

Key performance metrics

This page is from APP, the official source of professional practice for policing.

First published 21 March 2022

Written by College of Policing

2 mins read

This section outlines the minimum requirements and additional metrics or indicators that are relevant and suitable for collation and analysis for operational deployments. Additionally, as part of the force procurement process, due diligence should be given to expected algorithm performance (or accuracy).

The National Institute of Standards and Technology (NIST) regularly undertake large-scale facial recognition system tests. Although these provide a good starting point, it is incumbent upon the system owner to understand how these tests relate to the accuracy and equitability of their deployed algorithm. While publicly available test data from NIST can inform owners, it will usually be informative to measure accuracy of the specific operational algorithm on operationally realistic data.

There are two key metrics that determine the 'accuracy' of an LFR system and a third that details the time taken to generate an alert.

True recognition rate (TRR)

This is also referred to as the true positive identification rate.

The TRR is the number of times when individuals on a watchlist are known to have passed through the zone of recognition and the LFR system correctly generated an alert, as a proportion of the total number of times these individuals passed through the zone of recognition (regardless of whether an alert is generated).

This metric can only be generated by 'seeding' known subjects (for example, police officers or staff) into a **Blue Watchlist** and measuring the number of times those subjects are present in the zone of recognition against the number of alerts generated. Users of LFR systems (and vendors) should not focus so closely on maximising this metric, as it may increase the false alert rate to an extent that is

not possible to manage the number of false alerts.

False alert rate (FAR)

This is also referred to as false positive identification rate.

This is the number of individuals who are not on the watchlist but generate a false alert or confirmed false alert, as a proportion of the total number of people who pass through the zone of recognition.

Further information

The NPCC Facial Recognition Board are available to support decision making.

All of the TRR and FAR metrics should be recorded and reported to the SRO. Operational experience to date suggests that the FAR should be 0.1% or less (less than 1 in 1,000) in most scenarios. It should be noted that the number of false alerts generated is greatly affected by the number of subjects processed by the LFR system and, to a lesser extent, the size of the watchlist.

The configurable threshold (the point at which two images being compared will result in an alert) will have a direct impact on the TRR and FAR. The threshold needs to be set with care so as to maximise the probability of returning true alerts, while keeping the number of false alerts to acceptable levels, as determined by the SRO in light of the force's use case.

Recognition time (RT)

This is the average time taken between a subject on the watchlist passing before a camera and the generation of an alert. Note that the actual amount of time taken to act on an alert will always be longer than the RT, as additional time is needed for the LFR operator to assess the alert and to pass it to an LFR engagement officer, who will then make a final decision on whether to engage or not.

The RT should be sufficiently small that an effective response to an alert is possible before the subject has moved too far from the point where the initial alert occurred. High-resolution video cameras with multiple faces in each frame will require significant processing power if the RT is to be fast enough to enable a real-time response.

Tags

Digital intelligence and investigation