avenue holds promise. In essence, regression models convert many FOQA parameters into one or two principal risk metrics, which pilots already monitor (speed and sink rate). If those metrics are predicted to violate limits, the ML system would raise a flag early.

Feature Reduction and Explainability

One concern with any ML system is explainability—pilots and regulators want to know *why* the system is issuing a warning. The good news is that in this domain the ML-derived precursors align well with pilot intuition about unstable approaches. For example, in Ackley's analysis, the highest-ranked features (at 2000 ft) were things one would expect: distance-to-runway, glidepath error, ground speed, thrust setting, vertical speed, etc. (Source: www.researchgate.net). Likewise, Lepez Da Silva's feature importance agreed on airspeed, thrust and glide path. Even "black-box" neural methods like the one used by Chiu & Lai can be partially interpreted by analyzing input importances and clustering (they used HDBSCAN to isolate abnormal flights) (Source: www.mdpi.com). In practice, an AI warning tool could display the most significant indicator (e.g. "Approach speed too high" or "glideslope deviation increasing") alongside the alert, helping crews trust the system.

The studies often reduce thousands of FOQA parameters to a core handful. For example, Ackley *et al.* found that using just **5 features** still gave respectable accuracy (Source: www.researchgate.net), and that expanding to 20–35 features saturated performance (Source: www.researchgate.net). This validates that FOQA's depth can be distilled: a few key sensors carry most of the predictive power. In our Table 2 above, we listed some of these top predictors, drawn from multiple ML analyses (Source: www.researchgate.net) (Source: www.researchgate.net) (Source: jglobal.jst.go.jp). While the exact ranking can differ, they consistently emphasize energy/airspeed, path integrity (glide/localizer), and configuration (flaps/thrust).

Conceptually, the ML model acts as a multivariate filter: instead of watching *one* parameter against a threshold (e.g. "gear not down by 1000 ft"), it watches a combination (e.g. high speed *and* large descent rate *and* IDLE thrust). Thus, it is sensitive to more subtle instability patterns (e.g. slowly decaying airspeed or a creeping glideslope intercept from above). This allows an earlier alert: instead of waiting to miss a threshold, the system can say "the current trend is moving toward an unstable outcome."

Case Studies and Examples

To illustrate the problem and potential benefits of AI assistance, we present representative incidents of unstable approaches and consider how an AI system might have intervened earlier.

- **Learjet 35A, Teterboro NJ (2017):** As noted by the NTSB, the crew "attempt[ed] to salvage an unstabilized visual approach," resulting in a stall short of the runway (Source: flightsafety.org). This accident underscores how continuation bias can cause a stabilized-approach SOP to be ignored. An AI-based warning (e.g. "unstable approach – recommend go-around") triggered as early as 2000 ft could have overruled the bias.
- **A320, Muscat OMA (2019):** An Orange2Fly A320 on ILS at Muscat became unstabilized around 930 ft AGL due to **autothrust being inadvertently off** (Source: skybrary.aero) (Source: skybrary.aero). The autopilot was disconnected high on approach and the crew did not realize thrust was idle. Pitch rate increased, triggering "Alpha Floor" (automatic TO/GA). No accident occurred, but the approach was clearly unsafe. Notably, the aircraft's vertical speed (~1080 fpm) and glide-slope error (half-scale deflection) had already exceeded typical stab criteria by 600 ft (Source: skybrary.aero). An AI fuse might have recognized earlier that *energy was bleeding off (airspeed decreasing, insufficient thrust)* and *glide path error* was growing, and thus warned at, say, 1500–2000 ft. Indeed, the ML-identified features (thrust IDLE, high sink rate, glideslope deviation) exactly match the cues in this event (Source: www.researchgate.net) (Source: skybrary.aero).
- **B737-800, Lanzarote Spain (2019):** In this TUI Airways case, the crew descended well below the non-precision VOR profile once the runway was in sight, triggering EGPWS "CAUTION TERRAIN" and "PULL UP" alarms (Source: skybrary.aero). The autoland systems were disconnected and the crew levelled off, recovering safely. In effect, this was an unstabilized approach (normals would profile to the published descent path). QAR data analysis showed the aircraft was ~700 ft below the required altitude 2 NM before the FAF (Source: skybrary.aero). An ML system trained on similar approaches could have spotted the early descent lapse. For example, the combination of *below-gate altitude + near-limit descent + visual conditions* should raise a flag. The GA and collision with terrain alarms activated later, but a data-driven system could have warned earlier (hundreds of feet sooner).
- **Miscellaneous:** Other reported incidents (e.g. Saab 340 near CFIT, etc.) often share these elements: high speed, unusual pitch/energy, and misaligned glide paths at relatively high altitude. Even some runway excursion incidents start as unstable approaches (touchdowns too far down, too fast). In each, the FOQA-triggered "unstable approach" event likely occurs at or below 1000 ft – too late to recover. ML analysis could extend the warning envelope substantially earlier.

These cases highlight both the promise and the real-world context. In some cases, local automation (like Alpha-floor in the A320) partially rescued the flight. But relying on such last-second protections is risky. Ideally, the system would **augment the flight crew** by generating an advisory well before critical terrain or energy states occur. To study real-world applicability, some researchers have even proposed integrating ML outputs with Flight Management Systems or Heads-Up Displays, showing a “risk meter” or pop-up annotation. For example, if an ML classifier outputs a high probability of “unstable approach,” the interface could display “Energized/NRST!” (Near or Dangerous maneuver!). Prototypes might leverage electronic flight bag apps or EICAS messages.

Data Analysis: Evidence from Studies

The published studies provide quantitative evidence of feasibility. Key findings include:

- **Early detection at high altitude:** ML models consistently flag instability above 1000 ft. Ackley et al. stress detection around 2000 ft (Source: www.researchgate.net) (Source: www.researchgate.net). In one experiment, even at 3000 ft their classifier had “reasonable accuracy” (Source: www.researchgate.net). Chiu & Lai’s DNN, though using sparse ADS-B-derived features, achieved >80% accuracy on data segments well over the runway threshold (Source: www.mdpi.com) (though exact detection altitude is not stated, their training set appears to span entire approach segments).
- **Feature-based analysis matches intuition:** Statistical significance tests (p-values) for features confirm that ML-chosen features diverge markedly between stable vs unstable approaches. In Ackley’s Table 6 (Source: www.researchgate.net), almost all top features (distance, airspeed, glide dot error, vertical deviation, etc.) show extremely low p-values, meaning they robustly differ. Visualization of distributions (not cited here, but described in papers) show unstable flights clustering at higher speeds and larger errors. The ML models therefore rest on physically meaningful cues.
- **Model performance:** Accuracy metrics in the literature (when reported) are strong. For instance, Ackley’s best RF model at 2000 ft achieved ~83% test accuracy (Source: www.researchgate.net). Chiu’s neural net gave ~85% on one metric and ~82% on another (Source: www.mdpi.com). Even though these are not perfect, in a safety context an 80–85% alert with proper margin and crew judgement could still prevent more accidents than a 100% reactive rule. (Future work might ensemble models or fuse data to improve.)
- **Prediction vs. threshold rules:** Importantly, ML can catch unstable trends that simple threshold rules might miss. For example, consider an approach with airspeed +15 kt above V_{REF}, descent 800 fpm, within glideslope limits but heading slightly off. No single parameter may exceed a “hard limit”, so threshold monitors might pass the approach as “stable” at 1000 ft. A trained ML model, by contrast, has learned from historical data that this combination often *leads to* future exceedances. Thus it would flag it earlier. This “pattern detection” nature is why ML can outperform static gates.

The case studies and data evidence together suggest **substantial lead time gain**. If the typical approach gate is 1000 ft (allowing ~20 seconds out time to decide/go-around), an ML-based system that triggers at 2000–3000 ft could give an extra 20–40 seconds. Those extra seconds are critical for crews to recognize the issue, brief the maneuver, and initiate a clean escape path, especially in high-workload or single-pilot situations.

Implications and Future Directions

The demonstrated potential of ML-based early warning raises a number of broader considerations:

- **Human Factors and Trust:** Any AI assistance must be designed to ensure pilot trust and situational awareness. Pilots must see *why* an alert is issued. Ideally, the system would highlight key parameters (e.g. “High descent + excess speed”) in an intuitive display, rather than a cryptic probability. Studies in flight deck automation caution against unexpected or unexplained alerts. Therefore, explainability (grounding the alert in understandable terms) is as crucial as raw accuracy. SESAR and research groups emphasize the need for *human-centered AI*, where the pilot always has the final authority but is aided by predictive cues.
- **Real-Time Data Availability:** FOQA data is typically collected after flight, not continuously available onboard. For this concept to work, needed data must stream or be accessible during flight. Most required parameters are already available to pilots (airframe sensors, FMS). A prototype system could run onboard by tapping into existing avionics data buses or via an Electronic Flight Bag app using ACARS. One technical challenge is ensuring low latency: the models must predict and update at least every second or two during approach, as data arrives. The cited ML models are generally fast (decision trees and even DNNs can run in real time on modern hardware) (Source: www.researchgate.net).

- **Regulatory Acceptance:** Introducing AI for safety-critical decisions (like go-around) will require rigorous validation. Regulators will expect thorough testing, including in pilots-in-the-loop simulators. There are also liability considerations: ultimately the crew decides, but if an AI fails to warn when needed, or issues false alarms, how is responsibility handled? Early involvement of FAA/EASA will be important. The FAA's recent push for Performance-Based Navigation and ADS-B compliance suggests openness to integrating advanced systems, but an AI go-around advisor would be a new domain.
- **Integration with Cockpit Procedures:** Standard Operating Procedures would need to evolve. For instance, airlines might train crews to respond to an "AI Unstable Approach Warning" similarly to a GPWS pull-up but with planned go-around profile. The phrasing and alerting conventions matter. Would the system call for an immediate missed approach, or simply display a caution? This depends on conservatism – likely it would be advisory unless highly confident.
- **False Alarms versus Missed Detections:** The balance between sensitivity and nuisance alerts is critical. As with wind-shear or TCAS, too many false warnings can cause crews to disable the tool ("cry wolf" effect). Therefore, thresholding the ML output (e.g. only warn at very high risk level) must be carefully tuned. Research into this trade-off (analogous to ROC analysis) is needed. The studies cited mostly report overall accuracy; fewer discuss false-positive rates. Companies deploying such systems would likely build in adaptivity (e.g. weighting alerts by weather or phase of flight).
- **Extending to Other Events:** The concept of "precursor-based warnings" is broadly applicable. Unstable approaches share underlying patterns with limited go-around performance or runway excursions. A comprehensive AI safety system could also incorporate runway status (blocked runway ahead), weather hazards, or even crew performance metrics. One academic group (Lorente *et al.*, 2024) has already used multiple-instance learning to generate real-time risk scores for generic safety incidents, finding that pilots had an average of ~8 seconds to react after receiving a precursor alert (Source: www.sciencedirect.com). This hints at a future where FOQA+AI continuously monitors many risk vectors.
- **Future Research Directions:** Further work is needed on model robustness. For example, approaches to runway 28L may have different nominal profiles than to runway 10R; models should account for airport and aircraft-specific effects. Transfer learning between fleets, dealing with noisy or missing data, and combining FOQA with weather/traffic inputs are all active research areas. Additionally, field trials in simulators or actual flights would provide invaluable feedback. For instance, one could flight-test announcing "warning" messages to pilots in a simulator to see if they heed an AI go-around alert when they otherwise would have continued.

In summary, AI-assisted go-around decision tools appear technically feasible and promise meaningful safety benefits. They can turn FOQA from a post-flight audit tool into a real-time flight-deck aid. As machine learning models and on-board computing advance, the era of predictive flight safety may not be far off.

Tables

Table 1. *Summary of ML approaches for unstable-approach/go-around prediction and their performance.* (FOQA: Flight Operational Quality Assurance data; ADS-B: Mode S/ADS-B tracking data)

STUDY (YEAR)	DATA	ML TECHNIQUE	PREDICTION TASK	KEY RESULT / PERFORMANCE
Ackley et al., 2020 (Source: www.researchgate.net) (Source: www.researchgate.net)	FOQA (4500 flights; 1300 unstable)	Random Forest classifier (+ feature selection)	Classify unstable vs. normal approach at altitudes (3000', 2000', ...)	~82–83% accuracy at 2000' AGL with ≈20–35 features (Source: www.researchgate.net). Can detect unstable approaches ≥2000' before landing (Source: www.researchgate.net).
Lepez Duarte et al., 2024 (Source: jglobal.jst.go.jp)	FOQA (several thousands of flights)	Random Forest (LR, SVM, LSTM compared)	Classify unstable approaches	RF outperformed others (handling imbalance). Key predictors: airspeed, glideslope error, flaps, thrust (Source: jglobal.jst.go.jp) (consistently across fleets).
Chiu & Lai, 2023 (Source: www.mdpi.com)	ADS-B tracks (Taipei Airport)	Fully-connected Deep Neural Network	Detect unstable approach (energy-based features)	85.15% accuracy for energy deviations; 82.11% for trajectory deviation (Source: www.mdpi.com). Similar performance across weather conditions. (ADS-B limit: no cockpit data, but still high accuracy.)
Puranik et al., 2020 (Source: www.researchgate.net)	FOQA (6 airframe types, 70+ airports)	Random Forest regression	Predict landing metrics (TAS, GS)	RMSE 2.62 kn (TAS) and 2.98 kn (GS) (Source: www.researchgate.net). Prediction fast; can provide values at decision altitudes (e.g. ~2000') to inform go-around.
Monstein et al., 2022 (Source: www.mdpi.com)	ADS-B (9M landings, OpenSky 2019)	Statistical models, GLM	Analyze go-around rates	Introduced the largest open go-around dataset. Demonstrated using GLMs to model runway go-around rates. (Dataset now aids ML research.)
Lorente et al., 2024 (Source: www.sciencedirect.com)	Proprietary FOQA-like (23,549 flights, China)	Deep precursor model (TCN + MIL)	Real-time precursor identification	Global incident prediction accuracy 95.11% , up from ~91–94% for baselines (Source: www.sciencedirect.com). On average yields ~8 s crew reaction time post-warning (for general incident risk).

Table 2. Key flight parameters identified by ML as precursors to unstable approaches. These features were among those most predictive in FOQA-based ML models (Source: www.researchgate.net) (Source: jglobal.jst.go.jp) (and can be monitored continuously during approach).

PARAMETER	ROLE/SIGNIFICANCE
Airspeed	High approach speed implies excess kinetic energy. ML studies rank it as a top predictor (Source: jglobal.jst.go.jp) (Source: www.researchgate.net)—unstable approaches typically have mean speeds several knots above nominal.
Glideslope Error	Vertical de-tracking from the ILS glide path. Large positive error (below glideslope) indicates high descent rate or late descent. Appears as one of top two features (Source: www.researchgate.net) (Source: jglobal.jst.go.jp).
Lateral Deviation	Horizontal offset from runway centerline/localizer. Significant deviation may precede runway excursions. ML models include “distance from centerline” in the top features (Source: www.researchgate.net).
Throttle Setting (N1)	Engine thrust level. Idle or low-thrust approaches can cause uncontrolled sink. ML features “N1 Target” or similar often appear as critical (Source: www.researchgate.net) (Source: jglobal.jst.go.jp).
Flap Position	Landing flap setting. Late or improper flap extension is a risk. ML analysis flagged “flap position” as key (Source: jglobal.jst.go.jp)—e.g. flying long final-level with flaps up is a precursor.
Vertical Speed	Rate of descent. High V/S (over ~1000 fpm) often precedes instability. Though interrelated with glideslope error, it is sometimes used directly or inferred. FLT analysts note such approaches exceed SOP limits (Source: flightsafety.org).
Aircraft Pitch/Attitude	Lots of nose-up or trailing flare with insufficient thrust can indicate impending stall. Some models derive angle-of-attack or thrust deficiency; steep pitch coupled with decelerating speed can trigger instability alerts.
Total Energy Trend	A combined measure (altitude + kinetic energy). ML or clustering methods (like HDBSCAN) reveal outliers where total energy is inconsistent with normal approach profiles (Source: www.mdpi.com).
Distance to Runway	Context for other cues. Being far out with high energy might be recoverable, but remaining distance at decision altitudes amplifies risk. “Distance from airport” was a top feature for the RF model at 2000 ft (Source: www.researchgate.net).
Airport Elevation/Terrain	High-altitude airports (e.g. Aspen, La Paz) reduce engine performance and require more conservative profiles. Some models include terrain/elevation factors (Source: www.researchgate.net). ML can learn that the same speed might be safe at sea level but not at high field elevation.

Conclusion

Continued unsafe outcomes from unstabilized approaches have prompted exploration of novel decision-support tools. This report has shown that **AI-driven systems, trained on FOQA and flight trajectory data, can play a valuable role** in the go-around decision process. By analyzing historical flight data, ML models have learned to recognize subtle combinations of parameters that herald an unstable approach. Empirical evidence from multiple studies indicates that a well-designed ML classifier can label an approach as “at risk” **hundreds of feet ahead** of the conventional 1000-foot rule (Source: www.researchgate.net). In practical terms, this translates to several additional seconds of warning time for the crew – often enough to positively affect the outcome.

Our review highlights that the critical predictive features (airspeed, thrust, configuration, glideslope adherence, etc.) are ones pilots already monitor, but ML has the advantage of watching them all simultaneously and adaptively. In essence, an AI co-pilot is summarizing the collective wisdom of thousands of previous flights. The relative simplicity of some of the top features (as few as 5–20 variables) also bodes well for on-board implementation; it suggests that real-time algorithms would not be prohibitively complex or opaque.

The broader implications are significant. **Safety enhancement:** With early warning, crews can avoid continuing marginal approaches, likely preventing runway excursions, stall events, and CFIT. **Regulatory evolution:** If validated, such tools may lead regulators to formally revise stabilized approach guidelines, possibly introducing AI-enabled predictive gates or advisory modes. **Pilot training and SOPs:** Operators will likely incorporate AI-warning

response into training (for instance, procedures for when an “Unstable Approach” alert appears on the EFB or PFD). **System design:** There is an emerging need for flight deck interfaces that effectively deliver AI insights (e.g. composite risk indices, color-coded alerts, or recommended go-around maneuvers).

In conclusion, integrating machine learning with FOQA data represents a promising step toward proactive flight safety. By flagging deteriorating approaches in real time, such systems can extend the safety envelope beyond current static stab criteria (Source: www.researchgate.net) (Source: www.researchgate.net). While challenges remain in implementation, human factors, and certification, the research-to-date is clear: AI can detect flight instability patterns that humans alone may miss, and do so with ample lead time. Future work will refine these models, validate them in live operations, and ultimately potentially make the go-around decision **smarter and safer** than ever before.

Sources: This report synthesizes findings from aviation safety and machine learning literature (Source: www.mdpi.com) (Source: skybrary.aero) (Source: www.mdpi.com) (Source: www.researchgate.net) (Source: www.researchgate.net) (Source: flightsafety.org), including case reports and recent open-access research on FOQA-based predictive models (Source: jglobal.jst.go.jp) (Source: www.researchgate.net).

Tags: ai go-around decisions, unstable approaches, foqa data, machine learning aviation, flight data monitoring, predictive safety, aviation safety, approach gates

DISCLAIMER

This document is provided for informational purposes only. No representations or warranties are made regarding the accuracy, completeness, or reliability of its contents. Any use of this information is at your own risk. Landing Aero shall not be liable for any damages arising from the use of this document. This content may include material generated with assistance from artificial intelligence tools, which may contain errors or inaccuracies. Readers should verify critical information independently. All product names, trademarks, and registered trademarks mentioned are property of their respective owners and are used for identification purposes only. Use of these names does not imply endorsement. This document does not constitute professional or legal advice. For specific guidance related to your needs, please consult qualified professionals.