NISTIR 8238

Ongoing Face Recognition Vendor Test (FRVT) Part 2: Identification

Patrick Grother
Mei Ngan
Kayee Hanaoka

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8238

National Institute of Standards and Technology
U.S. Department of Commerce

NISTIR 8238

Ongoing Face Recognition Vendor Test (FRVT) Part 2: Identification

Patrick Grother
Mei Ngan
Kayee Hanaoka
Information Access Division
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8238

November 2018

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Under Secretary of Commerce for Standards and Technology

ACKNOWLEDGMENTS

The authors are grateful to Wayne Salamon and Greg Fiumara at NIST for designing robust software infrastructure for image and template storage and parallel execution of algorithms across our computers. Thanks also to Brian Cochran at NIST for providing highly available computers and network-attached storage.

DISCLAIMER

Specific hardware and software products identified in this report were used in order to perform the evaluations described in this document. In no case does identification of any commercial product, trade name, or vendor, imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

Executive Summary

This report documents performance of face recognition algorithms submitted for evaluation on image datasets maintained at NIST. The algorithms implement one-to-many identification of faces appearing in two-dimensional images. The primary dataset is comprised of 26.6 million reasonably well-controlled live portrait photos of 12.3 million individuals. Three smaller datasets containing more unconstrained photos are also used: 3.2 million webcam images; 2.5 million photojournalism and amateur photographer photos; and 90 thousand faces cropped from surveillance-style video clips. The report will be useful for comparison of face recognition algorithms, and assessment of absolute capability.

The report details recognition accuracy for 127 algorithms from 45 developers, associating performance with participant names. The algorithms are prototypes, submitted in February and June 2018 by research and development laboratories of commercial face recognition suppliers and one university. The algorithms were submitted to NIST as compiled libraries and are evaluated as black boxes behind a NIST-specified C++ testing interface. The report therefore does not describe how algorithms operate. The evaluation was run in two phases, starting Feburary and June 2018 respectively, with developers receiving technical feedback after each. A third phase commenced on October 30, 2018, results from which will reported in the first quarter of 2019.

The major result of the evaluation is that massive gains in accuracy have been achieved in the last five years (2013-2018) and these far exceed improvements made in the prior period (2010-2013). While the industry gains are broad - at least 28 developers' algorithms now outperform the most accurate algorithm from late 2013 - there remains a wide range of capabilities. With good quality portrait photos, the most accurate algorithms will find matching entries, when present, in galleries containing 12 million individuals, with error rates below 0.2%. The remaining errors are in large part attributable to long-run ageing and injury. However, for at least 10% of images - those with significant ageing or sub-standard quality - identification often succeeds but recognition confidence is diminished such that matches become indistinguishable from false positives, and human adjudication becomes necessary.

The accuracy gains stem from the integration, or complete replacement, of prior approaches with those based on deep convolutional neural networks. As such, face recognition has undergone an industrial revolution, with algorithms increasingly tolerant of poor quality images. Whether the revolution continues or has moved into a more evolutionary phase, further gains can be expected as machine learning architectures further develop, larger datasets are assembled and benchmarks are further utilized.

Overview

Audience: This report is intended for developers, integrators, end users, policy makers and others who have some familiarity with biometrics applications and performance metrics. The methods documented here will be of interest to organizations engaged in tests of face recognition algorithms.

Prior benchmarks: Automated face recognition accuracy has improved massively in the two decades since initial commercialization of the various technologies. NIST has tracked that improvement through its conduct of regular independent, free, open, and public evaluations. These have fostered improvements in the state of the art. This report serves as an update to the NIST Interagency Report 8009 - FRVT Performance of Face Identification Algorithms, published in April 2014. That report documented identification accuracy for portrait image searches into a database of 1.6 million identities.

Scope: This report documents recognition results for four databases containing in excess of 30.2 million still photographs of 14.4 million individuals. This constitutes the largest public and independent evaluation of face recognition ever conducted. It includes results for accuracy, speed, investigative vs. identification applications, scalability to large populations, use of multiple images per person, images of cooperative and non-cooperative subjects.

The report also includes results for ageing and recognition of twins. It otherwise does not address causes of recognition failure, neither image-specific problems nor subject-specific factors including demographics. A separate report on demographic dependencies in face recognition will be published in the future. Additionally out of scope are: performance of live human-in-the-loop transactional systems like automated border control gates; human recognition accuracy as used in forensic applications; and recognition of persons in video sequences (which NIST evaluated separately [7]). Some of those applications share core matching technologies that *are* tested in this report.

Images: Four kinds of images are employed. The primary dataset is a new set of law enforcement mugshot images (Fig. 2) which are enrolled and then searched with three kinds of images: 1) other mugshots (i.e. within-domain); 2) poor quality webcam images (Fig. 3) collected in similar detention operations (cross-domain); and 3) frames from surveillance videos (Figs. 7, 8); additionally wild images (Fig. 5) are searched against other wild images.

Participation and industry coverage: The report includes performance figures for 127 prototype algorithms from the research laboratories of 39 commercial developers and one university. This represents a substantial majority of the face recognition industry, but only a tiny minority of the academic community. Participation was open worldwide. While there is no charge for participation, developers incur some software engineering expense in implementing their algorithms behind NIST application programming interface (API). The test is a black-box test where the function of the algorithm, and the intellectual property associated with it, is hidden inside pre-compiled libraries.

While participation in the test was open to any organization worldwide a number of other companies who claim a capability to do face recognition did not participate. Most academic institutions active in face recognition also did not participate. This report therefore does not capture their technical capabilities except to the extent that those technologies have been adopted or licensed by FRVT participants.

Recent technology development: Most face recognition research with convolutional neural networks (CNNs) has been aimed at achieving invariance to pose, illumination and expression variations that characterize photojournalism and social media images. The initial research [12,17] employed large numbers of images of relatively few ($\sim 10^4$) individuals to learn invariance. Inevitably much larger populations ($\sim 10^7$) were employed for training [9,14] but the benchmark, Labeled Faces in the Wild with an Equal Error Rate metric [10], represents an easy task, one-to-one verification at very high false match rates. While a larger scale identification benchmark duly followed, Megaface [11], its primary metric, rank one hit rate, contrasts with the high threshold discrimination task required in many large-population applications of face recognition, namely credential de-duplication, background checks and intelligence searches. There, identification in galleries containing up to 10^8 individuals must be performed using a) very few images per individual and b) stringent thresholds to afford very low false positive identification rates. FRVT 2018 was launched to measure the capability of the new technologies, including in these two cases. FRVT has included open-set identification tests since 2002, reporting both false negative and positive identification rates [6].

Performance metrics for applications: This report documents the performance of one-to-many face recognition algorithms. The word "performance" here refers to recognition accuracy and computational resource usage, as measured by executing those algorithms on massive sequestered datasets.

Broadly, identification algorithms operate in, and are configured for, three applications:

- **Investigation**: Consider a crime scene at which a suspect or victim is photographed, and their identity is not known. Given a recognition algorithm, and an authoritative set of reference photos, investigators search the photo against that set. Generally there is no guarantee that the subject is in the reference set. The face algorithm is configured to produce either a fixed number of candidate identities, say 50, or a set of closely similar candidates. These are then presented to a human reviewer who compares the subject with the candidate photographs. If the human determines that one of the candidates is a match, then the subject can be identified e.g. by name or whatever biographic information resides in the database. This application is characterized by very low search volumes - perhaps just one photo - and availability of labor to review candidates. This application of face recognition was prominent in the news in June 2018[1].
- **Negative identification:** Consider a driving license administrator that daily receives tens of thousands of photographs. The goal is to detect whether the applicant is present in a database under another name, e.g. to evade a driving ban. This is referred to as negative identification because the default assumption is that subjects are not in the database[2]. A face recognition system would search submitted photographs against the reference database and produce candidate matches. In this case, given high volumes and limited labor availability, only that subset of searches that produce a strongly matching candidate will be sent for human review. The system operator establishes a threshold that balances candidate volumes with labor availability. Candidates matching with strength below threshold are not returned. Video surveillance likewise can have high search volumes far above availability of reviewer labor.
- **Positive identification:** In applications where most subjects are enrolled in the database, e.g. access control to a cruise ship, face recognition might be used to implement single-factor authentication: Subjects do not present an identity claim; instead the mere presentation of their face to the system is an implicit claim to be enrolled, and they are granted access if their face matches *any* enrolled identity. The security of such a system is specified in much the same way as a verification system, by limiting false positive outcomes to below a certain rate. This is more onerous than verification, however, because the incoming face will typically be compared to all N enrollees. Another application in this category is facilitation, where enrollees present to the system to record their presence, and where unenrolled individuals who happen to present do not match, and there is no consequence.

To support these, accuracy is stated in two ways: Rank-based metrics appropriate to investigational use and threshold-based metrics for identification tasks. Both sets of metrics include tradeoffs. In investigation, overall accuracy will be reduced if labor is only available to review few candidates from the automated system. In identification applications where false positives must be limited to satisfy reviewer labor availabiliy or a security objective, higher false negative rates are implied. This report includes extensive quantification of this tradeoff. See Sec. 3

Template diversity: The FRVT is designed to evaluate black-box technologies with the consequence that the templates that hold features extracted from face images are entirely proprietary opaque binary data that embed considerable intel-

[1]A suspect was identified in a murder investigation: *Newspaper Shooting Shows Widening Use of Facial Recognition by Authorities https://www.nytimes.com/2018/06/29/business/newspaper-shooting-facial-recognition.html*

[2]This terminology is taken from the ISO/IEC 2382-37:2017 standardized biometrics vocabulary.

lectual property of the developer. Despite migration to CNN-based technologies there is no consensus on the optimal template sizes, indicating a diversity of approaches. There is no prospect of a standard template which would require a common feature set to be extracted from faces. Interoperability in automated face recognition remains solidly based on images: The ICAO portrait [21] from the ISO/IEC 19794-5 Token frontal [18], and the ANSI/NIST Type 10 [20] versions.

Automated search and human review: Virtually all applications of automated face recognition require human involvement at some frequency: Always for investigational applications; rarely in positive identification applications, after rejection (false or otherwise); and rarely in negative identification applications, after an alarm (false or otherwise). The human role is usually to compare a reference image with a query image to render either a definitive decision on "exclusion" (different subjects), or "identification" (same subject), or a declaration that one or both images have "no value" and that no decision can be made. Note that automated face recognition algorithms are not built to do exclusion - low scores from a face comparison arise from different faces *and* poor quality images.

Human review is error prone [4, 13, 19] and is sensitive to image acquisition and quality. Accurate human review is supported by high resolution - as specified in the Type 50, 51 acquisition profiles of the ANSI/NIST Type 10 record [20], and by multiple non-frontal views as specified in the same standard. These often afford views of the ear. Organizations involved in image collection should consider supporting human adjudication by collecting high-resolution frontal and non-frontal views, preparing low resolution versions for automated face recognition [18], and retaining both for any subsequent resolution of candidate matches.

Next steps: In the first quarter of 2019, NIST expects to publish two further reports from FRVT 2018: The first is an update to this report with results obtained for 90 algorithms from 49 developers submitted to NIST at the end of October 2018. The second is a report on demographic dependencies in face recognition.

Technical Summary

Accuracy gains since 2013 In April 2014, NIST reported mugshot-based face recognition accuracy for algorithms submitted to NIST in October 2013. In an exact repeat of that test - searching mugshots in an enrolled gallery of 1.6 million subjects - the most accurate algorithm in June 2018 makes a factor of 20 fewer misses than the most accurate algorithm in 2013, NEC E30C. This means that about 95% of the searches that had failed now yield the correct result at rank 1. To put that into context, only modest gains were realized between 2010 and 2013: NEC's algorithms reduced misses by less about 30%, while the other active developers reduced their error rates by around 10%. See Tables 10 and 12, and Figure 19.

Application Mode	Metric	Num-subjects	Num-images	Algorithm Date	Algorithm Name	FNIR
Investigation	Miss rate Rank=20	1.6M	1.6M	2013-OCT	NEC-30	2.9%
Investigation	Miss rate Rank=20	1.6M	1.6M	2018-JUN	Microsoft-4	0.15%
Investigation	Miss rate Rank=1	1.6M	1.6M	2013-OCT	NEC-30	4.1%
Investigation	Miss rate Rank=1	1.6M	1.6M	2018-JUN	Microsoft-4	0.23%
Identification	Miss rate FPIR=0.001	1.6M	1.6M	2013-OCT	NEC-30	9.7%
Identification	Miss rate FPIR=0.001	1.6M	1.6M	2018-JUN	Yitu-2	1.6%

Table 1: Accuracy gains since 2013.

The massive reduction in error rates over the last five years stem from wholesale replacement of the old algorithms with those based on (deep) convolutional neural networks (CNN). This constitutes a revolution rather than the evolution that defined the period 2010-2013. The rapid innovations around CNNs including, for example, Resnets [9], Inception [16], very deep networks [12, 15], and spatial transformers, may yet produce further gains. Even without that possibility, the results imply that prospective end-users should establish whether installed algorithms predate the development of the prototypes evaluated here and inquire with suppliers on availability of the latest versions.

Absolute accuracy 2018: For the most accurate algorithms the proportion of searches that do not yield the correct mate in the top 50 hypothesized identities is close to zero (or, more precisely, it is close to the rate at which samples are mislabelled due to clerical errors). Moreover, the correct response is almost always at the top rank. Thus, for the Microsoft_4 algorithm executing searches into a database of 12 million adults, the proportion of mated-searches that do not yield the correct mate at rank 1 is 0.45%. However, this impressive achievement - close to perfect recognition - must be put in context: First, many algorithms are not close to achieving this; second, it only applies to mugshot images searched in mugshot galleries; third, in many cases, the correct response is at rank 1 but its similarity score is below typical operational thresholds; fourth, as the number of enrolled subjects grows, some mates are displaced from rank one by lookalike subjects. These aspects are detailed below.

Application Mode	Metric	Num-subjects	Enrollment type	Num-images	Algorithm	FNIR Raw	FNIR Corrected[3]
Investigation	Miss rate Rank-50	12M	Lifetime	26.1M	Microsoft-4	0.06%	0.06%
Investigation	Miss rate Rank-1	12M	Lifetime	26.1M	Microsoft-4	0.19%	0.19%
Investigation	Miss rate Rank-1	12M	Recent	12M	Microsoft-4	0.45%	0.27%

Table 2: Absolute accuracy 2018.

▷ **Accuracy across commercial providers:** Recognition accuracy is very strongly dependent on the algorithm, and more generally on the developer of the algorithm. Recognition error rates in a particular scenario range from a few tenths of one percent up to beyond fifty percent. Thus algorithms from some developers are quite un-competitive and should not be deployed. It also implies that technological diversity remains in face recognition, and that there is no consensus on approach and no commoditization of the technology. See Table 17.

▷ **Error rates at high threshold:** In positive or negative identification applications, a threshold is set to limit the rate at which non-mate searches produce false positives. This has the consequence that some mated searches will report the mate below threshold, i.e. a miss, even if it is at rank 1. The utility of this is that many non-mated

Application Mode	Metric	Num-subjects	Num-images	Algorithm	FNIR Raw	FNIR Corrected
Identification	Miss rate FPIR = 0.001	12M	12M	Microsoft-4	15.8%	15.6%
Identification	Miss rate FPIR = 0.001	12M	12M	SIAT-1	10.7%	10.5%
Identification	Miss rate FPIR = 0.001	12M	12M	Yitu-2	12.4%	12.2%

Table 3: Error rates at high threshold.

[3] See Section 3.8.2

searches will usually not return any candidate identities at all. As shown in the inset tables rank-one miss rates are very low but much higher when a stringent threshold is imposed - even with the most accurate algorithms, some mates score weakly such that 10% to 20% searches fail to return mates above threshold. Broadly this occurs for three reasons: poor image quality, ageing, and presence of lookalikes. See Table 16 and Figure 51.

▷ **Image Quality:** Poor quality photographs undermine recognition, either because the imaging system is poor (lighting, camera etc) or because the subject mis-presents to the camera (head orientation, facial expression, occlusion etc.). Imaging problems can be eliminated by design - i.e. by ensuring adherence to long-standing face image capture standards. Presentation problems, however, must be detected at capture time, either by the photographer, or by an automated system, and re-capture performed.

The most accurate algorithms in FRVT are highly tolerant of image quality problems. This derives from the invariance advantages possessed by CNN-based algorithms, and this is the reason why accuracy has improved since 2013. For example, the Microsoft algorithms are highly tolerant of non-frontal pose, to the point that the few profile-view images that remain in the FRVT frontal mugshot dataset are very often recognized correctly.

▷ **Ageing:** A larger source of error in long-run criminal justice applications is ageing. All faces age. While this usually proceeds in a graceful and progressive manner, drug use may expedite this, and surgery may be effective in delaying it - the effects on face recognition have not been quantified. The change in appearance causes face recognition similarity scores to decline such that over the longer term, accuracy will decline. This is essentially unavoidable, and can only be mitigated by scheduled re-capture, as in passport re-issuance. To quantify ageing effects, we used the more accurate algorithms to enroll the earliest image of 3.1 million adults and then search with 10.3 million newer photos taken up to 18 years after the the initial enrollment photo. Accuracy is seen to degrade progressively with time, as mate scores decline and non-mates displace mates from rank 1 position. More accurate algorithms tend to be less sensitive to ageing, although accuracy alone does not predict ageing tolerance perfectly. The more accurate algorithms give fewer errors after 18 years of ageing than middle tier algorithms give after four. Note also we do not quantify an ageing rate - more formal methods [1] borrowed from the longitudinal analysis literature have been published for doing so (given suitable data). See Figures 68, 73 and 78.

Algorithm	Investigational miss rate FNIR(N, 1, 0), N=3.1 million							
	2 YR	4 YR	6 YR	8 YR	10 YR	12 YR	14 YR	18 YR
Microsoft-4	0.32%	0.47%	0.60%	0.7%	0.9%	1.0%	1.3%	1.6%
Visionlabs-4	0.48%	0.70%	0.91%	1.1%	1.3%	1.5%	1.9%	2.4%
Yitu-2	0.66%	0.83%	0.94%	1.0%	1.2%	1.5%	2.2%	3.3%
Megvii-0	0.94%	1.57%	2.36%	3.4%	4.7%	6.1%	8.3%	11.1%
ISystems-2	1.01%	1.35%	1.69%	2.0%	2.3%	2.6%	3.0%	4.0%
Neurotechnology-4	1.04%	1.34%	1.56%	1.7%	1.9%	2.1%	2.4%	3.2%
Idemia-4	1.10%	1.51%	1.96%	2.4%	2.8%	3.1%	3.7%	5.4%
Cogent-1	1.28%	1.84%	2.50%	3.3%	4.1%	4.9%	6.1%	7.9%
Cognitec-1	1.49%	2.28%	3.12%	4.0%	4.8%	5.5%	6.6%	8.1%
NEC-0	1.95%	3.16%	4.45%	5.8%	7.0%	8.2%	10.0%	12.4%
RankOne-2	2.12%	3.13%	4.31%	5.6%	7.1%	8.8%	11.3%	15.4%

Table 4: Impact of ageing on accuracy.

▷ **Accuracy in large populations:** Prior NIST mugshot tests had run on enrolled populations of $N \leq 1.6$ million. Here we extend that to $N = 12$ million people. This new database is more difficult than the mugshot database used to gauge accuracy improvements since FRVT 2010 and FRVT 2014. See Figure 4

On the new database, termed FRVT 2018, identification miss rates climb very slowly as population size increases. For the most accurate algorithm when searching a database of size 640 000, about 0.27% of searches fail to produce the correct mate as its best hypothesized identity. In a database of 12 000 000 this rises to 0.45%. This benign growth in miss rates is fundamentally the reason for the utility of face recognition in large scale one-to-many search applications. See Table 14 and Figure 31.

The reason for this is that as more identities are enrolled into an database, the possibility of a false positive increases due to lookalike faces that yield extreme values from the right tail of the non-mate score distribution. However, these scores are lower than most mate scores such that when an identification algorithm is configured with a threshold of zero, and where human adjudication is always necessary, rank-one identification miss rates scale very favorably with population size, N, growing slowly, approximately as a power law, aN^b with $b \ll 1$. This dependency was first noted in 2010. Depending on the algorithm, the exponent b for mugshot searches is low, around 0.06 for the Cogent algorithms with up to 12 million identities. The most accurate algorithms have somewhat larger values $b = 0.17$ (Microsoft-4) and 0.08 (Yitu-2). See Table 14.

In any case, variations in accuracy with increasing population size are small relative to both ageing and algorithm choice. See Figure 22.

▷ **Twins:** One component of the residual errors is that which arises from incorrect association of twins. The more accurate face recognition algorithms tested here are incapable of distinguishing twins, not just identical (monozygotic) but also same-sex fraternal (dizygotic) twins. A twin, when present in an enrollment database will invariably produce a false positive if the twin is searched. Of the five algorithms tested, all incorrectly identify twins against eachother, except in many cases where the fraternal twins are of different sex. The inset table shows how often Twin A is not retrieved when Twin A, or Twin B, is searched. Twins constitute around 3.4% of all live infants in 2016[4] such that system operators might annotate twins in databases, and establish training and procedures to handle false positive outcomes. See Figure 23

N = 640 104	Investigational miss rate FNIR(N, 1, 0)			
Enrol Twin A	Search: Twin A		Search: Twin B	
Algorithm	Identical	Fraternal	Identical	Fraternal
Microsoft-4	0%	0%	0%	32%
Idemia-4	0%	0%	1%	35%
Siat-1	0%	0%	1%	33%
Visionlabs-4	0%	0%	0%	32%
Yitu-2	0%	0%	0%	36%
Desired result	0%	0%	100%	100%

Table 5: Accuracy on twins.

Accuracy within commercial providers: While results for up to five algorithms from each developer are reported here, the intra-provider accuracy variations are usually much smaller than the inter-provider variations. However from Phase 1 to 2, February to June 2018, some developers attained up to a five-fold reduction in misses. Such rapid gains imply that the revolution is not yet over, and further gains may be realized in Phase 3 starting October 30, 2018. Some developers submitted variants that explore an accuracy-speed tradespace. See Figure 19 and Table 17.

Utility of adjudicating long candidate lists: In the regime where a system is configured with a threshold of zero, and where human adjudication is always necessary, the reviewer will find some mates on candidate lists at ranks far above one. This usually occurs because either the probe image or its corresponding enrolled mate image have poor quality, or large time-lapse. The accuracy benefits of traversing say 50 candidates are broadly that the rank-1 miss rate is reduced by up to a factor of two. See Figure 39 and compare Tables 14 and 15.

However, accuracy from the leading algorithm is now so high - mates that in 2013 were placed at rank > 1, are now at rank 1 - such that reviewers can expect to review substantially fewer candidates. Note, however, for the proportion of searches where there is no mate, reviewers might still examine all candidates, fruitlessly.

Utility of enrolling multiple images per subject: We run three kinds of enrollment: First, by enrolling just the most recent image; second by create a single template from a person's full lifetime history of images; and third by enroling multiple images of a person separately (as though under different identities). The overall effect is that the enrollment of multiple images yields as much as a factor of two lower miss rates. This occurs because the most recent image may sometimes be of poorer quality than historical images. See Table 14.

Gains depend on the number of available images: FNIR drops steadily. However, a few algorithms give higher false positive rates. Figure 84.

Reduced template sizes: There has been a trend toward reduced template sizes, i.e. a smaller feature representation of an image. In 2014, the most accurate algorithm used a template of size 2.5KB; the figure in 2018 is 1 024 bytes. Close competitors produce templates of size 256, 364, 512, 4 136 and 4 442 bytes respectively. In 2014, the leading competitors

[4] This rate varies regionally, and has increased by a factor of two since 1980 due to fraternal twins being more common with in-vitro fertilization and as women have babies later in life.

had templates of size 4KB to 8KB. Some algorithms, when enrolling more than one image of a person, produce a template whose size is independent of the number of images given to the algorithm. This can be achieved by selecting a "best" image, or by integrating (fusing) information from the images. See Table 10.

Template generation times: Template generation times, as measured on a single circa-2016 server processor core [5], vary from 50 milliseconds upto nearly 1 second. This wide variation across developers may be relevant to end-users who have high-volume workflows. There has not been a wide downward trend since 2014. Note that speed may be expedited over the figure reported here by exploiting new vector instructions on recent chips. Note that GPUs were not used and, while indispenasble for training CNNs, are not necessary for feeding an image forward through a network. See Table 10.

Search times: Template search times, as measured on circa-2016 Intel server processor cores, vary massively across the industry. For a database of size 1 million subjects, and the more accurate implementations, durations range from 4 to 500 milliseconds, with other less accurate algorithms going much slower still. See Table 10.

Search time scalability: Several algorithms exhibit sublinear search time i.e. the duration does not double with a doubling of the enrolled population size, N. This was noted also in 2014. In 2018, however, logarithmic growth has been observed for one developer, and near logarithmic for one of the more accurate algorithms. The consequence of this is that as N increases even the fastest linear algorithm will quickly become much slower than the strongly sublinear algorithms. Figures 103 and 104.

Conclusions: As with other biometrics, accuracy of facial recognition implementations varies greatly across the industry. Absent other performance or economic parameters, users should prefer the most accurate algorithm. Note that accuracy, and algorithm rankings, vary somewhat with the kinds of images used and the mode of operation: investigation with zero threshold; or identification with high threshold.

[5] Intel Xeon CPU E5-2630 v4 running at 2.20GHz.

Release Notes

FRVT Activities: NIST initiated FRVT in February 2018, inviting participants to send up to seven one-to-many prototype algorithms. Since February 2017, NIST has been evaluating one-to-one verification algorithms on an ongoing basis. This allows developers to submit updated algorithms to NIST at any time but no more frequently than four calendar months. This more closely aligns development and evaluation schedules. Results are posted to the web within a few weeks of submission. Details and full report are linked from the Ongoing FRVT site.

FRVT Reports: The results of the FRVT appear in the series NIST Interagency Reports tabulated below. The reports were developed separately and released on different schedules. In prior years NIST has mostly reported FRVT results as a single report; this had the disadvantage that results from completed sub-studies were not published until all other studies were complete.

Date	Link	Title	No.
2014-03-20	PDF	FRVT Performance of Automated Age Estimation Algorithms	7995
2015-04-20	PDF	Face Recognition Vendor Test (FRVT) Performance of Automated Gender Classification Algorithms	8052
2014-05-21	PDF	FRVT Performance of face identification algorithms	8009
2017-03-07	PDF	Face In Video Evaluation (FIVE) Face Recognition of Non-Cooperative Subjects	8173
2017-11-23	PDF	The 2017 IARPA Face Recognition Prize Challenge (FRPC)	8197
2018-04-13	WWW	Ongoing Face Recognition Vendor Test (FRVT)	Draft

Details appear on pages linked from `https://www.nist.gov/programs-projects/face-projects`.

Appendices: This report is accompanied by appendices which present exhaustive results on a per-algorithm basis. These are machine-generated and are included because the authors believe that visualization of such data is broadly informative and vital to understanding the context of the report.

Typesetting: Virtually all of the tabulated content in this report was produced automatically. This involved the use of scripting tools to generate directly type-settable LaTeX content. This improves timeliness, flexibility, maintainability, and reduces transcription errors.

Graphics: Many of the Figures in this report were produced using the ggplot2 package running under R, the capabilities of which extend beyond those evident in this document.

1 Introduction

One-to-many identification represents the largest market for face recognition technology. Algorithms are used across the world in a diverse range of biometric applications: detection of duplicates in databases, detection of fraudulent applications for credentials such as passports and driving licenses, token-less access control, surveillance, social media tagging, lookalike discovery, criminal investigation, and forensic clustering.

This report contains a breadth of performance measurements relevant to many applications. Performance here refers to accuracy and resource consumption. In most applications, the core accuracy of a facial recognition algorithm is the most important performance variable. Resource consumption will be important also as it drives the amount of hardware, power, and cooling necessary to accomodate high volume workflows. Algorithms consume processing time, they require computer memory, and their static template data requires storage space. This report documents these variables.

1.1 Open-set searches

FRVT tested open-set identification algorithms. Real-world applications are almost always "open-set", meaning that some searches have an enrolled mate, but some do not. For example, some subjects have truly not been issued a visa or drivers license before; some law enforcement searches are from first-time arrestees[6]. In an "open-set" application, algorithms make no prior assumption about whether or not to return a high-scoring result, and for a mated search, the ideal behaviour is that the search produces the correct mate at high score and first rank. For a non-mate search, the ideal behavior is that the search produces zero high-scoring candidates.

Too many academic benchmarks execute only closed-set searches. The proportion of mates found in the rank one position is the default accuracy metric. This hit rate metric ignores the score with which a mate is found; weak hits count as much a strong hits. This ignores the real-world imperative that in many applications it is necessary to elevate a threshold to reduce the number of false positives.

2 Evaluation datasets

FRVT2018 used four kinds of images - mugshots, webcam, wild and surveillance - as described in the following sections.

2.1 Mugshot images

This is the third time that FRVT has employed large mugshot datasets. The main dataset used is refered to as the FRVT 2018 set. This set was extracted from a larger operational parent set, excluding all webcam images, profile images, and non-face images.

[6]Operationally closed-set applications are rare because it is usually not the case that all searches have an enrolled mate. One counter-example, however, is a cruise ship in which all passengers are enrolled and all searches should produce one, and only one, identity. Another example is forensic identification of dental records from an aircraft crash.

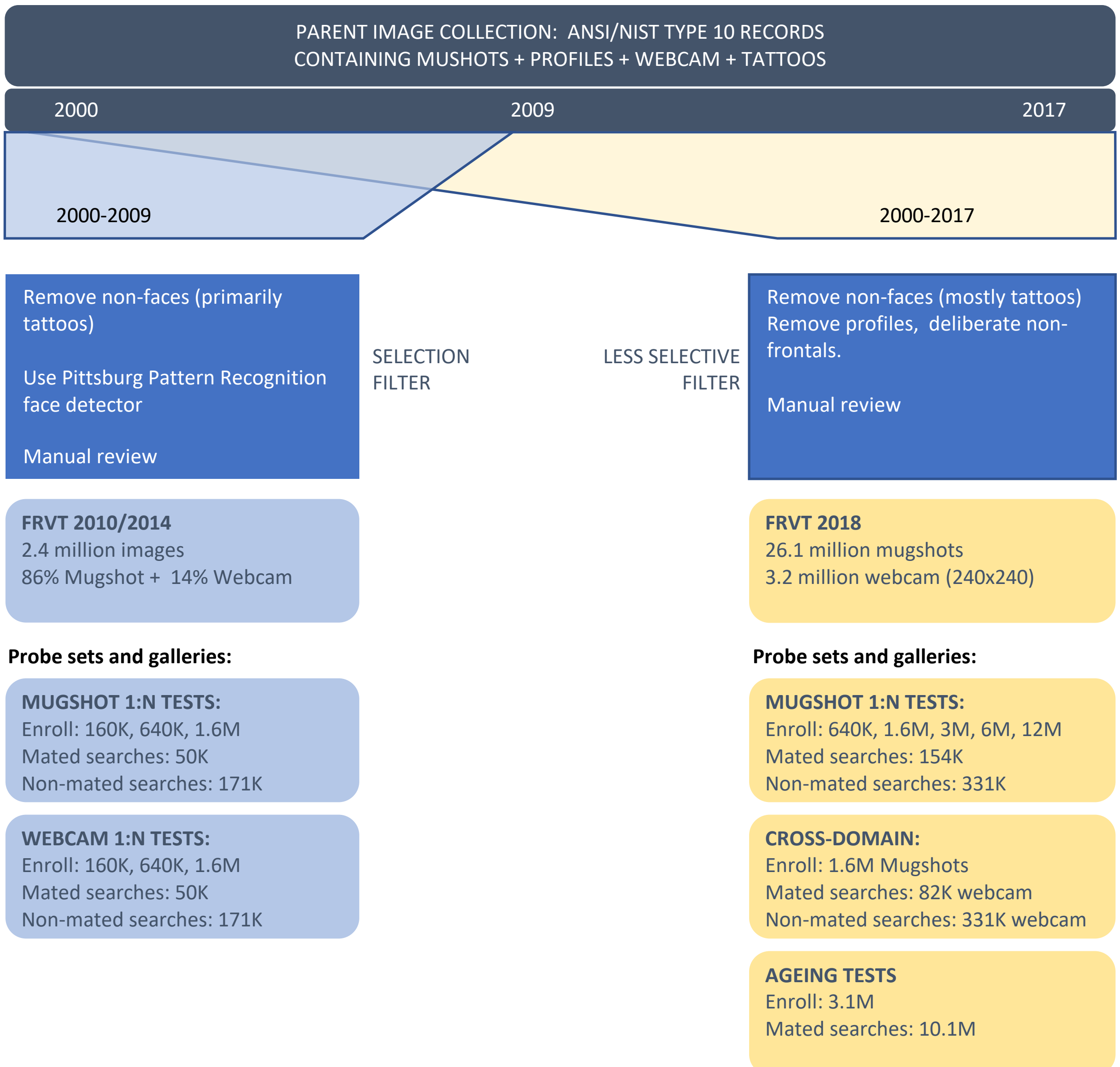

Figure 1: **Mugshot selection**. *The left branch of the figure applies to the mugshots used in FRVT 2014, then termed LEO. The right hand branch shows the much larger set used in FRVT 2018. The exact details of the image selection mean that recognition of images in the FRVT 2018 dataset is more difficult than in the FRVT 2014 (LEO) set - see Table 4.*

2.1.1 The FRVT 2014 partition

From the parent dataset we re-constituted the dataset employed in the NIST INTERAGENCY REPORT 8009 from 2014. That dataset is comprised of 86% mugshots and 14% webcam images. We use it here to exactly repeat the 2014 evaluation. It is refered to here as LEO and FRVT2014.

Example images are shown in Figures 2 and 3.

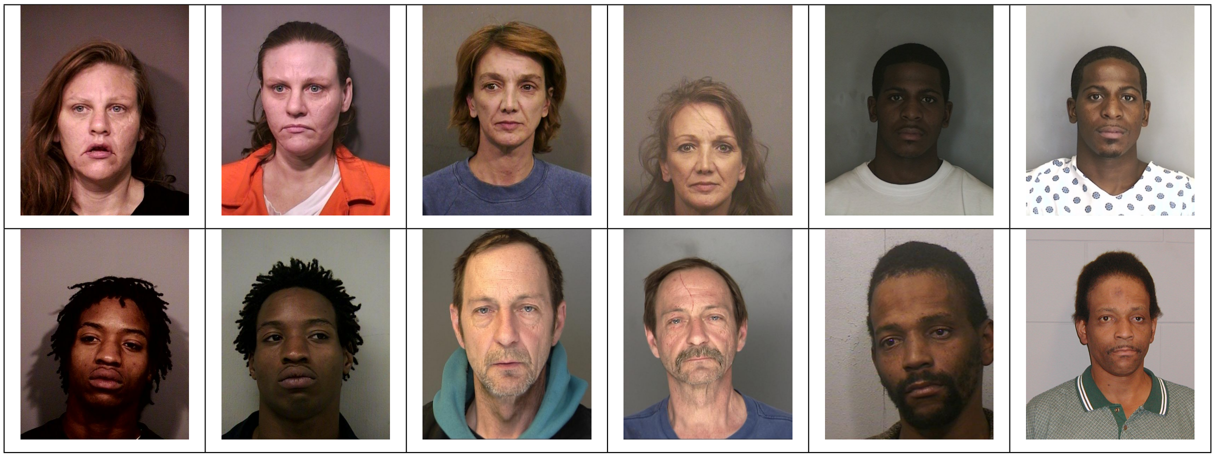

Figure 2: Six mated mugshot pairs representative of the FRVT-2014 (LEO) and FRVT-2018 datasets. The images are collected live, i.e. not scanned from paper. Image source: NIST Special Database 32

Figure 3: Twelve webcam images representative of probes against the FRVT-2018 mugshot gallery. The first eight images are four mated pairs. Such images present challenges to recognition including pose, non-uniform illumination, low contrast, compression, cropping, and low spatial sampling rate. Image source: NIST Special Database 32

- ▷ **Mugshots**: Comprising about 86% of the LEO database, are mugshots having reasonable compliance with the ANSI/NIST ITL1-2011 Type 10 standard's subject acquisition profiles levels 10-20 for frontal images [20]. The major departure from the standard's requirements is the presence of mild pose variations around frontal - the images of Figure 2 are typical. The images vary in size, with many being 480x600 pixels with JPEG compression applied to produce filesizes of between 18 and 36KB with many images outside this range, implying that about 0.5 bits are being encoded per pixel.

- ▷ **Webcam images**: The remaining 14% of the images were collected using an inexpensive webcam attached to a

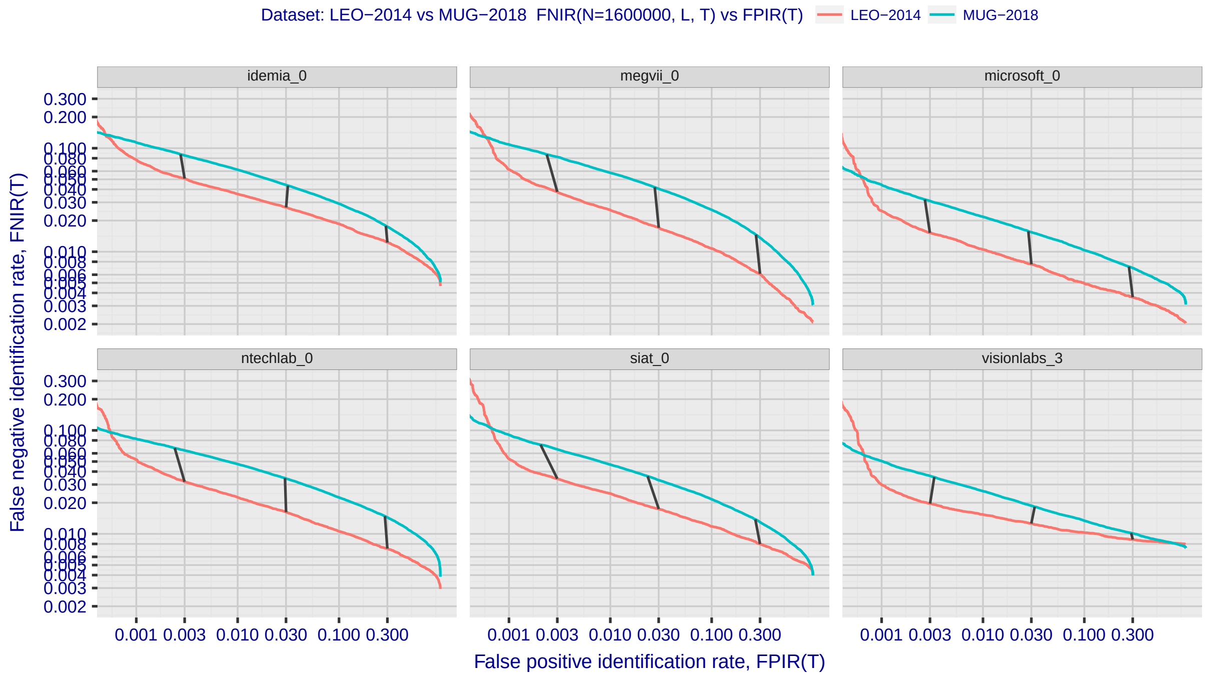

Figure 4: **[Relative difficuly of 2013, 2018 datasets]** *The figure shows results for 2018 algorithms running on two datasets: The* LEO *set used in* FRVT*2014 and the mugshots in the* FRVT*2018 dataset. The axes are identification miss rates vs. false positive rates. Across most of the range the new database is more difficult i.e. FNIR is roughly two times higher. However, at the right side - corresponding to low threshold, this gap reduces showing that algorithms can find weak mates in both databases about equally. At the left side FNIR reverses - this is thought to arise because of ground truth errors in the 2014 set, where a few subjects are present in the database under multiple IDs, giving rise to high non-mate scores that are actually mate scores.*

flexible operator-directed mount. These images are all of size 240x240 pixels, that are in considerable violation of most quality-related clauses of all face recognition standards. As evident in the figure, the most common defects are non-frontal pose (associated with the rotational degrees of freedom of the camera mount), low contrast (due to varying and intense background lights), and poor spatial resolution (due to inexpensive camera optics) - see examples in Fig 3. The images are overly JPEG compressed, to between 4 and 7KB, implying that only 0.5 to 1 bits are being encoded per color pixel.

The images are drawn from NIST Special Database 32 which may be downloaded here.

2.1.2 The FRVT 2018 partition

As shown in Figure 1 the main FRVT 2018 image set is comprised of 26.1 million mugshots and 3.2 million webcams, from which the enrollment and search sets of Table 6 are prepared. The images have broadly the same appearance and properties as those in the FRVT 2014 set. However, as part of the process to remove profile-view images and tattoo images, the FRVT 2014 set was assembled by using a face detector from Pittsburg Pattern Recognition that was used as a filter to exclude images for which a face could not be detected. The consequence of this is that poorly exposed photos are more likely to be absent from FRVT 2014 than they are in FRVT 2018, which used more permissive retention logic. Figure 4 shows that the newer FRVT 2018 database is more difficult than the earlier set.

Figure 5: Examples of "in the wild" stills. The top row gives the full original images; the second row gives the manually specified face region that is cropped and passed to the algorithms. The source images in this figure are published on the internet under Creative Commons licenses.

2.2 Unconstrained images

2.2.1 Wild images

In addition to portrait-styled mugshots, algorithms were also evaluated on a "wild" dataset composed of non-cooperative and unconstrained photojournalism and amateur photography imagery. The images are closely cropped from the parent images as shown in Figure 5. A portion of the images are collected by professional photographers and as such are captured, and selected, to not exhibit exposure and focus problems. Some of the photos were downloaded from websites with substantial amateur photographer imagery, which may contain images that do exhibit exposure and focus problems. Resolution varies widely as these images were downloaded from the internet with varying resampling and compression practices. The primary difficulties for face recognition is unconstrained yaw and pitch pose variation, with some images extending to profile view. Additionally faces can be occluded, including by hair and hands.

The images are cropped prior to passing them to the algorithm. The cropping is done per human-annotated rectangular bounding boxes. The algorithm must further localize the face and extract features. In many cases, there were multiple images of the subject provided to the algorithm, and the output was a single template representation of the subject.

$N_P = 332\,574$ subjects were searched against two galleries, where the number of enrolled subjects in each gallery were $N_{G1} = 1\,106\,777$ and $N_{G2} = 1\,107\,778$. Both gallery and search images were composed of unconstrained wild imagery.

2.2.2 Face Recognition Prize Challenge (FRPC) 2017 Dataset

The IARPA Face Recognition Prize Challenge (FRPC) 2017 was conducted to assess the capability of contemporary face recognition algorithms to recognize faces in photographs collected without tight quality constraints. The dataset consisted of images collected from individuals who are unaware of, and not cooperating with, the collection. Such images are characterized by variations in head orientation, facial expression, illumination, and also occlusion and reduced resolution.

Algorithms were run through the exact dataset used in the FRPC 2017 Identification track.

- ▷ **Enrolled portraits:** The enrollment database consisted of portrait images that were either visa images, mugshot images, or dedicated portraits collected from test subjects. These were collected typically using a digital single-lens reflex (DSLR) camera, ample two point light, and a standard uniform grey background. We defined five galleries containing, respectively, N = $\{16\,000, 48\,000, 160\,000, 320\,000, 691\,282\}$ images and people, i.e. exactly one image per person. These galleries include 825 portraits of the people who appear in the mated search sets described next. Examples of the portraits appear in Figure 6.
- ▷ **Mated search images:** The non-cooperative face images are faces cropped from video clips collected in surveillance settings. Examples of the cropped faces and the parent video frames are shown in Figures 7 and 8
- ▷ **Non-mated search images:** A separate set of $N_I = 79\,403$ faces cropped from video that are known not to contain any of the enrolled identities are used to estimate false positive accuracy.

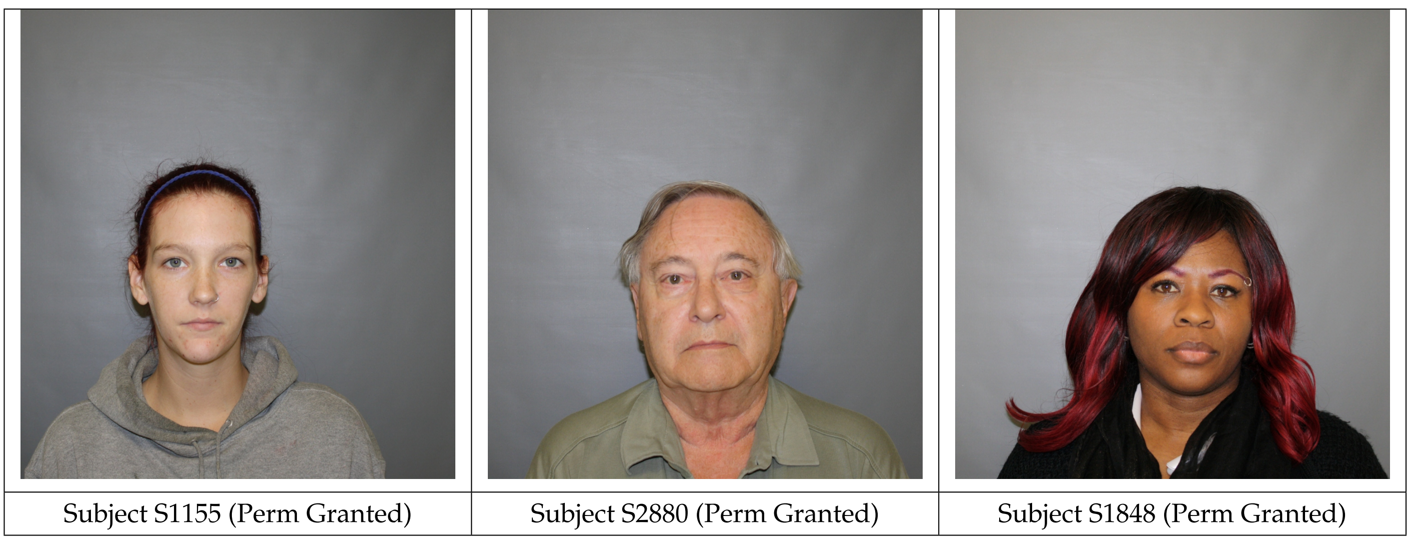

Figure 6: Examples of enrollment images collected with an SLR camera. The face images in this figure are from the DHS / S&T provided AEER dataset. The included subjects consented to release their images in public reports.

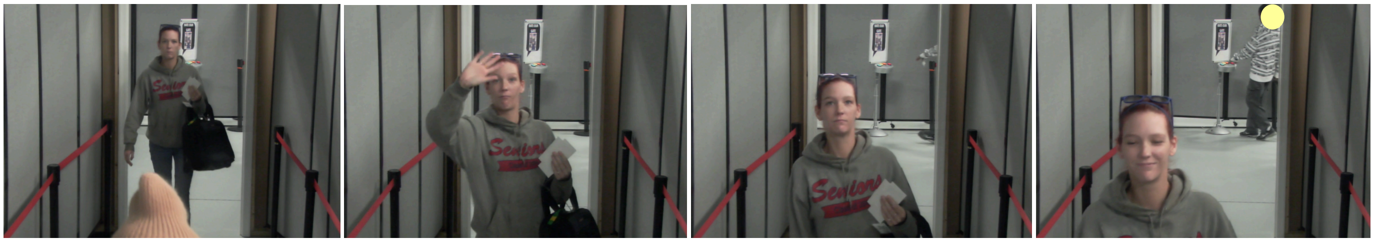

Figure 7: Example images from the ceiling mounted camera for the free movement scenarios from videos collected on an aircraft boarding ramp. The images in this table are from the subject S1115 in the DHS / S&T provided AEER dataset. The subject gave written opt-in permission to allow public release of all imagery. Where consent from individuals in the background was not obtained, their faces were masked (yellow circle).

Enrollment	T = 1 secs	T = 2 secs	T = 3 secs

Figure 8: Enrollment (left) and non-cooperative video-frame search examples from a boarding gate process. The algorithm received the enrollment image as is, and faces cropped from the video search frames. The images are from subject 79195746 in the DHS/ S&T AEER dataset. He consented to release of his images in public reports. For those individuals who did not consent to publication, their faces were masked (yellow circles).

2.3 Enrollment types

Many operational applications include collection and enrollment of biometric data from subjects on more than one occasion. This might be done on a regular basis, as might occur in credential (re-)issuance, or irregularly, as might happen in a criminal recidivist situation [3]. The number of images per person will depend on the application area: In civil identity credentialing (e.g. passports, driver's licenses), the images will be acquired approximately uniformly over time (e.g. ten years for a passport). While the distribution of dates for such images of a person might be assumed uniform, a number of factors might undermine this assumption[7]. In criminal applications, the number of images would depend on the number of arrests. The distribution of dates for arrest records for a person (i.e. the recidivism distribution) has been modeled using the exponential distribution but is recognized to be more complicated[8].

In any case, the 2010 NIST evaluation of face recognition showed that considerable accuracy benefits accrue with reten-

[7] For example, a person might skip applying for a passport for one cycle, letting it expire. In addition, a person might submit identical images (from the same photography session) to consecutive passport applications at five year intervals.
[8] A number of distributions have been considered to model recidivism, see for example [2].

Image				
Encounter	1	...	$K_i - 1$	K_i
Capture Time	T_1	...	$T_{K_i - 1}$	T_{K_i}
Role RECENT	Not used	Not used	Enrolled	Search
Role LIFETIME	Enrolled	Enrolled	Enrolled	Search

Figure 9: Depiction of the "recent" and "lifetime" enrollment types. Image source: NIST Special Database 32

tion and use of *all* historical images [5].

To this end, the FRVT API document provides $K \geq 1$ images of an individual to the enrollment software. The software is tasked with producing a single proprietary undocumented "black-box" template[9] from the K images. This affords the algorithm an ability to generate a *model* of the individual, rather than to simply extract features from each image on a sequential basis.

As depicted in Figure 9, the i-th individual in the LEO dataset has K_i images. These are labelled x_k for $k = 1 \ldots K_i$. To measure the utility of having multiple enrollment images, this report evaluates two kinds of enrollment:

▷ **Recent**: Only the second most recent image, $x_{K_i - 1}$ is enrolled. This type of enrollment mimics the operational policy of retaining the imagery from the most recent encounter. This might be done operationally to ameliorate the effects of face ageing. Obviously retaining only the most recent image should only be done if the identity of the person is trusted to be correct. For example, in an access control situation retention of the most recent successful *authentication* image would be hazardous if it could be a false positive.

▷ **Lifetime-consolidated**: All except the last image are enrolled, $x_1 \ldots x_{K_i - 1}$. This subject-centric strategy might be adopted if quality variations exist where an older image might be more suitable for matching, despite the ageing effect.

▷ **Lifetime-unconsolidated**: All except the last image are again enrolled, $x_1 \ldots x_{K_i - 1}$ but now separately, with different identifiers, such that the algorithm is not aware that the images are from the same face. This kind of event- or encounter-centric enrollment is very common when operational constraints preclude reliable consolidation of the historical encounters into a single identity. This also prevents the algorithm from a) building a holistic model of identity (as is common in speaker recognition systems) and b) implementing fusion, for example template-level fusion of feature vectors, or post-search score-level fusion. The result is that searches will typically yield more than one image of a person in the top ranks. This has consequences for appropriate metrics: The quantity "recall" expresses what fraction of the relevant faces are returned.

NIST first evaluated this kind of enrollment in mid 2018, and the results tables include some comparison of accuracy available from all three enrollment styles.

In all cases, the most recent image, x_{K_i}, is reserved as the search image. For the 1.6 million subject enrollment parition of the LEO data, $1 \leq K_i \leq 33$ with $K_i = 1$ in 80.1% of the individuals, $K_i = 2$ in 13.4%, $K_i = 3$ in 3.7%, $K_i = 4$ in 1.4%,

[9] There are no formal face template standards. Template standards only exist for fingerprint minutiae - see ISO/IEC 19794-2:2011.

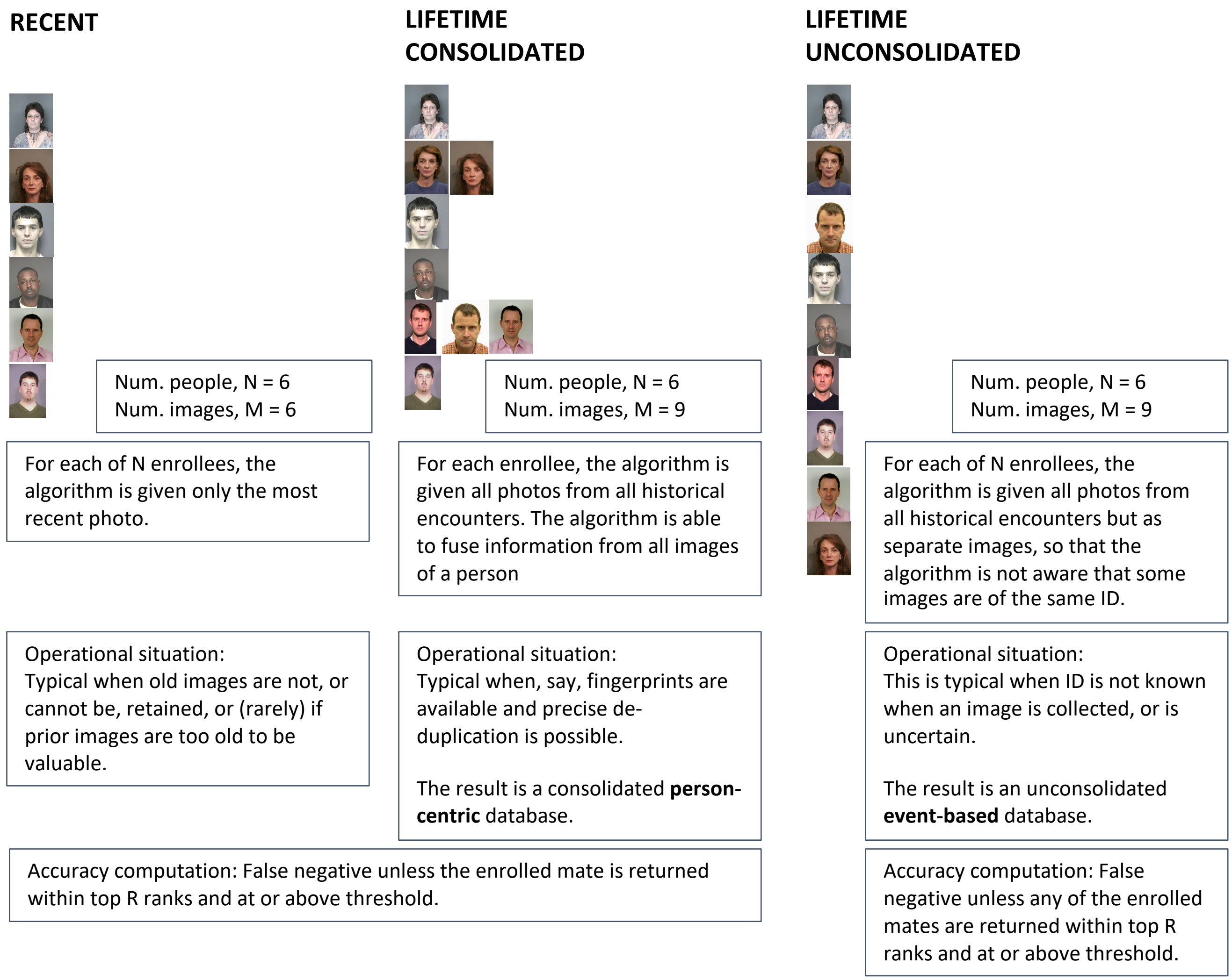

Figure 10: **Enrollment database types**. *The figure shows the three kinds of enrollment databases examined in this report. Image source: NIST Special Database 32*

	ENROLLMENT				SEARCH			
	TYPE SEE	POPULATION			MATE		NON-MATE	
	SECTION 2.3	FILTER	N-SUBJECTS	N-IMAGES	N-SUBJECTS	N-IMAGES	N-SUBJECTS	N-IMAGES
Mugshot trials from enrollment of single images								
1	RECENT	NATURAL	640 000	640 000	154 549	154 549	331 254	331 254
2	RECENT	NATURAL	1 600 000	1 600 000				
3	RECENT	NATURAL	3 000 000	3 000 000				
4	RECENT	NATURAL	6 000 000	6 000 000				
5	RECENT	NATURAL	12 000 000	12 000 000				
Mugshot trials from enrollment of lifetime images								
6	CONSOL	NATURAL	640 000	1 247 331				
7	CONSOL	NATURAL	1 600 000	3 351 206				
8	CONSOL	NATURAL	3 000 000	6 417 057				
9	CONSOL	NATURAL	6 000 000	12 976 185				
10	CONSOL	NATURAL	12 000 000	26 107 917				
11	UN-CONSOL	NATURAL	640 000	1 247 331				
12	UN-CONSOL	NATURAL	1 600 000	3 351 206				
Cross-domain								
13	MUGSHOTS AS ON ROW 2				82 106 WEBCAM	82 106 WEBCAM	331 254 WEBCAM	331 254 WEBCAM
Demographics								
14	RECENT	MALE, AGE21-40, ΔT ≤ 5 YR, BLACK AND WHITE BALANCED	800 000 B + 800 000 W	800 000 B + 800 000 W	100 000 B + 100 000 W	100 000 B + 100 000 W	100 000 B + 100 000 W	100 000 B + 100 000 W
15	RECENT	WHITE, AGE21-40, ΔT ≤ 5 YR, MALE AND FEMALE BALANCED	800 000 F + 800 000 M	800 000 F + 800 000 M	100 000 F + 100 000 M	100 000 F + 100 000 M	100 000 F + 100 000 M	100 000 F + 100 000 M
16	RECENT	BLACK, AGE21-40, ΔT ≤ 5 YR, MALE AND FEMALE BALANCED	500 000 F + 500 000 M	500 000 F + 500 000 M	97 000 F + 97 000 M	97 000 F + 97 000 M	100 000 F + 100 000 M	100 000 F + 100 000 M
Ageing								
17	OLDEST	NATURAL	3 068 801	3 068 801	2 853 221	10 951 064	0	0

Table 6: **Enrollment and search sets**. *Each row summarizes one identification trial. Unless stated otherwise, all entries refer to mugshot images. The term "natural" means that subjects were selected without heed to demographics, i.e. in the distribution native to this dataset. The probe images were collected in a different calendar year to the enrollment image.*

$K_i = 5$ in 0.6%, $K_i = 6$ in 0.3%, and $K_i > 6$ is 0.2% for everyone else. This distribution is substantially dependent on United States recidivism rates.

We did not evaluate the case of retaining only the highest quality image, since automated quality assessment is out of scope for this report. We do not anticipate that such strategies will prove beneficial when the quality assessment apparatus is imperfect and unvalidated.

3 Performance metrics

This section gives specific definitions for accuracy and timing metrics. Tests of open-set biometric algorithms must quantify frequency of two error conditions:

- ▷ **False positives**: Type I errors occur when search data from a person who has never been seen before is incorrectly associated with one or more enrollees' data.

▷ **Misses**: Type II errors arise when a search of an enrolled person's biometric does not return the correct identity.

Many practitioners prefer to talk about "hit rates" instead of "miss rates" - the first is simply one minus the other as detailed below. Sections 3.1 and 3.2 define metrics for the Type I and Type II performance variables.

Additionally, because recognition algorithms sometimes fail to produce a template from an image, or fail to execute a one-to-many search, the occurrence of such events must be recorded. Further because algorithms might elect to not produce a template from, for example, a poor quality image, these failure rates must be combined with the recognition error rates to support algorithm comparison. This is addressed in section 3.5.

Finally, section 3.7 discusses measurement of computation duration, and section 3.8 addresses the uncertainty associated with various measurements. Template size measurement is included with the results.

3.1 Quantifying false positives

It is typical for a search to be conducted into an enrolled population of N identities, and for the algorithm to be configured to return the closest L candidate identities. These candidates are ranked by their score, in descending order. A human analyst might examine either all L candidates, or just the top $R \leq L$ identities, or only those with score greater than threshold, T. The workload associated with such examination is discussed later, in 3.6.

False alarm performance is quantified in two related ways. These express how many searches produces false positives, and then, how many false positives are produced in a search.

False positive identification rate: The first quantity, FPIR, is the proportion of non-mate searches that produce an adverse outcome:

$$\text{FPIR}(N,T) = \frac{\text{Num. non-mate searches where one or more enrolled candidates are returned at or above threshold, T}}{\text{Num. non-mate searches attempted.}} \tag{1}$$

Under this definition, FPIR can be computed from the highest non-mate candidate produced in a search - it is not necessary to consider candidates at rank 2 and above. FPIR is the primary measure of Type I errors in this report.

Selectivity: However, note that in any given search, more than one non-mate may be returned above threshold. In order to quantify such events, a second quantity, selectivity (SEL), is defined as the *number* of non-mates returned on a candidate list, averaged over all searches.

$$\text{SEL}(N,T) = \frac{\text{Num. non-mate enrolled candidates returned at or above threshold, T}}{\text{Num. non-mate searches attempted.}} \tag{2}$$

Both of these metrics are useful operationally. FPIR is useful for targeting how often an adverse false positive outcome can occur, while SEL as a number is related to workload associated with adjudicating candidate lists. The relationship between the two quantities is complicated - it depends on whether an algorithm concentrates the false alarms in the results of a few searches or whether it disburses them across many. This was detailed in FRVT 2014, NISTIR 8009. It has not yet been detailed in FRVT 2018.

3.2 Quantifying hits and misses

If L candidates are returned in a search, a shorter candidate list can be prepared by taking the top $R \leq L$ candidates for which the score is above some threshold, $T \geq 0$. This reduction of the candidate list is done because thresholds may be applied, and only short lists might be reviewed (according to policy or labor availability, for example). It is useful then to state accuracy in terms of R and T, so we define a "miss rate" with the general name **false negative identification rate** (FNIR), as follows:

$$\text{FNIR}(N, R, T) = \frac{\text{Num. mate searches with enrolled mate found outside top R ranks or score below threshold, T}}{\text{Num. mate searches attempted.}} \quad (3)$$

This formulation is simple for evaluation in that it does not distinguish between causes of misses. Thus a mate that is not reported on a candidate list is treated the the same as a miss arising from face finding failure, algorithm intolerance of poor quality, or software crashes. Thus if the algorithm fails to produce a candidate list, either because the search failed, or because a search template was not made, the result is regarded as a miss, adding to FNIR.

Hit rates, and true positive identification rates: While FNIR states the "miss rate" as how often the correct candidate is either not above threshold or not at good rank, many communities prefer to talk of "hit rates". This is simply the **true positive identification rate**(TPIR) which is the complement of FNIR giving a positive statement of how often mated searches are successful:

$$\text{TPIR}(N, R, T) = 1 - \text{FNIR}(N, R, T) \quad (4)$$

This report does not report true positive "hit" rates, preferring false negative miss rates for two reasons. First, costs rise linearly with error rates. For example, if we double FNIR in an access control system, then we double user inconvenience and delay. If we express that as decrease of TPIR from, say 98.5% to 97%, then we mentally have to invert the scale to see a doubling in costs. More subtlely, readers don't perceive differences in numbers near 100% well, becoming inured to the "high nineties" effect where numbers close to 100 are perceived indifferently.

Reliability and **sensitivity** are corresponding terms, the former typically being identical to TPIR. This quantity is often cited in automated fingerprint identification system (AFIS) evaluations.

An important special case is the **cumulative match characteristic**(CMC) which summarizes accuracy of mated-searches only. It ignores similarity scores by relaxing the threshold requirement, and just reports the fraction of mated searches returning the mate at rank R or better.

$$\text{CMC}(N, R) = 1 - \text{FNIR}(N, R, 0) \quad (5)$$

We primarily cite the complement of this quantity, FNIR$(N, R, 0)$, the fraction of mates *not* in the top R ranks.

The **rank one hit rate** is the fraction of mated searches yielding the correct candidate at best rank, i.e. CMC(N, 1). While this quantity is the most common summary indicator of an algorithm's efficacy, it is not dependent on similarity scores, so it does not distinguish between strong (high scoring) and weak hits. It also ignores that an adjudicating reviewer is often willing to look at many candidates.

3.3 DET interpretation

In biometrics, a false negative occurs when an algorithm fails to match two samples of one person a Type II error. Correspondingly, a false positive occurs when samples from two persons are improperly associated a Type I error.

Matches are declared by a biometric system when the native comparison score from the recognition algorithm meets some threshold. Comparison scores can be either similarity scores, in which case higher values indicate that the samples are more likely to come from the same person, or dissimilarity scores, in which case higher values indicate different people. Similarity scores are traditionally computed by fingerprint and face recognition algorithms, while dissimilarities are used in iris recognition. In some cases, the dissimilarity score is a distance possessing metric properties. In any case, scores can be either mate scores, coming from a comparison of one persons samples, or nonmate scores, coming from comparison of different persons samples.

The words ”genuine” or ”authentic” are synonyms for mate, and the word ”impostor” is used a synonym for nonmate. The words ”mate” and ”nonmate” are traditionally used in identification applications (such as law enforcement search, or background checks) while genuine and impostor are used in verification applications (such as access control).

An error tradeoff characteristic represents the tradeoff between Type II and Type I classification errors. For identification this plots false negative vs. false positive identification rates i.e. FNIR vs. FPIR parametrically with T. Such plots are often called detection error tradeoff (DET) characteristics or receiver operating characteristic (ROC). These serve the same function error tradeoff but differ, for example, in plotting the complement of an error rate (e.g. TPIR = 1 FNIR) and in transforming the axes, most commonly using logarithms, to show multiple decades of FPIR. More rarely, the function might be the inverse of the Gaussian cumulative distribution function.

The slides of Figures 11 through 18 discuss presentation and interpretation of DETs used in this document for reporting face identification accuracy. Further detail is provided in formal biometrics testing standards, see the various parts of ISO/IEC 19795 Biometrics Testing and Reporting. More terms, including and beyond those to do with accuracy, appear in ISO/IEC 2382-37 Information technology – Vocabulary – Part 37: Harmonized biometric vocabulary

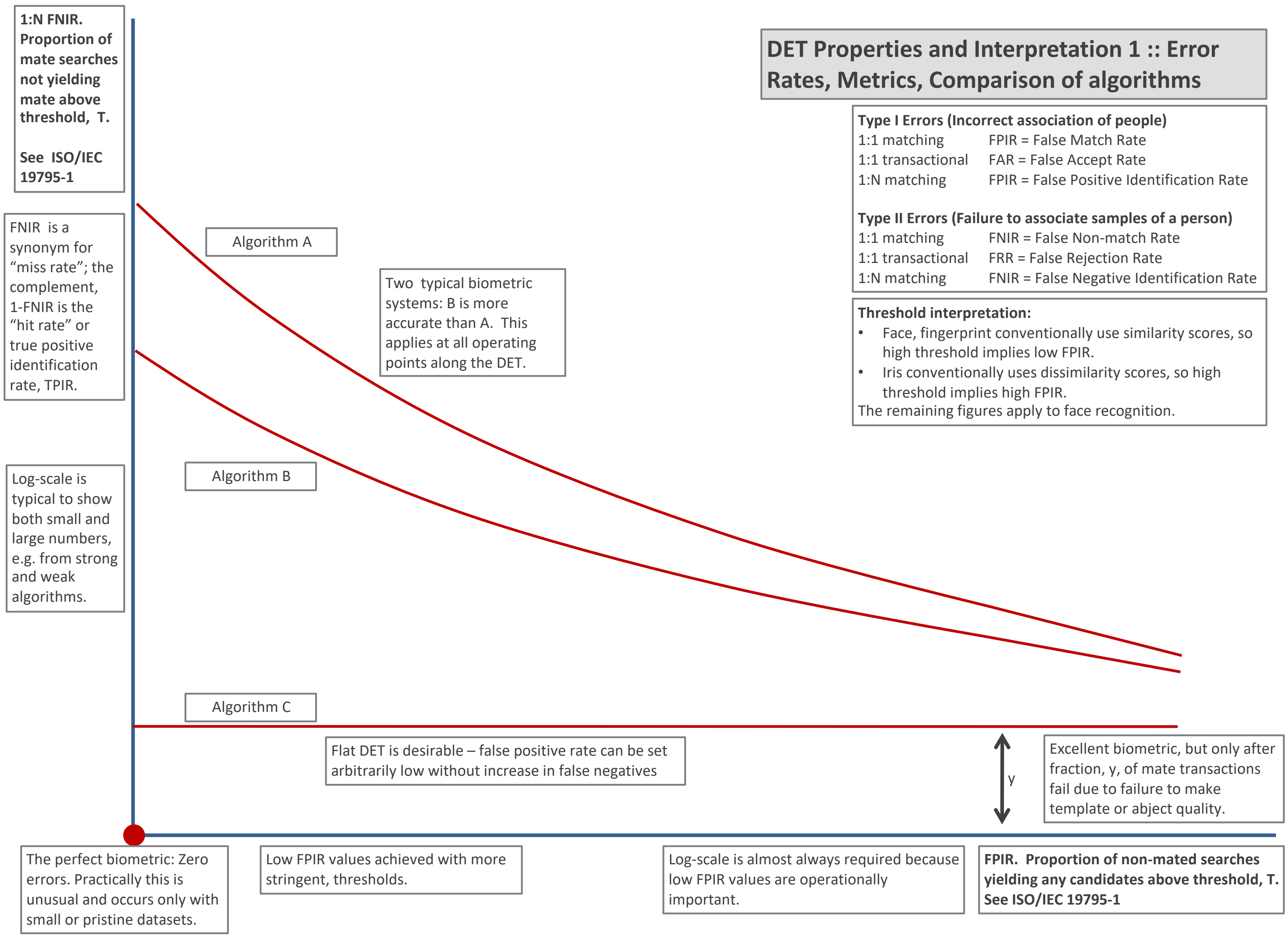

Figure 11: **DET as the primary performance reporting mechanism.**

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

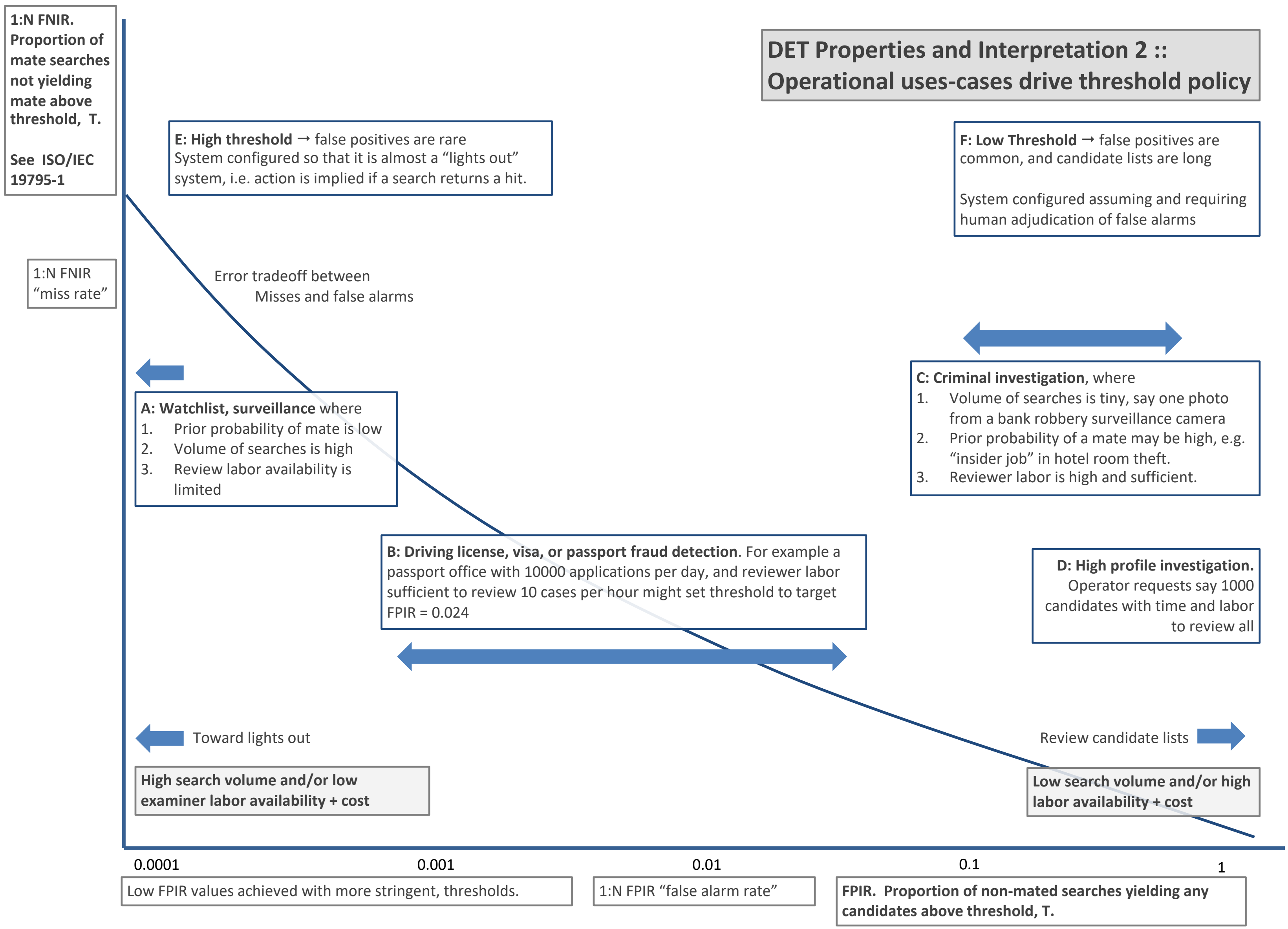

Figure 12: **DET as the primary performance reporting mechanism.**

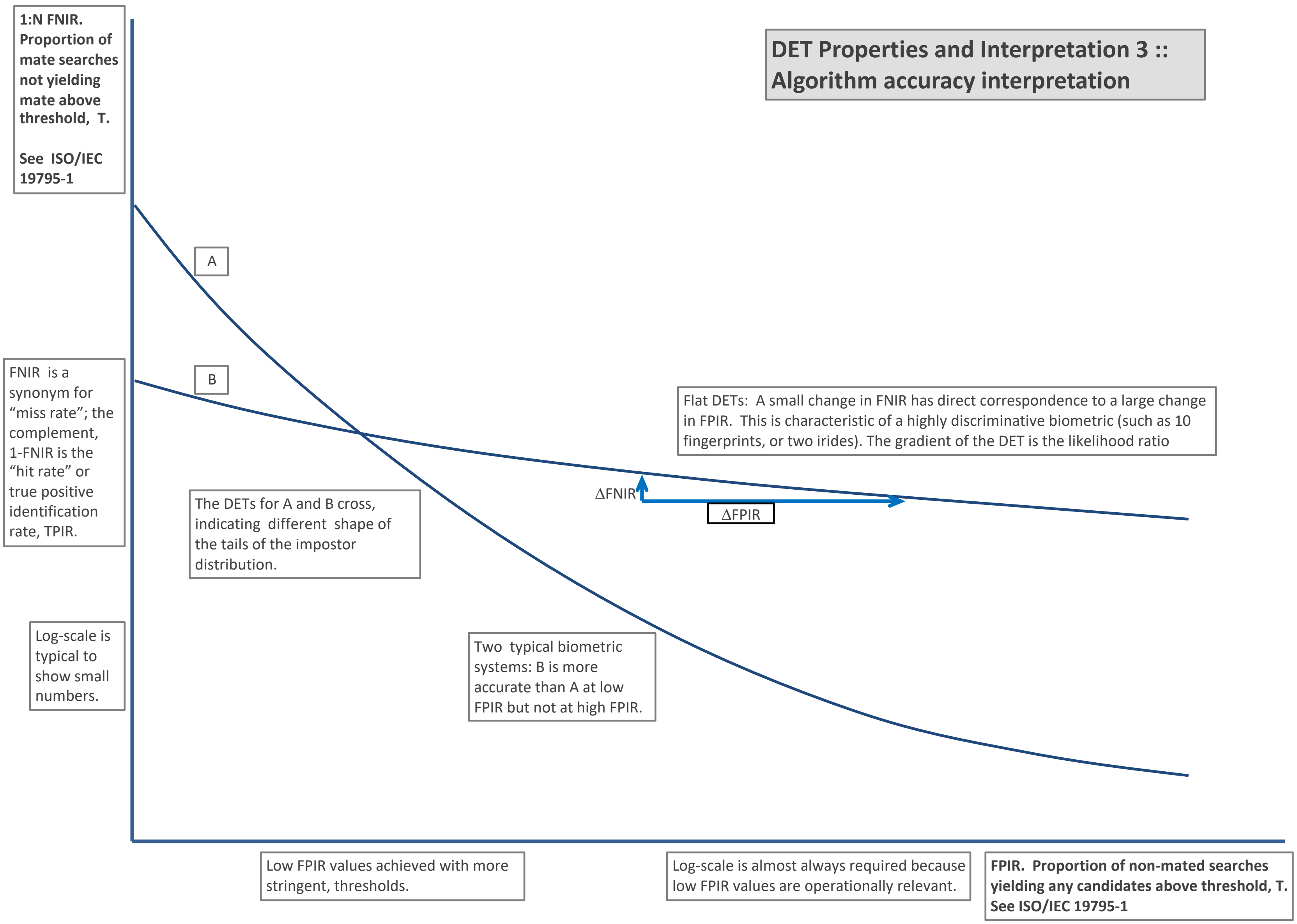

Figure 13: **DET as the primary performance reporting mechanism.**

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

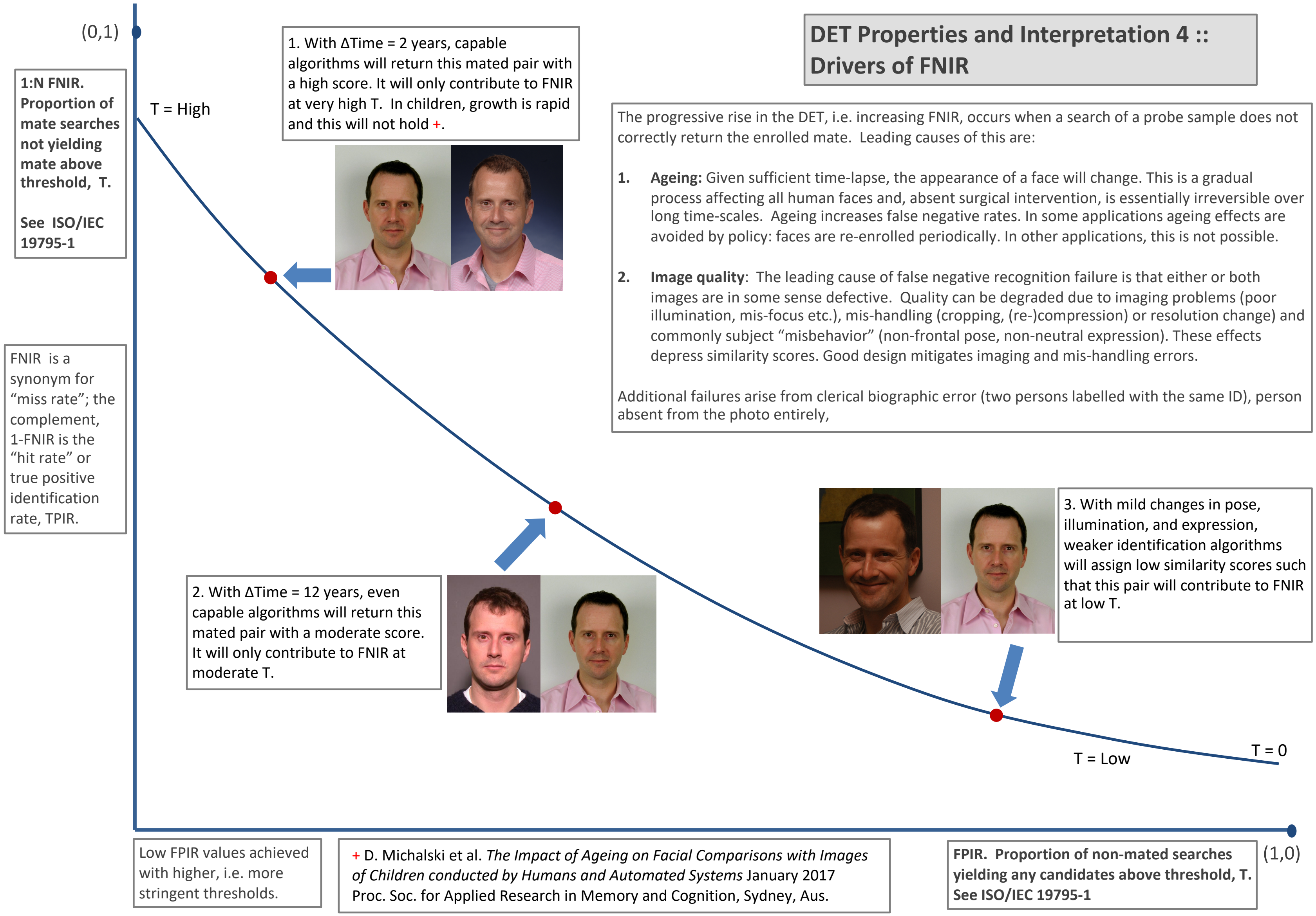

Figure 14: **DET as the primary performance reporting mechanism.**

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

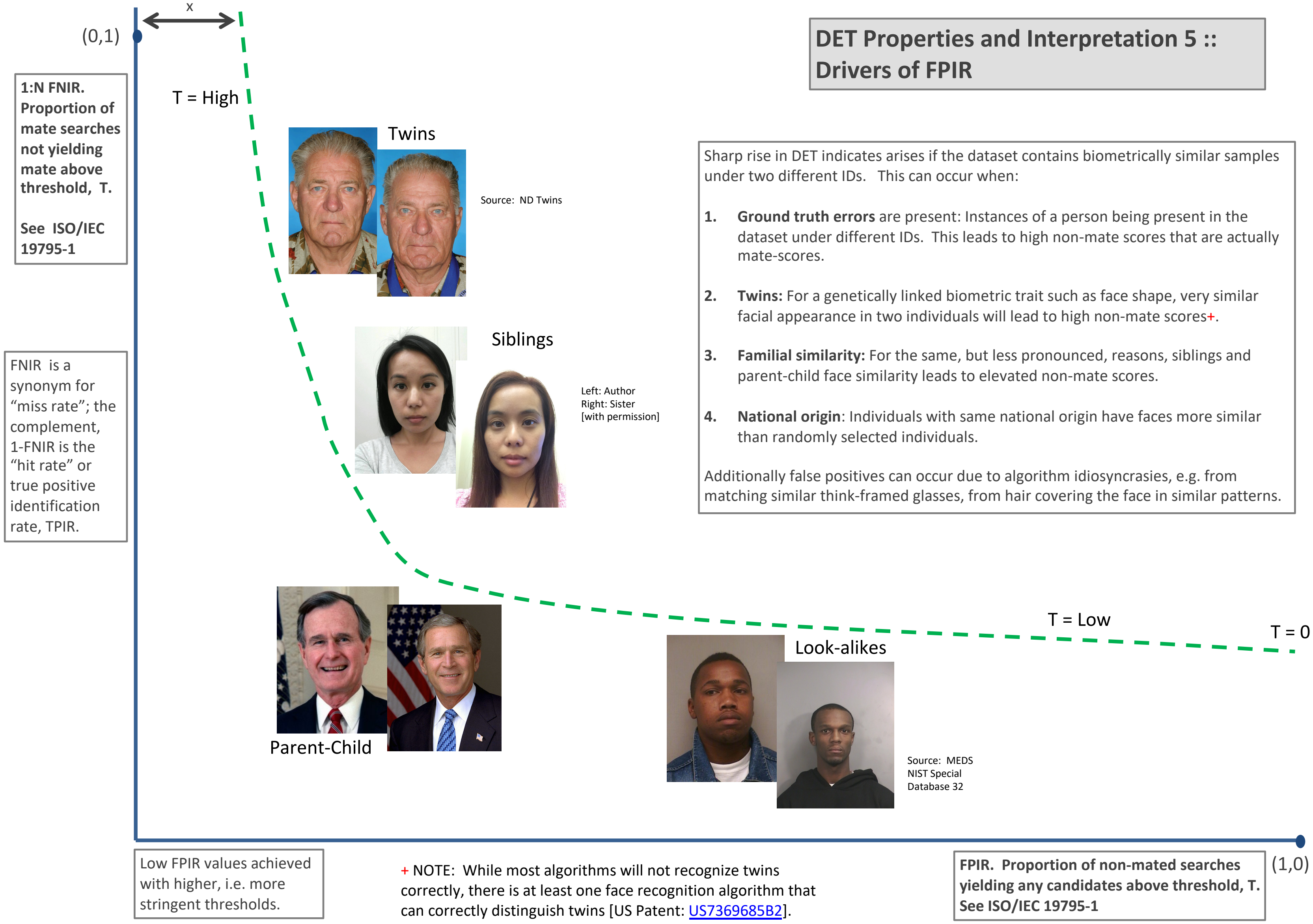

Figure 15: **DET as the primary performance reporting mechanism.**

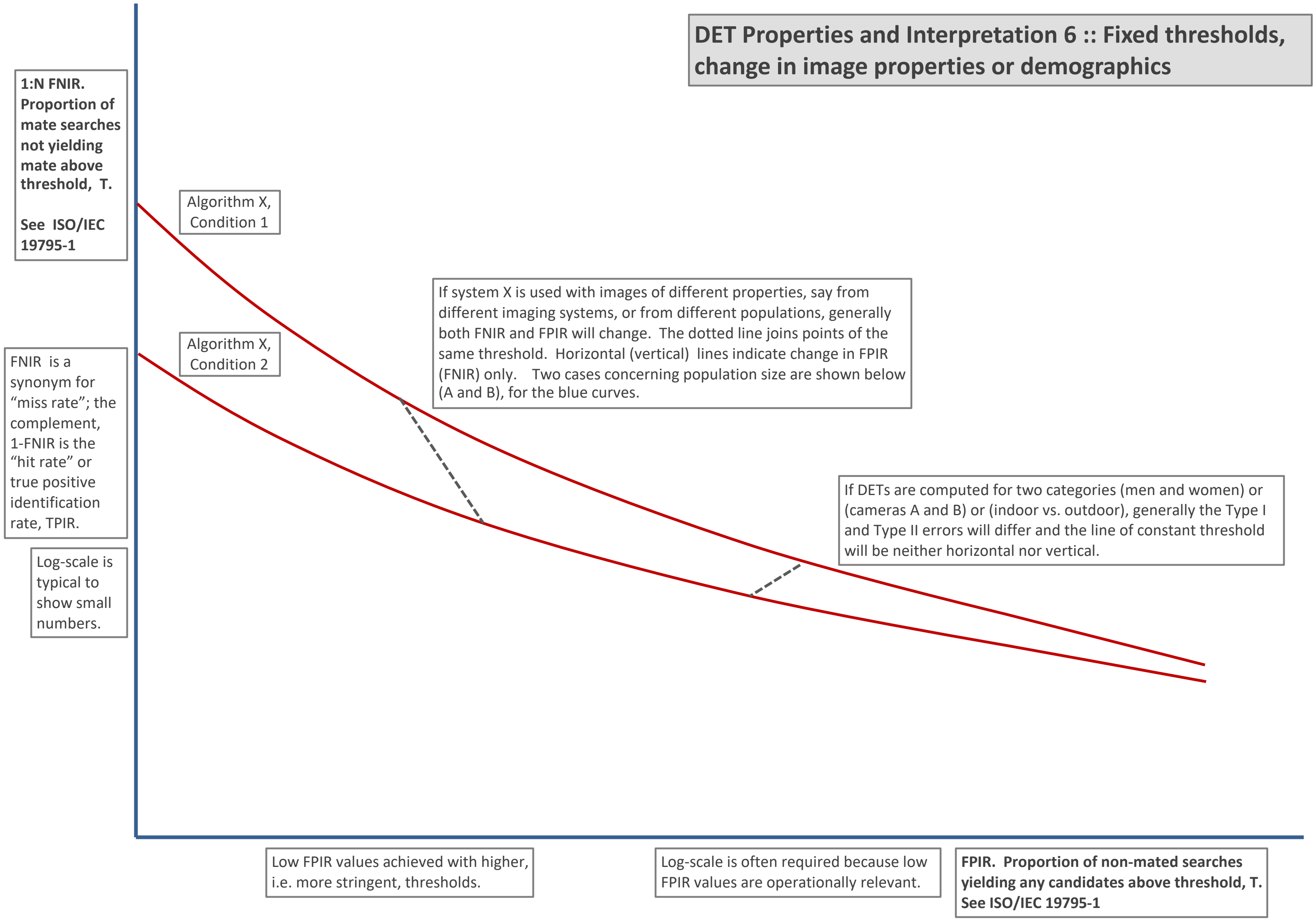

Figure 16: **DET as the primary performance reporting mechanism.**

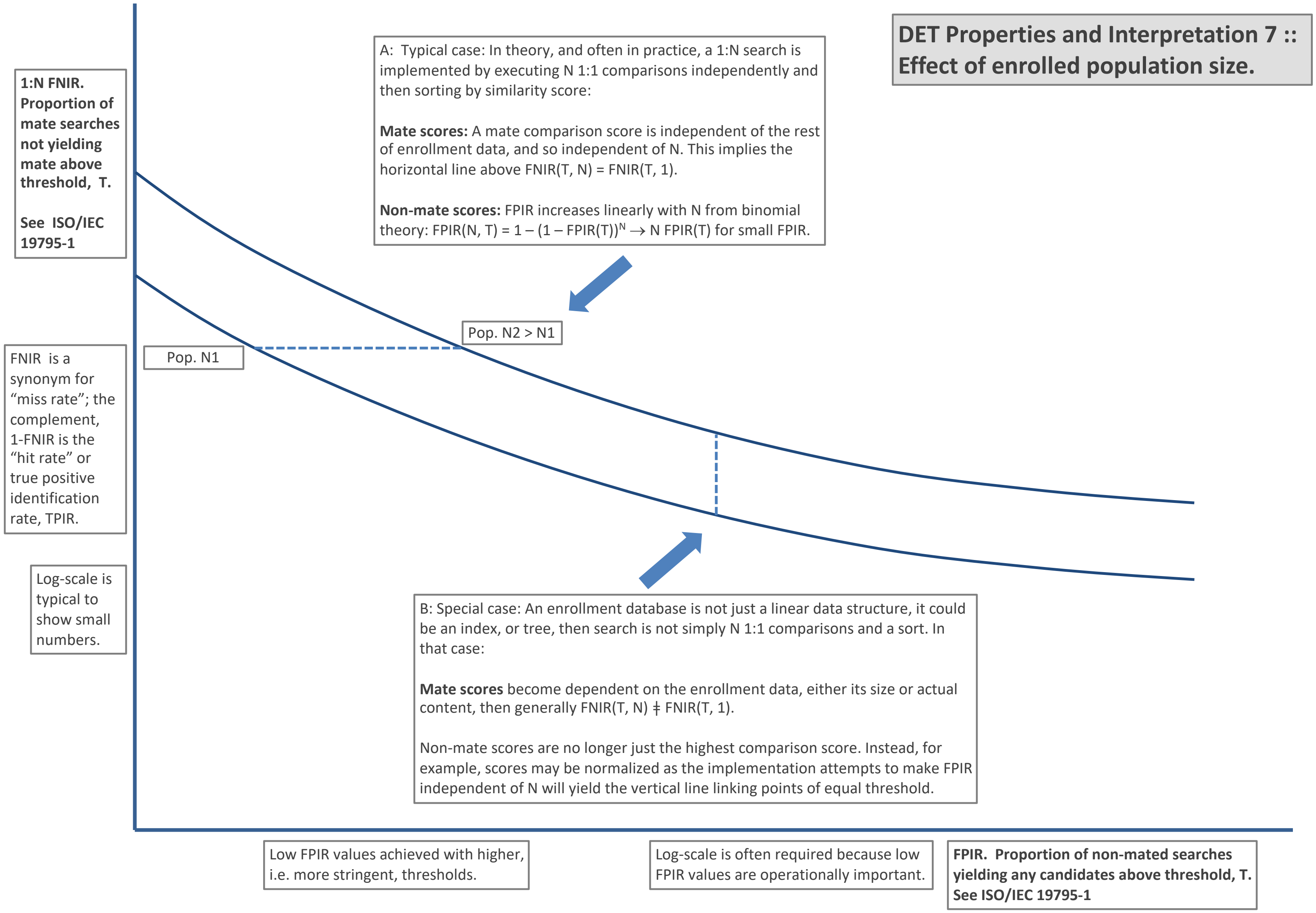

Figure 17: **DET as the primary performance reporting mechanism.**

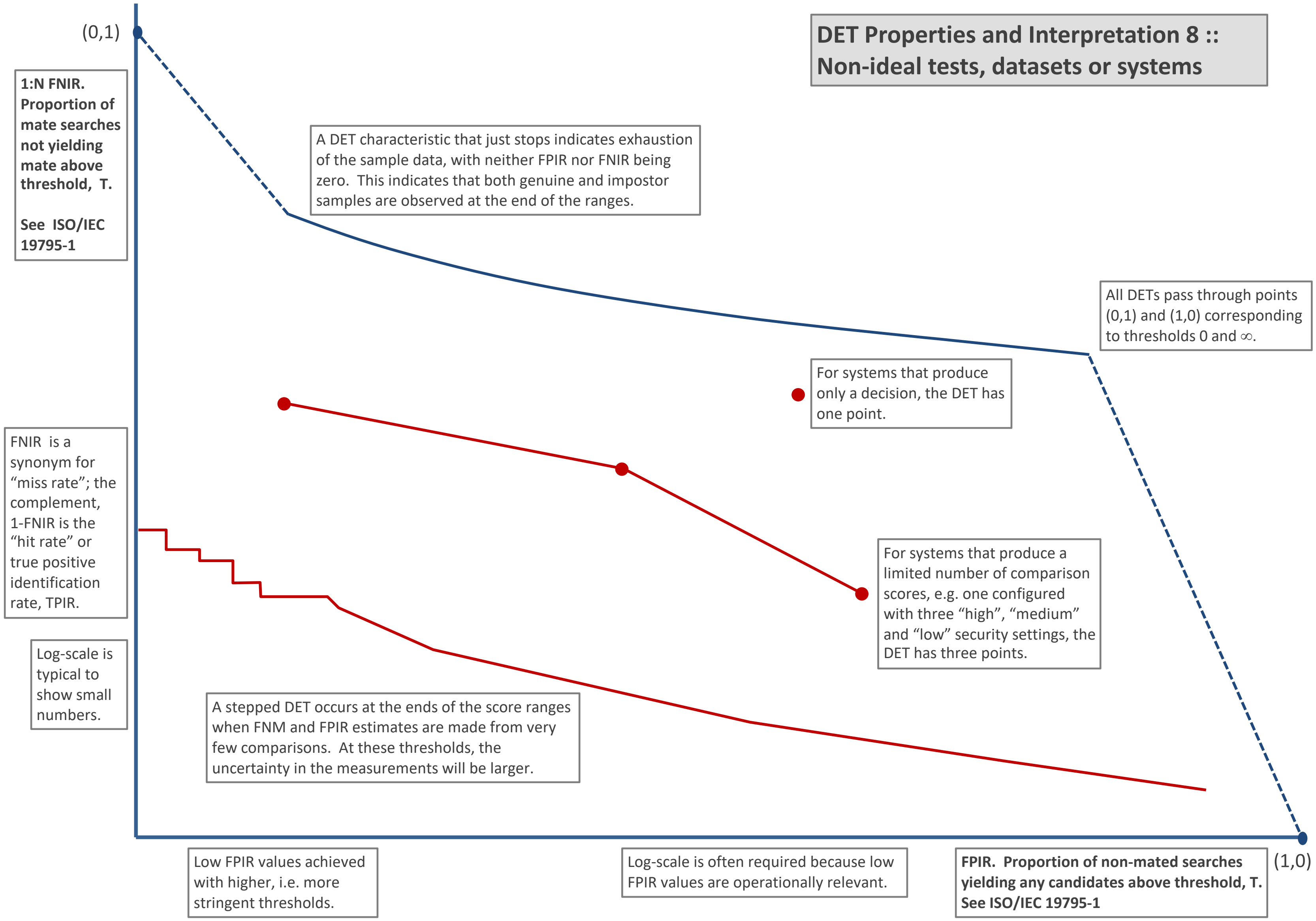

Figure 18: **DET as the primary performance reporting mechanism.**

3.4 Best practice testing requires execution of searches with and without mates

FRVT embeds 1:N searches of two kinds: Those for which there is an enrolled mate, and those for which there is not. The respective numbers for these types of searches appear in Table 6. However, it is common to conduct only mated searches[10]. The cumulative match characteristic is computed from candidate lists produced in mated searches. Even if the CMC is the only metric of interest, the actual trials executed in a test should nevertheless include searches for which no mate exists. As detailed in Table 6 the FRVT reserved disjoint populations of subjects for executing true non-mate searches.

3.5 Failure to extract features

During enrollment some algorithms fail to convert a face image to a template. The proportion of failures is the failure-to-enroll rate, denoted by FTE. Similarly, some search images are not converted to templates. The corresponding proportion is termed failure-to-extract, denoted by FTX.

We do not report FTX because we assume that the same underlying algorithm is used for template generation for enrollment and search.

Failure to extract rates are incorporated into FNIR and FPIR measurements as follows.

- ▷ **Enrollment templates**: Any failed enrollment is regarded as producing a zero length template. Algorithms are required by the API [8] to transparently process zero length templates. The effect of template generation failure on search accuracy depends on whether subsequent searches are mated, or non-mated: Mated searches will fail giving elevated FNIR; non-mated searches will not produce false positives so, to first order, FPIR will be reduced by a factor of $1-\mathrm{FTE}$.
- ▷ **Search templates and 1:N search**: In cases where the algorithm fails to produce a search template from input imagery, the result is taken to be a candidate list whose entries have no hypothesized identities and zero score. The effect of template generation failure on search accuracy depends on whether searches are mated, or non-mated: Mated searches will fail giving elevated FNIR; Non-mated searches will not produce false positives, so FPIR will be reduced.

$$\mathrm{FNIR}^{\dagger} = \mathrm{FTX} + (1 - \mathrm{FTX})\mathrm{FNIR} \tag{6}$$

$$\mathrm{FPIR}^{\dagger} = (1 - \mathrm{FTX})\mathrm{FPIR} \tag{7}$$

This approach is the correct treatment for positive-identification applications such as access control where cooperative users are enrolled and make attempts at recognition. This approach is not appropriate to negative identification applications, such as visa fraud detection, in which hostile individuals may attempt to evade detection by submitting poor quality samples. In those cases, template generation failures should be investigated as though a false alarm had occurred.

[10] For example, the Megaface benchmark. This is bad practice for several reasons: First, if a developer knows, or can reasonably assume, that a mate always exists, then unrealistic gaming of the test is possible. A second reason is that it does not put FPIR on equal footing with FNIR and that matters because in most applications, not all searches have mates - not everyone has been previously enrolled in a driving license issuance or a criminal justice system - so addressing between-class separation becomes necessary.

3.6 Fixed length candidate lists, threshold independent workload

Suppose an automated face identification algorithm returns L candidates, and a human reviewer is retained to examine up to R candidates, where $R \leq L$ might be set by policy, preference or labor availability. For now, assume also that the reviewer is not provided with, or ignores, similarity scores, and thresholds are not applied. Given the algorithm typically places mates at low (good) ranks, the number of candidates a reviewer can be expected to review can be derived as follows. Note that the reviewer will:

- ▷ Always inspect the first ranked image — Frac. reviewed = 1
- ▷ Then inspect those candidates where mate not confirmed at rank 1 — Frac. reviewed = 1-CMC(1)
- ▷ Then inspect those candidates where mate not confirmed at rank 1 or 2 — Frac. reviewed = 1-CMC(2)

etc. Thus if the reviewer will stop after a maximum of R candidates, the expected number of candidate reviews is

$$M(R) = 1 + (1 - CMC(1)) + (1 - CMC(2)) + \ldots + (1 - CMC(R-1)) \tag{8}$$

$$= R - \sum_{r=1}^{R-1} CMC(r) \tag{9}$$

A recognition algorithm that front-loads the cumulative match characteristic will offer reduced workload for the reviewer. This workload is defined only over the searches for which a mate exists. In the cases where there truly is no mate, the reviewer would review all R candidates. Thus, if the proportion of searches for which a mate does exist is β, which in the law enforcement context would be the recidivism rate [2], the full expression for workload becomes:

$$M(R) = \beta \left(R - \sum_{r=1}^{R-1} CMC(r) \right) \quad + \quad (1-\beta)R \tag{10}$$

$$= R - \beta \sum_{r=1}^{R-1} CMC(r) \tag{11}$$

3.7 Timing measurement

Algorithms were submitted to NIST as implementations of the application programming interface(API) specified by NIST in the Evaluation Plan [8]. The API includes functions for initialization, template generation, finalization, search, gallery insert, and gallery delete. Two template generation functions are required, one for the preparation of an enrollment template, and one for a search template.

In NIST's test harness, all functions were wrapped by calls to the C++ std::chrono::high resolution clock which on the dedicated timing machine counts 1ns clock ticks. Precision is somewhat worse than that however.

3.8 Uncertainty estimation

3.8.1 Random error

This study leverages operational datasets for measurement of recognition error rates. This affords several advantages. First, large numbers of searches are conducted (see Table 6) giving precision to the measurements. Moreover, for the two mugshot datasets, these do not involve reuse of individuals so binomial statistics can be expected to apply to recognition error counts. In that case, an observed count of a particular recognition outcome (i.e. a false negative or false positive) in M trials will sustain 95% confidence that the actual error rate is no larger than some value.

As an example, the minimum number of mugshot searches conducted in this report is M =154 549, and the observed FNIR is never below 0.002 so the measurement supports a conclusion that the actual FNIR is no higher than 0.00231 at 99% confidence level. On the false positive side, we tabulate FNIR at FPIR values as low as 0.001. Given estimates based on 331 254 non-mate trials, the actual FPIR values will be below 0.00115 at 99% confidence. In conclusion, large scale evaluation, without reuse of subjects, supports tight uncertainty bounds on the measured error rates.

3.8.2 Systematic error

The FRVT 2018 dataset includes anomalies discovered as a result of inspecting images involved in recognition failures from the most accurate algorithms. Two kinds of failure occur: False negatives (which, for the purpose here, include failures to make templates) and false positives.

False negative errors: We reviewed 600 false negative pairs for which either or both of the leading two algorithms did not put the correct mate in the top 50 candidates. Given 154 549 searches, this number represents 0.39% of the total, resulting in FNIR ~ 0.0039. Of the 600 pairs:

- ▷ **A: Poor quality**: About 20% of the pairs included images of very low quality, often greyscale, low resolution, blurred, low contrast, partially cropped, interlaced, or noisy scans of paper images. Additionally, in a few cases, the face is injured or occluded by bandages or heavy cosmetics.
- ▷ **B: Ground truth identity label bugs**: About 15% of the pairs are not actually mated. We only assigned this outcome when a pair is clearly not mated.
- ▷ **C: Profile views**: About 35% included an image of a profile (side) view of the face, or, more rarely, an image that was rotated 90 degrees in-plane (roll).
- ▷ **D: Tattoos**: About 30% included an image of a tattoo that contained a face image. These arise from mis-labelling in the parent dataset metadata.
- ▷ **E: Ageing**: There is considerable time-lapse between the two captures.

All these estimates are approximate. Of these, the tattoo and mislablled images can never be matched These consistute an accuracy floor in the sample implying that FNIR cannot be below 0.0018[11]. The profile-views and low-quality images could be successfully matched - indeed some algorithms do so. Likewise some poor quality images are matched.

[11] This value is the sum of two partial false negative rates: FNIR_B = 0.15 * 0.0039 plus FNIR_D = 0.3 * 0.0039

For the micrsoft-4 algorithm the lowest miss rate from (recent entry in Table 10) is FNIR(640 000, 50, 0) = 0.0018. This is close to the value estimated from the inspection of misses. It is below the 0.0039 figure because the algorithm does match some profile and poor quality images, that the yitu-2 algorithm does not.

For many tables (e.g. Table 10), the FNIR values obtained for the FRVT-2018 mugshots could be corrected by reducing them by 0.0018. The best values would then be indistinct from zero. The results in this report *were not* adjusted to account for this systematic error.

False positive errors: As depicted in Figure 18 some of the DET characteristics in this report exhibit a pronounced turn upward at low false positive rates. The shape can be caused by identity labelling errors in the ground truth of a dataset, specifically persons present in the database under two IDs such that some proportion of non-mate pairs are actually mated. For each of two algorithms, we reviewed all 330 non-mate pairs for which the first score on candidate lists was above the threshold that gives FPIR = 0.001. The pairs are categorized as follows:

- ▷ **A: Poor quality**: About 1% of image pairs has poor quality such that we cannot conclude anything about the ID of the persons.
- ▷ **B: Ground truth identity label bugs**: For another 44% we are confident that the same person is tagged under two IDs, so that the false positives are in fact not.
- ▷ **C: Same-session mates**: For about 2% we see that the pairs are mated and from the same photography session, yet the IDs are different due to some clerical or procedural mistake.
- ▷ **D: Inderminate ID**: For another 33% we are not confident; The pairs of images may be the same person, or twins, or naturally similar persons, we just cannot decide definitively.
- ▷ **E: Doppelgangers**: For about 20% of pairs we are confident that the probe is actually a different person (doppelganger). Our assessment is conservative - there may be more such pairs. This kind of error is expected from face recognition algorithms in large enough populations.

Of these categories, those in B and C, amounting to 46% of the observed false positives are actually not, such that the FPIR of 0.001 should be restated to about half of that. The results in this report have not been adjusted for this systematic error.

4 Results

This section details performance of the algorithms submitted to Phases 1 and 2 of FRVT 1:N 2018. Performance metrics were described in section 3. The main results are summarized in tabular form with more exhaustive data included as DET, CMC and related graphs in appendices as follows:

- ▷ The three tables 7-9 list algorithms alongside full developer names, acceptance date, size of the provided configuration data, template size and generation time, and search duration data.
  - The **template generation duration** is most important to applications that require fast response. For example, an eGate taking more than two seconds to produce a template might be unacceptable. Note that GPUs may

be of utility in expediting this operation for some algorithms, though at additional expense. Two additional factors should be considered[12][13].

– The **template size** is the size of the extracted feature vector (or vectors) and any needed header information. Large template sizes may be influential on bus or network bandwidth, storage requirements, and on search duration. While the template itself is an opaque data blob, the feature dimensionality might be estimated by assuming a four-bytes-per-float encoding. There is a wide range of encodings. For the more accurate algorithm, sizes range from 256 bytes to 4 138 bytes, indicating essentially no consensus on face modeling and template design.

– The **template size multiplier** column shows how, given k input images, the size of the template grows. Most implementations internally extract features from each image and concatenate them, and implement some score-level fusion logic during search. Other implementations, including many of the most accurate algorithms, produce templates whose size does not grow with k. This could be achieved via selection of the best quality image - but this is not optimal in handling ageing where the oldest image could be the best quality. Another mechanism would be feature-level fusion where information is fused from all k inputs. In any case, as a black-box test, the fusion scheme is proprietary and unknown.

– The size of the **configuration data** is the total size of all files resident in a vendor-provided directory that contains arbitrary read-only files such as parameters, recognition models (e.g caffe). Generally a large value for this quantity may prohibit the use of the algorithm on a resource-constrained device.

▷ Tables 10-11 report core rank-based accuracy for mugshot images. The population size is limited to N = 1.6 million identities because this is the largest gallery size on which all algorithms were executed. Notable observations from these tables are as follows:

– **Massive accuracy gains since 2014**: The FRVT 2014 columns show results for an exact repeat of the main identification experiment reported in the main FRVT 2014 report. The most accurate algorithm in 2018, microsoft-4, gives FNIR = 0.002 vs. the 2014 result for NEC of FNIR = 0.041. This constitutes almost a twenty-fold reduction in false negatives. Given 50 000 mated searches, there were 2 043 that did not yield a rank-1 mate in 2014. Of those, 1 929 now do because their score has been elevated to the top of the candidate list, above impostor scores. This reflects the algorithms' newfound ability to compensate for image quality problems and ageing.

– **Accuracy 2013-2018 vs. 2010-2013**: To put the accuracy gains into context, the gains in the period Feburary 2010 - October 2013[14] were very modest, a 1.1 fold reduction for Neurotechnology, Cognitec and Morpho and 1.4 fold reduction for NEC.

– The massive accuracy gains are consistent with an **industrial revolution** associated with the incorporation of convolutional neural network based techniques into the prototypes. This is distinct from the evolution measured in the prior period. We further note that the revolution is not over: Figure 19 shows that many developers have made great advances in the four months between Phases 1 and 2 of FRVT 2018, Feburary to

[12]The FRVT 2018 API prohibited threading, so some gains from parallelism may be available on multiple-cores or multiple processors, if the feature extraction code and be distributed across them.

[13]Note also that factors of two or more may be realizable by exploiting modern vector processing instructions on CPUs. It is not clear in our measurements whether all developers exploited Intel's AVX2 instructions, for example. Our machine was so equipped, but we insisted that the same compiled library should also run on older machines lacking that instruction. The more sophisticated implementations may have detected AVX2 presence and branched accordingly. The less sophisticated may be defaulted to the reduced instruction set. Readers should see the FRVT 2018 API document for the specific chip details.

[14]See NIST Interagency Reports 7709 and 8009.

June. Most developers saw a two-fold reduction in errors, with Neurotechnology seeing a five fold reduction. Given such rapid gains, the revolution is apparently on-going and we expect further gains in Phase 3 starting October 30, 2018. In particular, the developers who only participated in Phase 1 (e.g. Megvii) or Phase 2 (e.g. Cogent, Cognitec, NEC) may realize gains given knowledge of their initial FRVT results.

- The prevalance of green entries shows **broad accuracy gains since 2014** - around 28 developers now produce algorithms that give better FNIR(N, 1) values than the most accurate algorithm submitted to NIST in October 2013. For the developers who participated in both FRVT 2014 and FRVT 2018, the error rate reductions are plotted in Figure 20
- **Wide range in accuracy**: The rank-1 miss rates vary from FNIR(N, 1) = 0.002 for microsoft-4 up to about 0.5 for the very fast but inaccurate microfocus-x algorithms. Among the developers who are superior to NEC in 2013, the range is from 0.002 to 0.035 for camvi-3. This large accuracy range is consistent with the buyer-beware maxim, and indicates that face recognition software is far from being commoditized.
- **FRVT 2018 is more difficult than FRVT 2014**: Almost all FNIR values for the FRVT 2018 dataset are higher than those for the FRVT 2014 set. Both datasets come from the same source but differ in their preparation as depicted in Figure 1. Particularly, the earlier set employed a circa 2009 face detector to allow an image into the dataset. That would have excluded lower quality e.g. low-contrast or poorly posed faces.

▷ Tables 12-13 report threshold-based error rates, FNIR(N, L, T), for N = 1.6 million for mugshot-mugshot accuracy on FRVT 2014, FRVT 2018, and also (in pink) mugshot-webcam accuracy using FRVT 2018 enrollments. Notable observations from these tables are as follows:

- **Order of magnitude accuracy gains since 2014**: As with rank-based results, the gains in accuracy are substantial, though somewhat reduced. At FPIR = 0.01, the best improvement over NEC in 2014 is a nine-fold reduction in FNIR using the Microsoft_4 algorithm. At FPIR = 0.001, the largest gain is a six-fold reduction in FNIR via the Yitu_2 algorithm.
- **Broad gains across the industry**: About 19 companies realize accuracy better than the NEC benchmark from 2014. This is somewhat lower than the 28 developers who succeeded on the rank-1 metric. This may be due to the ubiquity of, and emphasis on, the rank-1 metric in many published algorithm development papers.
- **Webcam images**: Searches of webcam images give FNIR(N, T) values around 2 to 3 times higher than mugshot searches. Notably the leading developers with mugshots are approximately the same with poorer quality webcams. But some developers e.g. Camvi, Megvii, TongYi, and Neurotechnology do improve their relative rankings on webcams, perhaps indicating their algorithms were tailored to less constrained images.

▷ Tables 14, 15 and 16 show, respectively, rank 1, rank 50 and high-threshold FNIR values for all algorithms performing searches into five different gallery sizes, N = 640 000, N = 1 600 000, N = 3 000 000, N = 6 000 000 and 12 000 000. The Rank-1 table is included as a primary accuracy indicator. The Rank-50 table is included to inform agencies who routinely produce 50 candidates for human-review. The FPIR = 0.001 table is included to inform high-volume duplicate detection applications. The notable results are

- **Slow growth in rank-based miss rates:** FNIR(N, R) generally grows as a power law, aN^b. From the straight lines of many graphs of Figure 31 this is clearly a reasonable model for most, but not all, algorithms. The coefficient a can be interpreted as FNIR in a gallery of size 1. The more important coefficient b indicates scalability, and often, $b \ll 1$, implies very benign growth in FNIR. The coefficients of the models appear in the Tables 14 and 15.

– **Slow growth in threshold-based miss rates:** FNIR(N, T) also generally grows as a power law, aN^b except at the high threshold values corresponding to low FPIR values. This is visible in the plots of Figure 51 which show straight lines except for FPIR = 0.001, which increase more rapidly with N above 3 000 000. Each trace in those figures shows FNIR(N, T) at fixed FPIR with both N and T varying. Thus at large N, it is usually necessary to elevate T to maintain fixed FPIR. This causes increased FNIR. Why that would no-longer obey a power-law is not known. However, if we expect large galleries to contain individuals with familial relations to the non-mate search images - in the most extreme case, twins - then suppression of false positives becomes more difficult. This is discussed in the Figures starting at Fig. 18

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

	DEVELOPER	SHORT	SEQ.	VALIDATION	CONFIG[1]	TEMPLATE GENERATION			SEARCH DURATION[4] MILLISEC		
	FULL NAME	NAME	NUM.	DATE	DATA (MB)	SIZE (B)	MULT[2]	TIME (MS)[3]	N=1.6M		POWER LAW (μS)
									L=1	L=50	
1	3Divi	3divi	0	2018-02-09	186	[116]4096	k	[64]426	-	[71]553	[66]0.33 $N^{1.0}$
2	3Divi	3divi	1	2018-02-15	187	[124]4224	k	[68]428	-	[12]37	[37]0.03 $N^{1.0}$
3	3Divi	3divi	2	2018-02-15	187	[34]528	k	[66]428	-	[11]33	[73]0.02 $N^{1.0}$
4	3Divi	3divi	3	2018-06-19	165	[29]512	k	[91]625	[7]76	[15]76	[62]0.05 $N^{1.0}$
5	3Divi	3divi	4	2018-06-19	186	[114]4096	k	[92]628	[34]604	[82]801	[30]0.75 $N^{1.0}$
6	Alchera	alchera	0	2018-06-30	168	[94]2048	k	[34]263	[57]3296	[126]5420	[124]0.10 $N^{1.2}$
7	Alchera	alchera	1	2018-06-30	46	[80]2048	k	[6]66	[58]3516	[127]5489	[126]0.05 $N^{1.3}$
8	Aware	aware	0	2018-02-16	261	[68]1564	k	[98]653	-	[38]251	[40]0.19 $N^{1.0}$
9	Aware	aware	1	2018-02-16	232	[69]1564	k	[97]651	-	[39]251	[34]0.21 $N^{1.0}$
10	Aware	aware	2	2018-02-16	349	[105]2076	k	[128]912	-	[40]252	[42]0.19 $N^{1.0}$
11	Aware	aware	3	2018-06-22	350	[104]2076	k	[110]716	[54]2426	[111]2508	[117]0.50 $N^{1.1}$
12	Aware	aware	4	2018-06-22	349	[2]92	k	[108]712	[40]1232	[91]1187	[113]0.33 $N^{1.1}$
13	Ayonix	ayonix	0	2018-06-21	57	[57]1036	k	[1]10	[20]283	[46]298	[24]0.30 $N^{1.0}$
14	Camvi Technologies	camvitech	1	2018-02-16	94	[50]1024	1	[19]177	-	[9]23	[2]7066.90 $N^{0.1}$
15	Camvi Technologies	camvitech	2	2018-02-16	442	[54]1024	1	[114]774	-	[8]20	[1]7180.65 $N^{0.1}$
16	Camvi Technologies	camvitech	3	2018-06-30	233	[52]1024	1	[107]707	[4]10	[6]11	[5]857.59 $N^{0.2}$
17	Gemalto Cogent	cogent	0	2018-06-20	533	[33]525	k	[83]551	[31]494	[73]558	[38]0.46 $N^{1.0}$
18	Gemalto Cogent	cogent	1	2018-06-20	533	[32]525	k	[84]552	[32]498	[72]556	[46]0.39 $N^{1.0}$
19	Cognitec Systems GmbH	cognitec	0	2018-06-21	364	[100]2052	k	[18]176	[44]1748	[97]1780	[109]0.57 $N^{1.0}$
20	Cognitec Systems GmbH	cognitec	1	2018-06-21	412	[97]2052	k	[23]202	[46]1835	[96]1735	[115]0.45 $N^{1.1}$
21	Dermalog	dermalog	0	2018-02-16	0	[4]128	1	[48]344	-	[59]404	[88]0.19 $N^{1.0}$
22	Dermalog	dermalog	1	2018-02-16	0	[6]128	1	[17]171	-	[61]407	[99]0.17 $N^{1.0}$
23	Dermalog	dermalog	2	2018-02-16	0	[13]256	k	[47]344	-	[79]640	[63]0.40 $N^{1.0}$
24	Dermalog	dermalog	3	2018-06-21	0	[5]128	1	[25]211	[9]92	[16]92	[64]0.06 $N^{1.0}$
25	Dermalog	dermalog	4	2018-06-21	0	[3]128	1	[24]208	[8]91	[17]93	[85]0.05 $N^{1.0}$
26	Ever AI	everai	0	2018-06-21	142	[91]2048	1	[70]438	[3]4	[4]3	[8]42.41 $N^{0.3}$
27	Ever AI	everai	1	2018-06-21	200	[75]2048	1	[89]590	[24]336	[53]356	[123]0.03 $N^{1.1}$
28	Eyedea Recognition	eyedea	0	2018-02-16	644	[123]4152	k	[63]424	-	[80]640	[80]0.34 $N^{1.0}$
29	Eyedea Recognition	eyedea	1	2018-02-16	287	[60]1036	k	[42]311	-	[48]307	[87]0.15 $N^{1.0}$
30	Eyedea Recognition	eyedea	2	2018-02-16	287	[58]1036	k	[69]429	-	[47]305	[78]0.16 $N^{1.0}$
31	Eyedea Recognition	eyedea	3	2018-06-18	284	[59]1036	k	[51]385	[21]309	[50]311	[53]0.21 $N^{1.0}$
32	Glory Ltd	glory	0	2018-06-30	0	[22]418	k	[13]160	[33]575	[74]575	[93]0.26 $N^{1.0}$
33	Glory Ltd	glory	1	2018-06-30	0	[70]1726	k	[58]405	[47]1864	[99]1978	[91]0.93 $N^{1.0}$
34	Gorilla Technology	gorilla	0	2018-02-01	95	[130]8300	k	[65]427	-	[128]10426	[83]5.30 $N^{1.0}$
35	Gorilla Technology	gorilla	1	2018-06-19	91	[107]2156	k	[16]169	[64]5254	[123]5156	[57]3.31 $N^{1.0}$
36	loginface Corp	hbinno	0	2018-02-01	88	[31]520	-	[35]265	-	[62]419	[39]0.34 $N^{1.0}$
37	Hikvision Research Institute	hikvision	0	2018-02-12	378	[72]1808	1	[126]875	-	[108]2360	[102]0.97 $N^{1.0}$
38	Hikvision Research Institute	hikvision	1	2018-02-12	378	[74]1808	1	[118]820	-	[109]2403	[98]1.00 $N^{1.0}$
39	Hikvision Research Institute	hikvision	2	2018-02-12	378	[73]1808	1	[116]820	-	[110]2408	[97]1.00 $N^{1.0}$
40	Hikvision Research Institute	hikvision	3	2018-06-30	408	[63]1408	1	[94]633	[37]904	[89]1108	[36]0.91 $N^{1.0}$
41	Hikvision Research Institute	hikvision	4	2018-06-30	334	[62]1152	1	[73]510	[36]784	[85]1024	[33]0.86 $N^{1.0}$
42	Idemia	idemia	0	2018-02-16	371	[21]364	1	[60]416	-	[19]133	[69]0.08 $N^{1.0}$
43	Idemia	idemia	1	2018-02-16	371	[19]364	1	[61]417	-	[22]138	[84]0.07 $N^{1.0}$
44	Idemia	idemia	2	2018-02-16	371	[20]364	1	[62]417	-	[23]138	[79]0.07 $N^{1.0}$

Notes	
1	Configuration size does not capture static data present in libraries. Libraries are not counted because most implementations include common ancilliary libraries for image processing (e.g. openCV) or numerical computation (e.g. blas).
2	This multiplier expresses the increase in template size when k images are passed to the template generation function.
3	All durations are measured on Intel®Xeon®CPU E5-2630 v4 @ 2.20GHz processors. Estimates are made by wrapping the API function call in calls to std::chrono::high_resolution_clock which on the machine in (3) counts 1ns clock ticks. Precision is somewhat worse than that however.
4	Search durations are measured as in the prior note. The power-law model in the final column mostly fits the empirical results in Figure 103. However in certain cases the model is not correct and should not be used numerically.

Table 7: Summary of algorithms and properties included in this report. The blue superscripts give ranking for the quantity in that column. Missing search durations, denoted by "-", are absent because those runs were not executed.

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

	DEVELOPER	SHORT	SEQ.	VALIDATION	CONFIG[1]	TEMPLATE GENERATION			SEARCH DURATION[4] MILLISEC		
	FULL NAME	NAME	NUM.	DATE	DATA (MB)	SIZE (B)	MULT[2]	TIME (MS)[3]	N=1.6M		POWER LAW (μS)
									L=1	L=50	
45	Idemia	idemia	3	2018-06-21	472	[35]528	1	[103]689	[23]318	[54]361	[12]5.03 $N^{0.8}$
46	Idemia	idemia	4	2018-06-21	472	[36]528	1	[102]669	[12]168	[34]211	[121]0.02 $N^{1.1}$
47	Imagus Technology Pty Ltd	imagus	0	2018-02-14	35	[26]512	k	[3]43	-	[30]202	[31]0.19 $N^{1.0}$
48	Imagus Technology Pty Ltd	imagus	2	2018-06-21	35	[23]512	k	[7]76	[16]200	[33]208	[29]0.20 $N^{1.0}$
49	Imagus Technology Pty Ltd	imagus	3	2018-06-21	46	[28]512	k	[5]57	[17]201	[31]206	[25]0.21 $N^{1.0}$
50	Incode Technologies	incode	0	2018-06-29	23	[55]1024	k	[22]190	[41]1293	[119]3510	[127]0.00 $N^{1.5}$
51	Incode Technologies	incode	1	2018-06-29	151	[92]2048	k	[104]690	[42]1542	[121]4497	[125]0.06 $N^{1.3}$
52	Innovatrics	innovatrics	0	2018-02-16	0	[39]530	k	[71]455	-	[78]625	[27]0.61 $N^{1.0}$
53	Innovatrics	innovatrics	1	2018-02-16	0	[37]530	k	[44]316	-	[77]625	[26]0.62 $N^{1.0}$
54	Innovatrics	innovatrics	2	2018-06-21	0	[38]530	k	[32]255	[2]1	[2]2	[3]616.66 $N^{0.1}$
55	Innovatrics	innovatrics	3	2018-06-21	0	[40]530	k	[33]255	[49]2020	[98]1882	[48]1.30 $N^{1.0}$
56	Alivia / Innovation Sys.	isystems	0	2018-02-14	262	[86]2048	1	[27]222	-	[56]393	[81]0.21 $N^{1.0}$
57	Alivia / Innovation Sys.	isystems	1	2018-02-14	263	[45]1024	1	[26]222	-	[35]240	[60]0.15 $N^{1.0}$
58	Alivia / Innovation Sys.	isystems	2	2018-02-14	268	[82]2048	1	[45]316	[27]385	[66]484	[21]0.68 $N^{0.9}$
59	Megvii	megvii	0	2018-02-15	1327	[89]2048	1	[115]794	-	[45]284	[56]0.18 $N^{1.0}$
60	Microfocus	microfocus	0	2018-02-12	101	[15]256	k	[74]525	-	[27]184	[49]0.13 $N^{1.0}$
61	MicroFocus	microfocus	1	2018-02-16	101	[10]256	k	[75]527	-	[13]39	[120]0.00 $N^{1.1}$
62	Microfocus	microfocus	2	2018-02-16	101	[16]256	k	[76]529	-	[3]2	[11]0.61 $N^{0.6}$
63	Microfocus	microfocus	3	2018-06-22	101	[11]256	k	[36]269	[13]185	[28]188	[50]0.13 $N^{1.0}$
64	Microfocus	microfocus	4	2018-06-22	102	[14]256	k	[37]270	[14]186	[29]189	[45]0.13 $N^{1.0}$
65	Microsoft	microsoft	0	2018-01-30	126	[30]512	1	[38]283	-	[76]593	[106]0.22 $N^{1.0}$
66	Microsoft	microsoft	1	2018-02-12	165	[49]1024	1	[49]349	-	[84]869	[108]0.29 $N^{1.0}$
67	Microsoft	microsoft	2	2018-02-12	228	[53]1024	1	[85]555	-	[83]869	[107]0.32 $N^{1.0}$
68	Microsoft	microsoft	3	2018-06-20	230	[47]1024	1	[57]404	[43]1638	[95]1603	[110]0.51 $N^{1.1}$
69	Microsoft	microsoft	4	2018-06-20	437	[83]2048	1	[113]773	[56]2662	[113]2691	[111]0.83 $N^{1.1}$
70	NEC	nec	0	2018-06-21	131	[109]2592	k	[8]82	[22]317	[63]426	[18]0.73 $N^{0.9}$
71	NEC	nec	1	2018-06-29	131	[108]2592	k	[9]88	[15]193	[32]208	[28]0.21 $N^{1.0}$
72	Neurotechnology	neurotech	0	2018-02-16	331	[126]5214	k	[105]702	-	[116]3040	[70]1.79 $N^{1.0}$
73	Neurotechnology	neurotech	1	2018-02-16	331	[127]5214	k	[100]661	-	[118]3054	[67]1.82 $N^{1.0}$
74	Neurotechnology	neurotech	2	2018-02-16	331	[128]5214	k	[99]658	-	[117]3051	[65]1.85 $N^{1.0}$
75	Neurotechnology	neurotech	3	2018-06-27	265	[76]2048	k	[82]547	[39]1084	[86]1059	[61]0.73 $N^{1.0}$
76	Neurotechnology	neurotech	4	2018-06-27	265	[93]2048	k	[81]543	[38]1060	[87]1061	[22]1.22 $N^{1.0}$
77	N-Tech Lab	ntech	0	2018-02-16	2124	[125]4442	k	[111]730	-	[55]382	[41]0.27 $N^{1.0}$
78	N-Tech Lab	ntech	1	2018-02-16	851	[71]1736	k	[59]405	-	[24]161	[71]0.09 $N^{1.0}$
79	N-Tech Lab	ntech	3	2018-06-21	3664	[110]3484	k	[121]831	[26]384	[52]326	[43]0.24 $N^{1.0}$
80	N-Tech Lab	ntech	4	2018-06-21	3766	[111]3484	k	[129]929	[25]378	[51]312	[54]0.21 $N^{1.0}$
81	Rank One Computing	rankone	0	2018-02-07	0	[9]228	k	[4]50	-	[14]75	[20]0.12 $N^{0.9}$
82	Rank One Computing	rankone	1	2018-02-15	0	[18]324	k	[12]136	-	[26]169	[10]396.79 $N^{0.4}$
83	Rank One Computing	rankone	2	2018-06-19	0	[7]133	k	[10]113	[10]138	[20]137	[44]0.10 $N^{1.0}$
84	Rank One Computing	rankone	3	2018-06-19	0	[8]133	k	[11]114	[11]138	[21]137	[51]0.09 $N^{1.0}$
85	Rank One Computing	rankone	4	2018-10-09	0	[1]85	-	[2]36	-	[18]101	[130]-
86	RealNetworks	realnetworks	0	2018-06-21	96	[117]4100	1	[31]244	[60]4257	[114]2740	[75]1.51 $N^{1.0}$
87	RealNetworks	realnetworks	1	2018-06-21	105	[118]4104	k	[30]243	[59]3568	[102]2107	[74]1.16 $N^{1.0}$
88	Shaman Software	shaman	0	2018-02-12	0	[115]4096	k	[79]538	-	[67]523	[47]0.37 $N^{1.0}$

Notes	
1	Configuration size does not capture static data present in libraries. Libraries are not counted because most implementations include common ancilliary libraries for image processing (e.g. openCV) or numerical computation (e.g. blas).
2	This multiplier expresses the increase in template size when k images are passed to the template generation function.
3	All durations are measured on Intel®Xeon®CPU E5-2630 v4 @ 2.20GHz processors. Estimates are made by wrapping the API function call in calls to std::chrono::high_resolution_clock which on the machine in (3) counts 1ns clock ticks. Precision is somewhat worse than that however.
4	Search durations are measured as in the prior note. The power-law model in the final column mostly fits the empirical results in Figure 103. However in certain cases the model is not correct and should not be used numerically.

Table 8: Summary of algorithms and properties included in this report. The blue superscripts give ranking for the quantity in that column. Missing search durations, denoted by "-", are absent because those runs were not executed.

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

	DEVELOPER FULL NAME	SHORT NAME	SEQ. NUM.	VALIDATION DATE	CONFIG[1] DATA (MB)	TEMPLATE GENERATION SIZE (B)	MULT[2]	TIME (MS)[3]	SEARCH DURATION[4] MILLISEC N=1.6M L=1	L=50	POWER LAW (μS)
89	Shaman Software	shaman	1	2018-02-12	0	[113]4096	k	[86]557	-	[68]524	[55]0.35 $N^{1.0}$
90	Shaman Software	shaman	2	2018-02-12	0	[129]8192	k	[87]557	-	[81]688	[76]0.38 $N^{1.0}$
91	Shaman Software	shaman	3	2018-06-30	0	[81]2048	k	[106]704	[35]692	[49]310	[14]1.04 $N^{0.9}$
92	Shaman Software	shaman	4	2018-06-30	0	[88]2048	k	[96]642	[29]434	[42]267	[19]0.46 $N^{0.9}$
93	Shenzhen Inst. Adv. Tech. CAS	SIAT	0	2018-02-14	306	[61]1096	k	[50]358	-	[94]1343	[59]0.86 $N^{1.0}$
94	Shenzhen Inst. Adv. Tech. CAS	SIAT	1	2018-06-30	521	[95]2052	1	[123]842	[61]4512	[120]4402	[95]2.06 $N^{1.0}$
95	Shenzhen Inst. Adv. Tech. CAS	SIAT	2	2018-06-30	521	[99]2052	1	[127]906	[62]5101	[122]4884	[96]2.08 $N^{1.0}$
96	Smilart	smilart	0	2018-02-15	105	[46]1024	k	[15]168	-	[92]1285	[23]1.30 $N^{1.0}$
97	Smilart	smilart	1	2018-02-15	120	[51]1024	k	[101]662	-	[90]1135	[15]3.75 $N^{0.9}$
98	Smilart	smilart	2	2018-02-15	109	[48]1024	k	[88]560	-	[93]1302	[35]1.08 $N^{1.0}$
99	Smilart	smilart	4	2018-06-29	65	[25]512	-	[14]167	-	[129]15382	[128]-
100	Smilart	smilart	5	2018-06-29	562	[84]2048	-	[72]464	-	-	[129]-
101	Synesis	synesis	0	2018-02-15	332	[27]512	k	[29]237	-	[25]162	[68]0.09 $N^{1.0}$
102	Tevian	tevian	0	2018-02-16	666	[79]2048	1	[53]394	-	[60]405	[94]0.18 $N^{1.0}$
103	Tevian	tevian	1	2018-02-16	666	[87]2048	1	[56]398	-	[58]403	[86]0.20 $N^{1.0}$
104	Tevian	tevian	2	2018-02-16	666	[85]2048	1	[54]397	-	[57]402	[89]0.19 $N^{1.0}$
105	Tevian	tevian	3	2018-06-20	707	[77]2048	1	[40]300	[30]473	[70]539	[90]0.25 $N^{1.0}$
106	Tevian	tevian	4	2018-06-20	707	[90]2048	1	[39]299	[28]434	[69]537	[77]0.29 $N^{1.0}$
107	TigerIT Americas LLC	tiger	0	2018-06-29	333	[98]2052	k	[67]428	[45]1822	[115]2942	[72]1.63 $N^{1.0}$
108	TigerIT Americas LLC	tiger	1	2018-06-27	333	[96]2052	-	[55]398	[1]0	[1]1	[7]28.15 $N^{0.3}$
109	TongYi Transportation Technology	tongyi	0	2018-06-29	1701	[103]2070	k	[21]190	[53]2256	[106]2272	[105]0.85 $N^{1.0}$
110	TongYi Transportation Technology	tongyi	1	2018-06-29	1701	[102]2070	1	[20]189	[52]2238	[105]2257	[92]1.02 $N^{1.0}$
111	Visidon	visidon	0	2018-06-20	208	[56]1028	k	[46]337	[48]2006	[112]2566	[104]0.97 $N^{1.0}$
112	Vigilant Solutions	vigilant	0	2018-02-08	335	[66]1544	k	[119]823	-	[100]2058	[112]0.60 $N^{1.1}$
113	Vigilant Solutions	vigilant	1	2018-02-14	249	[101]2056	k	[112]739	-	[101]2075	[119]0.26 $N^{1.1}$
114	Vigilant Solutions	vigilant	2	2018-02-14	335	[67]1544	k	[117]820	-	[103]2121	[118]0.41 $N^{1.1}$
115	Vigilant Solutions	vigilant	3	2018-06-21	335	[65]1544	k	[122]832	[55]2453	[107]2307	[101]0.93 $N^{1.0}$
116	Vigilant Solutions	vigilant	4	2018-06-21	337	[64]1544	k	[120]830	[50]2050	[104]2251	[103]0.90 $N^{1.0}$
117	VisionLabs	visionlabs	3	2018-02-16	624	[12]256	1	[28]228	-	[5]5	[6]417.37 $N^{0.2}$
118	VisionLabs	visionlabs	4	2018-06-22	299	[17]256	1	[43]315	[5]19	[7]17	[4]2663.29 $N^{0.1}$
119	VisionLabs	visionlabs	5	2018-06-22	305	[24]512	1	[41]300	[6]54	[10]33	[9]166.84 $N^{0.4}$
120	Vocord	vocord	0	2018-02-16	872	[41]608	k	[77]536	-	[43]268	[58]0.17 $N^{1.0}$
121	Vocord	vocord	1	2018-02-16	872	[42]608	k	[78]536	-	[44]268	[52]0.18 $N^{1.0}$
122	Vocord	vocord	2	2018-02-16	924	[78]2048	k	[95]635	-	[37]248	[82]0.13 $N^{1.0}$
123	Vocord	vocord	3	2018-06-30	627	[43]896	k	[109]714	[18]215	[36]247	[13]0.81 $N^{0.9}$
124	Vocord	vocord	4	2018-06-30	627	[44]896	k	[80]538	[19]216	[41]253	[16]0.60 $N^{0.9}$
125	Zhuhai Yisheng Electronics Tech.	yisheng	0	2018-02-14	473	[106]2108	k	[90]615	-	[75]587	[100]0.24 $N^{1.0}$
126	Zhuhai Yisheng Electronics Tech.	yisheng	1	2018-06-19	474	[112]3704	k	[52]387	[51]2228	[88]1108	[32]0.99 $N^{1.0}$
127	Shanghai Yitu Technology	yitu	0	2018-02-12	1774	[120]4136	1	[93]633	-	[65]464	[114]0.12 $N^{1.1}$
128	Shanghai Yitu Technology	yitu	1	2018-02-12	1944	[119]4136	1	[130]930	-	[64]463	[122]0.04 $N^{1.1}$
129	Shanghai Yitu Technology	yitu	2	2018-06-21	2077	[122]4138	1	[124]870	[65]5516	[125]5417	[17]9.25 $N^{0.9}$
130	Shanghai Yitu Technology	yitu	3	2018-06-21	2077	[121]4138	1	[125]871	[63]5248	[124]5242	[116]1.08 $N^{1.1}$

Notes	
1	Configuration size does not capture static data present in libraries. Libraries are not counted because most implementations include common ancilliary libraries for image processing (e.g. openCV) or numerical computation (e.g. blas).
2	This multiplier expresses the increase in template size when k images are passed to the template generation function.
3	All durations are measured on Intel®Xeon®CPU E5-2630 v4 @ 2.20GHz processors. Estimates are made by wrapping the API function call in calls to std::chrono::high_resolution_clock which on the machine in (3) counts 1ns clock ticks. Precision is somewhat worse than that however.
4	Search durations are measured as in the prior note. The power-law model in the final column mostly fits the empirical results in Figure 103. However in certain cases the model is not correct and should not be used numerically.

Table 9: Summary of algorithms and properties included in this report. The blue superscripts give ranking for the quantity in that column. Missing search durations, denoted by "-", are absent because those runs were not executed.

MISSES OUTSIDE RANK R		RESOURCE USAGE		ENROL MOST RECENT, N = 1.6M							N = 1.6M, FRVT2018		
FNIR(N, T=0, R)		TEMPLATE		FRVT 2014			FRVT 2018				RECENT	LIFETIME	UNCONSOL
#	ALGORITHM	BYTES	MSEC	R=1	R=10	R=20	R=1	R=10	R=20	WORK-10	R =1		
1	3DIVI-0	^{113}4096	^{62}426	^{56}0.026	^{47}0.014	^{50}0.013	^{63}0.034	^{61}0.016	^{59}0.013	^{63}1.190	^{63}0.034		
2	3DIVI-1	^{121}4224	^{66}428	^{60}0.028	^{55}0.018	^{57}0.017	^{64}0.038	^{70}0.021	^{72}0.020	^{66}1.233	^{64}0.038		
3	3DIVI-2	^{32}528	^{64}428	^{63}0.030	^{62}0.020	^{63}0.019	^{68}0.040	^{74}0.024	^{76}0.023	^{71}1.259	^{68}0.040		
4	3DIVI-3	^{25}512	^{88}625	^{73}0.053	^{69}0.024	^{67}0.020	^{88}0.086	^{84}0.037	^{82}0.030	^{86}1.469	^{88}0.086	^{65}0.064	
5	3DIVI-4	^{110}4096	^{89}628				^{47}0.020	^{45}0.010	^{45}0.009	^{45}1.115	^{47}0.020	^{42}0.013	
6	ALCHERA-0	^{74}2048	^{32}263	^{45}0.021	^{58}0.018	^{59}0.018	^{44}0.019	^{57}0.014	^{62}0.013	^{52}1.138	^{44}0.019	^{39}0.012	
7	ALCHERA-1	^{83}2048	^{5}66				^{126}0.987	^{126}0.974	^{126}0.968	^{126}9.812	^{126}0.987	^{81}0.982	
8	AWARE-0	^{67}1564	^{95}653	^{72}0.053	^{77}0.040	^{78}0.038	^{84}0.064	^{87}0.042	^{87}0.039	^{85}1.439	^{84}0.064		
9	AWARE-1	^{66}1564	^{94}651	^{69}0.043	^{70}0.029	^{71}0.027	^{80}0.059	^{83}0.035	^{84}0.032	^{83}1.382	^{80}0.059		
10	AWARE-2	^{101}2076	^{125}912	^{74}0.056	^{79}0.044	^{80}0.043	^{81}0.060	^{86}0.040	^{86}0.038	^{84}1.416	^{81}0.060		
11	AWARE-3	^{102}2076	^{107}716	^{53}0.025	^{48}0.014	^{49}0.012	^{62}0.033	^{59}0.015	^{58}0.013	^{62}1.186	^{62}0.033	^{52}0.021	
12	AWARE-4	^{1}92	^{105}712				^{85}0.070	^{79}0.030	^{78}0.023	^{81}1.378	^{85}0.070	^{62}0.053	
13	AYONIX-0	^{58}1036	^{1}10	^{101}0.346	^{102}0.236	^{102}0.210	^{119}0.452	^{120}0.319	^{119}0.285	^{120}4.304	^{119}0.452	^{78}0.465	
14	CAMVI-1	^{45}1024	^{17}177	^{93}0.143	^{89}0.075	^{87}0.064	^{111}0.227	^{107}0.124	^{107}0.105	^{109}2.419	^{111}0.227		
15	CAMVI-2	^{53}1024	^{111}774	^{79}0.076	^{78}0.040	^{76}0.035	^{95}0.129	^{95}0.068	^{95}0.059	^{94}1.781	^{95}0.129		
16	CAMVI-3	^{49}1024	^{104}707	^{67}0.035	^{74}0.035	^{77}0.035	^{79}0.054	^{88}0.054	^{93}0.054	^{88}1.488	^{79}0.054	^{56}0.037	
17	COGENT-0	^{31}525	^{80}551	^{29}0.011	^{39}0.010	^{31}0.008	^{33}0.013	^{52}0.012	^{29}0.006	^{43}1.111	^{33}0.013	^{36}0.011	^{11}0.007
18	COGENT-1	^{30}525	^{81}552	^{28}0.011	^{38}0.010	^{30}0.008	^{32}0.013	^{51}0.012	^{28}0.006	^{42}1.111	^{32}0.013	^{35}0.011	^{10}0.007
19	COGNITEC-0	^{92}2052	^{16}176	^{44}0.020	^{45}0.013	^{46}0.012	^{59}0.029	^{58}0.014	^{57}0.012	^{59}1.167	^{59}0.029	^{50}0.021	
20	COGNITEC-1	^{95}2052	^{21}202	^{34}0.013	^{41}0.010	^{41}0.010	^{40}0.014	^{36}0.008	^{37}0.007	^{36}1.086	^{40}0.014	^{29}0.009	^{13}0.009
21	DERMALOG-0	^{4}128	^{46}344	^{78}0.075	^{75}0.037	^{74}0.030	^{96}0.131	^{93}0.065	^{92}0.053	^{93}1.778	^{96}0.131		
22	DERMALOG-1	^{3}128	^{15}171	^{83}0.096	^{80}0.051	^{79}0.042	^{98}0.156	^{97}0.080	^{97}0.066	^{98}1.945	^{98}0.156		
23	DERMALOG-2	^{9}256	^{45}344	^{80}0.079	^{76}0.039	^{75}0.031	^{97}0.138	^{94}0.068	^{94}0.055	^{96}1.817	^{97}0.138		
24	DERMALOG-3	^{5}128	^{23}211				^{93}0.128	^{92}0.063	^{91}0.050	^{92}1.752	^{93}0.128	^{69}0.097	
25	DERMALOG-4	^{2}128	^{22}208	^{75}0.071	^{73}0.034	^{73}0.028	^{92}0.127	^{91}0.062	^{90}0.050	^{91}1.748	^{92}0.127	^{67}0.096	
26	EVERAI-0	^{88}2048	^{68}438				^{48}0.021	^{65}0.019	^{69}0.018	^{60}1.174	^{48}0.021	^{46}0.017	^{16}0.025
27	EVERAI-1	^{76}2048	^{86}590	^{11}0.004	^{12}0.003	^{12}0.003	^{9}0.006	^{11}0.004	^{11}0.004	^{9}1.038	^{9}0.006	^{9}0.003	
28	EYEDEA-0	^{120}4152	^{61}424	^{99}0.201	^{97}0.100	^{96}0.081	^{115}0.300	^{115}0.160	^{113}0.130	^{115}2.864	^{115}0.300		
29	EYEDEA-1	^{57}1036	^{40}311	^{86}0.109	^{82}0.054	^{82}0.044	^{105}0.198	^{104}0.105	^{103}0.086	^{104}2.226	^{105}0.198		
30	EYEDEA-2	^{56}1036	^{67}429	^{87}0.110	^{81}0.054	^{83}0.044	^{106}0.200	^{105}0.107	^{105}0.089	^{105}2.246	^{106}0.200		
31	EYEDEA-3	^{55}1036	^{49}385	^{71}0.044	^{66}0.021	^{58}0.017	^{87}0.082	^{85}0.039	^{83}0.031	^{87}1.470	^{87}0.082	^{64}0.061	
32	GLORY-0	^{21}418	^{12}160				^{102}0.180	^{109}0.129	^{110}0.118	^{106}2.318	^{102}0.180	^{71}0.133	
33	GLORY-1	^{68}1726	^{57}405	^{85}0.109	^{93}0.083	^{94}0.078	^{94}0.129	^{98}0.089	^{98}0.080	^{97}1.925	^{94}0.129	^{66}0.093	
34	GORILLA-0	^{127}8300	^{63}427							^{127}10.000			
35	GORILLA-1	^{104}2156	^{14}169				^{82}0.063	^{75}0.025	^{75}0.020	^{78}1.331	^{82}0.063	^{58}0.041	
36	HBINNO-0	^{29}520	^{33}265	^{98}0.191	^{98}0.102	^{97}0.086	^{114}0.275	^{113}0.152	^{112}0.126	^{114}2.743	^{114}0.275		
37	HIK-0	^{70}1808	^{123}875	^{57}0.026	^{68}0.023	^{69}0.023	^{55}0.024	^{64}0.018	^{68}0.017	^{61}1.176	^{55}0.024		
38	HIK-1	^{71}1808	^{114}820	^{76}0.073	^{87}0.071	^{90}0.070	^{43}0.017	^{47}0.011	^{51}0.010	^{46}1.116	^{43}0.017		
39	HIK-2	^{72}1808	^{115}820	^{33}0.013	^{42}0.010	^{42}0.010	^{42}0.017	^{46}0.011	^{49}0.010	^{44}1.115	^{42}0.017	^{47}0.019	
40	HIK-3	^{61}1408	^{90}633				^{39}0.014	^{32}0.007	^{35}0.006	^{35}1.082	^{39}0.014	^{37}0.011	
41	HIK-4	^{60}1152	^{70}510	^{22}0.008	^{18}0.005	^{18}0.004	^{37}0.014	^{33}0.007	^{33}0.006	^{33}1.081	^{37}0.014	^{34}0.010	^{12}0.009
42	IDEMIA-0	^{19}364	^{58}416	^{24}0.008	^{20}0.005	^{21}0.005	^{28}0.011	^{27}0.006	^{27}0.006	^{28}1.070	^{28}0.011	^{21}0.006	
43	IDEMIA-1	^{20}364	^{60}417	^{25}0.008	^{21}0.005	^{22}0.005	^{30}0.012	^{31}0.007	^{34}0.006	^{30}1.072	^{30}0.012	^{22}0.006	
44	IDEMIA-2	^{18}364	^{59}417	^{32}0.013	^{37}0.010	^{40}0.010	^{31}0.013	^{34}0.008	^{38}0.007	^{34}1.081	^{31}0.013	^{31}0.010	
45	IDEMIA-3	^{34}528	^{100}689	^{30}0.011	^{29}0.008	^{29}0.007	^{24}0.010	^{29}0.006	^{32}0.006	^{26}1.066	^{24}0.010	^{18}0.005	^{9}0.005
46	IDEMIA-4	^{33}528	^{99}669	^{23}0.008	^{19}0.005	^{17}0.004	^{21}0.009	^{21}0.006	^{22}0.005	^{21}1.061	^{21}0.009	^{17}0.005	^{8}0.005
47	IMAGUS-0	^{28}512	^{2}43	^{100}0.216	^{100}0.124	^{100}0.105	^{116}0.305	^{116}0.175	^{115}0.146	^{116}2.977	^{116}0.305		
48	IMAGUS-2	^{27}512	^{6}76	^{95}0.145	^{86}0.069	^{84}0.056	^{109}0.222	^{106}0.111	^{106}0.090	^{107}2.329	^{109}0.222	^{72}0.183	
49	IMAGUS-3	^{24}512	^{4}57				^{118}0.358	^{117}0.215	^{117}0.181	^{117}3.380	^{118}0.358	^{76}0.301	
50	INCODE-0	^{50}1024	^{19}190				^{78}0.051	^{71}0.023	^{70}0.019	^{74}1.285	^{78}0.051	^{57}0.038	
51	INCODE-1	^{73}2048	^{101}690	^{31}0.012	^{28}0.008	^{28}0.007	^{45}0.019	^{38}0.009	^{40}0.008	^{40}1.106	^{45}0.019	^{41}0.013	
52	INNOVATRICS-0	^{38}530	^{69}455	^{62}0.029	^{53}0.017	^{54}0.016	^{70}0.042	^{67}0.019	^{66}0.016	^{68}1.234	^{70}0.042		
53	INNOVATRICS-1	^{35}530	^{42}316	^{61}0.029	^{52}0.017	^{53}0.016	^{69}0.042	^{66}0.019	^{65}0.016	^{67}1.234	^{69}0.042		
54	INNOVATRICS-2	^{36}530	^{30}255				^{76}0.048	^{82}0.035	^{85}0.033	^{80}1.343	^{76}0.048	^{61}0.050	
55	INNOVATRICS-3	^{37}530	^{31}255	^{37}0.015	^{27}0.006	^{26}0.005	^{60}0.029	^{48}0.012	^{48}0.010	^{56}1.151	^{60}0.029	^{53}0.030	
56	ISYSTEMS-0	^{89}2048	^{25}222	^{48}0.023	^{64}0.020	^{66}0.020	^{36}0.014	^{41}0.010	^{46}0.009	^{38}1.098	^{36}0.014	^{28}0.009	
57	ISYSTEMS-1	^{47}1024	^{24}222	^{49}0.023	^{63}0.020	^{65}0.020	^{35}0.014	^{40}0.010	^{47}0.009	^{39}1.098	^{35}0.014	^{27}0.009	
58	ISYSTEMS-2	^{84}2048	^{43}316	^{20}0.008	^{26}0.006	^{27}0.006	^{19}0.009	^{26}0.006	^{30}0.006	^{24}1.062	^{19}0.009	^{13}0.005	
59	MEGVII-0	^{79}2048	^{112}794	^{12}0.004	^{8}0.002	^{6}0.002	^{22}0.009	^{14}0.004	^{14}0.004	^{16}1.052	^{22}0.009	^{32}0.010	
60	MICROFOCUS-0	^{13}256	^{71}525	^{105}0.472	^{104}0.309	^{104}0.269	^{123}0.597	^{123}0.425	^{123}0.378	^{123}5.397	^{123}0.597		
61	MICROFOCUS-1	^{10}256	^{72}527	^{104}0.472	^{105}0.309	^{105}0.270	^{124}0.597	^{124}0.425	^{124}0.378	^{124}5.398	^{124}0.597		
62	MICROFOCUS-2	^{16}256	^{73}529	^{106}0.508	^{106}0.377	^{106}0.348	^{125}0.627	^{125}0.488	^{125}0.453	^{125}5.839	^{125}0.627		
63	MICROFOCUS-3	^{15}256	^{34}269	^{103}0.469	^{103}0.305	^{103}0.265	^{122}0.595	^{122}0.422	^{122}0.374	^{122}5.373	^{122}0.595	^{80}0.539	
64	MICROFOCUS-4	^{14}256	^{35}270				^{121}0.577	^{121}0.404	^{121}0.358	^{121}5.212	^{121}0.577	^{79}0.519	

Table 10: **Relative difficulty of FRVT 2014 and 2018 image sets**. *In columns 3 and 4 are template size and template generation duration. Thereafter values are rank-based FNIR, with T = 0. In columns 5, 6 and 7, green indicates FNIR below the best reported in NISTIR 8009 in 2014-04, for NEC CORP E30C, on identical images. These values are FNIR(N, 1) = 0.041, FNIR(N, 10) = 0.030 and FNIR(N, 20) = 0.029 Columns 8 and 9 show FRVT 2018 is slightly more difficult than FRVT 2014 (columns 5, 6). Column 10 is a workload statistic, a small value shows an algorithm front-loads mates into the first 10 candidates. The last three columns compare the enrollment styles in Figure 10. Throughout, blue superscripts indicate the rank of the algorithm for that column, and the best value is hightlighted in yellow.*

MISSES OUTSIDE RANK R		RESOURCE USAGE		ENROL MOST RECENT, N = 1.6M							N = 1.6M, FRVT2018		
FNIR(N, T=0, R)		TEMPLATE		FRVT 2014			FRVT 2018				RECENT	LIFETIME	UNCONSOL
#	ALGORITHM	BYTES	MSEC	R=1	R=10	R=20	R=1	R=10	R=20	WORK-10	R =1		
65	MICROSOFT-0	^{26}512	^{36}283	^{7}0.003	^{5}0.002	^{4}0.002	^{11}0.006	^{7}0.004	^{9}0.003	^{10}1.038	^{11}0.006	^{8}0.003	
66	MICROSOFT-1	^{51}1024	^{47}349	^{6}0.003	^{3}0.002	^{3}0.002	^{10}0.006	^{8}0.004	^{6}0.003	^{8}1.038	^{10}0.006	^{7}0.003	
67	MICROSOFT-2	^{46}1024	^{82}555	^{8}0.004	^{4}0.002	^{5}0.002	^{12}0.006	^{12}0.004	^{10}0.003	^{12}1.041	^{12}0.006	^{10}0.003	
68	MICROSOFT-3	^{43}1024	^{55}404	^{2}0.002	^{1}0.002	^{1}0.001	^{2}0.003	^{2}0.002	^{2}0.002	^{2}1.022	^{2}0.003	^{2}0.001	
69	MICROSOFT-4	^{87}2048	^{110}773	^{1}0.002	^{2}0.002	^{2}0.001	^{1}0.003	^{1}0.002	^{1}0.002	^{1}1.022	^{1}0.003	^{1}0.001	^{1}0.001
70	NEC-0	^{105}2592	^{7}82	^{36}0.014	^{33}0.009	^{35}0.008	^{46}0.020	^{39}0.009	^{41}0.008	^{41}1.110	^{46}0.020	^{40}0.013	^{14}0.013
71	NEC-1	^{106}2592	^{8}88	^{52}0.025	^{67}0.021	^{68}0.021	^{54}0.024	^{60}0.015	^{63}0.014	^{58}1.158	^{54}0.024	^{45}0.016	
72	NEUROTECHNOLOGY-0	^{125}5214	^{102}702	^{64}0.031	^{54}0.018	^{52}0.016	^{77}0.050	^{72}0.023	^{71}0.019	^{73}1.278	^{77}0.050		
73	NEUROTECHNOLOGY-1	^{124}5214	^{97}661	^{59}0.028	^{49}0.014	^{47}0.012	^{75}0.047	^{69}0.020	^{67}0.016	^{70}1.250	^{75}0.047		
74	NEUROTECHNOLOGY-2	^{123}5214	^{96}658	^{58}0.028	^{50}0.014	^{48}0.012	^{74}0.047	^{68}0.020	^{64}0.016	^{69}1.249	^{74}0.047		
75	NEUROTECHNOLOGY-3	^{82}2048	^{79}547	^{43}0.019	^{44}0.012	^{44}0.011	^{57}0.025	^{55}0.013	^{54}0.010	^{54}1.148	^{57}0.025	^{49}0.020	
76	NEUROTECHNOLOGY-4	^{90}2048	^{78}543	^{35}0.014	^{43}0.012	^{45}0.011	^{16}0.008	^{22}0.006	^{26}0.006	^{19}1.058	^{16}0.008	^{19}0.006	^{6}0.004
77	NTECHLAB-0	^{122}4442	^{108}730	^{15}0.006	^{13}0.003	^{13}0.003	^{29}0.012	^{20}0.005	^{15}0.005	^{25}1.064	^{29}0.012	^{24}0.008	
78	NTECHLAB-1	^{69}1736	^{56}405	^{19}0.008	^{15}0.004	^{14}0.003	^{38}0.014	^{25}0.006	^{19}0.005	^{31}1.074	^{38}0.014	^{30}0.010	
79	NTECHLAB-3	^{107}3484	^{118}831				^{17}0.008	^{13}0.004	^{13}0.004	^{14}1.047	^{17}0.008	^{16}0.005	^{7}0.005
80	NTECHLAB-4	^{108}3484	^{126}929	^{10}0.004	^{7}0.002	^{8}0.002	^{13}0.007	^{9}0.004	^{8}0.003	^{11}1.041	^{13}0.007	^{11}0.004	^{5}0.004
81	RANKONE-0	^{8}228	^{3}50	^{70}0.043	^{71}0.030	^{72}0.027	^{73}0.045	^{73}0.024	^{74}0.020	^{72}1.275	^{73}0.045	^{54}0.032	
82	RANKONE-1	^{17}324	^{11}136	^{65}0.032	^{65}0.021	^{64}0.019	^{56}0.025	^{54}0.012	^{52}0.010	^{53}1.145	^{56}0.025	^{48}0.019	
83	RANKONE-2	^{7}133	^{9}113	^{55}0.025	^{57}0.018	^{56}0.016	^{50}0.022	^{50}0.012	^{56}0.010	^{51}1.135	^{50}0.022	^{44}0.015	
84	RANKONE-3	^{6}133	^{10}114	^{54}0.025	^{56}0.018	^{55}0.016	^{49}0.022	^{49}0.012	^{55}0.010	^{50}1.135	^{49}0.022	^{43}0.015	^{15}0.015
85	REALNETWORKS-0	^{114}4100	^{29}244	^{46}0.023	^{34}0.010	^{36}0.008	^{72}0.043	^{63}0.017	^{61}0.013	^{65}1.222	^{72}0.043	^{59}0.044	
86	REALNETWORKS-1	^{115}4104	^{28}243				^{71}0.043	^{62}0.017	^{60}0.013	^{64}1.222	^{71}0.043	^{55}0.033	
87	SHAMAN-0	^{111}4096	^{76}538	^{89}0.119	^{90}0.076	^{89}0.069	^{100}0.171	^{100}0.098	^{102}0.085	^{100}2.092	^{100}0.171		
88	SHAMAN-1	^{112}4096	^{83}557	^{88}0.118	^{88}0.072	^{88}0.064	^{101}0.172	^{99}0.095	^{100}0.081	^{99}2.078	^{101}0.172		
89	SHAMAN-2	^{126}8192	^{84}557	^{97}0.180	^{99}0.105	^{98}0.092	^{113}0.262	^{114}0.154	^{114}0.131	^{113}2.710	^{113}0.262		
90	SHAMAN-3	^{85}2048	^{103}704	^{82}0.094	^{84}0.063	^{85}0.058	^{90}0.127	^{96}0.073	^{96}0.064	^{95}1.811	^{90}0.127	^{68}0.097	
91	SHAMAN-4	^{78}2048	^{93}642				^{110}0.224	^{108}0.126	^{108}0.107	^{110}2.431	^{110}0.224	^{73}0.187	
92	SIAT-0	^{59}1096	^{48}358	^{18}0.007	^{16}0.005	^{16}0.004	^{26}0.010	^{17}0.005	^{16}0.005	^{20}1.059	^{26}0.010		
93	SIAT-1	^{94}2052	^{120}842	^{9}0.004	^{11}0.003	^{11}0.003	^{3}0.004	^{5}0.003	^{5}0.003	^{5}1.031	^{3}0.004	^{75}0.264	^{2}0.001
94	SIAT-2	^{97}2052	^{124}906	^{81}0.081	^{91}0.080	^{95}0.080	^{4}0.004	^{6}0.003	^{7}0.003	^{6}1.032	^{4}0.004	^{74}0.213	
95	SMILART-0	^{52}1024	^{13}168	^{92}0.142	^{94}0.085	^{92}0.075	^{103}0.193	^{103}0.105	^{104}0.087	^{102}2.204	^{103}0.193		
96	SMILART-1	^{48}1024	^{98}662	^{94}0.144	^{92}0.082	^{91}0.071	^{108}0.219	^{110}0.130	^{109}0.113	^{111}2.435	^{108}0.219		
97	SMILART-2	^{44}1024	^{85}560	^{91}0.132	^{85}0.069	^{86}0.058	^{104}0.195	^{102}0.102	^{101}0.084	^{101}2.196	^{104}0.195		
98	SYNESIS-0	^{23}512	^{27}237	^{84}0.108	^{96}0.100	^{99}0.100	^{99}0.162	^{112}0.151	^{116}0.151	^{108}2.380	^{99}0.162		
99	TEVIAN-0	^{75}2048	^{51}394	^{39}0.017	^{35}0.010	^{37}0.009	^{52}0.022	^{43}0.010	^{43}0.008	^{48}1.122	^{52}0.022		
100	TEVIAN-1	^{91}2048	^{54}398	^{40}0.017	^{36}0.010	^{38}0.009	^{53}0.022	^{44}0.010	^{44}0.008	^{49}1.122	^{53}0.022		
101	TEVIAN-2	^{81}2048	^{52}397	^{42}0.017	^{40}0.010	^{39}0.009	^{51}0.022	^{42}0.010	^{42}0.008	^{47}1.121	^{51}0.022		
102	TEVIAN-3	^{77}2048	^{39}300				^{41}0.017	^{37}0.008	^{36}0.006	^{37}1.093	^{41}0.017	^{33}0.010	
103	TEVIAN-4	^{86}2048	^{37}299	^{26}0.009	^{22}0.005	^{20}0.005	^{34}0.013	^{30}0.006	^{25}0.005	^{32}1.076	^{34}0.013	^{25}0.008	
104	TIGER-0	^{93}2052	^{65}428	^{66}0.033	^{46}0.014	^{43}0.011	^{83}0.064	^{76}0.026	^{73}0.020	^{79}1.334	^{83}0.064	^{60}0.048	
105	TIGER-1	^{96}2052	^{53}398				^{117}0.308	^{119}0.296	^{120}0.295	^{118}3.691	^{117}0.308		
106	TONGYITRANS-0	^{100}2070	^{20}190				^{25}0.010	^{24}0.006	^{24}0.005	^{22}1.062	^{25}0.010	^{20}0.006	
107	TONGYITRANS-1	^{99}2070	^{18}189	^{21}0.008	^{25}0.006	^{24}0.005	^{23}0.010	^{23}0.006	^{23}0.005	^{23}1.062	^{23}0.010	^{38}0.011	
108	VD-0	^{54}1028	^{44}337	^{102}0.363	^{101}0.187	^{101}0.152	^{120}0.475	^{118}0.271	^{118}0.224	^{119}4.074	^{120}0.475	^{77}0.430	
109	VIGILANTSOLUTIONS-0	^{65}1544	^{116}823	^{77}0.073	^{72}0.033	^{70}0.027	^{89}0.125	^{89}0.058	^{88}0.046	^{89}1.712	^{89}0.125		
110	VIGILANTSOLUTIONS-1	^{98}2056	^{109}739	^{90}0.120	^{83}0.054	^{81}0.043	^{107}0.204	^{101}0.100	^{99}0.080	^{103}2.210	^{107}0.204		
111	VIGILANTSOLUTIONS-2	^{62}1544	^{113}820	^{96}0.159	^{95}0.090	^{93}0.077	^{112}0.239	^{111}0.139	^{111}0.118	^{112}2.555	^{112}0.239		
112	VIGILANTSOLUTIONS-3	^{64}1544	^{119}832	^{68}0.038	^{51}0.017	^{51}0.013	^{86}0.072	^{77}0.029	^{77}0.023	^{82}1.378	^{86}0.072	^{63}0.055	
113	VIGILANTSOLUTIONS-4	^{63}1544	^{117}830				^{91}0.127	^{90}0.058	^{89}0.046	^{90}1.721	^{91}0.127	^{70}0.099	
114	VISIONLABS-3	^{11}256	^{26}228	^{27}0.009	^{30}0.008	^{34}0.008	^{20}0.009	^{35}0.008	^{39}0.007	^{29}1.072	^{20}0.009	^{15}0.005	
115	VISIONLABS-4	^{12}256	^{41}315	^{4}0.003	^{9}0.002	^{9}0.002	^{6}0.004	^{4}0.003	^{4}0.003	^{4}1.031	^{6}0.004	^{5}0.002	
116	VISIONLABS-5	^{22}512	^{38}300	^{3}0.003	^{6}0.002	^{7}0.002	^{5}0.004	^{3}0.003	^{3}0.003	^{3}1.029	^{5}0.004	^{4}0.002	^{4}0.002
117	VOCORD-0	^{40}608	^{75}536	^{51}0.025	^{60}0.019	^{62}0.018	^{67}0.040	^{81}0.031	^{81}0.029	^{77}1.301	^{67}0.040		
118	VOCORD-1	^{39}608	^{74}536	^{50}0.025	^{59}0.019	^{61}0.018	^{66}0.040	^{80}0.031	^{80}0.029	^{76}1.299	^{66}0.040		
119	VOCORD-2	^{80}2048	^{92}635	^{47}0.023	^{61}0.019	^{60}0.018	^{65}0.038	^{78}0.030	^{79}0.029	^{75}1.290	^{65}0.038		
120	VOCORD-3	^{41}896	^{106}714	^{14}0.006	^{14}0.004	^{15}0.004	^{18}0.008	^{16}0.005	^{17}0.005	^{18}1.054	^{18}0.008	^{23}0.007	
121	VOCORD-4	^{42}896	^{77}538				^{27}0.010	^{28}0.006	^{31}0.006	^{27}1.068	^{27}0.010	^{26}0.008	
122	YISHENG-0	^{103}2108	^{87}615	^{38}0.016	^{32}0.009	^{32}0.008	^{58}0.027	^{53}0.012	^{50}0.010	^{55}1.149	^{58}0.027		
123	YISHENG-1	^{109}3704	^{50}387	^{41}0.017	^{31}0.009	^{33}0.008	^{61}0.029	^{56}0.013	^{53}0.010	^{57}1.156	^{61}0.029	^{51}0.021	
124	YITU-0	^{116}4136	^{91}633	^{17}0.007	^{24}0.006	^{25}0.005	^{15}0.007	^{19}0.005	^{21}0.005	^{17}1.053	^{15}0.007	^{14}0.005	
125	YITU-1	^{117}4136	^{127}930	^{16}0.007	^{23}0.005	^{23}0.005	^{14}0.007	^{18}0.005	^{20}0.005	^{15}1.052	^{14}0.007	^{12}0.005	
126	YITU-2	^{118}4138	^{121}870	^{5}0.003	^{10}0.003	^{10}0.003	^{7}0.004	^{10}0.004	^{12}0.004	^{7}1.035	^{7}0.004	^{3}0.001	^{3}0.002
127	YITU-3	^{119}4138	^{122}871	^{13}0.005	^{17}0.005	^{19}0.005	^{8}0.005	^{15}0.005	^{18}0.005	^{13}1.044	^{8}0.005	^{6}0.002	

Table 11: **Relative difficulty of FRVT 2014 and 2018 image sets**. *In columns 3 and 4 are template size and template generation duration. Thereafter values are rank-based FNIR, with* T *= 0. In columns 5, 6 and 7, green indicates FNIR below the best reported in NISTIR 8009 in 2014-04, for NEC CORP E30C, on identical images. These values are FNIR(N, 1) = 0.041, FNIR(N, 10) = 0.030 and FNIR(N, 20) = 0.029 Columns 8 and 9 show FRVT 2018 is slightly more difficult than FRVT 2014 (columns 5, 6). Column 10 is a workload statistic, a small value shows an algorithm front-loads mates into the first 10 candidates. The last three columns compare the enrollment styles in Figure 10. Throughout, blue superscripts indicate the rank of the algorithm for that column, and the best value is hightlighted in yellow.*

MISSES BELOW THRESHOLD, T		ENROL MOST RECENT MUGSHOT, N = 1.6M								
FNIR(N, T > 0, R > L)		DATASET: FRVT 2014 MUGSHOTS			DATASET: FRVT 2018 MUGSHOTS			DATASET: WEBCAM PROBES		
#	ALGORITHM	FPIR=0.001	FPIR=0.01	FPIR=0.1	FPIR=0.001	FPIR=0.01	FPIR=0.1	FPIR=0.001	FPIR=0.01	FPIR=0.1
1	3DIVI-0	^{61}0.175	^{66}0.103	^{60}0.055	^{68}0.256	^{72}0.160	^{73}0.086	^{56}0.425	^{57}0.302	^{56}0.180
2	3DIVI-1	^{60}0.175	^{65}0.103	^{63}0.056	^{67}0.256	^{73}0.160	^{74}0.087			
3	3DIVI-2	^{62}0.176	^{67}0.105	^{66}0.058	^{64}0.255	^{74}0.164	^{75}0.089			
4	3DIVI-3	^{74}0.287	^{77}0.183	^{76}0.105	^{84}0.402	^{89}0.284	^{88}0.168	^{68}0.626	^{70}0.497	^{66}0.343
5	3DIVI-4				^{53}0.171	^{53}0.096	^{52}0.047	^{51}0.343	^{51}0.237	^{51}0.138
6	ALCHERA-0	^{42}0.095	^{40}0.047	^{39}0.029	^{50}0.140	^{48}0.073	^{45}0.035	^{35}0.216	^{36}0.146	^{36}0.087
7	ALCHERA-1				^{126}0.999	^{125}0.999	^{126}0.995	^{109}1.000	^{109}1.000	^{90}1.000
8	AWARE-0	^{99}0.775	^{61}0.092	^{69}0.065	^{123}0.983	^{66}0.128	^{71}0.085	^{79}0.817	^{52}0.253	^{55}0.178
9	AWARE-1	^{102}0.863	^{57}0.084	^{59}0.055	^{124}0.996	^{65}0.127	^{70}0.081			
10	AWARE-2	^{98}0.757	^{60}0.090	^{70}0.067	^{122}0.977	^{64}0.120	^{68}0.078			
11	AWARE-3	^{43}0.096	^{43}0.056	^{50}0.035	^{49}0.131	^{51}0.085	^{54}0.051	^{45}0.298	^{46}0.204	^{50}0.132
12	AWARE-4				^{69}0.271	^{77}0.177	^{82}0.107	^{61}0.509	^{63}0.375	^{62}0.253
13	AYONIX-0	^{96}0.723	^{101}0.624	^{101}0.488	^{114}0.811	^{118}0.725	^{119}0.598	^{84}0.939	^{86}0.892	^{86}0.802
14	CAMVI-1	^{90}0.557	^{94}0.409	^{95}0.255	^{107}0.684	^{111}0.549	^{111}0.375	^{76}0.770	^{80}0.648	^{80}0.488
15	CAMVI-2	^{84}0.408	^{85}0.265	^{82}0.147	^{96}0.537	^{99}0.402	^{96}0.242			
16	CAMVI-3	^{24}0.046	^{36}0.038	^{52}0.036	^{26}0.074	^{41}0.060	^{60}0.055	^{17}0.132	^{31}0.108	^{38}0.094
17	COGENT-0	^{15}0.033	^{21}0.021	^{27}0.015	^{21}0.056	^{23}0.032	^{28}0.020	^{20}0.140	^{26}0.100	^{32}0.069
18	COGENT-1	^{14}0.033	^{20}0.021	^{26}0.015	^{20}0.056	^{22}0.032	^{27}0.020	^{19}0.140	^{24}0.100	^{31}0.069
19	COGNITEC-0	^{44}0.108	^{42}0.054	^{41}0.031	^{51}0.163	^{55}0.098	^{56}0.053	^{46}0.303	^{44}0.200	^{42}0.115
20	COGNITEC-1	^{32}0.063	^{32}0.031	^{32}0.018	^{37}0.105	^{36}0.055	^{35}0.027	^{36}0.230	^{35}0.135	^{33}0.071
21	DERMALOG-0	^{77}0.348	^{79}0.233	^{78}0.136	^{91}0.488	^{94}0.364	^{95}0.233	^{71}0.657	^{75}0.528	^{71}0.362
22	DERMALOG-1	^{80}0.397	^{86}0.279	^{85}0.172	^{94}0.528	^{101}0.405	^{100}0.268			
23	DERMALOG-2	^{78}0.362	^{81}0.248	^{81}0.147	^{93}0.503	^{96}0.378	^{97}0.244			
24	DERMALOG-3				^{90}0.484	^{93}0.362	^{93}0.231	^{69}0.655	^{74}0.526	^{70}0.361
25	DERMALOG-4	^{76}0.346	^{78}0.228	^{77}0.132	^{89}0.481	^{92}0.360	^{92}0.230	^{70}0.657	^{72}0.526	^{69}0.359
26	EVERAI-0				^{34}0.092	^{32}0.047	^{37}0.028	^{30}0.170	^{25}0.100	^{25}0.060
27	EVERAI-1	^{25}0.052	^{11}0.012	^{10}0.006	^{16}0.052	^{10}0.023	^{9}0.010	^{16}0.128	^{13}0.074	^{11}0.039
28	EYEDEA-0	^{97}0.724	^{99}0.549	^{99}0.357	^{115}0.812	^{117}0.679	^{116}0.484	^{83}0.914	^{83}0.783	^{82}0.619
29	EYEDEA-1	^{85}0.459	^{88}0.324	^{88}0.207	^{103}0.632	^{104}0.480	^{105}0.335			
30	EYEDEA-2	^{93}0.570	^{89}0.327	^{89}0.208	^{112}0.794	^{107}0.490	^{107}0.338			
31	EYEDEA-3	^{70}0.253	^{74}0.154	^{74}0.089	^{81}0.389	^{87}0.267	^{86}0.160	^{63}0.543	^{64}0.404	^{63}0.264
32	GLORY-0				^{78}0.369	^{90}0.297	^{94}0.233	^{64}0.547	^{67}0.470	^{74}0.390
33	GLORY-1	^{69}0.240	^{76}0.182	^{79}0.140	^{74}0.307	^{84}0.238	^{89}0.179	^{62}0.537	^{65}0.448	^{67}0.352
34	GORILLA-0									
35	GORILLA-1				^{85}0.408	^{85}0.248	^{84}0.136	^{57}0.453	^{59}0.314	^{58}0.191
36	HBINNO-0	^{95}0.632	^{98}0.498	^{98}0.336	^{111}0.766	^{115}0.632	^{115}0.458			
37	HIK-0	^{40}0.078	^{41}0.049	^{51}0.035	^{42}0.114	^{47}0.070	^{47}0.040	^{24}0.155	^{28}0.103	^{28}0.061
38	HIK-1	^{54}0.131	^{62}0.095	^{72}0.081	^{46}0.120	^{45}0.067	^{44}0.034			
39	HIK-2	^{37}0.076	^{35}0.037	^{35}0.022	^{47}0.121	^{46}0.067	^{43}0.034			
40	HIK-3				^{38}0.105	^{40}0.060	^{41}0.030	^{26}0.158	^{29}0.105	^{27}0.061
41	HIK-4	^{27}0.053	^{29}0.027	^{25}0.015	^{35}0.101	^{38}0.056	^{39}0.029	^{23}0.153	^{27}0.101	^{23}0.059
42	IDEMIA-0	^{38}0.077	^{34}0.036	^{33}0.019	^{41}0.114	^{42}0.062	^{38}0.029	^{37}0.240	^{37}0.156	^{35}0.085
43	IDEMIA-1	^{19}0.041	^{19}0.021	^{23}0.013	^{18}0.054	^{21}0.031	^{23}0.018			
44	IDEMIA-2	^{21}0.043	^{25}0.023	^{29}0.016	^{19}0.054	^{24}0.032	^{24}0.019			
45	IDEMIA-3	^{18}0.041	^{22}0.021	^{28}0.015	^{12}0.050	^{14}0.024	^{17}0.014	^{29}0.165	^{16}0.079	^{21}0.050
46	IDEMIA-4	^{17}0.034	^{16}0.019	^{24}0.013	^{7}0.040	^{13}0.024	^{18}0.014	^{14}0.118	^{15}0.079	^{20}0.050
47	IMAGUS-0	^{94}0.592	^{97}0.468	^{97}0.329	^{109}0.734	^{114}0.608	^{114}0.453	^{81}0.872	^{82}0.779	^{83}0.635
48	IMAGUS-2	^{91}0.561	^{95}0.410	^{94}0.253	^{110}0.751	^{112}0.566	^{112}0.377	^{78}0.816	^{79}0.645	^{78}0.460
49	IMAGUS-3				^{113}0.808	^{116}0.670	^{118}0.512	^{82}0.909	^{84}0.809	^{84}0.667
50	INCODE-0				^{75}0.313	^{81}0.201	^{81}0.107	^{55}0.420	^{58}0.304	^{57}0.191
51	INCODE-1	^{49}0.127	^{48}0.061	^{36}0.027	^{58}0.214	^{58}0.114	^{53}0.050	^{42}0.296	^{42}0.198	^{40}0.110
52	INNOVATRICS-0	^{59}0.171	^{64}0.100	^{62}0.055	^{66}0.255	^{76}0.165	^{77}0.089	^{52}0.361	^{53}0.258	^{54}0.159
53	INNOVATRICS-1	^{58}0.171	^{63}0.100	^{61}0.055	^{65}0.255	^{75}0.165	^{76}0.089			
54	INNOVATRICS-2				^{63}0.237	^{71}0.142	^{69}0.079	^{47}0.310	^{47}0.209	^{46}0.126
55	INNOVATRICS-3	^{55}0.133	^{54}0.068	^{46}0.033	^{60}0.224	^{68}0.134	^{64}0.068	^{43}0.297	^{45}0.203	^{43}0.116
56	ISYSTEMS-0	^{34}0.072	^{38}0.040	^{38}0.028	^{33}0.091	^{30}0.047	^{33}0.023	^{31}0.173	^{32}0.110	^{29}0.065
57	ISYSTEMS-1	^{35}0.072	^{37}0.040	^{37}0.028	^{31}0.090	^{28}0.047	^{32}0.023			
58	ISYSTEMS-2	^{23}0.045	^{18}0.020	^{19}0.011	^{28}0.081	^{26}0.035	^{21}0.015	^{15}0.126	^{17}0.080	^{18}0.046
59	MEGVII-0	^{31}0.062	^{28}0.025	^{18}0.011	^{40}0.109	^{39}0.058	^{34}0.025	^{11}0.116	^{9}0.067	^{6}0.034
60	MICROFOCUS-0	^{105}0.877	^{105}0.793	^{105}0.641	^{120}0.933	^{123}0.867	^{123}0.749	^{89}0.985	^{89}0.950	^{89}0.877
61	MICROFOCUS-1	^{104}0.877	^{104}0.793	^{104}0.641	^{119}0.933	^{122}0.867	^{122}0.749			
62	MICROFOCUS-2	^{106}0.878	^{106}0.796	^{106}0.654	^{121}0.934	^{124}0.870	^{124}0.758			
63	MICROFOCUS-3	^{103}0.872	^{103}0.791	^{103}0.640	^{118}0.931	^{121}0.866	^{121}0.748	^{88}0.979	^{88}0.948	^{88}0.876
64	MICROFOCUS-4				^{125}0.999	^{126}0.999	^{125}0.994	^{87}0.975	^{87}0.940	^{87}0.862

Table 12: **Threshold-based accuracy**. *Values are FNIR(N, T, L) with N = 1.6 million with thresholds set to produce FPIR = 0.001, 0.01, and 0.1 in non-mate searches. Columns 3-5 apply to FRVT-2014 mugshots: Green indicates FNIR below the best reported in NISTIR 8009 2014-04, for NEC CORP E30C, on identical images. These values are 0.097, 0.063 and 0.048 respectively. Columns 6-8 show the corresponding FNIR values for mugshots from new FRVT-2018 dataset. Finally, the three rightmost columns show FNIR for webcam images searched against the FRVT-2018 mugshot gallery. Throughout blue superscripts indicate the rank of the algorithm for that column.*

MISSES BELOW THRESHOLD, T		ENROL MOST RECENT MUGSHOT, N = 1.6M								
FNIR(N, T> 0, R >L)		DATASET: FRVT 2014 MUGSHOTS			DATASET: FRVT 2018 MUGSHOTS			DATASET: WEBCAM PROBES		
#	ALGORITHM	FPIR=0.001	FPIR=0.01	FPIR=0.1	FPIR=0.001	FPIR=0.01	FPIR=0.1	FPIR=0.001	FPIR=0.01	FPIR=0.1
65	MICROSOFT-0	[6] 0.025	[7] 0.010	[5] 0.005	[9] 0.044	[7] 0.022	[10] 0.010	[10] 0.115	[11] 0.071	[12] 0.040
66	MICROSOFT-1	[8] 0.026	[8] 0.011	[7] 0.005	[10] 0.045	[8] 0.022	[11] 0.011			
67	MICROSOFT-2	[10] 0.030	[12] 0.013	[12] 0.006	[14] 0.050	[16] 0.026	[14] 0.012			
68	MICROSOFT-3	[5] 0.019	[4] 0.007	[2] 0.004	[6] 0.030	[6] 0.014	[4] 0.006	[5] 0.091	[5] 0.056	[4] 0.028
69	MICROSOFT-4	[3] 0.017	[1] 0.007	[1] 0.004	[5] 0.029	[5] 0.013	[3] 0.005	[3] 0.087	[3] 0.053	[3] 0.026
70	NEC-0	[30] 0.059	[31] 0.030	[34] 0.019	[29] 0.082	[33] 0.049	[40] 0.029	[21] 0.140	[21] 0.093	[24] 0.059
71	NEC-1	[39] 0.078	[39] 0.043	[40] 0.030	[39] 0.108	[43] 0.063	[46] 0.035	[34] 0.197	[34] 0.133	[34] 0.083
72	NEUROTECHNOLOGY-0	[66] 0.204	[70] 0.110	[67] 0.060	[70] 0.295	[80] 0.196	[83] 0.108	[58] 0.465	[60] 0.317	[60] 0.196
73	NEUROTECHNOLOGY-1	[64] 0.197	[68] 0.107	[64] 0.057	[72] 0.299	[79] 0.195	[80] 0.105			
74	NEUROTECHNOLOGY-2	[63] 0.197	[69] 0.107	[65] 0.057	[73] 0.299	[78] 0.195	[79] 0.105			
75	NEUROTECHNOLOGY-3	[47] 0.114	[45] 0.060	[47] 0.034	[106] 0.665	[56] 0.101	[55] 0.052	[40] 0.266	[38] 0.164	[37] 0.088
76	NEUROTECHNOLOGY-4	[22] 0.045	[26] 0.024	[30] 0.018	[24] 0.066	[19] 0.030	[19] 0.014	[12] 0.117	[12] 0.073	[13] 0.040
77	NTECHLAB-0	[26] 0.052	[24] 0.023	[17] 0.011	[30] 0.083	[31] 0.047	[31] 0.023	[28] 0.162	[30] 0.105	[26] 0.061
78	NTECHLAB-1	[29] 0.057	[30] 0.027	[22] 0.013	[36] 0.102	[37] 0.056	[36] 0.027			
79	NTECHLAB-3				[22] 0.056	[20] 0.030	[20] 0.015	[13] 0.118	[14] 0.075	[15] 0.043
80	NTECHLAB-4	[7] 0.025	[9] 0.011	[9] 0.006	[8] 0.043	[12] 0.024	[13] 0.012	[7] 0.105	[8] 0.065	[9] 0.036
81	RANKONE-0	[65] 0.200	[59] 0.090	[68] 0.061	[59] 0.219	[67] 0.129	[67] 0.078	[54] 0.391	[56] 0.291	[59] 0.195
82	RANKONE-1	[57] 0.150	[55] 0.073	[56] 0.042	[52] 0.168	[52] 0.087	[50] 0.043			
83	RANKONE-2	[46] 0.109	[47] 0.060	[54] 0.039	[44] 0.120	[50] 0.073	[49] 0.042	[39] 0.261	[41] 0.190	[45] 0.126
84	RANKONE-3	[45] 0.109	[46] 0.060	[53] 0.039	[43] 0.120	[49] 0.073	[48] 0.042	[38] 0.255	[40] 0.187	[44] 0.122
85	REALNETWORKS-0	[68] 0.226	[56] 0.080	[55] 0.042	[62] 0.236	[70] 0.140	[66] 0.077	[49] 0.319	[49] 0.209	[48] 0.129
86	REALNETWORKS-1				[61] 0.236	[69] 0.140	[65] 0.077	[48] 0.319	[48] 0.209	[47] 0.129
87	SHAMAN-0	[79] 0.373	[83] 0.260	[86] 0.174	[88] 0.474	[95] 0.370	[99] 0.259	[67] 0.621	[71] 0.507	[72] 0.375
88	SHAMAN-1	[83] 0.405	[87] 0.283	[87] 0.183	[95] 0.532	[102] 0.406	[102] 0.274			
89	SHAMAN-2	[92] 0.567	[96] 0.444	[96] 0.298	[108] 0.700	[113] 0.582	[113] 0.424			
90	SHAMAN-3	[75] 0.343	[80] 0.244	[84] 0.156	[87] 0.453	[91] 0.348	[91] 0.225	[66] 0.597	[68] 0.472	[64] 0.317
91	SHAMAN-4				[100] 0.616	[106] 0.490	[109] 0.344	[75] 0.754	[78] 0.639	[79] 0.480
92	SIAT-0	[28] 0.053	[27] 0.025	[21] 0.012	[32] 0.091	[29] 0.047	[30] 0.022	[8] 0.107	[7] 0.064	[8] 0.035
93	SIAT-1	[4] 0.018	[3] 0.007	[6] 0.005	[1] 0.020	[1] 0.009	[2] 0.005	[53] 0.365	[61] 0.348	[65] 0.337
94	SIAT-2	[41] 0.093	[58] 0.084	[73] 0.082	[4] 0.024	[2] 0.009	[1] 0.005	[59] 0.478	[66] 0.460	[77] 0.451
95	SMILART-0	[87] 0.502	[92] 0.375	[92] 0.237	[101] 0.620	[105] 0.486	[103] 0.322			
96	SMILART-1	[89] 0.517	[93] 0.385	[93] 0.243	[105] 0.641	[110] 0.505	[108] 0.342			
97	SMILART-2	[88] 0.514	[91] 0.375	[91] 0.233	[102] 0.629	[108] 0.492	[104] 0.325			
98	SYNESIS-0	[82] 0.404	[84] 0.262	[80] 0.143	[99] 0.554	[97] 0.378	[90] 0.213	[74] 0.734	[77] 0.598	[76] 0.431
99	TEVIAN-0	[51] 0.127	[51] 0.065	[42] 0.032	[56] 0.203	[60] 0.114	[58] 0.054	[50] 0.331	[50] 0.227	[49] 0.132
100	TEVIAN-1	[52] 0.127	[52] 0.065	[43] 0.032	[57] 0.203	[61] 0.114	[59] 0.054			
101	TEVIAN-2	[50] 0.127	[53] 0.065	[44] 0.032	[55] 0.202	[59] 0.114	[57] 0.054			
102	TEVIAN-3				[54] 0.180	[54] 0.098	[51] 0.044	[44] 0.298	[43] 0.198	[41] 0.113
103	TEVIAN-4	[36] 0.074	[33] 0.035	[31] 0.018	[45] 0.120	[44] 0.066	[42] 0.031	[33] 0.176	[33] 0.115	[30] 0.065
104	TIGER-0	[71] 0.257	[73] 0.151	[71] 0.076	[82] 0.392	[86] 0.263	[85] 0.142	[60] 0.500	[62] 0.366	[61] 0.211
105	TIGER-1				[92] 0.491	[100] 0.404	[106] 0.337	[65] 0.580	[69] 0.487	[75] 0.396
106	TONGYITRANS-0				[27] 0.077	[27] 0.041	[25] 0.019	[9] 0.112	[10] 0.069	[10] 0.038
107	TONGYITRANS-1	[20] 0.043	[17] 0.020	[20] 0.011	[25] 0.069	[25] 0.035	[22] 0.016	[6] 0.101	[6] 0.062	[7] 0.034
108	VD-0	[101] 0.851	[102] 0.733	[102] 0.555	[117] 0.917	[120] 0.828	[120] 0.668	[85] 0.946	[85] 0.871	[85] 0.725
109	VIGILANTSOLUTIONS-0	[81] 0.397	[82] 0.260	[83] 0.154	[97] 0.539	[98] 0.394	[98] 0.247	[73] 0.695	[76] 0.557	[73] 0.389
110	VIGILANTSOLUTIONS-1	[86] 0.500	[90] 0.354	[90] 0.226	[104] 0.637	[109] 0.502	[110] 0.348			
111	VIGILANTSOLUTIONS-2	[100] 0.810	[100] 0.623	[100] 0.370	[116] 0.876	[119] 0.731	[117] 0.489			
112	VIGILANTSOLUTIONS-3	[73] 0.279	[75] 0.169	[75] 0.092	[86] 0.410	[88] 0.283	[87] 0.163	[72] 0.660	[73] 0.526	[68] 0.356
113	VIGILANTSOLUTIONS-4				[98] 0.550	[103] 0.424	[101] 0.268	[80] 0.817	[81] 0.709	[81] 0.523
114	VISIONLABS-3	[11] 0.030	[14] 0.015	[16] 0.010	[15] 0.051	[17] 0.026	[16] 0.013	[18] 0.137	[19] 0.091	[22] 0.051
115	VISIONLABS-4	[16] 0.034	[10] 0.012	[8] 0.005	[23] 0.060	[18] 0.026	[8] 0.010	[27] 0.159	[23] 0.097	[16] 0.045
116	VISIONLABS-5	[9] 0.030	[6] 0.010	[4] 0.005	[17] 0.053	[9] 0.022	[7] 0.008	[22] 0.147	[18] 0.087	[14] 0.041
117	VOCORD-0	[56] 0.133	[50] 0.063	[49] 0.034	[83] 0.399	[63] 0.116	[63] 0.062	[41] 0.285	[39] 0.181	[39] 0.108
118	VOCORD-1	[53] 0.130	[49] 0.062	[48] 0.034	[71] 0.299	[62] 0.116	[62] 0.062			
119	VOCORD-2	[48] 0.120	[44] 0.057	[45] 0.033	[77] 0.366	[57] 0.107	[61] 0.057			
120	VOCORD-3	[33] 0.065	[23] 0.022	[13] 0.009	[48] 0.126	[34] 0.050	[26] 0.020	[25] 0.155	[22] 0.093	[19] 0.048
121	VOCORD-4				[79] 0.380	[35] 0.054	[29] 0.021	[32] 0.173	[20] 0.093	[17] 0.046
122	YISHENG-0	[72] 0.258	[72] 0.116	[57] 0.046	[80] 0.380	[83] 0.209	[72] 0.086	[86] 0.974	[55] 0.275	[53] 0.146
123	YISHENG-1	[67] 0.223	[71] 0.115	[58] 0.047	[76] 0.348	[82] 0.208	[78] 0.090	[77] 0.808	[54] 0.269	[52] 0.144
124	YITU-0	[13] 0.031	[15] 0.016	[15] 0.010	[13] 0.050	[15] 0.025	[15] 0.012	[4] 0.090	[4] 0.054	[5] 0.030
125	YITU-1	[12] 0.030	[13] 0.015	[14] 0.009	[11] 0.047	[11] 0.023	[12] 0.011			
126	YITU-2	[1] 0.016	[2] 0.007	[3] 0.005	[2] 0.020	[3] 0.011	[5] 0.006	[1] 0.049	[1] 0.028	[1] 0.016
127	YITU-3	[2] 0.017	[5] 0.009	[11] 0.006	[3] 0.021	[4] 0.011	[6] 0.007	[2] 0.052	[2] 0.033	[2] 0.021

Table 13: **Threshold-based accuracy**. *Values are FNIR(N, T, L) with N = 1.6 million with thresholds set to produce FPIR = 0.001, 0.01, and 0.1 in non-mate searches. Columns 3-5 apply to FRVT-2014 mugshots: Green indicates FNIR below the best reported in NISTIR 8009 2014-04, for NEC CORP E30C, on identical images. These values are 0.097, 0.063 and 0.048 respectively. Columns 6-8 show the corresponding FNIR values for mugshots from new FRVT-2018 dataset. Finally, the three rightmost columns show FNIR for webcam images searched against the FRVT-2018 mugshot gallery. Throughout blue superscripts indicate the rank of the algorithm for that column.*

MISSES NOT AT RANK 1		ENROL LIFETIME						ENROL MOST RECENT					
FNIR(N, T= 0, R >1)		DATASET: FRVT 2018						DATASET: FRVT 2018					
#	ALGORITHM	N=0.64M	N=1.6M	N=3.0M	N=6.0M	N=12.0M	aN^b	N=0.64M	N=1.6M	N=3.0M	N=6.0M	N=12.0M	aN^b
1	3DIVI-3	[83] 0.0494	[65] 0.0645	[42] 0.0759	[38] 0.0898		[51] 0.0014 $N^{0.267}$ [35]	[88] 0.0680	[88] 0.0857				[48] 0.0023 $N^{0.252}$ [52]
2	ALCHERA-0	[45] 0.0106	[39] 0.0121	[28] 0.0135	[26] 0.0170		[40] 0.0006 $N^{0.207}$ [20]	[48] 0.0167	[44] 0.0186				[54] 0.0035 $N^{0.117}$ [11]
3	AWARE-3	[57] 0.0165	[52] 0.0209	[33] 0.0247	[31] 0.0297		[37] 0.0005 $N^{0.263}$ [34]	[63] 0.0264	[62] 0.0332	[35] 0.0387	[33] 0.0456	[32] 0.0532	[39] 0.0011 $N^{0.239}$ [47]
4	AYONIX-0	[114] 0.4198	[78] 0.4649	[49] 0.4969	[44] 0.5318		[62] 0.1021 $N^{0.106}$ [4]	[121] 0.4095	[119] 0.4519				[63] 0.0973 $N^{0.108}$ [9]
5	CAMVI-3	[49] 0.0144	[56] 0.0368	[38] 0.0528	[41] 0.1791		[2] 0.0000 $N^{1.076}$ [62]	[60] 0.0224	[79] 0.0544				[1] 0.0000 $N^{0.969}$ [64]
6	COGENT-0	[42] 0.0103	[36] 0.0106	[22] 0.0109	[17] 0.0114	[15] 0.0122	[55] 0.0047 $N^{0.057}$ [2]	[41] 0.0127	[33] 0.0131	[23] 0.0136	[19] 0.0141	[18] 0.0151	[57] 0.0058 $N^{0.058}$ [2]
7	COGENT-1	[41] 0.0103	[35] 0.0106				[57] 0.0074 $N^{0.025}$ [1]	[40] 0.0127	[32] 0.0131	[22] 0.0136	[18] 0.0141	[17] 0.0151	[56] 0.0058 $N^{0.058}$ [1]
8	COGNITEC-0	[50] 0.0146	[50] 0.0205				[19] 0.0001 $N^{0.376}$ [58]	[58] 0.0221	[59] 0.0286	[33] 0.0339	[32] 0.0378	[31] 0.0443	[36] 0.0010 $N^{0.233}$ [44]
9	COGNITEC-1	[27] 0.0069	[29] 0.0089	[21] 0.0106	[19] 0.0128	[17] 0.0154	[27] 0.0002 $N^{0.275}$ [39]	[37] 0.0116	[40] 0.0143	[27] 0.0165	[25] 0.0192	[24] 0.0225	[25] 0.0006 $N^{0.226}$ [41]
10	DERMALOG-4	[85] 0.0759	[67] 0.0961	[45] 0.1105	[40] 0.1260		[54] 0.0037 $N^{0.227}$ [25]	[92] 0.1040	[92] 0.1274				[55] 0.0054 $N^{0.221}$ [39]
11	EVERAI-0	[26] 0.0065	[46] 0.0166				[1] 0.0000 $N^{1.029}$ [61]	[31] 0.0102	[48] 0.0209	[34] 0.0348			[2] 0.0000 $N^{0.795}$ [63]
12	EVERAI-1	[9] 0.0022	[9] 0.0027				[23] 0.0001 $N^{0.222}$ [24]	[8] 0.0047	[9] 0.0056	[9] 0.0061			[22] 0.0005 $N^{0.166}$ [20]
13	EYEDEA-3	[82] 0.0480	[64] 0.0613	[41] 0.0717	[37] 0.0831		[52] 0.0018 $N^{0.246}$ [29]	[87] 0.0663	[87] 0.0824				[52] 0.0028 $N^{0.238}$ [46]
14	GLORY-1	[91] 0.0818	[66] 0.0932	[43] 0.1007	[39] 0.1091		[59] 0.0147 $N^{0.129}$ [6]	[97] 0.1154	[94] 0.1291				[61] 0.0223 $N^{0.123}$ [14]
15	HIK-2	[55] 0.0155	[47] 0.0185	[31] 0.0208	[29] 0.0240	[24] 0.0272	[49] 0.0012 $N^{0.193}$ [12]	[43] 0.0147	[42] 0.0172				[44] 0.0015 $N^{0.173}$ [23]
16	HIK-3	[36] 0.0085	[37] 0.0107				[31] 0.0003 $N^{0.255}$ [31]	[36] 0.0115	[39] 0.0141	[26] 0.0164	[26] 0.0194	[25] 0.0228	[19] 0.0005 $N^{0.235}$ [45]
17	HIK-4	[35] 0.0083	[34] 0.0104	[25] 0.0121	[22] 0.0146	[18] 0.0177	[29] 0.0003 $N^{0.260}$ [33]	[35] 0.0112	[37] 0.0138	[25] 0.0159	[24] 0.0188	[23] 0.0220	[21] 0.0005 $N^{0.230}$ [43]
18	IDEMIA-0	[19] 0.0048	[21] 0.0063	[14] 0.0076	[12] 0.0095	[12] 0.0116	[17] 0.0001 $N^{0.304}$ [48]	[29] 0.0093	[28] 0.0113	[20] 0.0131	[20] 0.0153	[20] 0.0182	[16] 0.0004 $N^{0.227}$ [42]
19	IDEMIA-1	[21] 0.0049	[22] 0.0065	[16] 0.0080	[14] 0.0100	[16] 0.0124	[14] 0.0001 $N^{0.320}$ [53]	[30] 0.0096	[30] 0.0116	[21] 0.0135	[21] 0.0162	[21] 0.0194	[15] 0.0004 $N^{0.243}$ [49]
20	IDEMIA-2	[33] 0.0075	[31] 0.0099	[24] 0.0119	[24] 0.0149	[21] 0.0183	[24] 0.0001 $N^{0.304}$ [49]	[33] 0.0105	[31] 0.0126				[31] 0.0008 $N^{0.194}$ [29]
21	IDEMIA-3	[17] 0.0041	[18] 0.0054				[16] 0.0001 $N^{0.294}$ [47]	[22] 0.0080	[24] 0.0095	[17] 0.0110	[16] 0.0127	[16] 0.0148	[18] 0.0005 $N^{0.212}$ [36]
22	IDEMIA-4	[18] 0.0042	[17] 0.0052	[11] 0.0061	[10] 0.0074	[11] 0.0088	[25] 0.0001 $N^{0.257}$ [32]	[23] 0.0080	[21] 0.0092	[16] 0.0106	[15] 0.0124	[15] 0.0143	[23] 0.0005 $N^{0.202}$ [31]
23	IMAGUS-2	[102] 0.1470	[72] 0.1833	[46] 0.2086	[42] 0.2379		[58] 0.0083 $N^{0.215}$ [21]	[109] 0.1838	[109] 0.2223				[60] 0.0115 $N^{0.208}$ [35]
24	INCODE-1	[39] 0.0098	[41] 0.0131	[35] 0.0286	[34] 0.0466		[3] 0.0000 $N^{0.729}$ [60]	[45] 0.0151	[45] 0.0190				[24] 0.0005 $N^{0.250}$ [50]
25	ISYSTEMS-0	[31] 0.0074	[28] 0.0085	[19] 0.0095	[16] 0.0105	[14] 0.0118	[45] 0.0009 $N^{0.160}$ [8]	[39] 0.0122	[36] 0.0136				[50] 0.0025 $N^{0.119}$ [13]
26	ISYSTEMS-1	[32] 0.0074	[27] 0.0085	[18] 0.0094	[15] 0.0105	[13] 0.0118	[46] 0.0009 $N^{0.158}$ [7]	[38] 0.0122	[35] 0.0136				[51] 0.0025 $N^{0.118}$ [12]
27	ISYSTEMS-2	[15] 0.0039	[13] 0.0046	[9] 0.0052			[36] 0.0004 $N^{0.175}$ [10]	[21] 0.0076	[19] 0.0088	[15] 0.0096	[13] 0.0108	[13] 0.0121	[34] 0.0009 $N^{0.156}$ [16]
28	MEGVII-0	[30] 0.0072	[32] 0.0099	[26] 0.0123	[25] 0.0150	[20] 0.0182	[21] 0.0001 $N^{0.317}$ [51]	[20] 0.0075	[22] 0.0094	[18] 0.0111	[17] 0.0134	[19] 0.0162	[6] 0.0002 $N^{0.264}$ [55]
29	MICROFOCUS-3	[116] 0.4791	[80] 0.5389	[50] 0.5771			[61] 0.0951 $N^{0.121}$ [5]	[123] 0.5417	[122] 0.5953				[64] 0.1370 $N^{0.103}$ [8]
30	MICROSOFT-0	[7] 0.0021	[8] 0.0026	[5] 0.0031	[5] 0.0040	[5] 0.0048	[11] 0.0000 $N^{0.280}$ [41]	[10] 0.0051	[11] 0.0058	[10] 0.0066	[9] 0.0077	[9] 0.0090	[14] 0.0003 $N^{0.199}$ [30]
31	MICROSOFT-1	[6] 0.0020	[7] 0.0026	[4] 0.0031	[4] 0.0038	[4] 0.0047	[9] 0.0000 $N^{0.286}$ [44]	[9] 0.0049	[10] 0.0056				[26] 0.0006 $N^{0.158}$ [18]
32	MICROSOFT-2	[10] 0.0023	[10] 0.0029	[6] 0.0035	[6] 0.0042	[6] 0.0051	[13] 0.0001 $N^{0.272}$ [38]	[12] 0.0052	[12] 0.0061				[20] 0.0005 $N^{0.174}$ [24]
33	MICROSOFT-3	[2] 0.0009	[2] 0.0011				[7] 0.0000 $N^{0.255}$ [30]	[2] 0.0028	[2] 0.0032	[2] 0.0035	[2] 0.0039	[2] 0.0045	[12] 0.0003 $N^{0.166}$ [21]
34	MICROSOFT-4	[1] 0.0008	[1] 0.0010	[1] 0.0013	[1] 0.0015	[1] 0.0019	[6] 0.0000 $N^{0.285}$ [43]	[1] 0.0027	[1] 0.0031	[1] 0.0034	[1] 0.0038	[1] 0.0045	[8] 0.0003 $N^{0.174}$ [25]
35	NEC-0	[38] 0.0097	[40] 0.0127	[29] 0.0154	[27] 0.0185	[22] 0.0223	[28] 0.0002 $N^{0.284}$ [42]	[46] 0.0157	[46] 0.0196	[28] 0.0229	[27] 0.0270	[26] 0.0320	[27] 0.0006 $N^{0.243}$ [48]
36	NEC-1	[48] 0.0136	[45] 0.0164				[48] 0.0009 $N^{0.202}$ [18]	[55] 0.0206	[54] 0.0235	[31] 0.0259	[30] 0.0292	[27] 0.0329	[49] 0.0024 $N^{0.160}$ [19]
37	NEUROTECHNOLOGY-3	[56] 0.0161	[49] 0.0199				[43] 0.0007 $N^{0.234}$ [27]	[54] 0.0204	[57] 0.0250	[32] 0.0288	[31] 0.0331	[30] 0.0386	[40] 0.0011 $N^{0.216}$ [37]
38	NEUROTECHNOLOGY-4	[22] 0.0049	[19] 0.0058	[12] 0.0065	[11] 0.0075	[10] 0.0087	[35] 0.0004 $N^{0.195}$ [14]	[19] 0.0072	[16] 0.0082	[13] 0.0090	[12] 0.0100	[12] 0.0114	[33] 0.0009 $N^{0.156}$ [15]
39	NTECHLAB-0	[24] 0.0056	[24] 0.0077	[17] 0.0094	[18] 0.0114		[15] 0.0001 $N^{0.323}$ [54]	[28] 0.0092	[29] 0.0115	[24] 0.0137	[22] 0.0164	[22] 0.0196	[10] 0.0003 $N^{0.261}$ [53]
40	NTECHLAB-1	[29] 0.0070	[30] 0.0097	[23] 0.0119	[21] 0.0146	[19] 0.0179	[20] 0.0001 $N^{0.317}$ [52]	[34] 0.0108	[38] 0.0139				[9] 0.0003 $N^{0.278}$ [58]
41	NTECHLAB-3	[13] 0.0037	[16] 0.0051				[8] 0.0000 $N^{0.351}$ [57]	[15] 0.0065	[17] 0.0082	[14] 0.0096	[14] 0.0115	[14] 0.0135	[7] 0.0002 $N^{0.251}$ [51]
42	NTECHLAB-4	[11] 0.0030	[11] 0.0040	[7] 0.0049	[8] 0.0060	[9] 0.0075	[10] 0.0000 $N^{0.315}$ [50]	[13] 0.0056	[13] 0.0068	[11] 0.0078	[10] 0.0092	[11] 0.0107	[11] 0.0003 $N^{0.220}$ [38]
43	RANKONE-0	[68] 0.0255	[54] 0.0319	[36] 0.0366	[33] 0.0425	[26] 0.0486	[50] 0.0014 $N^{0.220}$ [23]	[77] 0.0375	[73] 0.0455	[36] 0.0514	[34] 0.0564	[33] 0.0654	[53] 0.0032 $N^{0.186}$ [27]
44	RANKONE-1	[53] 0.0152	[48] 0.0194	[32] 0.0224	[30] 0.0260	[25] 0.0302	[41] 0.0007 $N^{0.232}$ [26]	[61] 0.0226	[56] 0.0247				[58] 0.0062 $N^{0.097}$ [5]
45	RANKONE-2	[47] 0.0117	[44] 0.0149				[34] 0.0003 $N^{0.268}$ [36]	[53] 0.0181	[50] 0.0221	[30] 0.0250	[29] 0.0288	[29] 0.0330	[42] 0.0012 $N^{0.204}$ [34]
46	RANKONE-3	[46] 0.0117	[43] 0.0149	[30] 0.0172	[28] 0.0200	[23] 0.0236	[38] 0.0005 $N^{0.237}$ [28]	[52] 0.0181	[49] 0.0221	[29] 0.0250	[28] 0.0288	[28] 0.0330	[41] 0.0012 $N^{0.204}$ [33]
47	REALNETWORKS-0	[75] 0.0337	[59] 0.0443	[37] 0.0527			[42] 0.0007 $N^{0.290}$ [45]	[68] 0.0330	[72] 0.0426				[30] 0.0008 $N^{0.280}$ [60]
48	SHAMAN-3	[90] 0.0808	[68] 0.0969	[44] 0.1091			[56] 0.0060 $N^{0.195}$ [15]	[94] 0.1074	[90] 0.1266				[59] 0.0097 $N^{0.180}$ [26]
49	SIAT-1	[112] 0.2638	[75] 0.2639	[47] 0.2640			-	[4] 0.0037	[3] 0.0039	[3] 0.0041	[3] 0.0044	[4] 0.0049	[35] 0.0010 $N^{0.098}$ [6]
50	SIAT-2	[110] 0.2127	[74] 0.2128				-	[5] 0.0037	[4] 0.0040	[4] 0.0042	[4] 0.0045	[3] 0.0049	[37] 0.0011 $N^{0.092}$ [4]
51	TEVIAN-4	[25] 0.0058	[25] 0.0080	[20] 0.0097			[12] 0.0001 $N^{0.341}$ [56]	[32] 0.0105	[34] 0.0134				[13] 0.0003 $N^{0.264}$ [54]
52	TIGER-0	[78] 0.0364	[60] 0.0480	[39] 0.0565	[36] 0.0678		[47] 0.0009 $N^{0.278}$ [40]	[81] 0.0494	[83] 0.0638				[43] 0.0012 $N^{0.279}$ [59]
53	TONGYITRANS-1	[37] 0.0096	[38] 0.0114	[27] 0.0127	[23] 0.0148		[44] 0.0007 $N^{0.193}$ [11]	[24] 0.0080	[23] 0.0095				[28] 0.0006 $N^{0.189}$ [28]
54	VD-0	[113] 0.3583	[77] 0.4303	[48] 0.4776	[43] 0.5281		[60] 0.0355 $N^{0.174}$ [9]	[120] 0.4073	[120] 0.4751				[62] 0.0431 $N^{0.168}$ [22]
55	VIGILANTSOLUTIONS-3	[80] 0.0410	[63] 0.0549	[40] 0.0654	[35] 0.0654		[53] 0.0023 $N^{0.219}$ [22]	[86] 0.0561	[86] 0.0719				[45] 0.0015 $N^{0.271}$ [57]
56	VISIONLABS-3	[12] 0.0037	[15] 0.0050	[13] 0.0076	[20] 0.0130		[4] 0.0000 $N^{0.563}$ [59]	[18] 0.0070	[20] 0.0089	[19] 0.0124	[23] 0.0185		[3] 0.0000 $N^{0.434}$ [62]
57	VISIONLABS-4	[5] 0.0016	[5] 0.0020				[22] 0.0001 $N^{0.203}$ [19]	[6] 0.0037	[6] 0.0044	[7] 0.0049	[8] 0.0062	[8] 0.0088	[4] 0.0001 $N^{0.282}$ [61]
58	VISIONLABS-5	[4] 0.0015	[4] 0.0018	[3] 0.0020	[3] 0.0028	[3] 0.0040	[5] 0.0000 $N^{0.332}$ [55]	[3] 0.0035	[5] 0.0041	[5] 0.0046	[6] 0.0054	[7] 0.0068	[5] 0.0002 $N^{0.223}$ [40]
59	VOCORD-3	[23] 0.0053	[23] 0.0067	[15] 0.0080	[13] 0.0096		[26] 0.0001 $N^{0.271}$ [37]	[17] 0.0070	[18] 0.0085				[17] 0.0005 $N^{0.204}$ [32]
60	YISHENG-1	[54] 0.0155	[51] 0.0208	[34] 0.0248	[32] 0.0298		[33] 0.0003 $N^{0.294}$ [46]	[62] 0.0227	[61] 0.0290				[29] 0.0006 $N^{0.266}$ [56]
61	YITU-0	[16] 0.0040	[14] 0.0047	[10] 0.0053	[9] 0.0061	[8] 0.0071	[30] 0.0003 $N^{0.200}$ [17]	[16] 0.0066	[15] 0.0074	[12] 0.0082	[11] 0.0092	[10] 0.0103	[32] 0.0008 $N^{0.156}$ [17]
62	YITU-1	[14] 0.0039	[12] 0.0046	[8] 0.0051	[7] 0.0059	[7] 0.0069	[32] 0.0003 $N^{0.194}$ [13]	[14] 0.0065	[14] 0.0072				[46] 0.0015 $N^{0.110}$ [10]
63	YITU-2	[3] 0.0013	[3] 0.0015	[2] 0.0017	[2] 0.0019	[2] 0.0023	[18] 0.0001 $N^{0.196}$ [16]	[7] 0.0041	[7] 0.0044	[6] 0.0047	[5] 0.0050	[5] 0.0055	[38] 0.0011 $N^{0.099}$ [7]
64	YITU-3	[8] 0.0021	[6] 0.0023				[39] 0.0006 $N^{0.098}$ [3]	[11] 0.0052	[8] 0.0054	[8] 0.0057	[7] 0.0061	[6] 0.0065	[47] 0.0017 $N^{0.081}$ [3]

Table 14: Effect of N on FNIR at rank 1, for five enrollment population sizes, N. The left five columns apply for consolidated enrollment of a variable number of lifetime images from each subject. The right five columns apply for enrollment of one recent image. Missing entries usually apply because another algorithm from the same developer was run instead. Some developers are missing because less accurate algorithms were not run on galleries with $N > 1\,600\,000$. Throughout blue superscripts indicate the rank of the algorithm for that column, and yellow highlighting indicates the most accurate value.

MISSES NOT AT RANK 50		ENROL LIFETIME						ENROL MOST RECENT					
FNIR(N, T= 0, R >50)		DATASET: FRVT 2018						DATASET: FRVT 2018					
#	ALGORITHM	N=0.64M	N=1.6M	N=3.0M	N=6.0M	N=12.0M	aN^b	N=0.64M	N=1.6M	N=3.0M	N=6.0M	N=12.0M	aN^b
1	3DIVI-3	[74] 0.0103	[62] 0.0151	[40] 0.0192	[36] 0.0241		[20] 0.0001 $N^{0.382}$ [54]	[77] 0.0159	[79] 0.0217				[9] 0.0002 $N^{0.343}$ [57]
2	ALCHERA-0	[67] 0.0073	[51] 0.0076	[33] 0.0079	[30] 0.0101		[53] 0.0012 $N^{0.133}$ [14]	[72] 0.0125	[66] 0.0129				[62] 0.0079 $N^{0.034}$ [5]
3	AWARE-3	[50] 0.0039	[43] 0.0050	[29] 0.0061	[28] 0.0077		[24] 0.0001 $N^{0.299}$ [40]	[57] 0.0081	[60] 0.0101	[32] 0.0118	[31] 0.0139	[32] 0.0170	[17] 0.0003 $N^{0.248}$ [50]
4	AYONIX-0	[112] 0.1723	[77] 0.2142	[48] 0.2467	[44] 0.2850		[62] 0.0085 $N^{0.225}$ [30]	[121] 0.1967	[119] 0.2402				[63] 0.0107 $N^{0.218}$ [46]
5	CAMVI-3	[79] 0.0142	[68] 0.0367	[44] 0.0527	[42] 0.1789		[3] 0.0000 $N^{1.080}$ [60]	[82] 0.0221	[96] 0.0541				[2] 0.0000 $N^{0.980}$ [64]
6	COGENT-0	[24] 0.0021	[23] 0.0024	[14] 0.0027	[15] 0.0031	[16] 0.0045	[22] 0.0001 $N^{0.253}$ [36]	[27] 0.0047	[25] 0.0050	[19] 0.0054	[21] 0.0062	[26] 0.0122	[8] 0.0001 $N^{0.288}$ [53]
7	COGENT-1	[23] 0.0021	[22] 0.0024				[40] 0.0002 $N^{0.189}$ [25]	[26] 0.0047	[24] 0.0050	[18] 0.0054	[20] 0.0062	[25] 0.0122	[7] 0.0001 $N^{0.288}$ [52]
8	COGNITEC-0	[49] 0.0039	[50] 0.0067				[5] 0.0000 $N^{0.599}$ [58]	[52] 0.0076	[59] 0.0099	[33] 0.0120	[30] 0.0123	[30] 0.0148	[25] 0.0004 $N^{0.218}$ [45]
9	COGNITEC-1	[30] 0.0024	[31] 0.0028	[21] 0.0032	[19] 0.0037	[15] 0.0044	[41] 0.0002 $N^{0.200}$ [26]	[39] 0.0056	[37] 0.0060	[26] 0.0066	[24] 0.0072	[22] 0.0081	[40] 0.0010 $N^{0.128}$ [26]
10	DERMALOG-4	[84] 0.0186	[65] 0.0272	[42] 0.0340	[39] 0.0427		[36] 0.0001 $N^{0.372}$ [52]	[86] 0.0262	[90] 0.0365				[12] 0.0002 $N^{0.363}$ [58]
11	EVERAI-0	[58] 0.0050	[61] 0.0150				[2] 0.0000 $N^{1.185}$ [61]	[53] 0.0077	[77] 0.0182	[36] 0.0317			[1] 0.0000 $N^{0.919}$ [63]
12	EVERAI-1	[12] 0.0013	[11] 0.0014				[45] 0.0004 $N^{0.096}$ [10]	[12] 0.0031	[13] 0.0033	[10] 0.0034			[43] 0.0012 $N^{0.070}$ [15]
13	EYEDEA-3	[76] 0.0113	[63] 0.0160	[41] 0.0209	[37] 0.0252		[30] 0.0001 $N^{0.364}$ [50]	[80] 0.0175	[80] 0.0236				[14] 0.0002 $N^{0.326}$ [55]
14	GLORY-1	[99] 0.0415	[70] 0.0490	[45] 0.0539	[40] 0.0600		[60] 0.0047 $N^{0.164}$ [17]	[107] 0.0604	[105] 0.0698				[61] 0.0073 $N^{0.158}$ [32]
15	HIK-2	[73] 0.0084	[56] 0.0090	[34] 0.0097	[31] 0.0106	[25] 0.0118	[56] 0.0018 $N^{0.115}$ [12]	[62] 0.0087	[55] 0.0093				[56] 0.0035 $N^{0.068}$ [14]
16	HIK-3	[25] 0.0023	[30] 0.0028				[33] 0.0001 $N^{0.230}$ [32]	[21] 0.0044	[28] 0.0051	[22] 0.0058	[23] 0.0066	[21] 0.0076	[21] 0.0003 $N^{0.189}$ [38]
17	HIK-4	[28] 0.0023	[32] 0.0028	[22] 0.0033	[20] 0.0039	[18] 0.0048	[29] 0.0001 $N^{0.246}$ [33]	[23] 0.0045	[29] 0.0051	[23] 0.0058	[22] 0.0065	[20] 0.0076	[24] 0.0004 $N^{0.175}$ [34]
18	IDEMIA-0	[16] 0.0016	[17] 0.0019	[11] 0.0023	[10] 0.0026	[9] 0.0031	[27] 0.0001 $N^{0.226}$ [31]	[25] 0.0045	[26] 0.0051	[20] 0.0055	[18] 0.0060	[19] 0.0067	[38] 0.0008 $N^{0.134}$ [27]
19	IDEMIA-1	[18] 0.0019	[21] 0.0024	[20] 0.0029	[18] 0.0036	[17] 0.0046	[13] 0.0000 $N^{0.307}$ [42]	[34] 0.0049	[36] 0.0058	[25] 0.0065	[25] 0.0076	[23] 0.0089	[20] 0.0003 $N^{0.201}$ [42]
20	IDEMIA-2	[40] 0.0031	[37] 0.0040	[24] 0.0048	[24] 0.0058	[22] 0.0074	[21] 0.0001 $N^{0.290}$ [38]	[45] 0.0061	[43] 0.0069				[41] 0.0010 $N^{0.135}$ [28]
21	IDEMIA-3	[19] 0.0019	[19] 0.0022				[42] 0.0002 $N^{0.175}$ [19]	[31] 0.0049	[32] 0.0053	[21] 0.0057	[19] 0.0062	[18] 0.0067	[42] 0.0011 $N^{0.109}$ [23]
22	IDEMIA-4	[15] 0.0015	[13] 0.0017	[8] 0.0020	[8] 0.0023	[8] 0.0028	[31] 0.0001 $N^{0.207}$ [27]	[20] 0.0043	[20] 0.0046	[15] 0.0051	[15] 0.0055	[16] 0.0062	[39] 0.0008 $N^{0.121}$ [25]
23	IMAGUS-2	[97] 0.0348	[71] 0.0510	[46] 0.0641	[41] 0.0804		[43] 0.0002 $N^{0.375}$ [53]	[100] 0.0468	[101] 0.0657				[19] 0.0003 $N^{0.371}$ [59]
24	INCODE-1	[32] 0.0026	[34] 0.0033	[39] 0.0167	[38] 0.0323		[1] 0.0000 $N^{1.217}$ [62]	[37] 0.0055	[38] 0.0063				[34] 0.0007 $N^{0.153}$ [31]
25	ISYSTEMS-0	[55] 0.0048	[42] 0.0050	[26] 0.0053	[23] 0.0056	[21] 0.0060	[54] 0.0017 $N^{0.076}$ [6]	[59] 0.0086	[54] 0.0089				[57] 0.0048 $N^{0.044}$ [7]
26	ISYSTEMS-1	[56] 0.0048	[44] 0.0050	[25] 0.0053	[22] 0.0056	[20] 0.0060	[55] 0.0017 $N^{0.075}$ [5]	[60] 0.0086	[53] 0.0089				[58] 0.0049 $N^{0.041}$ [6]
27	ISYSTEMS-2	[34] 0.0026	[28] 0.0027	[18] 0.0029			[52] 0.0012 $N^{0.061}$ [3]	[36] 0.0054	[35] 0.0056	[24] 0.0058	[17] 0.0060	[17] 0.0063	[53] 0.0027 $N^{0.051}$ [10]
28	MEGVII-0	[11] 0.0012	[16] 0.0019	[12] 0.0025	[17] 0.0032	[14] 0.0041	[6] 0.0000 $N^{0.422}$ [56]	[5] 0.0026	[9] 0.0031	[9] 0.0034	[11] 0.0039	[10] 0.0048	[10] 0.0002 $N^{0.204}$ [43]
29	MICROFOCUS-3	[114] 0.2047	[79] 0.2625	[50] 0.3017			[61] 0.0070 $N^{0.252}$ [34]	[123] 0.2518	[122] 0.3113				[64] 0.0114 $N^{0.232}$ [48]
30	MICROSOFT-0	[3] 0.0008	[6] 0.0010	[5] 0.0011	[4] 0.0012	[3] 0.0014	[28] 0.0001 $N^{0.174}$ [18]	[8] 0.0028	[8] 0.0031	[6] 0.0032	[7] 0.0035	[6] 0.0037	[35] 0.0007 $N^{0.101}$ [20]
31	MICROSOFT-1	[4] 0.0008	[4] 0.0009	[4] 0.0011	[3] 0.0012	[4] 0.0014	[25] 0.0001 $N^{0.177}$ [21]	[7] 0.0028	[7] 0.0030				[37] 0.0007 $N^{0.098}$ [19]
32	MICROSOFT-2	[5] 0.0008	[5] 0.0010	[3] 0.0011	[5] 0.0012	[5] 0.0014	[23] 0.0001 $N^{0.186}$ [23]	[10] 0.0029	[10] 0.0032				[36] 0.0007 $N^{0.101}$ [21]
33	MICROSOFT-3	[2] 0.0004	[2] 0.0004				[16] 0.0001 $N^{0.153}$ [16]	[2] 0.0018	[2] 0.0019	[2] 0.0021	[2] 0.0022	[2] 0.0023	[32] 0.0006 $N^{0.078}$ [17]
34	MICROSOFT-4	[1] 0.0004	[1] 0.0004	[1] 0.0005	[1] 0.0005	[1] 0.0006	[19] 0.0001 $N^{0.140}$ [15]	[1] 0.0018	[1] 0.0019	[1] 0.0020	[1] 0.0021	[1] 0.0022	[33] 0.0007 $N^{0.070}$ [16]
35	NEC-0	[26] 0.0023	[33] 0.0030	[23] 0.0038	[21] 0.0047	[19] 0.0059	[11] 0.0000 $N^{0.324}$ [45]	[38] 0.0055	[39] 0.0064	[27] 0.0074	[26] 0.0085	[24] 0.0100	[22] 0.0003 $N^{0.205}$ [44]
36	NEC-1	[69] 0.0076	[52] 0.0080				[59] 0.0038 $N^{0.051}$ [2]	[75] 0.0135	[67] 0.0138	[34] 0.0142	[32] 0.0147	[31] 0.0154	[60] 0.0073 $N^{0.046}$ [8]
37	NEUROTECHNOLOGY-3	[48] 0.0038	[45] 0.0051				[15] 0.0000 $N^{0.326}$ [47]	[47] 0.0068	[49] 0.0083	[28] 0.0097	[29] 0.0116	[29] 0.0137	[16] 0.0003 $N^{0.243}$ [49]
38	NEUROTECHNOLOGY-4	[21] 0.0020	[20] 0.0024	[15] 0.0027	[14] 0.0031	[12] 0.0035	[39] 0.0002 $N^{0.189}$ [24]	[28] 0.0048	[27] 0.0051	[17] 0.0054	[16] 0.0057	[15] 0.0060	[46] 0.0016 $N^{0.081}$ [18]
39	NTECHLAB-0	[13] 0.0013	[12] 0.0016	[9] 0.0021	[9] 0.0026		[10] 0.0000 $N^{0.320}$ [43]	[14] 0.0033	[15] 0.0039	[13] 0.0043	[13] 0.0051	[13] 0.0058	[15] 0.0002 $N^{0.193}$ [41]
40	NTECHLAB-1	[14] 0.0013	[15] 0.0018	[10] 0.0022	[11] 0.0029	[13] 0.0038	[8] 0.0000 $N^{0.366}$ [51]	[15] 0.0034	[17] 0.0040				[18] 0.0003 $N^{0.177}$ [35]
41	NTECHLAB-3	[10] 0.0010	[10] 0.0012				[17] 0.0001 $N^{0.219}$ [29]	[9] 0.0028	[12] 0.0032	[11] 0.0035	[10] 0.0039	[9] 0.0044	[23] 0.0004 $N^{0.149}$ [30]
42	NTECHLAB-4	[7] 0.0009	[8] 0.0010	[6] 0.0012	[6] 0.0014	[6] 0.0016	[18] 0.0001 $N^{0.208}$ [28]	[6] 0.0027	[5] 0.0030	[7] 0.0032	[6] 0.0035	[7] 0.0039	[27] 0.0005 $N^{0.120}$ [24]
43	RANKONE-0	[68] 0.0074	[58] 0.0100	[36] 0.0120	[34] 0.0146	[26] 0.0176	[37] 0.0001 $N^{0.297}$ [39]	[74] 0.0127	[72] 0.0159	[35] 0.0185	[34] 0.0206	[33] 0.0252	[29] 0.0006 $N^{0.226}$ [47]
44	RANKONE-1	[51] 0.0042	[47] 0.0055	[32] 0.0067	[29] 0.0082	[24] 0.0100	[26] 0.0001 $N^{0.300}$ [41]	[56] 0.0078	[50] 0.0086				[48] 0.0020 $N^{0.103}$ [22]
45	RANKONE-2	[47] 0.0037	[40] 0.0047				[35] 0.0001 $N^{0.253}$ [35]	[51] 0.0075	[52] 0.0087	[30] 0.0098	[28] 0.0111	[28] 0.0128	[31] 0.0006 $N^{0.184}$ [37]
46	RANKONE-3	[46] 0.0037	[39] 0.0047	[27] 0.0055	[25] 0.0067	[23] 0.0079	[34] 0.0001 $N^{0.258}$ [37]	[50] 0.0075	[51] 0.0087	[29] 0.0098	[27] 0.0111	[27] 0.0128	[30] 0.0006 $N^{0.184}$ [36]
47	REALNETWORKS-0	[61] 0.0059	[54] 0.0083	[35] 0.0108			[12] 0.0000 $N^{0.393}$ [55]	[55] 0.0077	[58] 0.0098				[13] 0.0002 $N^{0.267}$ [51]
48	SHAMAN-3	[94] 0.0344	[69] 0.0404	[43] 0.0452			[58] 0.0032 $N^{0.177}$ [20]	[101] 0.0468	[97] 0.0544				[59] 0.0053 $N^{0.163}$ [33]
49	SIAT-1	[118] 0.2635	[80] 0.2635	[49] 0.2636			-	[11] 0.0029	[6] 0.0030	[5] 0.0031	[3] 0.0032	[3] 0.0033	[45] 0.0016 $N^{0.046}$ [9]
50	SIAT-2	[116] 0.2124	[76] 0.2124				-	[13] 0.0031	[11] 0.0032	[8] 0.0032	[4] 0.0033	[4] 0.0034	[49] 0.0020 $N^{0.032}$ [4]
51	TEVIAN-4	[20] 0.0019	[18] 0.0022	[13] 0.0025			[38] 0.0002 $N^{0.185}$ [22]	[19] 0.0041	[19] 0.0046				[28] 0.0006 $N^{0.143}$ [29]
52	TIGER-0	[62] 0.0061	[57] 0.0097	[37] 0.0125	[35] 0.0164		[9] 0.0000 $N^{0.444}$ [57]	[65] 0.0098	[68] 0.0139				[5] 0.0001 $N^{0.384}$ [61]
53	TONGYITRANS-1	[60] 0.0057	[48] 0.0060	[30] 0.0062	[26] 0.0067		[57] 0.0020 $N^{0.076}$ [7]	[32] 0.0049	[30] 0.0052				[51] 0.0022 $N^{0.061}$ [12]
54	VD-0	[110] 0.1006	[75] 0.1421	[47] 0.1752	[43] 0.2147		[50] 0.0011 $N^{0.340}$ [48]	[118] 0.1248	[118] 0.1699				[44] 0.0014 $N^{0.336}$ [56]
55	VIGILANTSOLUTIONS-3	[66] 0.0072	[59] 0.0110	[38] 0.0143	[33] 0.0143		[32] 0.0001 $N^{0.322}$ [44]	[71] 0.0118	[74] 0.0166				[6] 0.0001 $N^{0.373}$ [60]
56	VISIONLABS-3	[38] 0.0030	[38] 0.0042	[31] 0.0066	[32] 0.0119		[4] 0.0000 $N^{0.612}$ [59]	[43] 0.0057	[44] 0.0073	[31] 0.0106	[33] 0.0166		[3] 0.0000 $N^{0.481}$ [62]
57	VISIONLABS-4	[9] 0.0010	[9] 0.0011				[44] 0.0002 $N^{0.103}$ [11]	[4] 0.0025	[4] 0.0027	[4] 0.0030	[9] 0.0039	[14] 0.0059	[4] 0.0000 $N^{0.290}$ [54]
58	VISIONLABS-5	[8] 0.0009	[7] 0.0010	[7] 0.0012	[7] 0.0016	[7] 0.0026	[7] 0.0000 $N^{0.341}$ [49]	[3] 0.0025	[3] 0.0026	[3] 0.0029	[5] 0.0033	[8] 0.0044	[11] 0.0002 $N^{0.192}$ [40]
59	VOCORD-3	[27] 0.0023	[24] 0.0025	[16] 0.0028	[12] 0.0031		[47] 0.0004 $N^{0.123}$ [13]	[18] 0.0040	[18] 0.0042				[47] 0.0017 $N^{0.063}$ [13]
60	YISHENG-1	[42] 0.0035	[41] 0.0047	[28] 0.0058	[27] 0.0072		[14] 0.0000 $N^{0.325}$ [46]	[49] 0.0069	[47] 0.0082				[26] 0.0005 $N^{0.191}$ [39]
61	YITU-0	[33] 0.0026	[29] 0.0027	[19] 0.0029	[16] 0.0031	[11] 0.0034	[49] 0.0008 $N^{0.090}$ [8]	[30] 0.0048	[23] 0.0049	[16] 0.0052	[14] 0.0054	[12] 0.0057	[50] 0.0021 $N^{0.060}$ [11]
62	YITU-1	[31] 0.0026	[27] 0.0027	[17] 0.0029	[13] 0.0031	[10] 0.0034	[48] 0.0008 $N^{0.090}$ [9]	[29] 0.0048	[22] 0.0049				[55] 0.0033 $N^{0.029}$ [3]
63	YITU-2	[6] 0.0008	[3] 0.0009	[2] 0.0009	[2] 0.0010	[2] 0.0010	[46] 0.0004 $N^{0.063}$ [4]	[16] 0.0034	[14] 0.0035	[12] 0.0036	[8] 0.0036	[5] 0.0037	[52] 0.0024 $N^{0.027}$ [1]
64	YITU-3	[17] 0.0018	[14] 0.0018				[51] 0.0011 $N^{0.036}$ [1]	[24] 0.0045	[21] 0.0047	[14] 0.0047	[12] 0.0048	[11] 0.0049	[54] 0.0031 $N^{0.029}$ [2]

Table 15: Effect of N on FNIR at rank 50, for five enrollment population sizes, N. The left five columns apply for consolidated enrollment of a variable number of lifetime images from each subject. The right five columns apply for enrollment of one recent image. Missing entries usually apply because another algorithm from the same developer was run instead. Some developers are missing because less accurate algorithms were not run on galleries with $N > 1\,600\,000$. Throughout blue superscripts indicate the rank of the algorithm for that column, and yellow highlighting indicates the most accurate value.

MISSES BELOW THRESHOLD, T		ENROL LIFETIME					ENROL MOST RECENT				
FNIR(N, T> 0, R >L)		DATASET: FRVT 2018					DATASET: FRVT 2018				
#	ALGORITHM	N=0.64M	N=1.6M	N=3.0M	N=6.0M	N=12.0M	N=0.64M	N=1.6M	N=3.0M	N=6.0M	N=12.0M
1	3DIVI-3	[80] 0.3000	[66] 0.3499	[43] 0.3859	[40] 0.4344		[84] 0.3550	[84] 0.4023			
2	ALCHERA-0	[42] 0.0852	[43] 0.1105	[31] 0.1361	[29] 0.1913		[50] 0.1128	[50] 0.1405			
3	AWARE-3	[41] 0.0846	[39] 0.0991	[28] 0.1148	[24] 0.1459		[49] 0.1122	[49] 0.1306	[33] 0.1471	[30] 0.1793	[25] 0.2395
4	AYONIX-0	[110] 0.8262	[77] 0.8490	[48] 0.8640	[43] 0.8809		[116] 0.7795	[114] 0.8114			
5	CAMVI-3	[16] 0.0281	[20] 0.0509	[16] 0.0680	[28] 0.1871		[18] 0.0413	[26] 0.0736			
6	COGENT-0	[20] 0.0387	[19] 0.0434	[13] 0.0523	[13] 0.0784	[7] 0.1559	[24] 0.0455	[21] 0.0557	[17] 0.0734	[18] 0.1194	[18] 0.2029
7	COGENT-1	[31] 0.0598	[21] 0.0513				[23] 0.0455	[20] 0.0557	[18] 0.0734	[17] 0.1194	[17] 0.2029
8	COGNITEC-0	[45] 0.0989	[44] 0.1256				[51] 0.1345	[51] 0.1626	[34] 0.1892	[31] 0.2205	[30] 0.2859
9	COGNITEC-1	[30] 0.0597	[31] 0.0777	[22] 0.0946	[21] 0.1315	[21] 0.2552	[37] 0.0832	[37] 0.1045	[27] 0.1244	[24] 0.1561	[23] 0.2338
10	DERMALOG-4	[84] 0.3405	[69] 0.3892	[45] 0.4181	[41] 0.4533		[90] 0.4380	[89] 0.4813			
11	EVERAI-0	[24] 0.0460	[30] 0.0676				[31] 0.0681	[34] 0.0921	[25] 0.1223		
12	EVERAI-1	[12] 0.0255	[15] 0.0360				[13] 0.0383	[16] 0.0518	[14] 0.0686		
13	EYEDEA-3	[79] 0.2911	[64] 0.3283	[42] 0.3673	[39] 0.4154		[83] 0.3498	[81] 0.3893			
14	GLORY-1	[66] 0.2160	[55] 0.2447	[37] 0.2618	[34] 0.2884		[76] 0.2790	[74] 0.3067			
15	HIK-2	[50] 0.1104	[48] 0.1363	[32] 0.1610	[30] 0.2061	[24] 0.3067	[46] 0.0985	[47] 0.1212			
16	HIK-3	[43] 0.0885	[42] 0.1097				[38] 0.0853	[38] 0.1054	[26] 0.1228	[23] 0.1552	[26] 0.2500
17	HIK-4	[40] 0.0839	[41] 0.1031	[29] 0.1225	[27] 0.1518	[22] 0.2618	[36] 0.0821	[35] 0.1013	[24] 0.1173	[22] 0.1498	[27] 0.2503
18	IDEMIA-0	[33] 0.0645	[32] 0.0802	[23] 0.0986	[20] 0.1237	[15] 0.1872	[41] 0.0920	[41] 0.1135	[30] 0.1332	[27] 0.1628	[20] 0.2208
19	IDEMIA-1	[18] 0.0304	[16] 0.0377	[11] 0.0465	[8] 0.0623	[8] 0.1578	[19] 0.0444	[18] 0.0540	[12] 0.0647	[10] 0.0856	[9] 0.1618
20	IDEMIA-2	[23] 0.0453	[23] 0.0564	[14] 0.0668	[14] 0.0896	[11] 0.1706	[21] 0.0449	[19] 0.0543			
21	IDEMIA-3	[8] 0.0238	[8] 0.0308				[12] 0.0373	[12] 0.0497	[20] 0.0927	[32] 0.2887	[32] 0.4442
22	IDEMIA-4	[7] 0.0223	[5] 0.0276	[3] 0.0338	[3] 0.0478	[5] 0.1556	[7] 0.0326	[7] 0.0399	[7] 0.0472	[7] 0.0644	[11] 0.1659
23	IMAGUS-2	[105] 0.6616	[75] 0.7143	[47] 0.7503	[42] 0.7867		[111] 0.7092	[110] 0.7510			
24	INCODE-1	[54] 0.1400	[51] 0.1796	[35] 0.2159	[33] 0.2741		[58] 0.1763	[58] 0.2143			
25	ISYSTEMS-0	[28] 0.0485	[28] 0.0633	[20] 0.0795	[18] 0.1057	[16] 0.2072	[33] 0.0707	[33] 0.0912			
26	ISYSTEMS-1	[26] 0.0480	[27] 0.0627	[19] 0.0784	[17] 0.1054	[17] 0.2081	[32] 0.0702	[31] 0.0903			
27	ISYSTEMS-2	[21] 0.0394	[22] 0.0545	[15] 0.0679			[28] 0.0612	[28] 0.0814	[22] 0.1006	[21] 0.1405	[24] 0.2374
28	MEGVII-0	[39] 0.0822	[40] 0.1023	[30] 0.1228	[25] 0.1489	[19] 0.2348	[40] 0.0895	[40] 0.1086	[29] 0.1287	[26] 0.1606	[21] 0.2288
29	MICROFOCUS-3	[117] 0.9002	[80] 0.9213	[50] 0.9342			[120] 0.9119	[118] 0.9310			
30	MICROSOFT-0	[5] 0.0208	[6] 0.0292	[4] 0.0361	[4] 0.0536	[4] 0.1502	[8] 0.0329	[9] 0.0443	[8] 0.0544	[9] 0.0767	[12] 0.1733
31	MICROSOFT-1	[6] 0.0214	[7] 0.0299	[5] 0.0373	[5] 0.0542	[9] 0.1585	[10] 0.0339	[10] 0.0449			
32	MICROSOFT-2	[10] 0.0252	[11] 0.0345	[6] 0.0425	[6] 0.0600	[6] 0.1558	[14] 0.0387	[14] 0.0503			
33	MICROSOFT-3	[4] 0.0133	[4] 0.0193				[6] 0.0223	[6] 0.0304	[6] 0.0384	[6] 0.0570	[7] 0.1603
34	MICROSOFT-4	[3] 0.0128	[3] 0.0179	[2] 0.0241	[2] 0.0405	[10] 0.1628	[5] 0.0209	[5] 0.0288	[5] 0.0360	[5] 0.0550	[6] 0.1576
35	NEC-0	[27] 0.0483	[25] 0.0604	[18] 0.0726	[16] 0.0989	[20] 0.2378	[29] 0.0662	[29] 0.0815	[21] 0.0961	[19] 0.1199	[16] 0.1994
36	NEC-1	[36] 0.0711	[36] 0.0899				[39] 0.0889	[39] 0.1081	[28] 0.1276	[25] 0.1565	[22] 0.2311
37	NEUROTECHNOLOGY-3	[100] 0.5809	[74] 0.6390				[107] 0.5959	[106] 0.6649	[36] 0.7217	[34] 0.7852	[33] 0.8336
38	NEUROTECHNOLOGY-4	[22] 0.0427	[24] 0.0575	[17] 0.0711	[15] 0.0954	[14] 0.1845	[25] 0.0493	[24] 0.0656	[19] 0.0810	[16] 0.1167	[19] 0.2138
39	NTECHLAB-0	[29] 0.0518	[29] 0.0666	[21] 0.0850	[19] 0.1158		[30] 0.0677	[30] 0.0830	[23] 0.1029	[20] 0.1306	[15] 0.1948
40	NTECHLAB-1	[32] 0.0634	[33] 0.0818	[24] 0.1006	[23] 0.1337	[18] 0.2162	[35] 0.0803	[36] 0.1021			
41	NTECHLAB-3	[19] 0.0329	[18] 0.0434				[20] 0.0445	[22] 0.0561	[15] 0.0699	[14] 0.0933	[8] 0.1609
42	NTECHLAB-4	[11] 0.0253	[9] 0.0337	[7] 0.0433	[12] 0.0692	[13] 0.1845	[9] 0.0337	[8] 0.0431	[9] 0.0545	[8] 0.0749	[5] 0.1528
43	RANKONE-0	[55] 0.1485	[50] 0.1788	[36] 0.2210	[35] 0.3260	[26] 0.4758	[60] 0.1899	[59] 0.2192	[35] 0.2635	[33] 0.2992	[31] 0.4301
44	RANKONE-1	[51] 0.1211	[49] 0.1549	[34] 0.1804	[32] 0.2371	[25] 0.3530	[54] 0.1542	[52] 0.1683			
45	RANKONE-2	[38] 0.0744	[38] 0.0943				[48] 0.0998	[44] 0.1200	[32] 0.1382	[29] 0.1744	[29] 0.2636
46	RANKONE-3	[37] 0.0744	[37] 0.0943	[27] 0.1120	[26] 0.1490	[23] 0.2946	[47] 0.0998	[43] 0.1200	[31] 0.1382	[28] 0.1744	[28] 0.2636
47	REALNETWORKS-0	[64] 0.2098	[57] 0.2476	[39] 0.2837			[63] 0.2003	[62] 0.2362			
48	SHAMAN-3	[87] 0.3506	[70] 0.3921	[46] 0.4295			[88] 0.4179	[87] 0.4527			
49	SIAT-1	[74] 0.2695	[60] 0.2727	[38] 0.2758			[2] 0.0160	[1] 0.0201	[2] 0.0260	[1] 0.0380	[1] 0.1069
50	SIAT-2	[68] 0.2198	[54] 0.2239				[4] 0.0179	[4] 0.0242	[4] 0.0301	[4] 0.0434	[4] 0.1377
51	TEVIAN-4	[35] 0.0685	[35] 0.0878	[26] 0.1029			[43] 0.0952	[45] 0.1201			
52	TIGER-0	[78] 0.2859	[65] 0.3361	[41] 0.3659	[38] 0.4139		[82] 0.3452	[82] 0.3921			
53	TONGYITRANS-1	[34] 0.0659	[34] 0.0835	[25] 0.1017	[22] 0.1328		[26] 0.0545	[25] 0.0693			
54	VD-0	[114] 0.8686	[79] 0.9048	[49] 0.9242	[44] 0.9381		[118] 0.8892	[117] 0.9171			
55	VIGILANTSOLUTIONS-3	[81] 0.3061	[67] 0.3568	[44] 0.3861	[36] 0.3861		[86] 0.3648	[86] 0.4097			
56	VISIONLABS-3	[13] 0.0260	[12] 0.0347	[10] 0.0444	[11] 0.0678		[16] 0.0394	[15] 0.0506	[11] 0.0629	[13] 0.0902	
57	VISIONLABS-4	[17] 0.0294	[17] 0.0402				[22] 0.0452	[23] 0.0604	[16] 0.0733	[15] 0.0982	[13] 0.1893
58	VISIONLABS-5	[9] 0.0250	[13] 0.0353	[9] 0.0441	[9] 0.0628	[12] 0.1727	[17] 0.0396	[17] 0.0531	[13] 0.0654	[12] 0.0878	[14] 0.1894
59	VOCORD-3	[44] 0.0969	[45] 0.1295	[33] 0.1627	[31] 0.2361		[45] 0.0973	[48] 0.1258			
60	YISHENG-1	[73] 0.2539	[61] 0.3002	[40] 0.3366	[37] 0.3892		[78] 0.3026	[76] 0.3483			
61	YITU-0	[15] 0.0279	[14] 0.0358	[12] 0.0468	[10] 0.0636	[3] 0.1389	[15] 0.0388	[13] 0.0502	[10] 0.0622	[11] 0.0862	[10] 0.1621
62	YITU-1	[14] 0.0261	[10] 0.0341	[8] 0.0434	[7] 0.0611	[2] 0.1361	[11] 0.0366	[11] 0.0472			
63	YITU-2	[1] 0.0096	[1] 0.0133	[1] 0.0174	[1] 0.0274	[1] 0.1180	[1] 0.0156	[2] 0.0204	[1] 0.0258	[2] 0.0382	[2] 0.1241
64	YITU-3	[2] 0.0103	[2] 0.0139				[3] 0.0165	[3] 0.0213	[3] 0.0266	[3] 0.0389	[3] 0.1248

Table 16: Effect of N. Values are threshold-based FNIR, *at* FPIR *= 0.001 for five enrollment population sizes, N. The left six columns apply for enrollment of a variable number of images per subject. The right six columns apply for enrollment of one image. Missing entries usually apply because another algorithm from the same developer was run instead. Some developers are missing because less accurate algorithms were not run on galleries with $N \geq 3\,000\,000$. Throughout blue superscripts indicate the rank of the algorithm for that column. Caution: The Power-low models are mostly intended to draw attention to the kind of behavior, not as a model to be used prediction*

MISSES BELOW		INVESTIGATION: RANK ONE MISS RATE, FNIR(N, 0, 1)					IDENTIFICATION: HIGH T → FPIR = 0.01, FNIR(N, T, L)					FAILURE TO EXTRACT FEATURES				
THRESHOLD, T		N=1.6M	N=1.6M	N=1.6M	N=0.7M	N=1.1M	N=1.6M	N=1.6M	N=1.6M	N=0.7M	N=1.1M	N=1.6M	N=0.6M	N=0.6M	N=0.7M	N=16K
#	ALGORITHM	FRVT-14	FRVT-18	WEBCAM	FRPC	WILD	FRVT-14	FRVT-18	WEBCAM	FRPC	WILD^{+}	FRVT-14	FRVT-18	WEBCAM	FRPC	WILD
1	3DIVI-0	^{56}0.026	^{63}0.034	^{53}0.086	^{60}0.191	^{30}0.071	^{66}0.103	^{72}0.160	^{57}0.302	^{54}0.435	^{32}0.095	0.004	0.003	0.007	0.011	0.013
2	3DIVI-1	^{60}0.028	^{64}0.038		^{63}0.217	^{33}0.074	^{65}0.103	^{73}0.160		^{55}0.435	^{33}0.095	0.004	0.003		0.011	0.013
3	3DIVI-2	^{63}0.030	^{68}0.040		^{65}0.225	^{35}0.076	^{67}0.105	^{74}0.164		^{58}0.439	^{34}0.096	0.004	0.003		0.011	0.013
4	3DIVI-3	^{73}0.053	^{88}0.086	^{66}0.206	^{79}0.328	^{49}0.094	^{77}0.183	^{89}0.284	^{70}0.497	^{63}0.508	^{51}0.136	0.003	0.002	0.005	0.007	0.009
5	3DIVI-4		^{47}0.020	^{44}0.062				^{53}0.096	^{51}0.237				0.002	0.005		
6	ALCHERA-0	^{45}0.021	^{44}0.019	^{37}0.047	^{42}0.132	^{46}0.092	^{40}0.047	^{48}0.073	^{36}0.146	^{29}0.208	^{25}0.089	0.010	0.006	0.014	0.093	0.030
7	ALCHERA-1		^{126}0.987	^{91}1.000				^{125}0.999	^{109}1.000				0.006	0.013		
8	AWARE-0	^{72}0.053	^{84}0.064	^{61}0.138	^{74}0.286	^{83}0.588	^{61}0.092	^{66}0.128	^{52}0.253	^{52}0.421	^{83}0.587	0.013	0.006	0.054	0.129	0.143
9	AWARE-1	^{69}0.043	^{80}0.059		^{72}0.276	^{82}0.580	^{57}0.084	^{65}0.127		^{53}0.424	^{81}0.580	0.013	0.006		0.129	0.143
10	AWARE-2	^{74}0.056	^{81}0.060		^{75}0.287		^{60}0.090	^{64}0.120		^{51}0.415		0.013	0.006		0.129	0.143
11	AWARE-3	^{53}0.025	^{62}0.033	^{54}0.090	^{48}0.165	^{81}0.503	^{43}0.056	^{51}0.085	^{46}0.204	^{41}0.305	^{80}0.505	0.003	0.004	0.003	0.027	0.014
12	AWARE-4		^{85}0.070	^{65}0.176				^{77}0.177	^{63}0.375				0.003	0.003		
13	AYONIX-0	^{101}0.346	^{119}0.452	^{87}0.685	^{93}0.626	^{80}0.400	^{101}0.624	^{118}0.725	^{86}0.892	^{87}0.815	^{82}0.586	0.016	0.010	0.031	0.082	0.068
14	CAMVI-1	^{93}0.143	^{111}0.227	^{79}0.337	^{81}0.349	^{61}0.148	^{94}0.409	^{111}0.549	^{80}0.648	^{85}0.771	^{61}0.196	0.006	0.005	0.009	0.050	0.058
15	CAMVI-2	^{79}0.076	^{95}0.129		^{69}0.243	^{57}0.130	^{85}0.265	^{99}0.402		^{72}0.608	^{56}0.157	0.006	0.005		0.050	0.058
16	CAMVI-3	^{67}0.035	^{79}0.054	^{55}0.090	^{47}0.160	^{60}0.139	^{36}0.038	^{41}0.060	^{31}0.108	^{21}0.179	^{44}0.130	0.008	0.006	0.013	0.072	0.074
17	COGENT-0	^{29}0.011	^{33}0.013	^{35}0.046	^{66}0.232	^{47}0.093	^{21}0.021	^{23}0.032	^{26}0.100	^{43}0.318	^{40}0.110	0.000	0.000	0.000	0.000	0.000
18	COGENT-1	^{28}0.011	^{32}0.013	^{34}0.046			^{20}0.021	^{22}0.032	^{24}0.100			0.000	0.000	0.000		
19	COGNITEC-0	^{44}0.020	^{59}0.029	^{41}0.059			^{42}0.054	^{55}0.098	^{44}0.200			0.002	0.003	0.002		
20	COGNITEC-1	^{34}0.013	^{40}0.014	^{27}0.034	^{32}0.087	^{32}0.074	^{32}0.031	^{36}0.055	^{35}0.135	^{37}0.296	^{18}0.072	0.002	0.003	0.002	0.037	0.025
21	DERMALOG-0	^{78}0.075	^{96}0.131	^{70}0.218	^{68}0.237	^{34}0.075	^{79}0.233	^{94}0.364	^{75}0.528	^{61}0.492	^{38}0.104	0.004	0.003	0.002	0.011	0.020
22	DERMALOG-1	^{83}0.096	^{98}0.156		^{71}0.264	^{44}0.089	^{86}0.279	^{101}0.405		^{65}0.537	^{48}0.131	0.004	0.003		0.011	0.020
23	DERMALOG-2	^{80}0.079	^{97}0.138		^{67}0.236	^{37}0.076	^{81}0.248	^{96}0.378		^{62}0.507	^{39}0.105	0.004	0.003		0.011	0.020
24	DERMALOG-3		^{93}0.128	^{69}0.217				^{93}0.362	^{74}0.526				0.002	0.002		
25	DERMALOG-4	^{75}0.071	^{92}0.127	^{68}0.215	^{64}0.224	^{26}0.066	^{78}0.228	^{92}0.360	^{72}0.526	^{56}0.437	^{30}0.095	0.003	0.001	0.002	0.004	0.013
26	EVERAI-0		^{48}0.021	^{30}0.038				^{32}0.047	^{25}0.100				0.000	0.000		
27	EVERAI-1	^{11}0.004	^{9}0.006	^{10}0.020	^{7}0.034	^{84}0.928	^{11}0.012	^{10}0.023	^{13}0.074	^{6}0.100	^{84}0.927	0.000	0.000	0.000	0.000	0.000
28	EYEDEA-0	^{99}0.201	^{115}0.300	^{82}0.443	^{84}0.369	^{58}0.131	^{99}0.549	^{117}0.679	^{83}0.783	^{84}0.757	^{68}0.249	0.001	0.001	0.003	0.008	0.008
29	EYEDEA-1	^{86}0.109	^{105}0.198		^{49}0.172	^{31}0.072	^{88}0.324	^{104}0.480		^{64}0.534	^{47}0.131	0.001	0.001		0.008	0.008
30	EYEDEA-2	^{87}0.110	^{106}0.200		^{55}0.184	^{28}0.070	^{89}0.327	^{107}0.490		^{67}0.548	^{45}0.130	0.001	0.000		0.007	0.005
31	EYEDEA-3	^{71}0.044	^{87}0.082	^{62}0.148	^{37}0.100	^{23}0.064	^{74}0.154	^{87}0.267	^{64}0.404	^{38}0.299	^{26}0.091	0.001	0.001	0.003	0.008	0.008
32	GLORY-0		^{102}0.180	^{76}0.320				^{90}0.297	^{67}0.470				0.011	0.013		
33	GLORY-1	^{85}0.109	^{94}0.129	^{73}0.267	^{88}0.453	^{75}0.315	^{76}0.182	^{84}0.238	^{65}0.448	^{66}0.547	^{73}0.353	0.014	0.011	0.013	0.207	0.114
34	GORILLA-0				^{77}0.293	^{87}0.994				^{80}0.708	^{87}0.994	0.001	0.001		0.004	0.008
35	GORILLA-1		^{82}0.063	^{56}0.095		^{15}0.057		^{85}0.248	^{59}0.314		^{20}0.076		0.001	0.001		0.007
36	HBINNO-0	^{98}0.191	^{114}0.275		^{87}0.437	^{78}0.335	^{98}0.498	^{115}0.632		^{93}0.975	^{75}0.411	0.022	0.007		0.043	0.151
37	HIK-0	^{57}0.026	^{55}0.024	^{25}0.033	^{14}0.042	^{62}0.153	^{41}0.049	^{47}0.070	^{28}0.103	^{18}0.160	^{55}0.155	0.013	0.010	0.004	0.017	0.027
38	HIK-1	^{76}0.073	^{43}0.017		^{12}0.039	^{65}0.162	^{62}0.095	^{45}0.067		^{16}0.159	^{58}0.166	0.002	0.003		0.008	0.013
39	HIK-2	^{33}0.013	^{42}0.017		^{8}0.035	^{50}0.094	^{35}0.037	^{46}0.067		^{15}0.158	^{37}0.103	0.002	0.001		0.001	0.008
40	HIK-3		^{39}0.014	^{21}0.027				^{40}0.060	^{29}0.105				0.000	0.000		
41	HIK-4	^{22}0.008	^{37}0.014	^{20}0.027	^{10}0.037	^{18}0.062	^{29}0.027	^{38}0.056	^{27}0.101	^{13}0.143	^{19}0.075	0.001	0.000	0.000	0.002	0.008
42	IDEMIA-0	^{24}0.008	^{28}0.011	^{28}0.034	^{36}0.096	^{68}0.166	^{34}0.036	^{42}0.062	^{37}0.156	^{39}0.302	^{71}0.288	0.003	0.003	0.000	0.003	0.002
43	IDEMIA-1	^{25}0.008	^{30}0.012		^{35}0.095	^{64}0.157	^{19}0.021	^{21}0.031		^{24}0.191	^{63}0.205	0.003	0.003		0.003	0.002
44	IDEMIA-2	^{32}0.013	^{31}0.013		^{54}0.183	^{71}0.198	^{25}0.023	^{24}0.032		^{32}0.242	^{66}0.242	0.008	0.005		0.146	0.031
45	IDEMIA-3	^{30}0.011	^{24}0.010	^{26}0.034			^{22}0.021	^{14}0.024	^{16}0.079			0.000	0.000	0.000		
46	IDEMIA-4	^{23}0.008	^{21}0.009	^{24}0.032	^{27}0.086	^{10}0.051	^{16}0.019	^{13}0.024	^{15}0.079	^{20}0.177	^{15}0.064	0.000	0.000	0.000	0.001	0.003
47	IMAGUS-0	^{100}0.216	^{116}0.305	^{84}0.482	^{90}0.496	^{73}0.222	^{97}0.468	^{114}0.608	^{82}0.779	^{83}0.746	^{72}0.311	0.011	0.009	0.013	0.089	0.049
48	IMAGUS-2	^{95}0.145	^{109}0.222	^{74}0.301	^{83}0.353	^{63}0.154	^{95}0.410	^{112}0.566	^{79}0.645	^{76}0.652	^{70}0.252	0.004	0.004	0.008	0.052	0.023
49	IMAGUS-3		^{118}0.358	^{85}0.513				^{116}0.670	^{84}0.809				0.004	0.008		
50	INCODE-0		^{78}0.051	^{58}0.100				^{81}0.201	^{58}0.304				0.001	0.004		
51	INCODE-1	^{31}0.012	^{45}0.019	^{36}0.046	^{29}0.086	^{12}0.052	^{48}0.061	^{58}0.114	^{42}0.198	^{31}0.230	^{11}0.062	0.003	0.001	0.004	0.021	0.009
52	INNOVATRICS-0	^{62}0.029	^{70}0.042	^{50}0.076	^{43}0.134	^{69}0.188	^{64}0.100	^{76}0.165	^{53}0.258	^{49}0.400	^{67}0.245	0.002	0.002	0.008	0.012	0.093
53	INNOVATRICS-1	^{61}0.029	^{69}0.042		^{44}0.134	^{70}0.193	^{63}0.100	^{75}0.165		^{50}0.401	^{65}0.221	0.002	0.002		0.012	0.093
54	INNOVATRICS-2		^{76}0.048	^{49}0.074				^{71}0.142	^{47}0.209				0.000	0.001		
55	INNOVATRICS-3	^{37}0.015	^{60}0.029	^{39}0.055	^{33}0.089	^{29}0.071	^{54}0.068	^{68}0.134	^{45}0.203		^{22}0.081	0.000	0.000	0.001	0.003	0.007
56	ISYSTEMS-0	^{48}0.023	^{36}0.014	^{31}0.038	^{58}0.187	^{67}0.163	^{38}0.040	^{30}0.047	^{32}0.110	^{35}0.285	^{60}0.169	0.003	0.003	0.013	0.033	0.065
57	ISYSTEMS-1	^{49}0.023	^{35}0.014		^{57}0.187	^{66}0.162	^{37}0.040	^{28}0.047		^{36}0.286	^{59}0.169	0.003	0.003		0.033	0.065
58	ISYSTEMS-2	^{20}0.008	^{19}0.009	^{19}0.026	^{20}0.061	^{8}0.049	^{18}0.020	^{26}0.035	^{17}0.080	^{14}0.146	^{8}0.051	0.003	0.002	0.002	0.009	0.009
59	MEGVII-0	^{12}0.004	^{22}0.009	^{5}0.017	^{13}0.041	^{17}0.061	^{28}0.025	^{39}0.058	^{9}0.067	^{30}0.227	^{29}0.094	0.000	0.000	0.000	0.001	0.005
60	MICROFOCUS-0	^{105}0.472	^{123}0.597	^{90}0.782	^{96}0.760	^{76}0.316	^{105}0.793	^{123}0.867	^{89}0.950	^{91}0.924	^{77}0.434	0.011	0.005	0.030	0.094	0.065
61	MICROFOCUS-1	^{104}0.472	^{124}0.597		^{95}0.760	^{77}0.316	^{104}0.793	^{122}0.867		^{90}0.924	^{78}0.434	0.011	0.005		0.094	0.065
62	MICROFOCUS-2	^{106}0.508	^{125}0.627		^{97}0.774	^{79}0.342	^{106}0.796	^{124}0.870		^{92}0.925	^{79}0.447	0.011	0.005		0.094	0.065
63	MICROFOCUS-3	^{103}0.469	^{122}0.595	^{89}0.781	^{94}0.753	^{74}0.279	^{103}0.791	^{121}0.866	^{88}0.948	^{89}0.904	^{76}0.412	0.003	0.001	0.005	0.016	0.014
64	MICROFOCUS-4		^{121}0.577	^{88}0.758				^{126}0.999	^{87}0.940				0.001	0.005		

Table 17: Miss rates by dataset. At left, rank 1 miss rates relevant to investigations; at right, with threshold set to target FPIR = 0.01 for higher volume, low prior, uses. $^{+}$For the WILD set, FPIR = 0.1 Yellow indicates most accurate algorithm. Green means better than NISTIR 8009 in 2014-04 for NEC CORP E30C (0.041 and 0.063, respectively)on identical mugshots, and than NTechLab / Yitu in FRPC NISTIR 8197 in 2017-11 (values 0.031 and 0.133) for travel concourse frames. Throughout blue superscripts indicate the rank of the algorithm for that column.

MISSES BELOW THRESHOLD, T		INVESTIGATION: RANK ONE MISS RATE, FNIR(N, 0, 1)					IDENTIFICATION: HIGH T → FPIR = 0.01, FNIR(N, T, L)					FAILURE TO EXTRACT FEATURES				
		N=1.6M	N=1.6M	N=1.6M	N=0.7M	N=1.1M	N=1.6M	N=1.6M	N=1.6M	N=0.7M	N=1.1M	N=1.6M	N=0.6M	N=0.6M	N=0.7M	N=16K
#	ALGORITHM	FRVT-14	FRVT-18	WEBCAM	FRPC	WILD	FRVT-14	FRVT-18	WEBCAM	FRPC	WILD+	FRVT-14	FRVT-18	WEBCAM	FRPC	WILD
65	MICROSOFT-0	[7] 0.003	[11] 0.006	[12] 0.021	[21] 0.061	[24] 0.065	[7] 0.010	[7] 0.022	[11] 0.071	[28] 0.206	[16] 0.065	0.000	0.000	0.001	0.006	0.019
66	MICROSOFT-1	[6] 0.003	[10] 0.006		[18] 0.052	[20] 0.062	[8] 0.011	[8] 0.022		[27] 0.204	[10] 0.061	0.000	0.000		0.006	0.019
67	MICROSOFT-2	[8] 0.004	[12] 0.006		[19] 0.057	[21] 0.063	[12] 0.013	[16] 0.026		[26] 0.200	[14] 0.063	0.000	0.000		0.006	0.019
68	MICROSOFT-3	[2] 0.002	[2] 0.003	[3] 0.012			[4] 0.007	[6] 0.014	[5] 0.056			0.000	0.000	0.001		
69	MICROSOFT-4	[1] 0.002	[1] 0.003	[2] 0.012	[2] 0.015	[1] 0.039	[1] 0.007	[5] 0.013	[3] 0.053	[2] 0.055	[3] 0.043	0.000	0.000	0.001	0.006	0.004
70	NEC-0	[36] 0.014	[46] 0.020	[32] 0.041	[22] 0.069	[88] 0.999	[31] 0.030	[33] 0.049	[21] 0.093	[9] 0.110	[88] 0.999	0.001	0.001	0.002	0.016	0.064
71	NEC-1	[52] 0.025	[54] 0.024	[40] 0.056			[39] 0.043	[43] 0.063	[34] 0.133			0.005	0.005	0.003		
72	NEUROTECHNOLOGY-0	[64] 0.031	[77] 0.050	[59] 0.104	[39] 0.125	[89] 1.000	[70] 0.110	[80] 0.196	[60] 0.317	[44] 0.332	[89] 1.000	0.004	0.004	0.022	0.050	0.091
73	NEUROTECHNOLOGY-1	[59] 0.028	[75] 0.047		[28] 0.086	[85] 0.954	[68] 0.107	[79] 0.195		[42] 0.306	[85] 0.953	0.001	0.001		0.018	0.028
74	NEUROTECHNOLOGY-2	[58] 0.028	[74] 0.047		[25] 0.082	[86] 0.983	[69] 0.107	[78] 0.195		[40] 0.304	[86] 0.983	0.001	0.001		0.013	0.028
75	NEUROTECHNOLOGY-3	[43] 0.019	[57] 0.025	[33] 0.042			[45] 0.060	[56] 0.101	[38] 0.164			0.001	0.000	0.001		
76	NEUROTECHNOLOGY-4	[35] 0.014	[16] 0.008	[9] 0.020	[31] 0.087	[45] 0.090	[26] 0.024	[19] 0.030	[12] 0.073	[17] 0.159	[42] 0.122	0.001	0.000	0.001	0.009	0.007
77	NTECHLAB-0	[15] 0.006	[29] 0.012	[23] 0.031	[6] 0.026	[3] 0.041	[24] 0.023	[31] 0.047	[30] 0.105	[7] 0.100	[2] 0.043	0.001	0.000	0.001	0.001	0.005
78	NTECHLAB-1	[19] 0.008	[38] 0.014		[11] 0.038	[6] 0.045	[30] 0.027	[37] 0.056		[8] 0.110	[7] 0.049	0.001	0.000		0.001	0.005
79	NTECHLAB-3		[17] 0.008	[17] 0.023				[20] 0.030	[14] 0.075				0.000	0.000		
80	NTECHLAB-4	[10] 0.004	[13] 0.007	[7] 0.019	[5] 0.024	[5] 0.043	[9] 0.011	[12] 0.024	[8] 0.065	[4] 0.070	[6] 0.048	0.000	0.000	0.000	0.002	0.003
81	RANKONE-0	[70] 0.043	[73] 0.045	[60] 0.117	[78] 0.302	[55] 0.114	[59] 0.090	[67] 0.129	[56] 0.291	[71] 0.584	[57] 0.161	0.000	0.000	0.000	0.002	0.000
82	RANKONE-1	[65] 0.032	[56] 0.025		[56] 0.185	[39] 0.077	[55] 0.073	[52] 0.087		[60] 0.468	[36] 0.102	0.000	0.000		0.002	0.000
83	RANKONE-2	[55] 0.025	[50] 0.022	[48] 0.071			[47] 0.060	[50] 0.073	[41] 0.190			0.000	0.000	0.000		
84	RANKONE-3	[54] 0.025	[49] 0.022	[46] 0.068	[61] 0.191	[40] 0.078	[46] 0.060	[49] 0.073	[40] 0.187	[45] 0.364	[31] 0.095	0.000	0.000	0.000	0.002	0.000
85	REALNETWORKS-0	[46] 0.023	[72] 0.043	[52] 0.078	[30] 0.087	[36] 0.076	[56] 0.080	[70] 0.140	[49] 0.209	[23] 0.184	[23] 0.084	0.001	0.001	0.000	0.001	0.004
86	REALNETWORKS-1		[71] 0.043	[51] 0.078				[69] 0.140	[48] 0.209				0.001	0.000		
87	SHAMAN-0	[89] 0.119	[100] 0.171	[72] 0.262	[80] 0.338	[56] 0.115	[83] 0.260	[95] 0.370	[71] 0.507	[73] 0.628	[52] 0.146	0.020	0.020	0.011	0.098	0.043
88	SHAMAN-1	[88] 0.118	[101] 0.172		[73] 0.283	[54] 0.113	[87] 0.283	[102] 0.406		[70] 0.576	[54] 0.153	0.020	0.020		0.098	0.043
89	SHAMAN-2	[97] 0.180	[113] 0.262		[82] 0.351	[59] 0.132	[96] 0.444	[113] 0.582		[77] 0.681	[62] 0.201	0.020	0.020		0.098	0.043
90	SHAMAN-3	[82] 0.094	[90] 0.127	[64] 0.172	[70] 0.258	[52] 0.109	[80] 0.244	[91] 0.348	[68] 0.472	[59] 0.465	[49] 0.132	0.020	0.020	0.011	0.097	0.043
91	SHAMAN-4		[110] 0.224	[75] 0.319				[106] 0.490	[78] 0.639				0.020	0.011		
92	SIAT-0	[18] 0.007	[26] 0.010	[14] 0.021	[3] 0.019	[41] 0.078	[27] 0.025	[29] 0.047	[7] 0.064	[5] 0.090	[69] 0.250	0.000	0.000	0.000	0.001	0.008
93	SIAT-1	[9] 0.004	[3] 0.004	[78] 0.333	[1] 0.009	[2] 0.040	[3] 0.007	[1] 0.009	[61] 0.348	[1] 0.033	[1] 0.041	0.000	0.000	0.000	0.001	0.003
94	SIAT-2	[81] 0.081	[4] 0.004	[83] 0.446			[58] 0.084	[2] 0.009	[66] 0.460			0.077	0.000	0.000		
95	SMILART-0	[92] 0.142	[103] 0.193	[77] 0.325	[89] 0.468	[121] 1.000	[92] 0.375	[105] 0.486		[82] 0.717	[121] 1.000	0.015	0.008		0.203	0.121
96	SMILART-1	[94] 0.144	[108] 0.219		[85] 0.398	[110] 1.000	[93] 0.385	[110] 0.505		[79] 0.700	[110] 1.000	0.012	0.021		0.003	0.006
97	SMILART-2	[91] 0.132	[104] 0.195		[86] 0.408	[99] 1.000	[91] 0.375	[108] 0.492		[78] 0.686	[99] 1.000	0.002	0.000		0.008	0.048
98	SYNESIS-0	[84] 0.108	[99] 0.162	[81] 0.361	[92] 0.608		[84] 0.262	[97] 0.378	[77] 0.598	[81] 0.713		0.004	0.002	0.009	0.042	0.081
99	TEVIAN-0	[39] 0.017	[52] 0.022	[45] 0.066	[50] 0.172	[13] 0.054	[51] 0.065	[60] 0.114	[50] 0.227	[46] 0.389	[17] 0.072	0.003	0.002	0.005	0.055	0.007
100	TEVIAN-1	[40] 0.017	[53] 0.022		[52] 0.172	[19] 0.062	[52] 0.065	[61] 0.114		[48] 0.389	[21] 0.078	0.003	0.002		0.055	0.007
101	TEVIAN-2	[42] 0.017	[51] 0.022		[51] 0.172	[48] 0.093	[53] 0.065	[59] 0.114		[47] 0.389	[41] 0.118	0.003	0.002		0.055	0.008
102	TEVIAN-3		[41] 0.017	[38] 0.052				[54] 0.098	[43] 0.198				0.001	0.002		
103	TEVIAN-4	[26] 0.009	[34] 0.013	[29] 0.038	[26] 0.085	[9] 0.050	[33] 0.035	[44] 0.066	[33] 0.115	[25] 0.193	[13] 0.063	0.002	0.001	0.002	0.004	0.005
104	TIGER-0	[66] 0.033	[83] 0.064	[57] 0.095	[23] 0.074	[106] 1.000	[73] 0.151	[86] 0.263	[62] 0.366	[34] 0.256	[106] 1.000	0.001	0.000	0.000	0.001	0.005
105	TIGER-1		[117] 0.308	[80] 0.351				[100] 0.404	[69] 0.487				0.000	0.000		
106	TONGYITRANS-0		[25] 0.010	[16] 0.022				[27] 0.041	[10] 0.069				0.003	0.001		
107	TONGYITRANS-1	[21] 0.008	[23] 0.010	[15] 0.022	[17] 0.049	[53] 0.112	[17] 0.020	[25] 0.035	[6] 0.062	[11] 0.130	[50] 0.134	0.002	0.003	0.001	0.006	0.009
108	VD-0	[102] 0.363	[120] 0.475	[86] 0.551	[91] 0.505	[72] 0.217	[102] 0.733	[120] 0.828	[85] 0.871	[88] 0.819	[74] 0.362	0.012	0.011	0.013	0.075	0.026
109	VIGILANTSOLUTIONS-0	[77] 0.073	[89] 0.125	[67] 0.212	[59] 0.188	[38] 0.076	[82] 0.260	[98] 0.394	[76] 0.557	[68] 0.552	[53] 0.152	0.001	0.000	0.001	0.005	0.003
110	VIGILANTSOLUTIONS-1	[90] 0.120	[107] 0.204		[76] 0.288	[51] 0.103	[90] 0.354	[109] 0.502		[75] 0.651	[64] 0.209	0.001	0.000		0.005	0.003
111	VIGILANTSOLUTIONS-2	[96] 0.159	[112] 0.239		[62] 0.195	[22] 0.064	[100] 0.623	[119] 0.731		[74] 0.639	[43] 0.129	0.001	0.000		0.005	0.003
112	VIGILANTSOLUTIONS-3	[68] 0.038	[86] 0.072	[63] 0.151	[53] 0.175	[25] 0.065	[75] 0.169	[88] 0.283	[73] 0.526	[69] 0.553	[46] 0.131	0.001	0.000	0.001	0.005	0.003
113	VIGILANTSOLUTIONS-4		[91] 0.127	[71] 0.244				[103] 0.424	[81] 0.709				0.000	0.001		
114	VISIONLABS-3	[27] 0.009	[20] 0.009	[22] 0.030	[34] 0.093	[11] 0.051	[14] 0.015	[17] 0.026	[19] 0.091	[33] 0.246	[4] 0.046	0.004	0.002	0.003	0.014	0.014
115	VISIONLABS-4	[4] 0.003	[6] 0.004	[8] 0.020			[10] 0.012	[18] 0.026	[23] 0.097			0.001	0.001	0.001		
116	VISIONLABS-5	[3] 0.003	[5] 0.004	[6] 0.019	[9] 0.036	[4] 0.043	[6] 0.010	[9] 0.022	[18] 0.087	[12] 0.133	[5] 0.046	0.001	0.001	0.001	0.005	0.006
117	VOCORD-0	[51] 0.025	[67] 0.040	[47] 0.068	[41] 0.129		[50] 0.063	[63] 0.116	[39] 0.181	[104] 1.000		0.014	0.015	0.025	0.008	0.019
118	VOCORD-1	[50] 0.025	[66] 0.040		[40] 0.129		[49] 0.062	[62] 0.116		[94] 0.998		0.013	0.015		0.016	0.018
119	VOCORD-2	[47] 0.023	[65] 0.038		[45] 0.144		[44] 0.057	[57] 0.107		[117] 1.000		0.013	0.015		0.016	0.015
120	VOCORD-3	[14] 0.006	[18] 0.008	[18] 0.024	[24] 0.074	[14] 0.057	[23] 0.022	[34] 0.050	[22] 0.093	[10] 0.127	[12] 0.062	0.001	0.001	0.011	0.052	0.006
121	VOCORD-4		[27] 0.010	[13] 0.021				[35] 0.054	[20] 0.093				0.000	0.000		
122	YISHENG-0	[38] 0.016	[58] 0.027	[42] 0.060	[46] 0.145	[27] 0.067	[72] 0.116	[83] 0.209	[55] 0.275	[86] 0.787	[35] 0.100	0.002	0.002	0.005	0.013	0.014
123	YISHENG-1	[41] 0.017	[61] 0.029	[43] 0.060	[38] 0.121	[16] 0.061	[71] 0.115	[82] 0.208	[54] 0.269	[57] 0.438	[24] 0.087	0.002	0.002	0.005	0.013	0.014
124	YITU-0	[17] 0.007	[15] 0.007	[11] 0.020	[16] 0.044	[43] 0.086	[15] 0.016	[15] 0.025	[4] 0.054	[22] 0.182	[28] 0.094	0.002	0.003	0.001	0.006	0.026
125	YITU-1	[16] 0.007	[14] 0.007		[15] 0.042	[42] 0.086	[13] 0.015	[11] 0.023		[19] 0.174	[27] 0.092	0.002	0.003		0.006	0.026
126	YITU-2	[5] 0.003	[7] 0.004	[1] 0.010	[4] 0.019	[7] 0.046	[2] 0.007	[3] 0.011	[1] 0.028	[3] 0.055	[9] 0.051	0.000	0.000	0.000	0.000	0.000
127	YITU-3	[13] 0.005	[8] 0.005	[4] 0.016			[5] 0.009	[4] 0.011	[2] 0.033			0.002	0.003	0.001		

Table 18: Miss rates by dataset. At left, rank 1 miss rates relevant to investigations; at right, with threshold set to target FPIR = 0.01 for higher volume, low prior, uses. +For the WILD set, FPIR = 0.1 Yellow indicates most accurate algorithm. Green means better than NISTIR 8009 in 2014-04 for NEC CORP E30C (0.041 and 0.063, respectively)on identical mugshots, and than NTechLab / Yitu in FRPC NISTIR 8197 in 2017-11 (values 0.031 and 0.133) for travel concourse frames. Throughout blue superscripts indicate the rank of the algorithm for that column.

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

*Figure 19: . **[Mugshot Dataset] Error rate reductions in 2018**. For each FRVT2018 participant, the plot shows accuracy gains between Phase 1 (Feb 2018) and Phase 2 (Jun 2018) according to two metrics: rank one miss rate, FNIR(N, 1, 0), and high threshold, FNIR(N, N, T), set to achieve FPIR = 0.003. The text "Red=" gives the best reduction multiplier for the given metric on the recent enrollment type - a smaller value is better.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

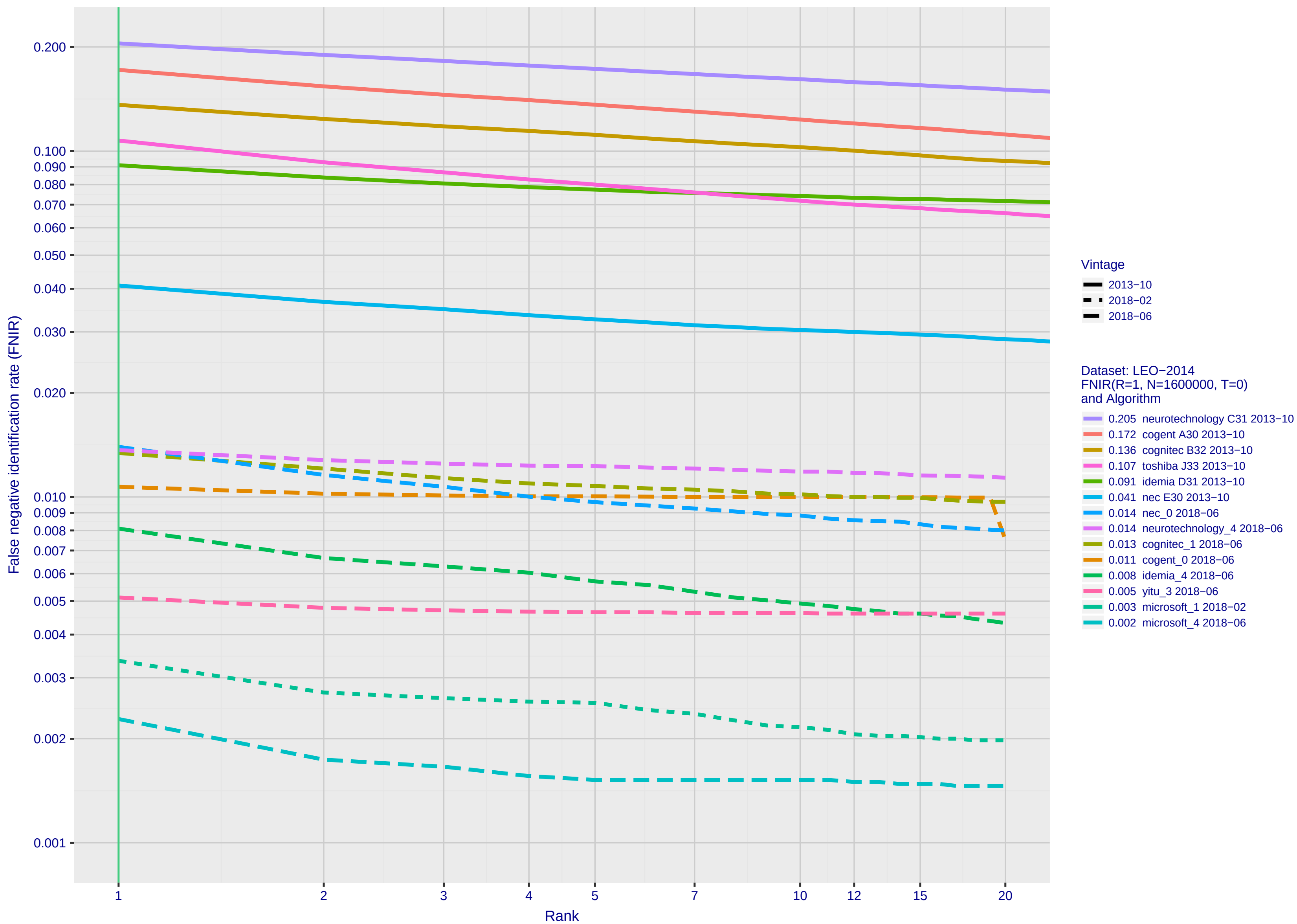

Figure 20: **[Gains 2013-2018]** *On the* LEO *set used in* FRVT*2014, the figure shows investigational miss rates vs. rank for the most accurate algorithms submitted to NIST in October 2013 and in February/June 2018. The reduction in error rates is an order of magnitude. For the most accurate algorithms, miss rates fell approximately twelvefold from 4.1% to 0.34%.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

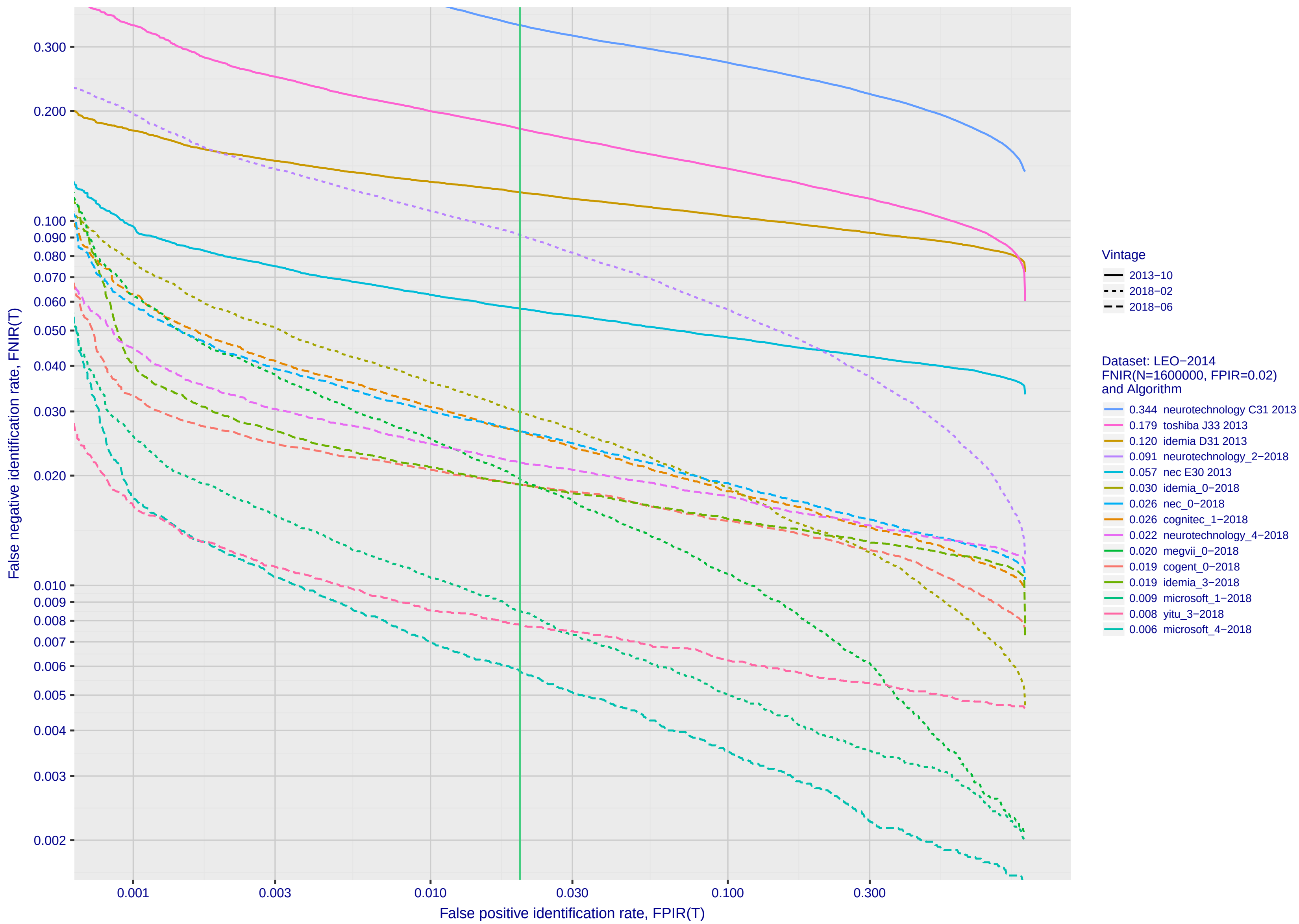

Figure 21: **[Gains 2013-2018]** *On the* LEO *set used in* FRVT*2014, the figure shows identification miss rates vs. false positive rates for the most accurate algorithms submitted to NIST in October 2013 and February/June 2018. The reduction in error rates is not as large as for rank-based miss rates but, for the most accurate algorithms, miss rates fell tenfold from 5.7% to 0.6% at* FPIR = *0.02 as tabulated, and shown along the green vertical line.*

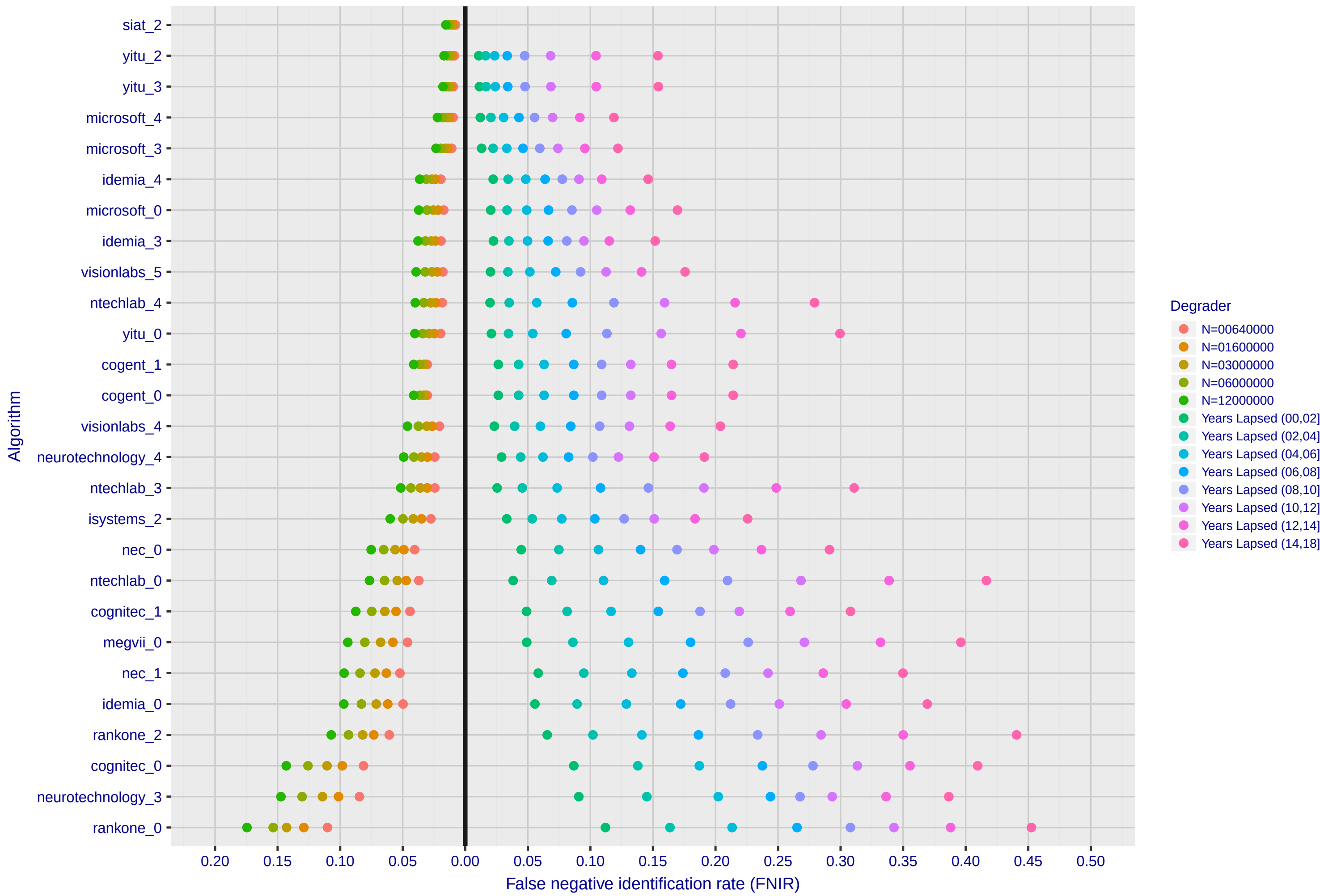

Figure 22: **[FRVT-2018 Mugshot Ageing Dataset] Contrast of ageing and population size dependency..** *The Figure shows, at left, the dependence FNIR(N) for the FRVT-2018, as tabulated in Table 14. At right, is FNIR(N = 3 000 000, ΔT) from Figure:68. Ageing miss rates are computed over all searches binned by number of years between search and initial enrollment. In all cases, FPIR = 0.01.*

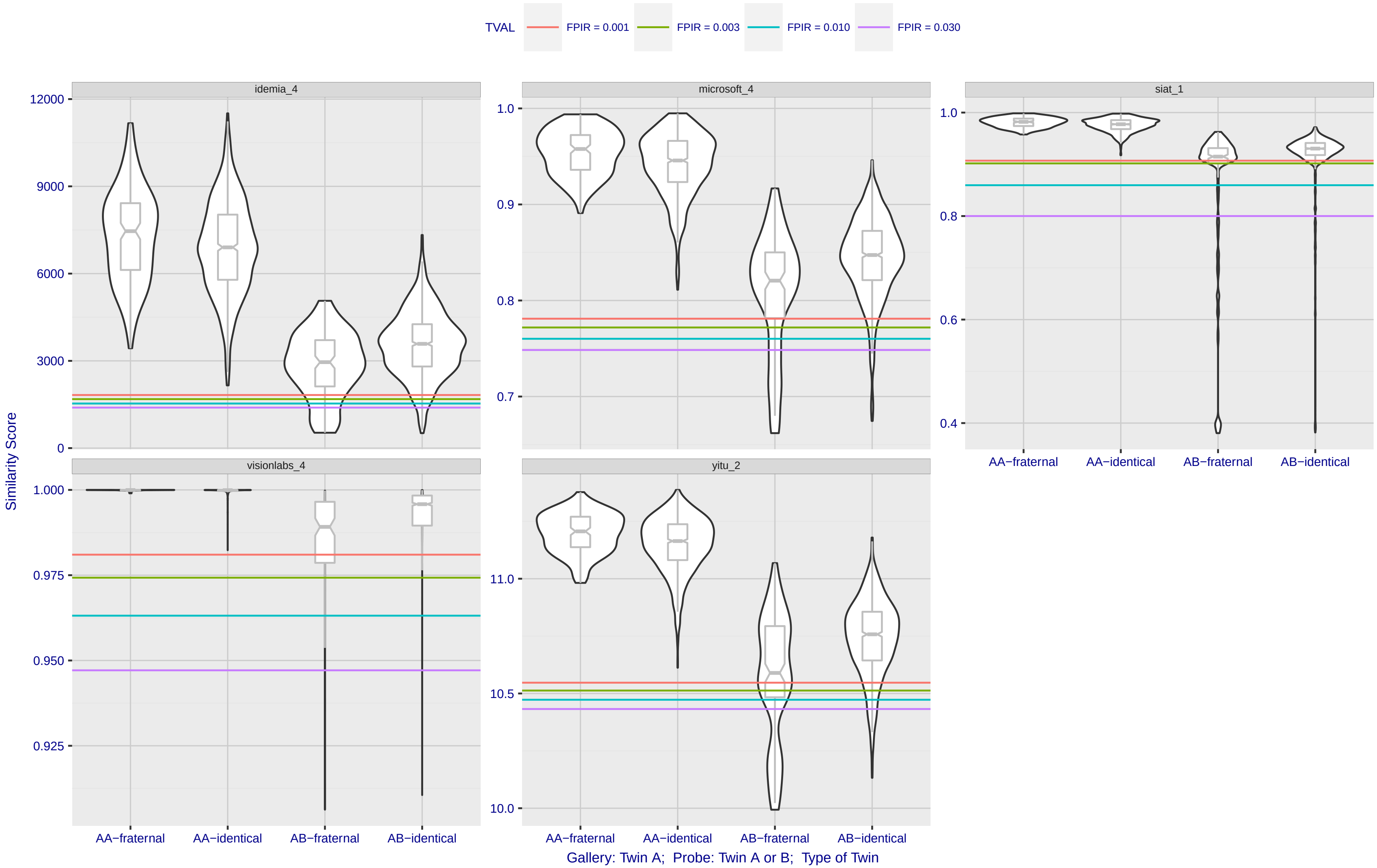

Figure 23: **[Notre Dame Twins Dataset] High scores from twins.**. *The Figure shows native similarity scores from searches into a dataset of N = 640 000 background mugshot images plus 104 portrait images, one from each of one of a pair of twins. Two distributions of scores are plotted for each of monozygotic (identical) and dizygotic (fraternal) twins. The first distribution ("AA") shows the mate score from Twin A against their own enrollment. The second ("AB") shows scores from searches of Twin B against the Twin A enrollment: As these are non-mate scores they should be below the various thresholds shown as horizontal lines. That they usually are not is an indication that twins produce very high non-mate scores. Note in theory half of dyzygotic (fraternal) twins are different sex. In the sample used here some fraternal twins are correctly rejected.*

Appendices

Appendix A Accuracy on large-population FRVT 2018 mugshots

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

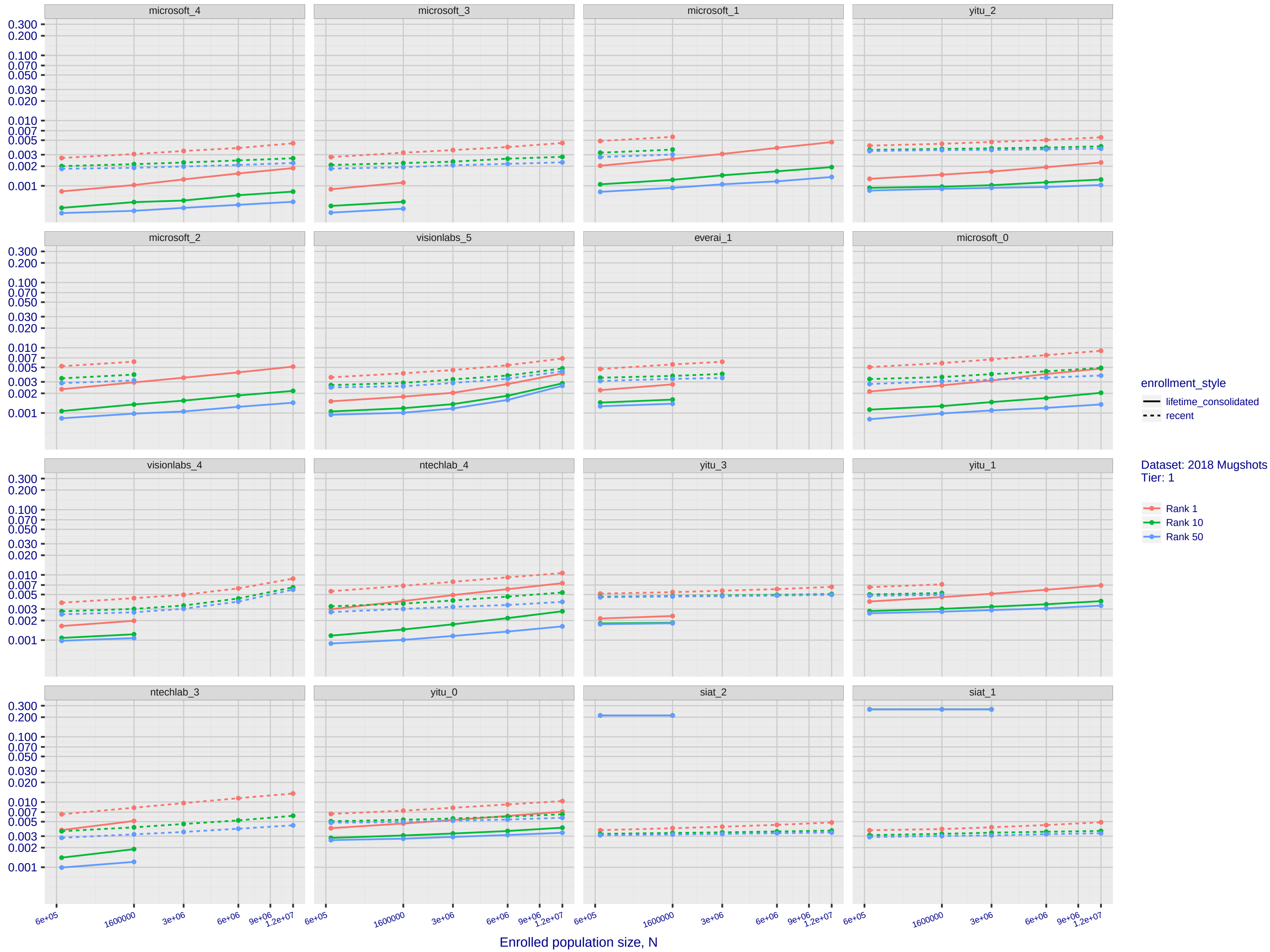

Figure 24: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric rewards even weak scoring rank 1 mates. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria being rank 1 hit rate on a gallery size of 640 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

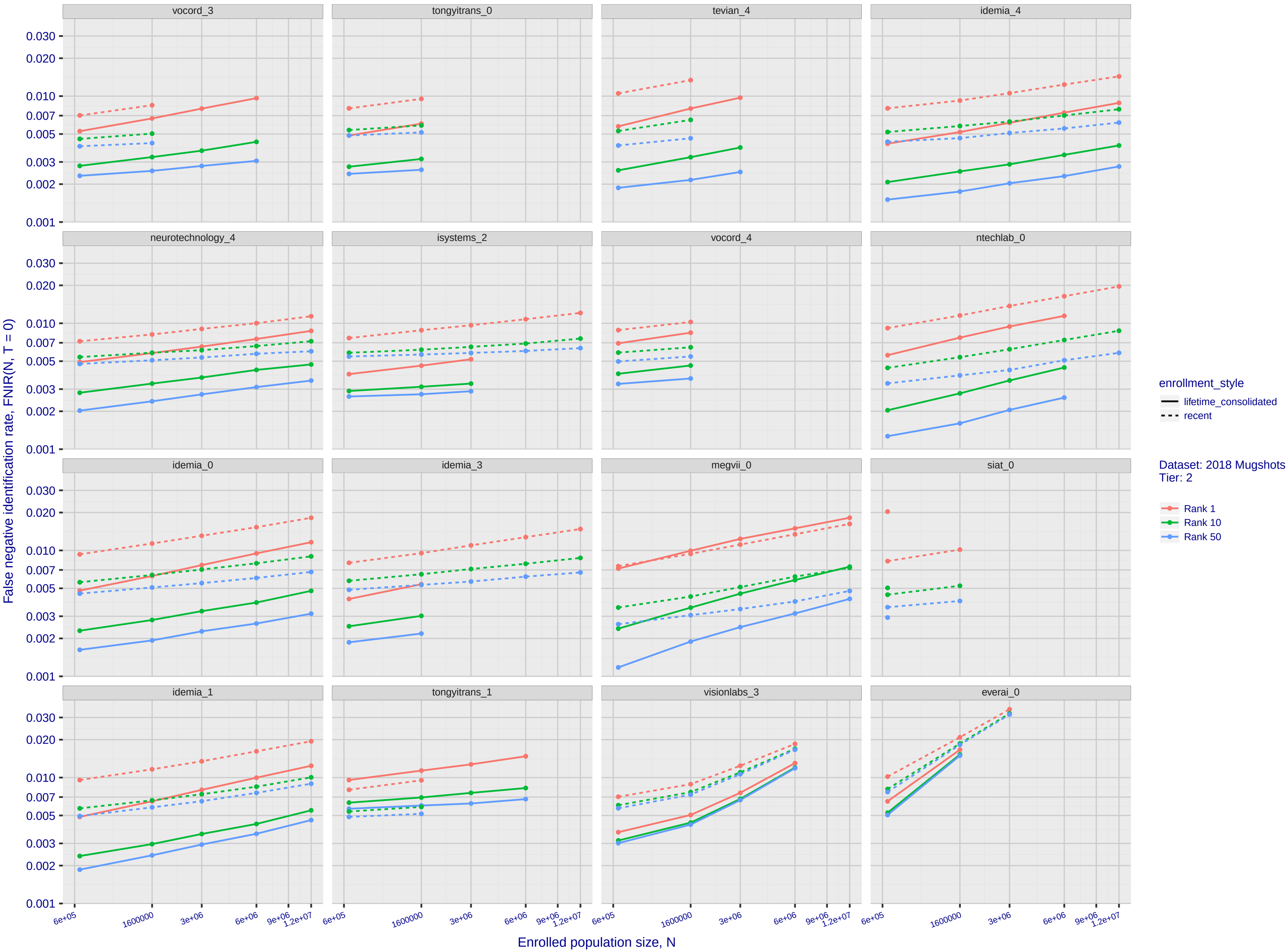

Figure 25: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric rewards even weak scoring rank 1 mates. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria being rank 1 hit rate on a gallery size of 640 000.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

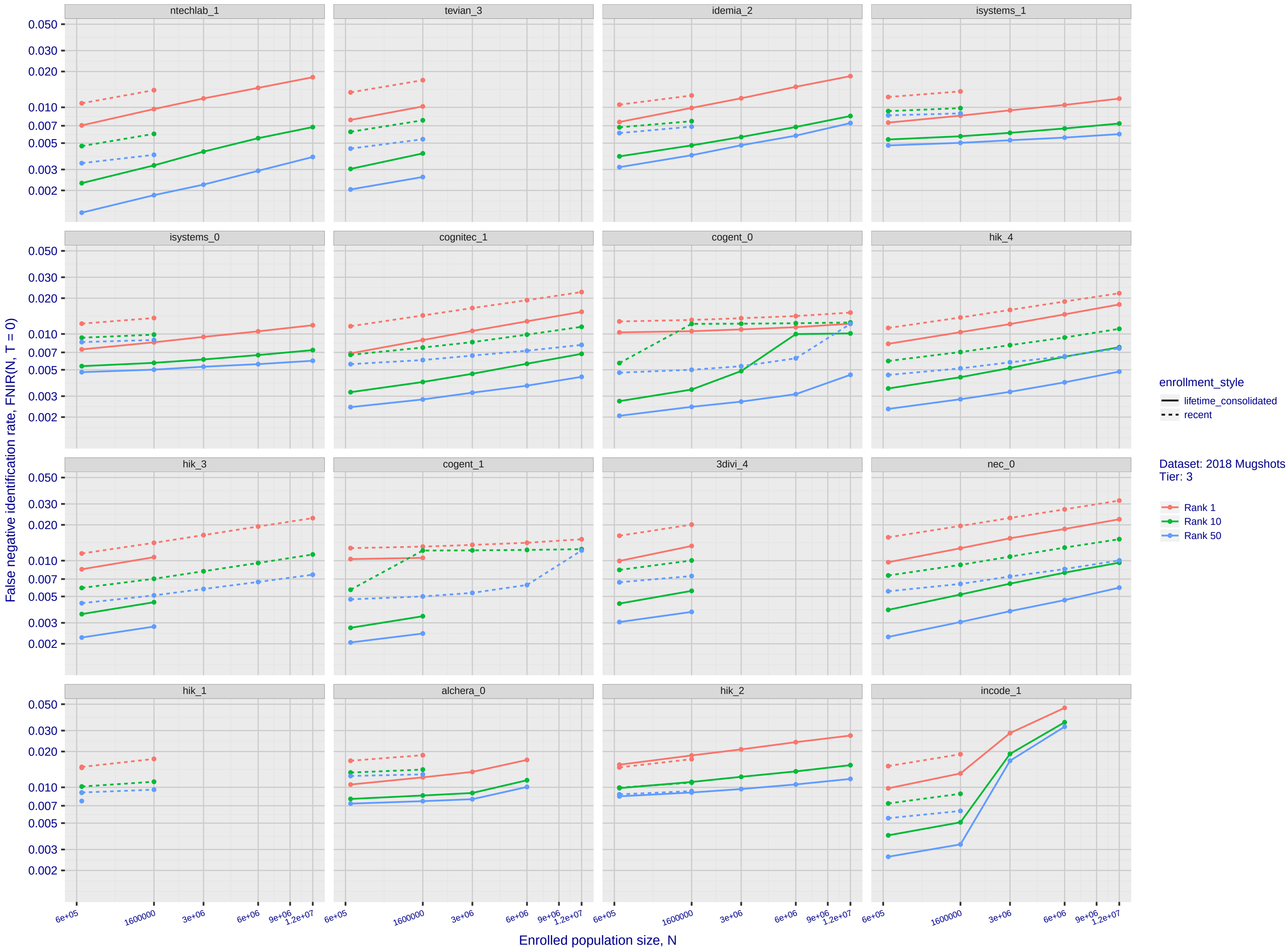

Figure 26: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric rewards even weak scoring rank 1 mates. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria being rank 1 hit rate on a gallery size of 640 000.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

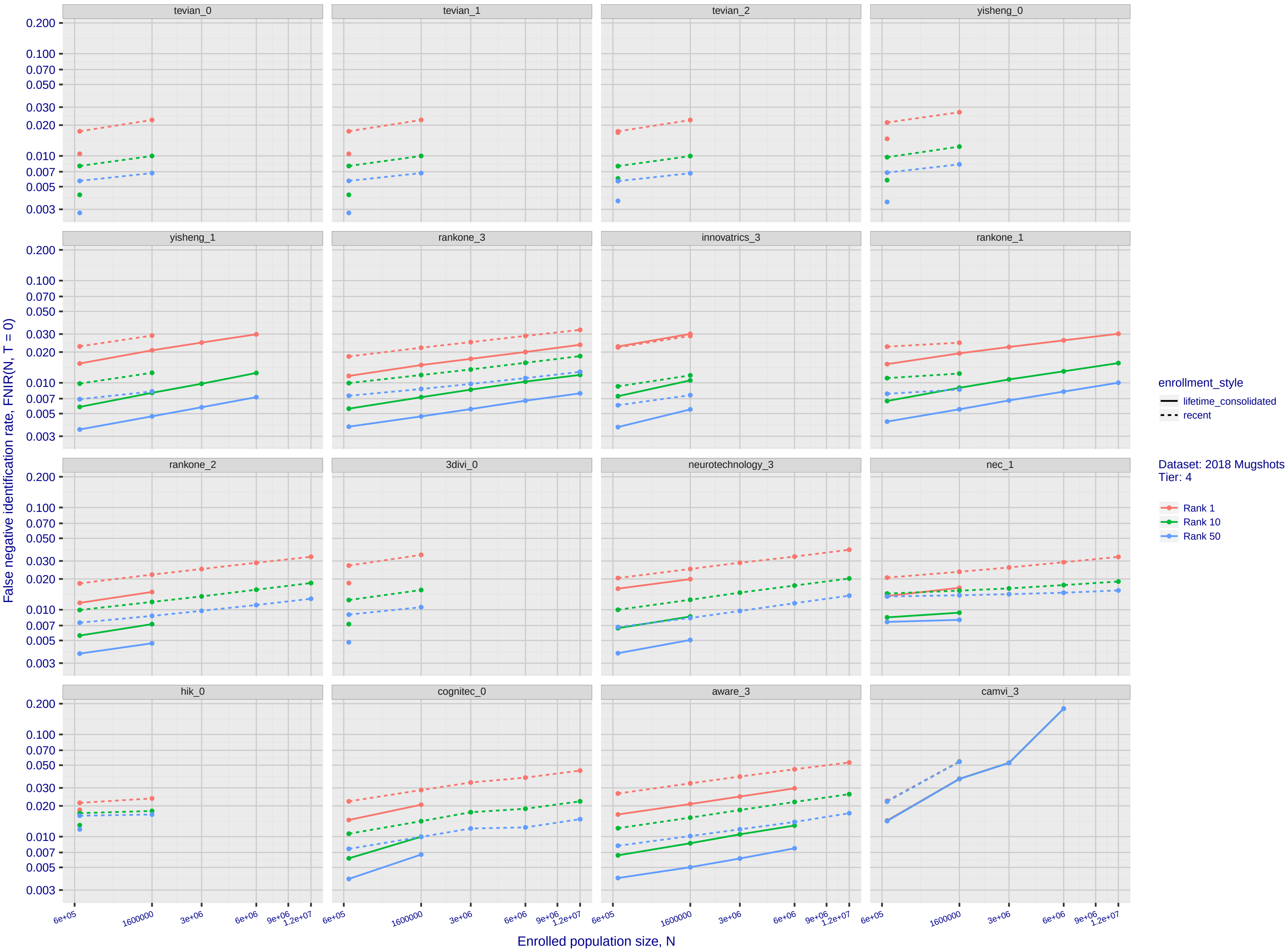

Figure 27: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric rewards even weak scoring rank 1 mates. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria being rank 1 hit rate on a gallery size of 640 000.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

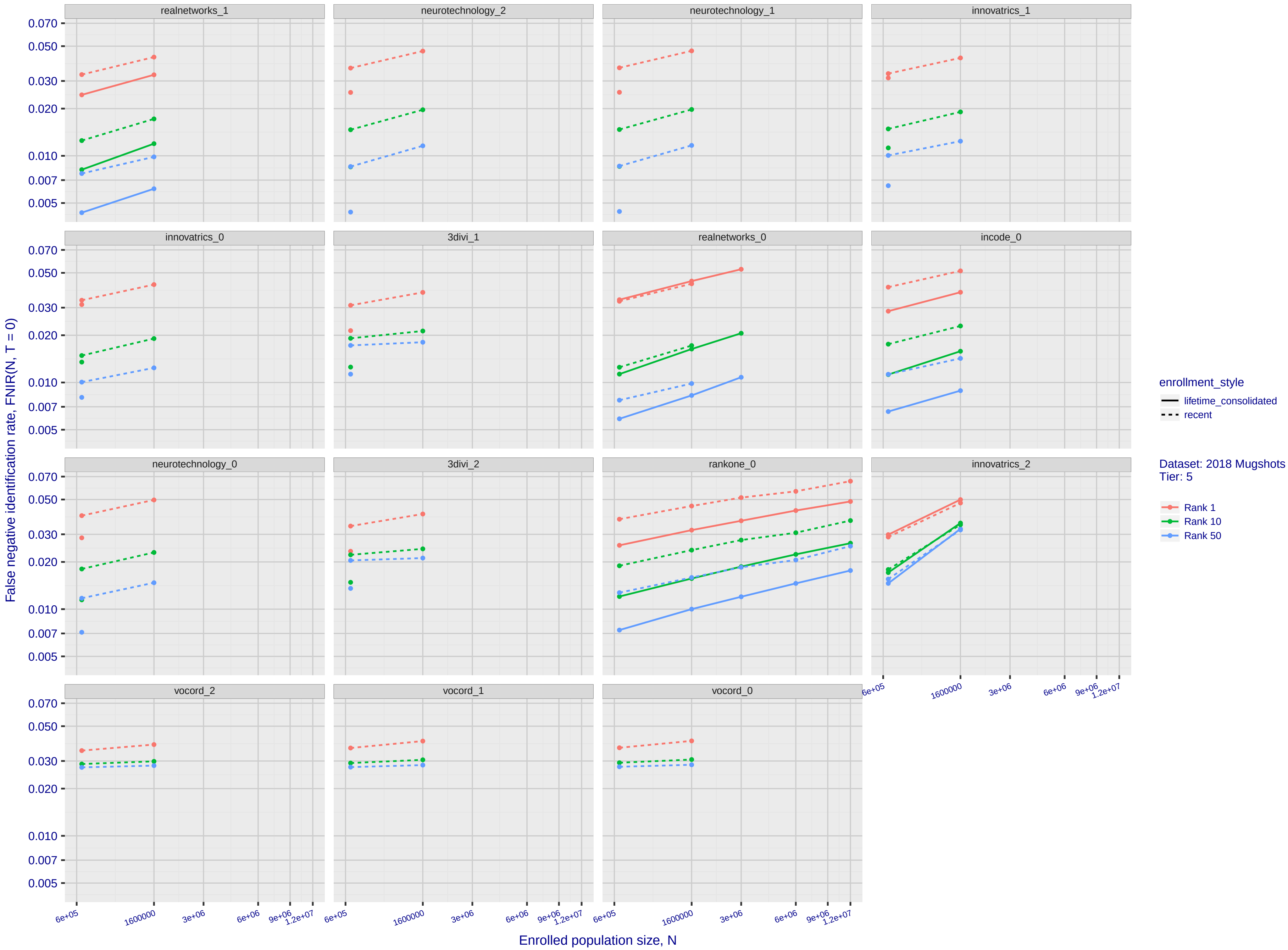

Figure 28: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric rewards even weak scoring rank 1 mates. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria being rank 1 hit rate on a gallery size of 640 000.*

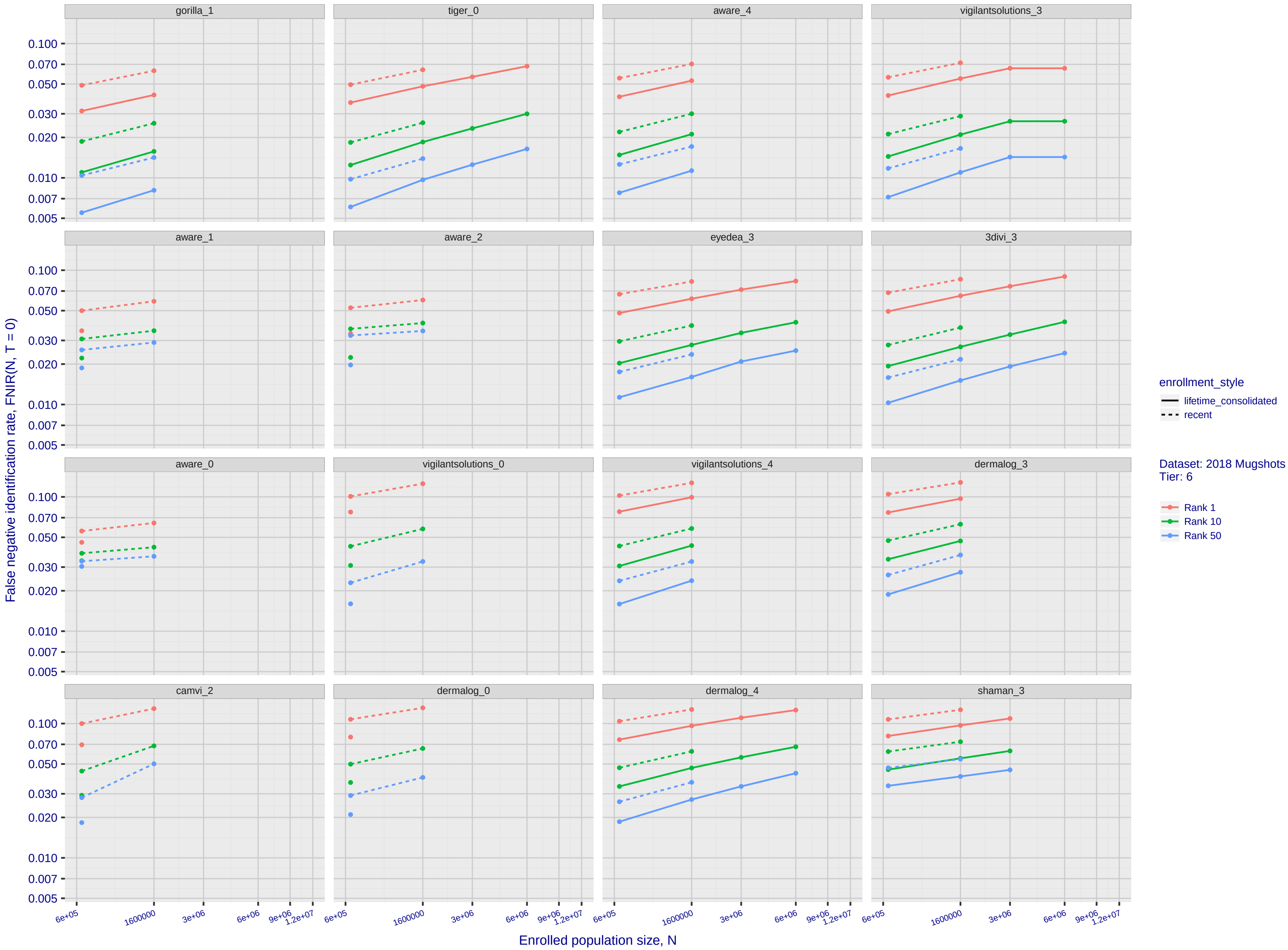

Figure 29: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric rewards even weak scoring rank 1 mates. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria being rank 1 hit rate on a gallery size of 640 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

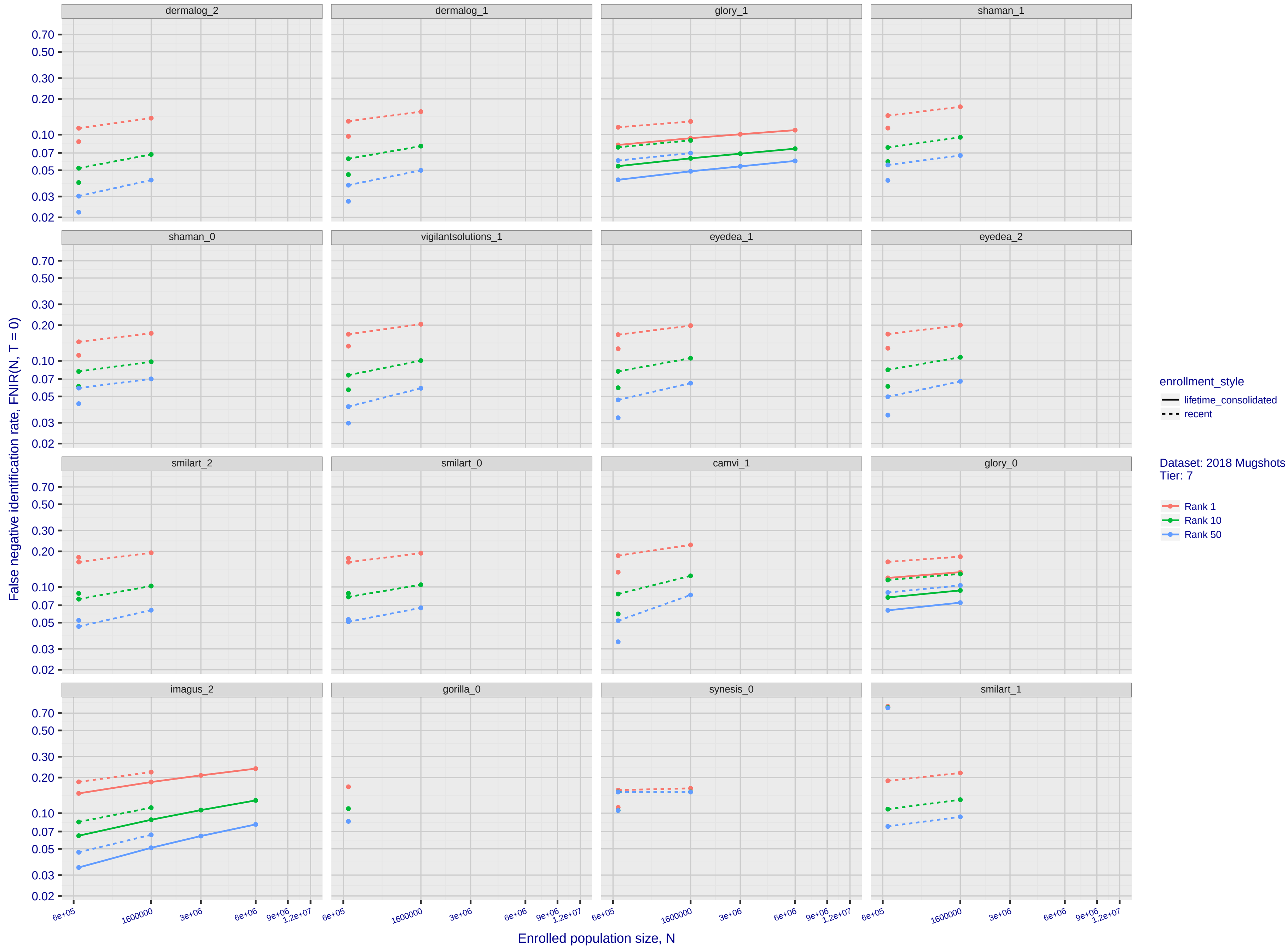

Figure 30: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric rewards even weak scoring rank 1 mates. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria being rank 1 hit rate on a gallery size of 640 000.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

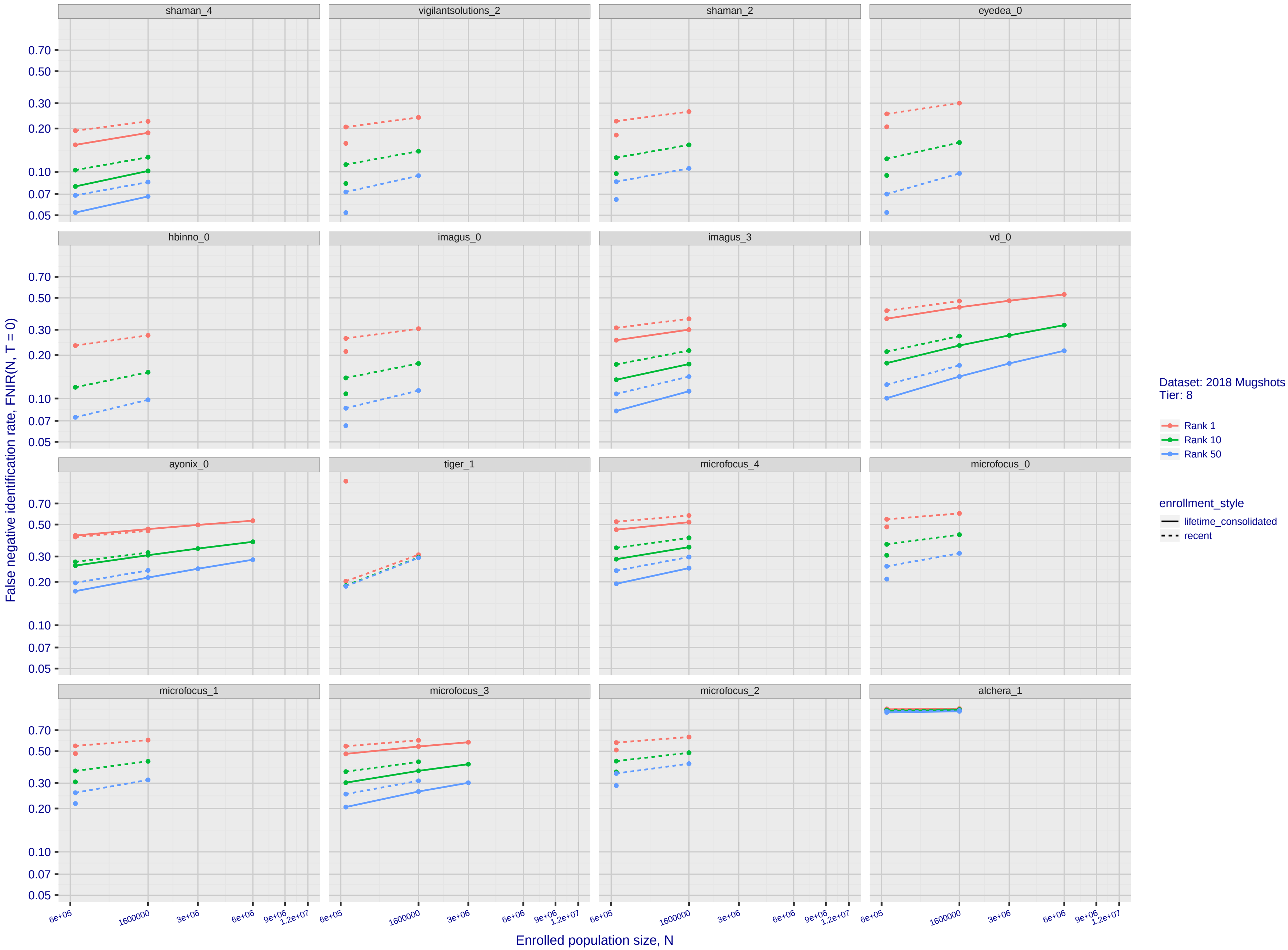

Figure 31: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric rewards even weak scoring rank 1 mates. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria being rank 1 hit rate on a gallery size of 640 000.*

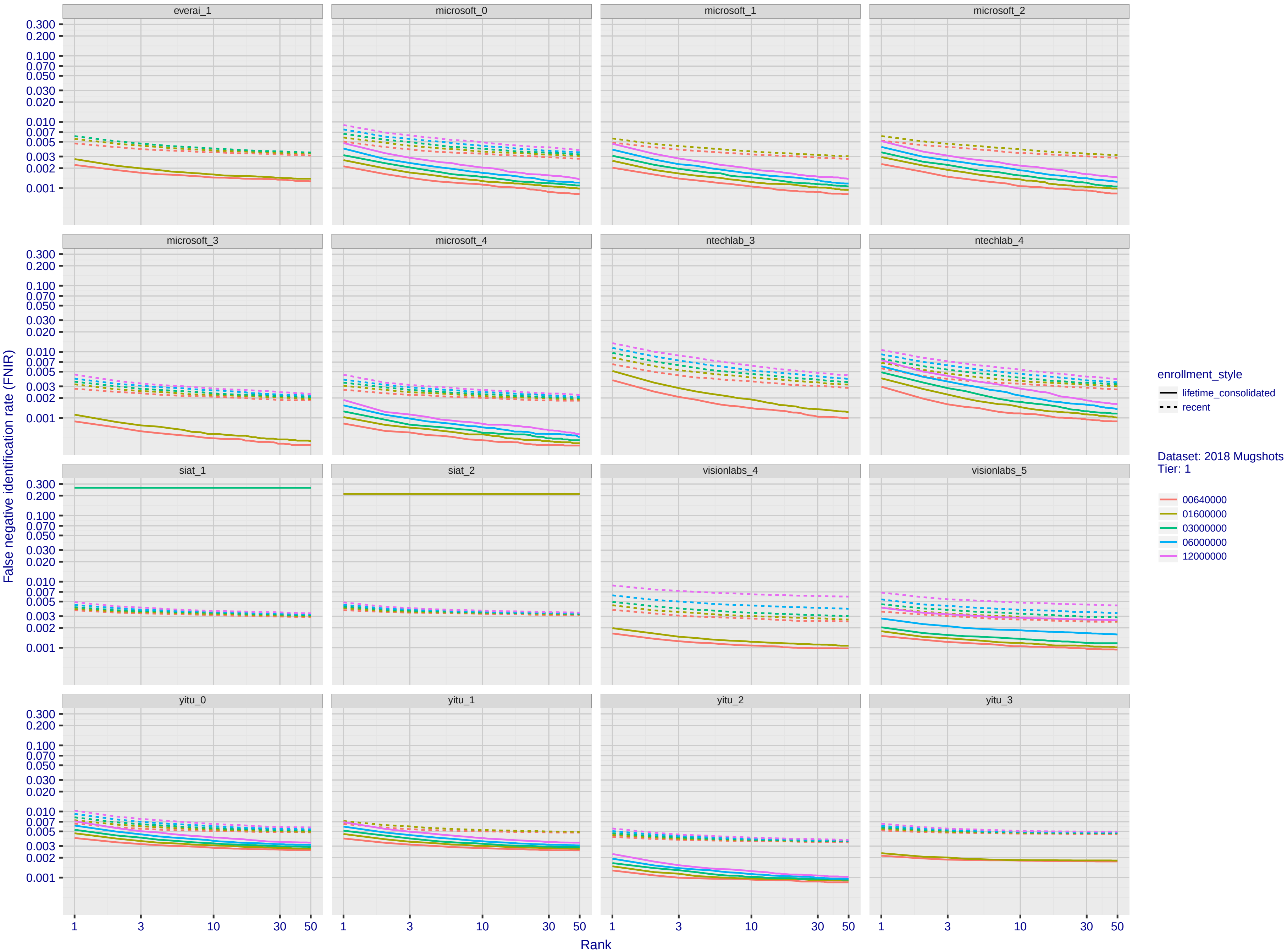

Figure 32: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank**. *For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted candidate lists. Results are sorted and reported into tiers for clarity, with the tiering criteria being rank 1 hit rate on a gallery size of N = 640 000 subjects.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

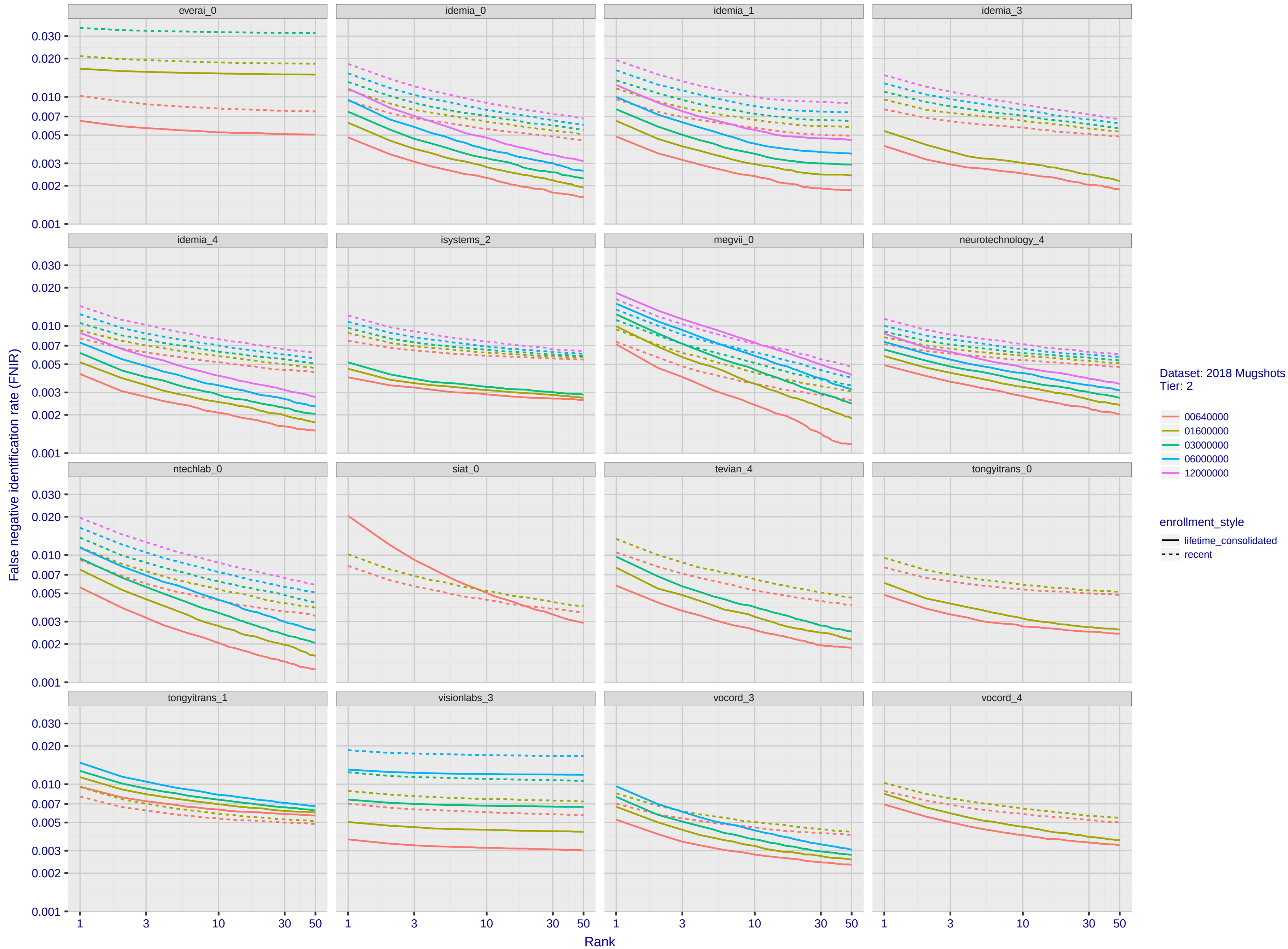

Figure 33: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank**. *For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted candidate lists. Results are sorted and reported into tiers for clarity, with the tiering criteria being rank 1 hit rate on a gallery size of N = 640 000 subjects.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

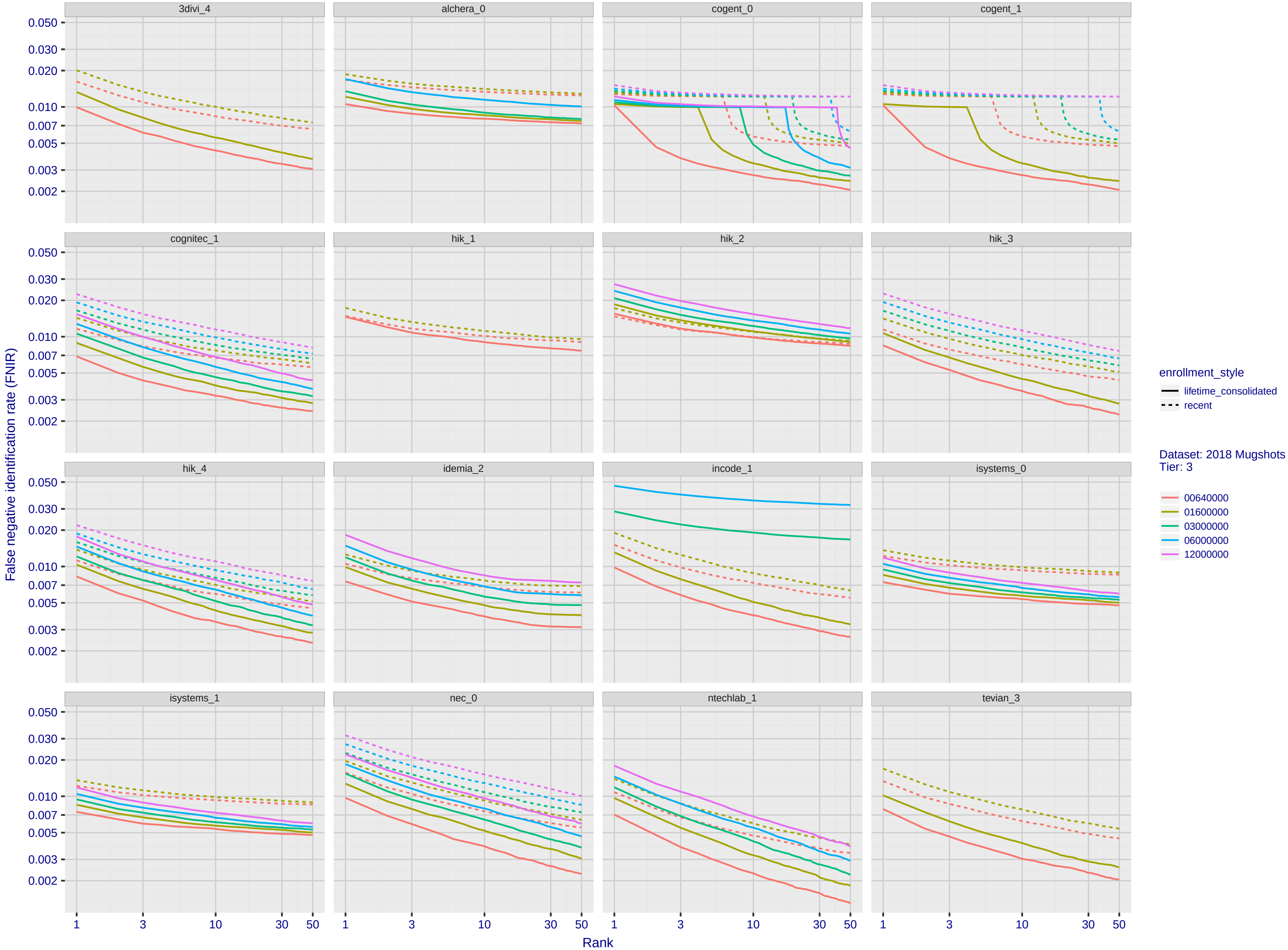

Figure 34: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank**. *For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted candidate lists. Results are sorted and reported into tiers for clarity, with the tiering criteria being rank 1 hit rate on a gallery size of N = 640 000 subjects.*

FNIR(N, R, T) = False neg. identification rate	N = Num. enrolled subjects	T = Threshold	T = 0 → Investigation
FPIR(N, T) = False pos. identification rate	R = Num. candidates examined		T > 0 → Identification

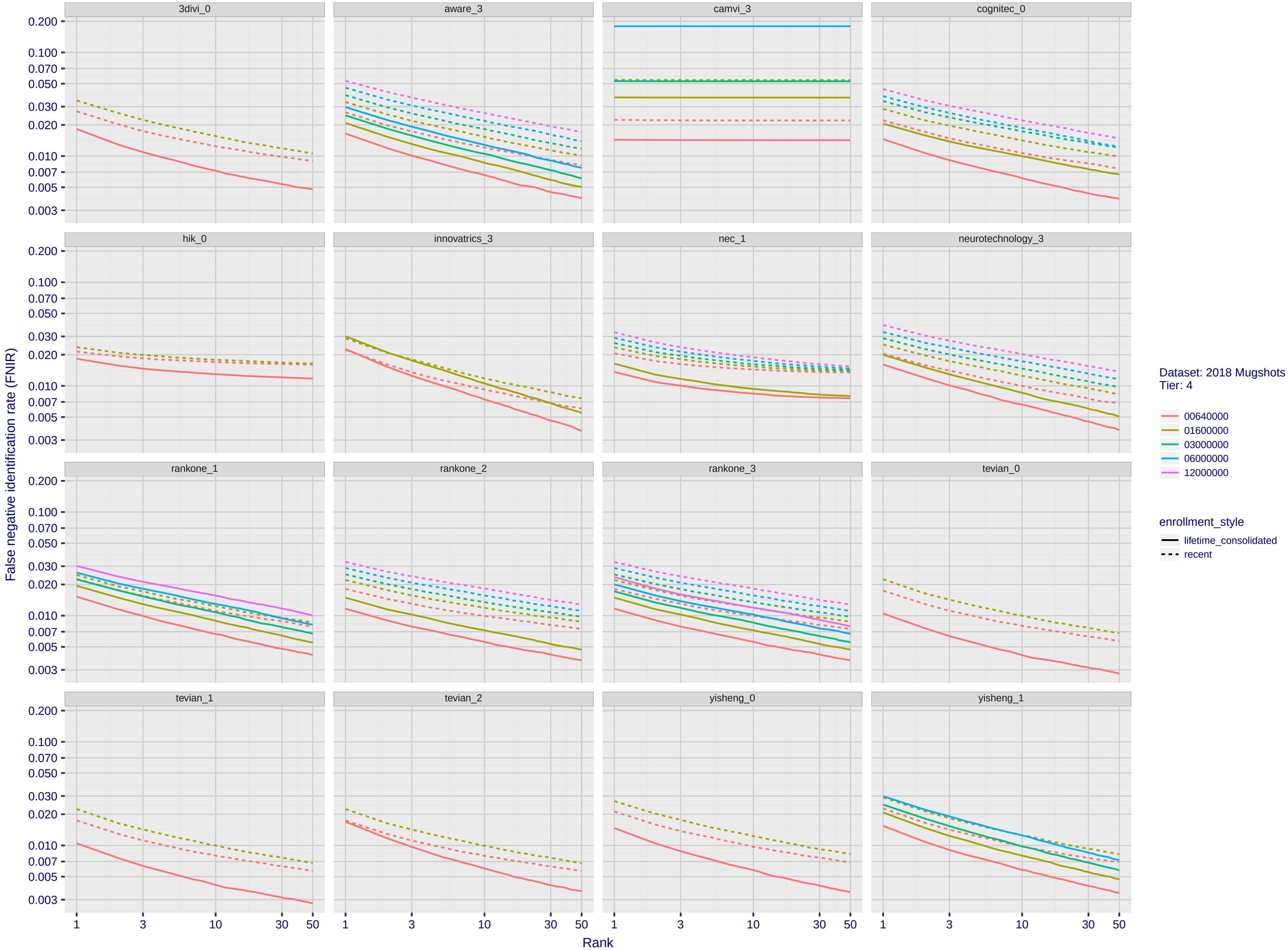

Figure 35: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank**. *For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted candidate lists. Results are sorted and reported into tiers for clarity, with the tiering criteria being rank 1 hit rate on a gallery size of N = 640 000 subjects.*

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8238

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

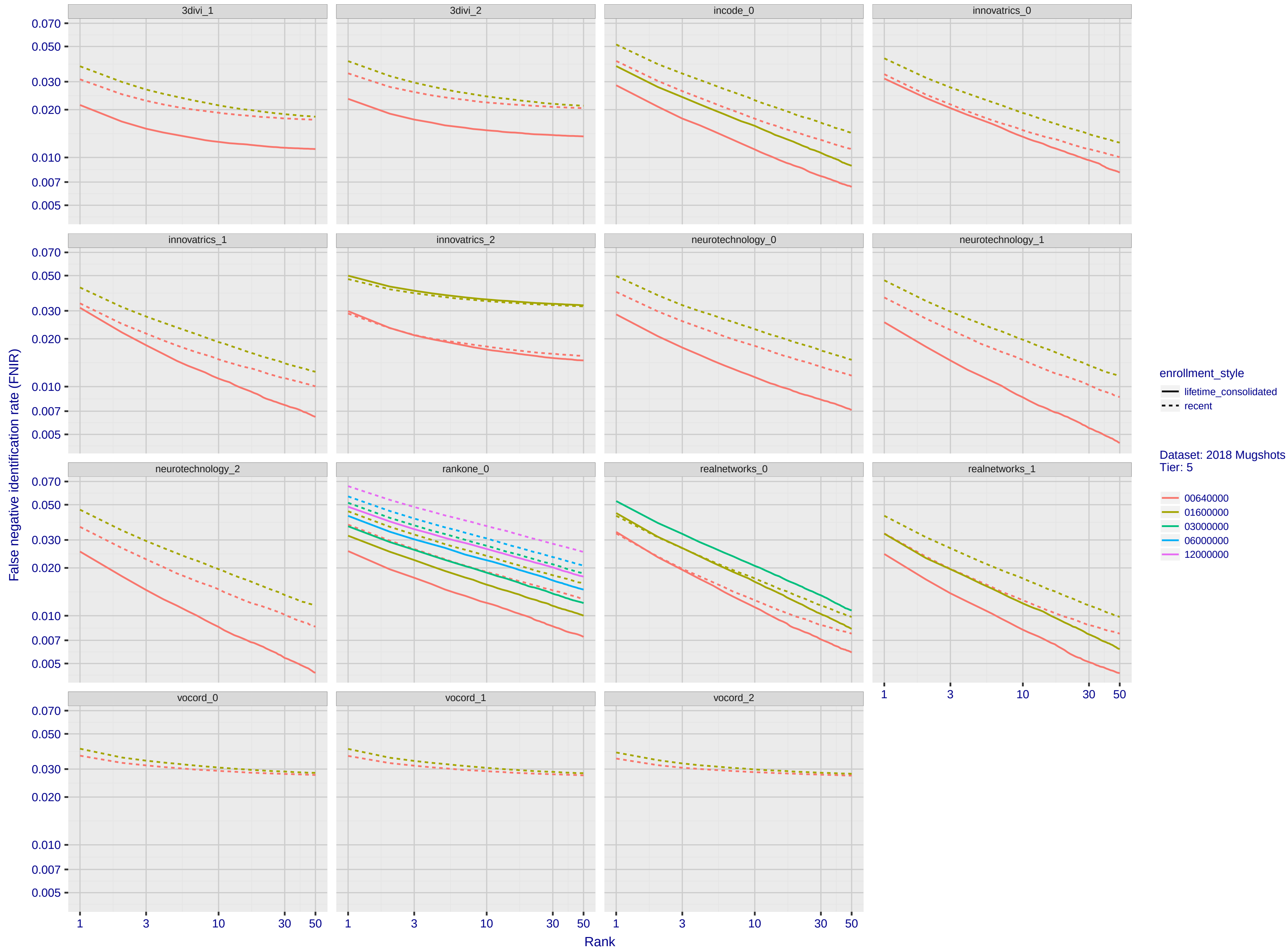

Figure 36: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank**. *For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted candidate lists. Results are sorted and reported into tiers for clarity, with the tiering criteria being rank 1 hit rate on a gallery size of N = 640 000 subjects.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

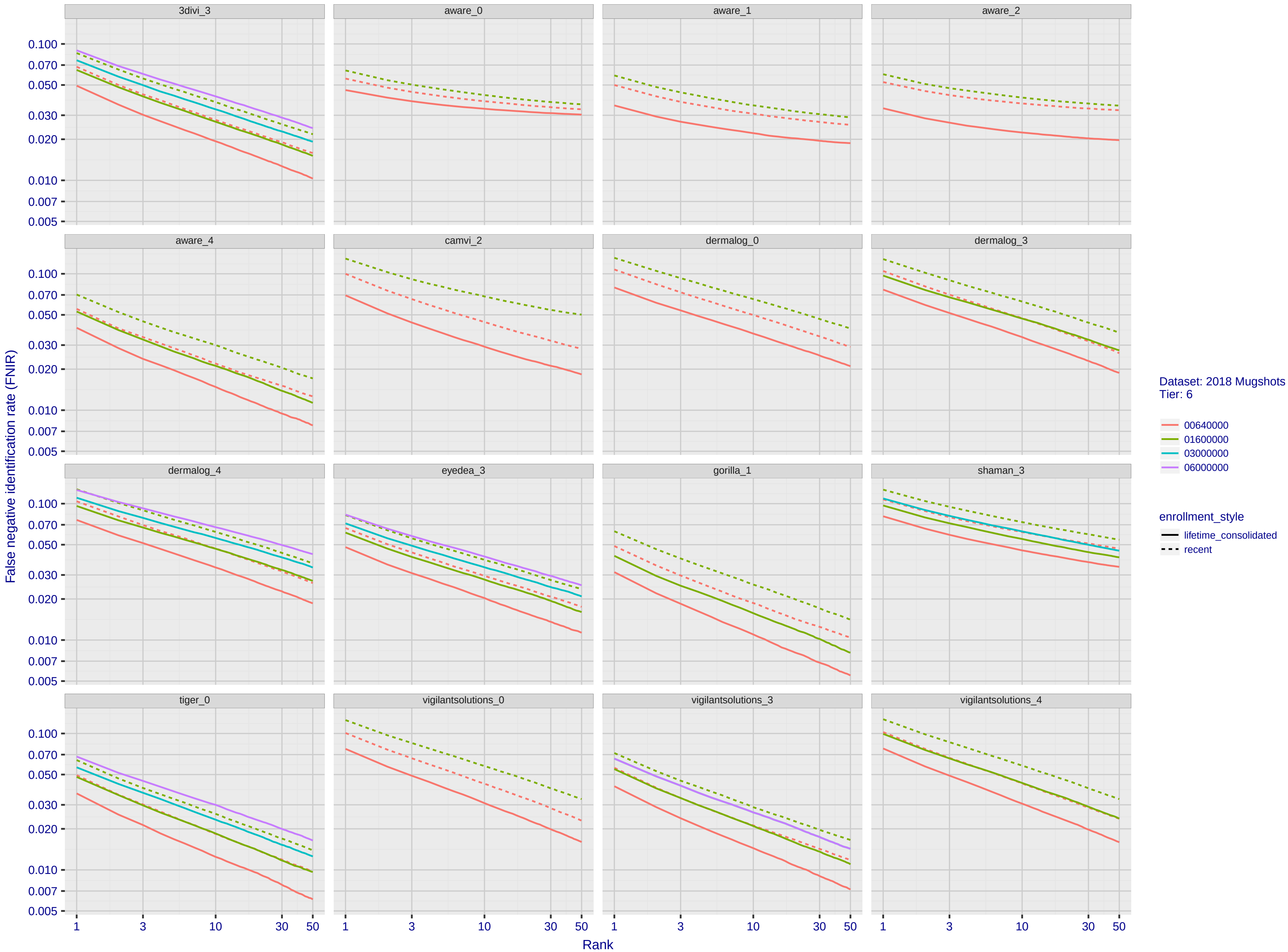

Figure 37: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank**. *For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted candidate lists. Results are sorted and reported into tiers for clarity, with the tiering criteria being rank 1 hit rate on a gallery size of N = 640 000 subjects.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

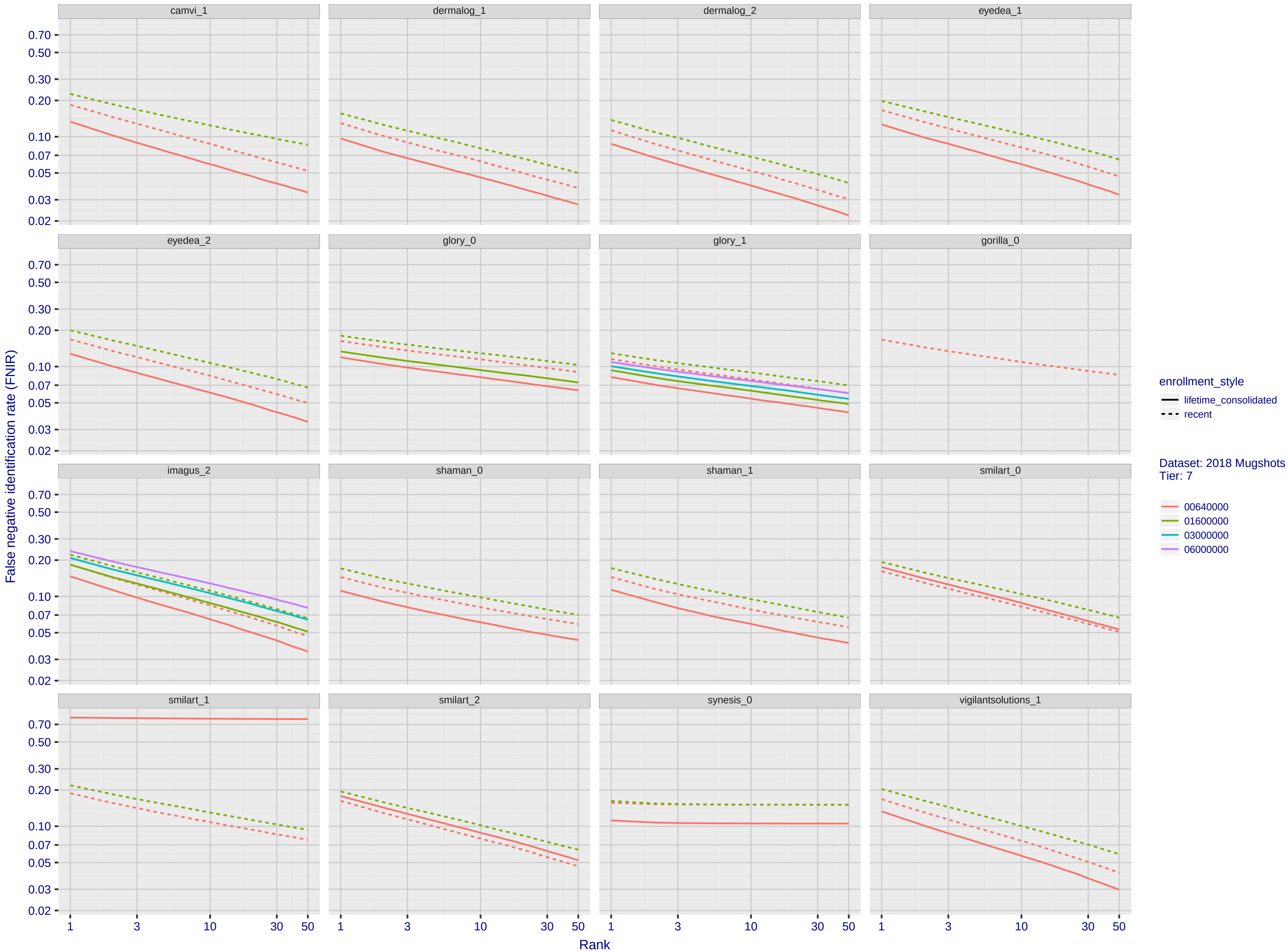

Figure 38: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank**. *For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted candidate lists. Results are sorted and reported into tiers for clarity, with the tiering criteria being rank 1 hit rate on a gallery size of N = 640 000 subjects.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

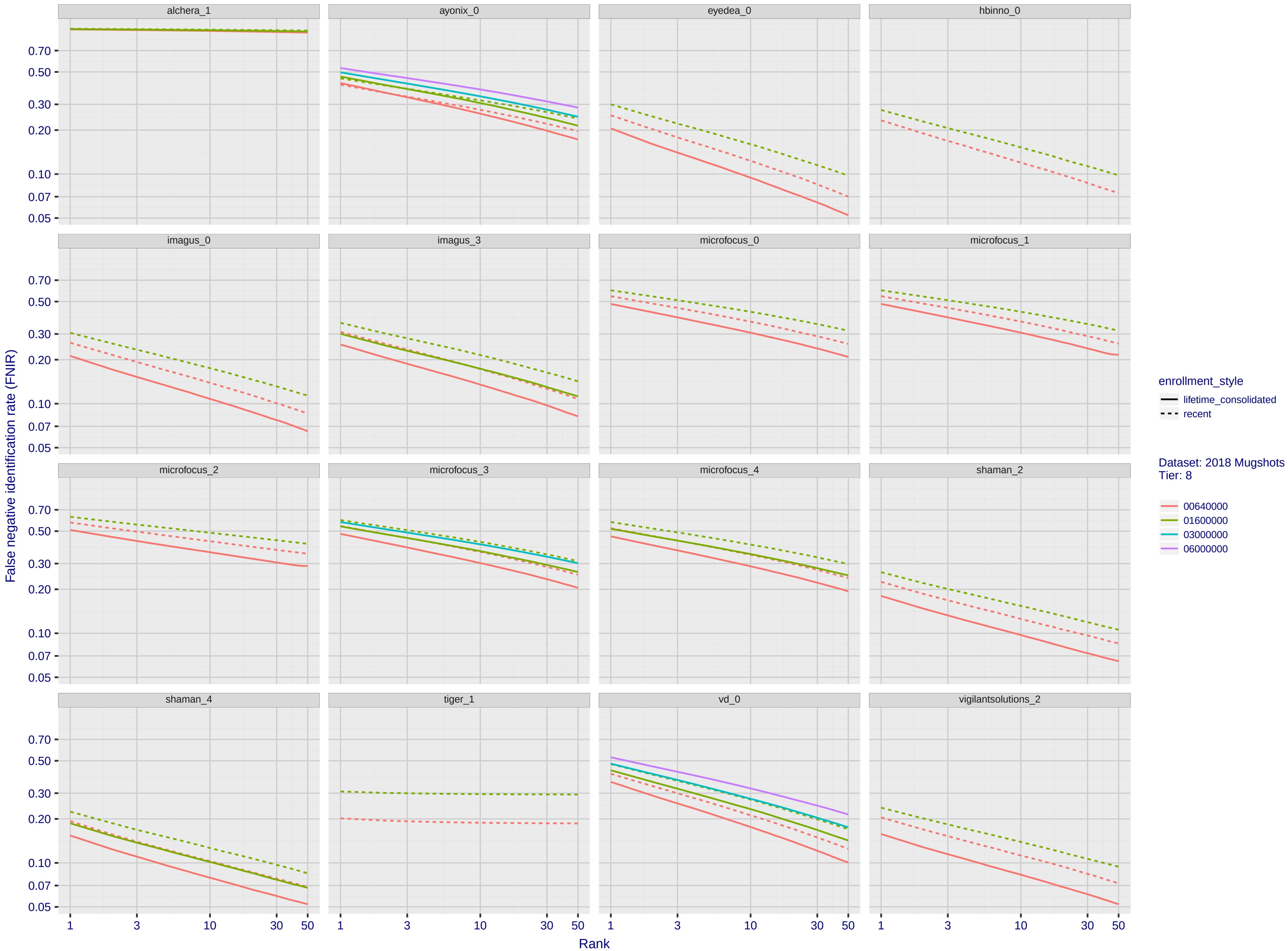

Figure 39: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank**. *For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted candidate lists. Results are sorted and reported into tiers for clarity, with the tiering criteria being rank 1 hit rate on a gallery size of N = 640 000 subjects.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

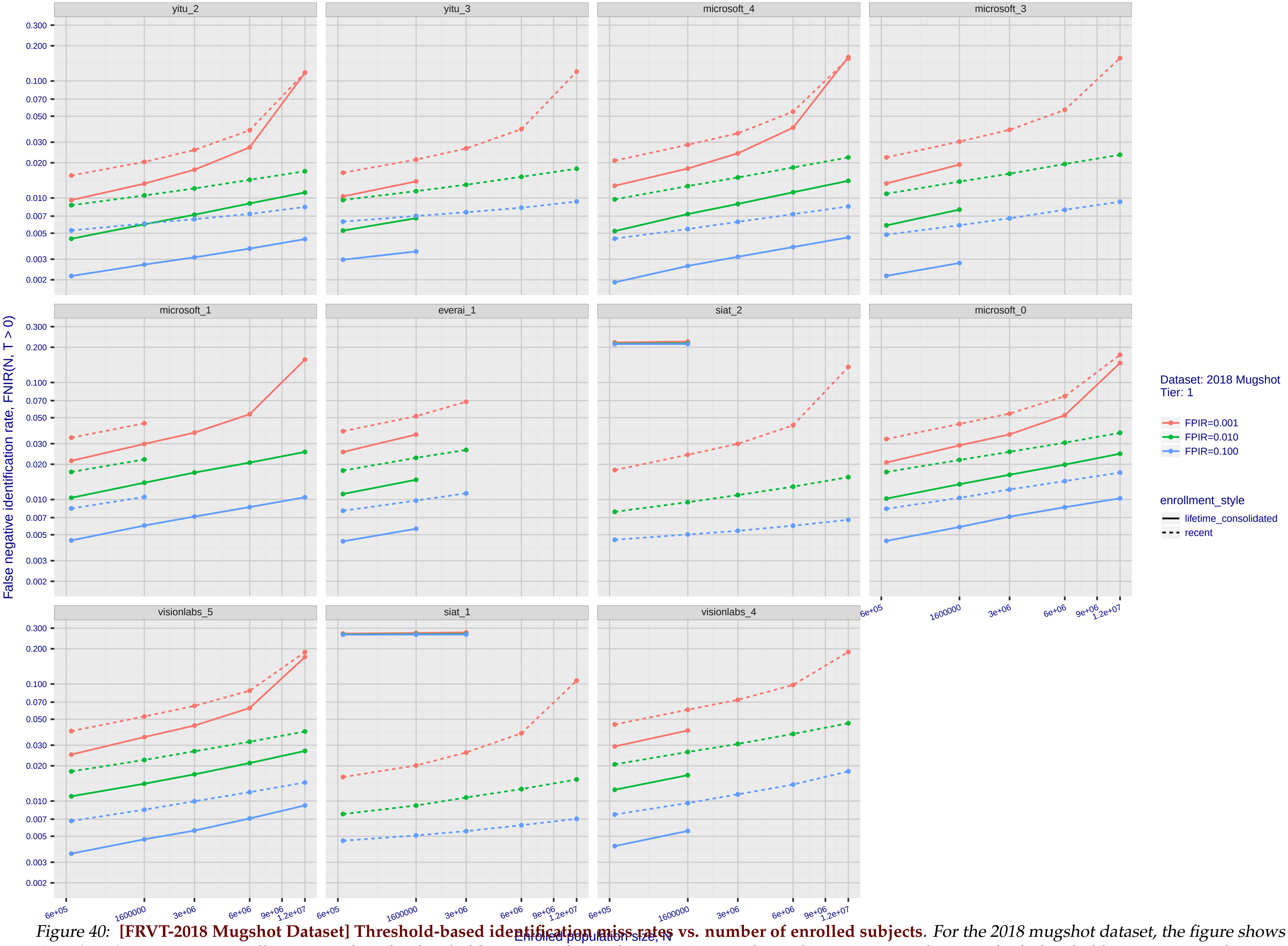

Figure 40: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

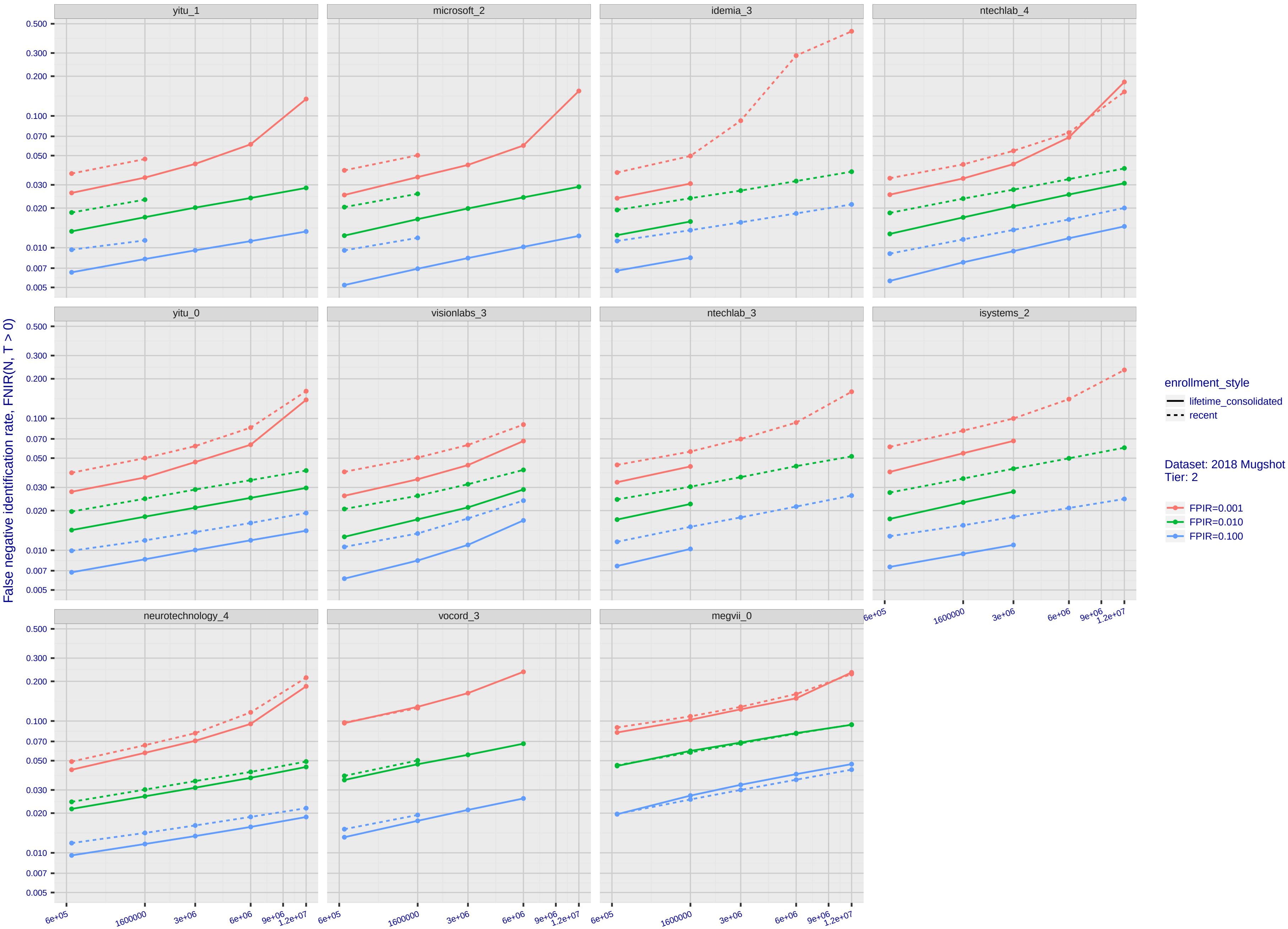

Figure 41: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

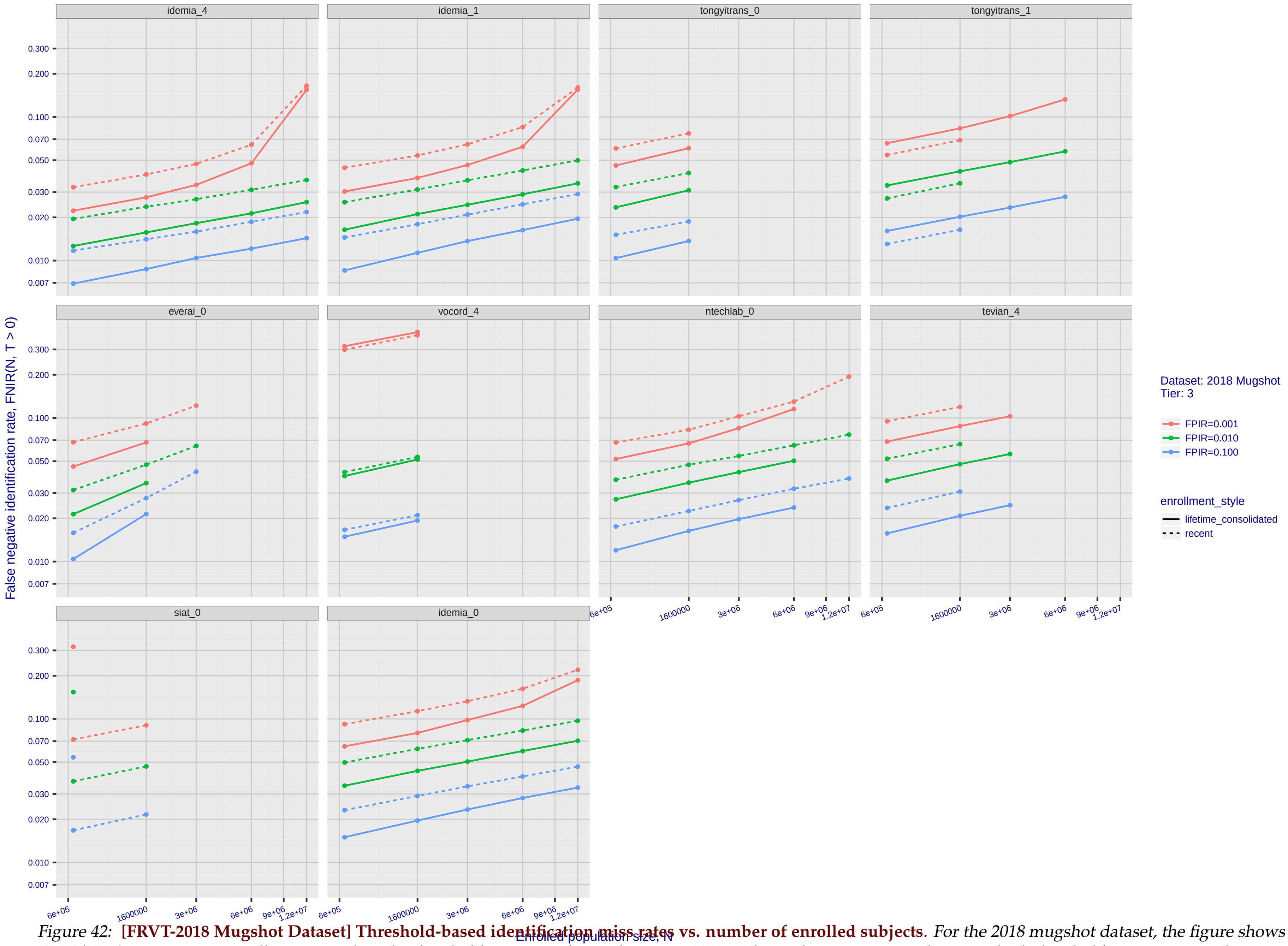

Figure 42: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

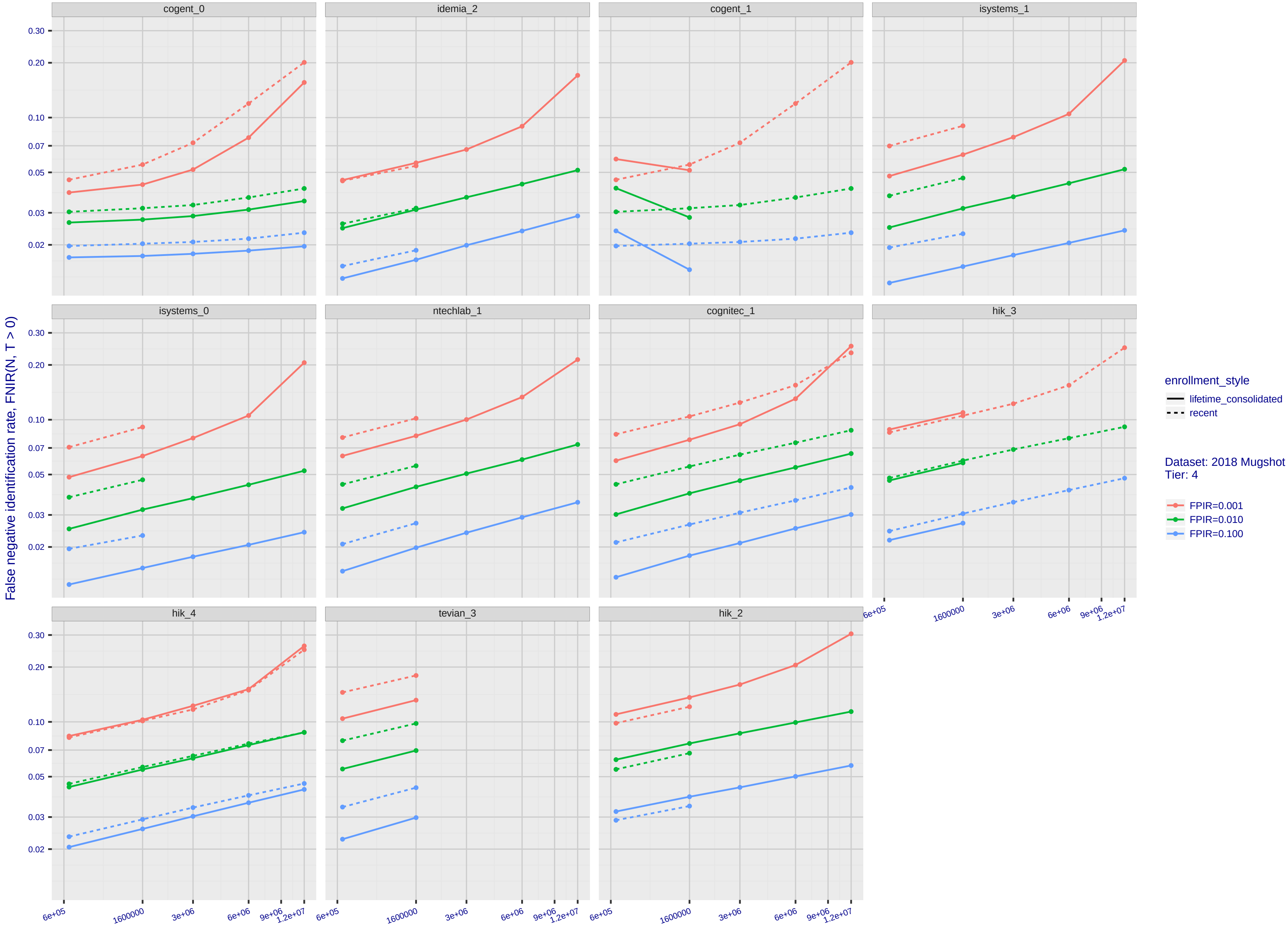

Figure 43: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

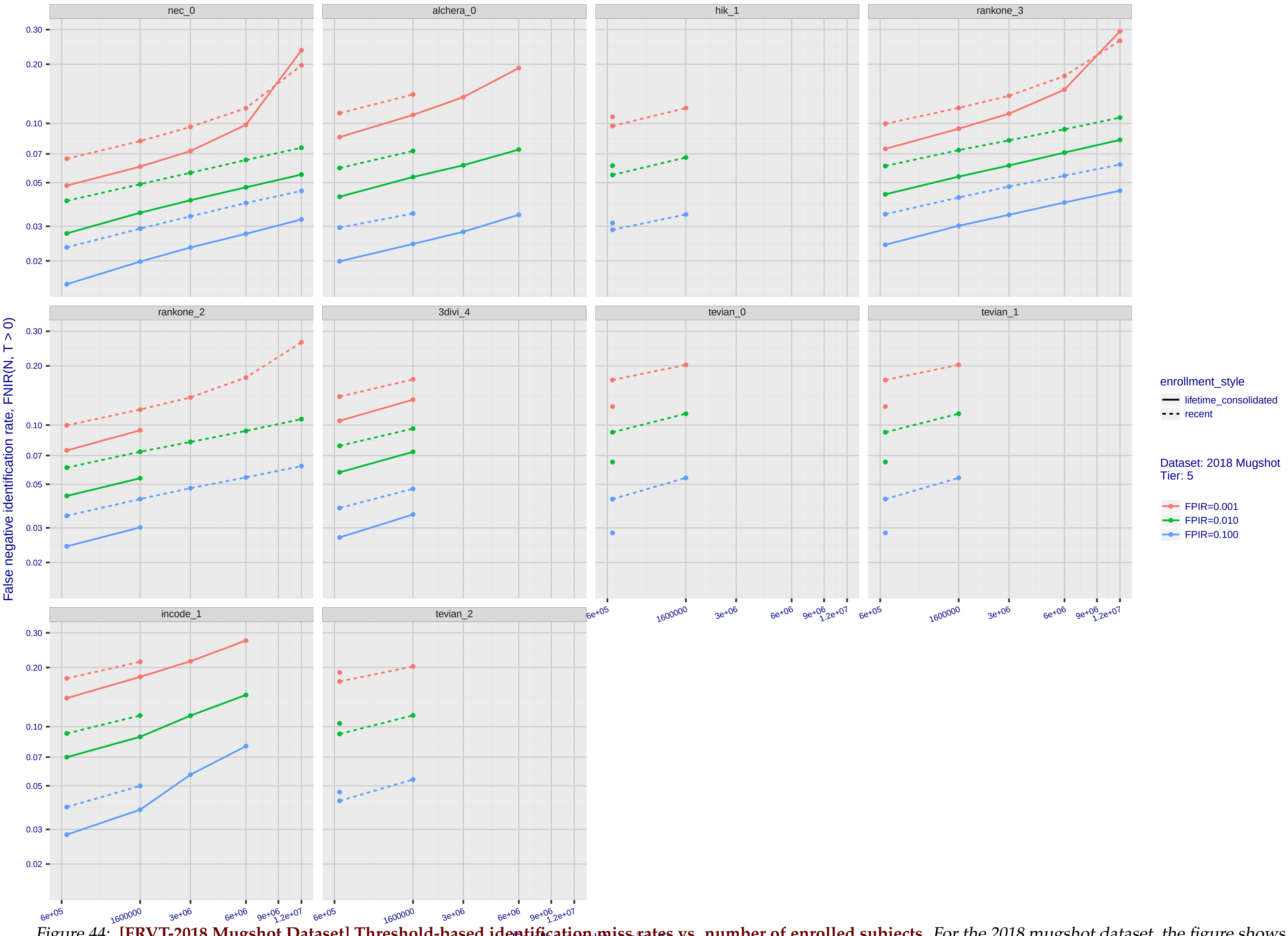

Figure 44: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

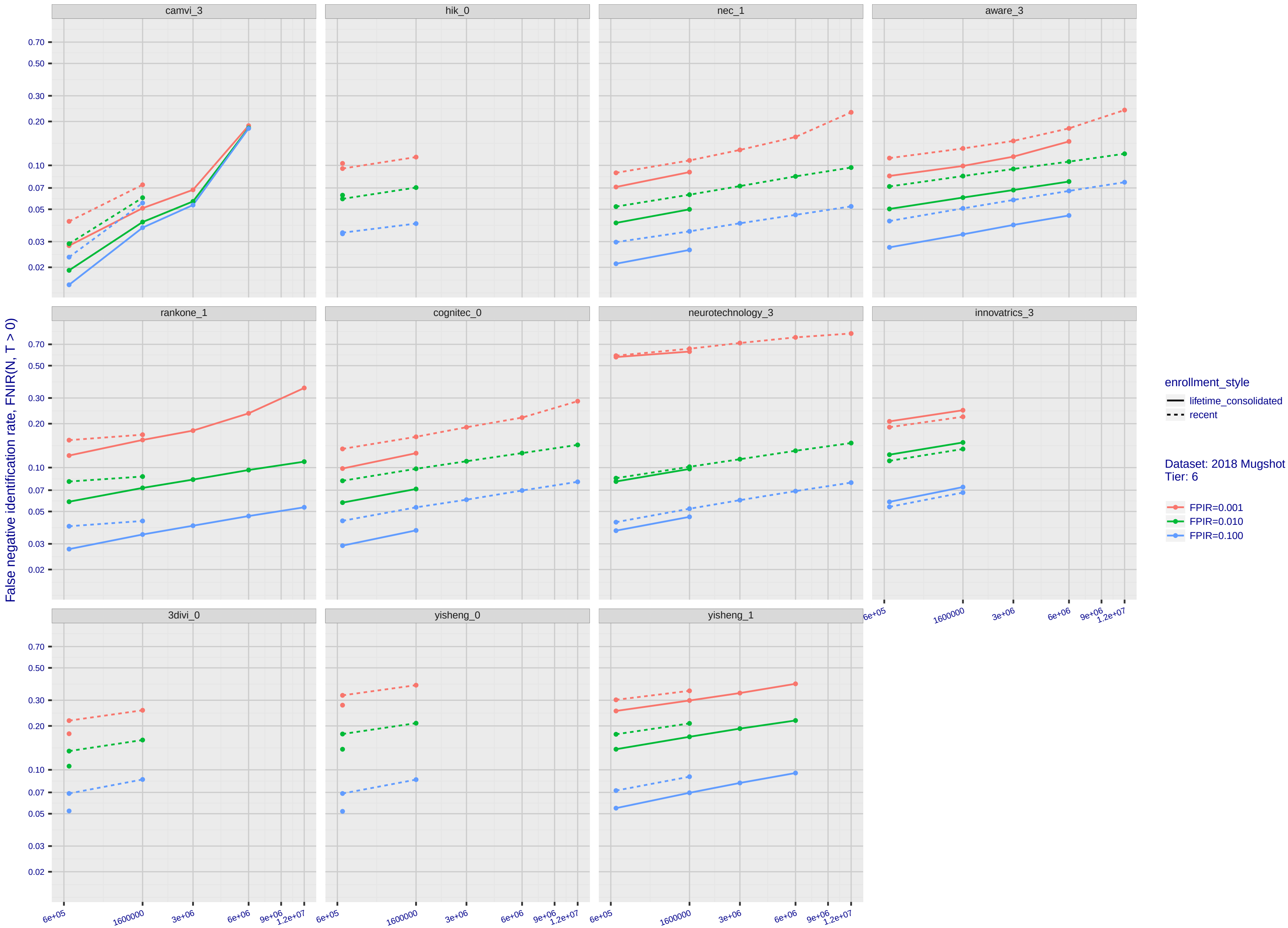

Figure 45: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

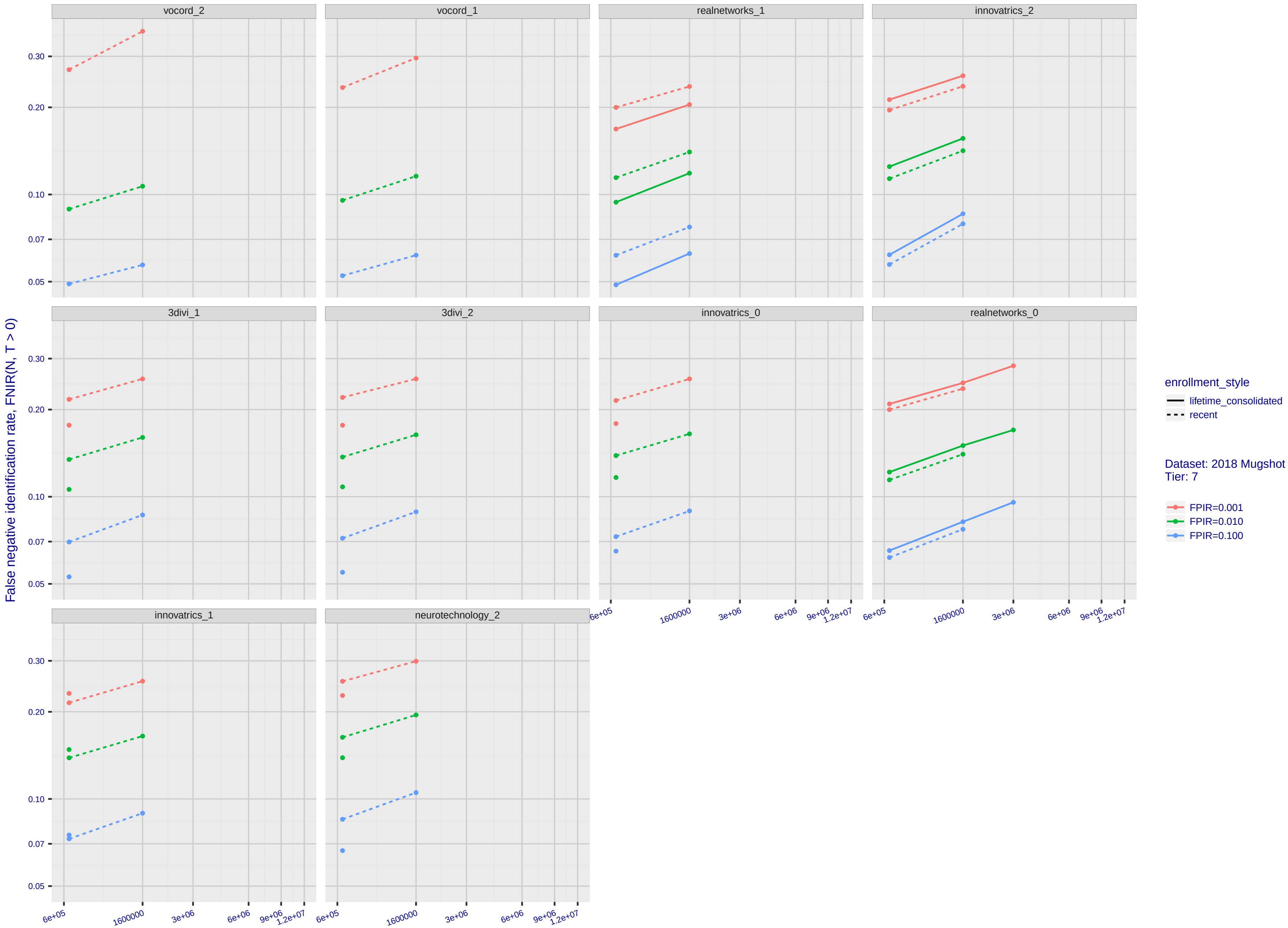

Figure 46: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

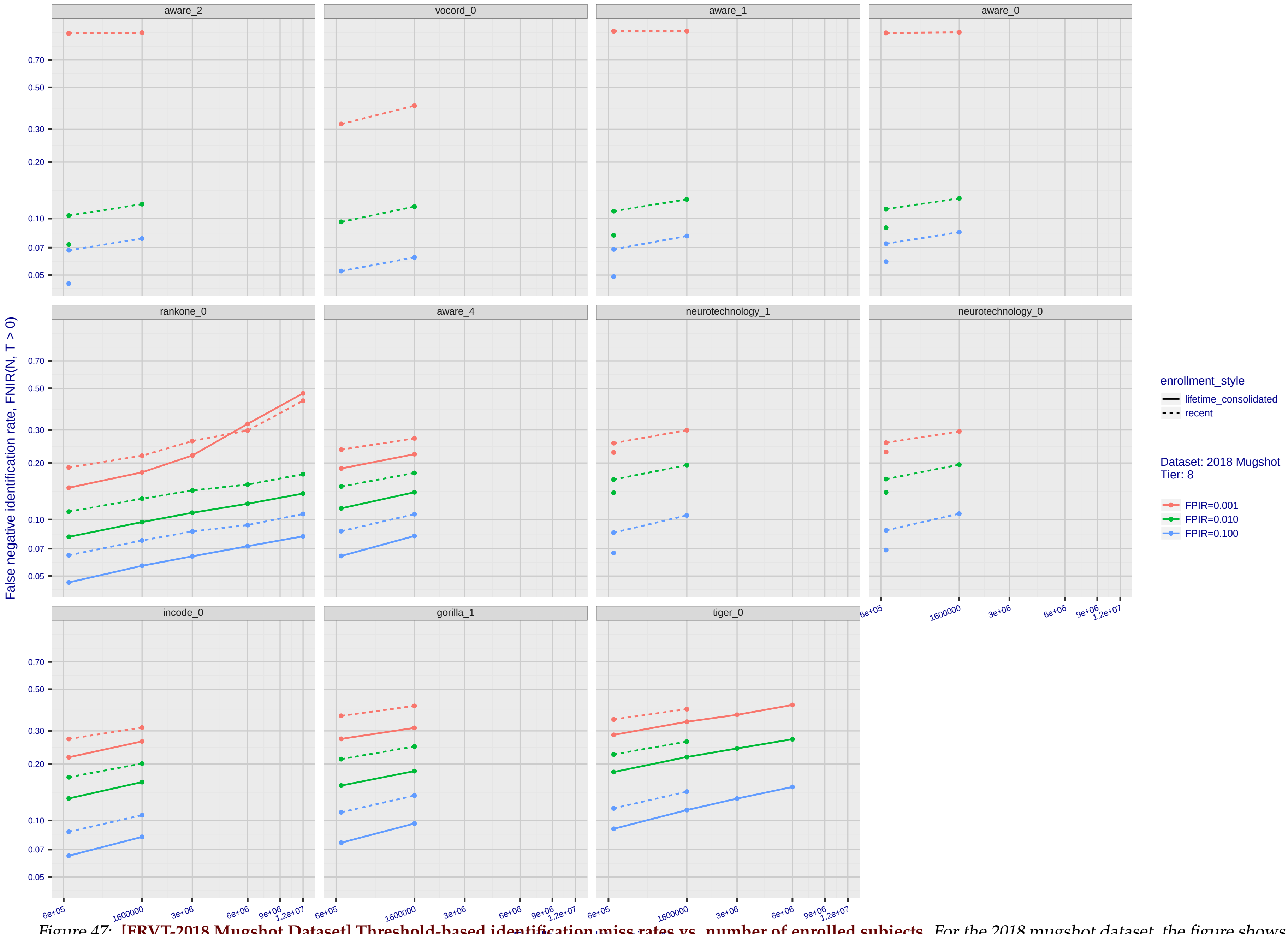

Figure 47: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

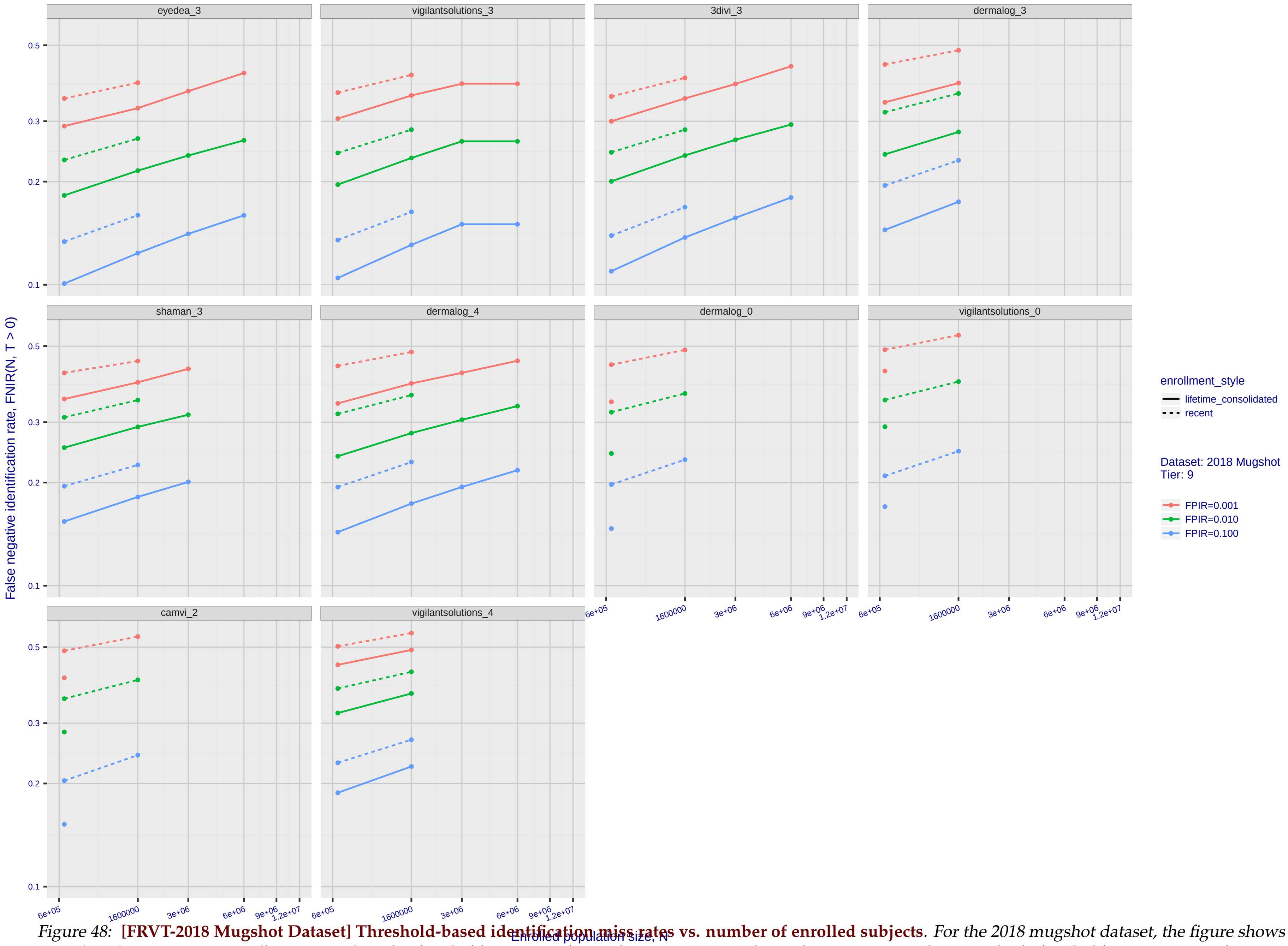

Figure 48: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

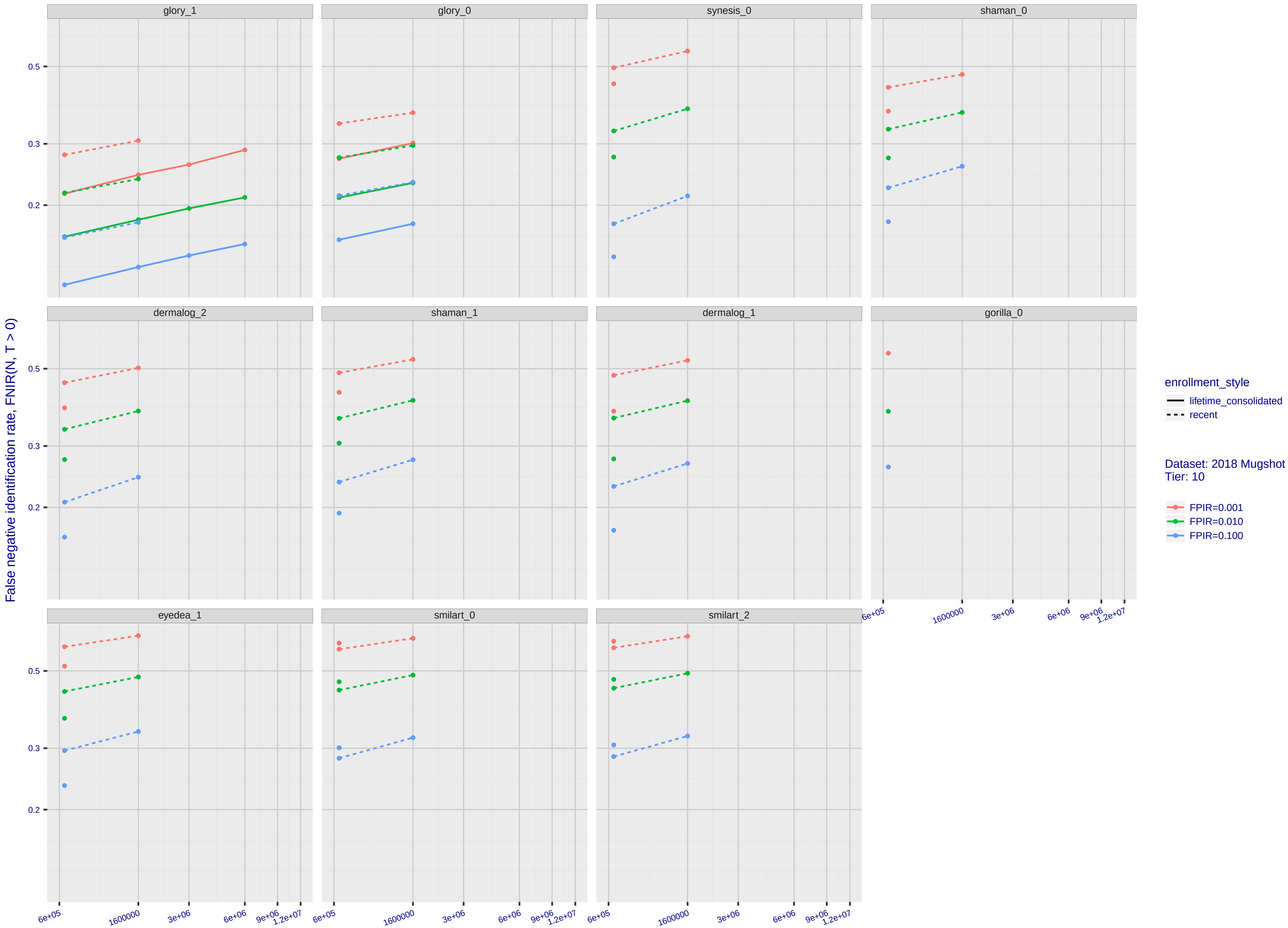

Figure 49: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

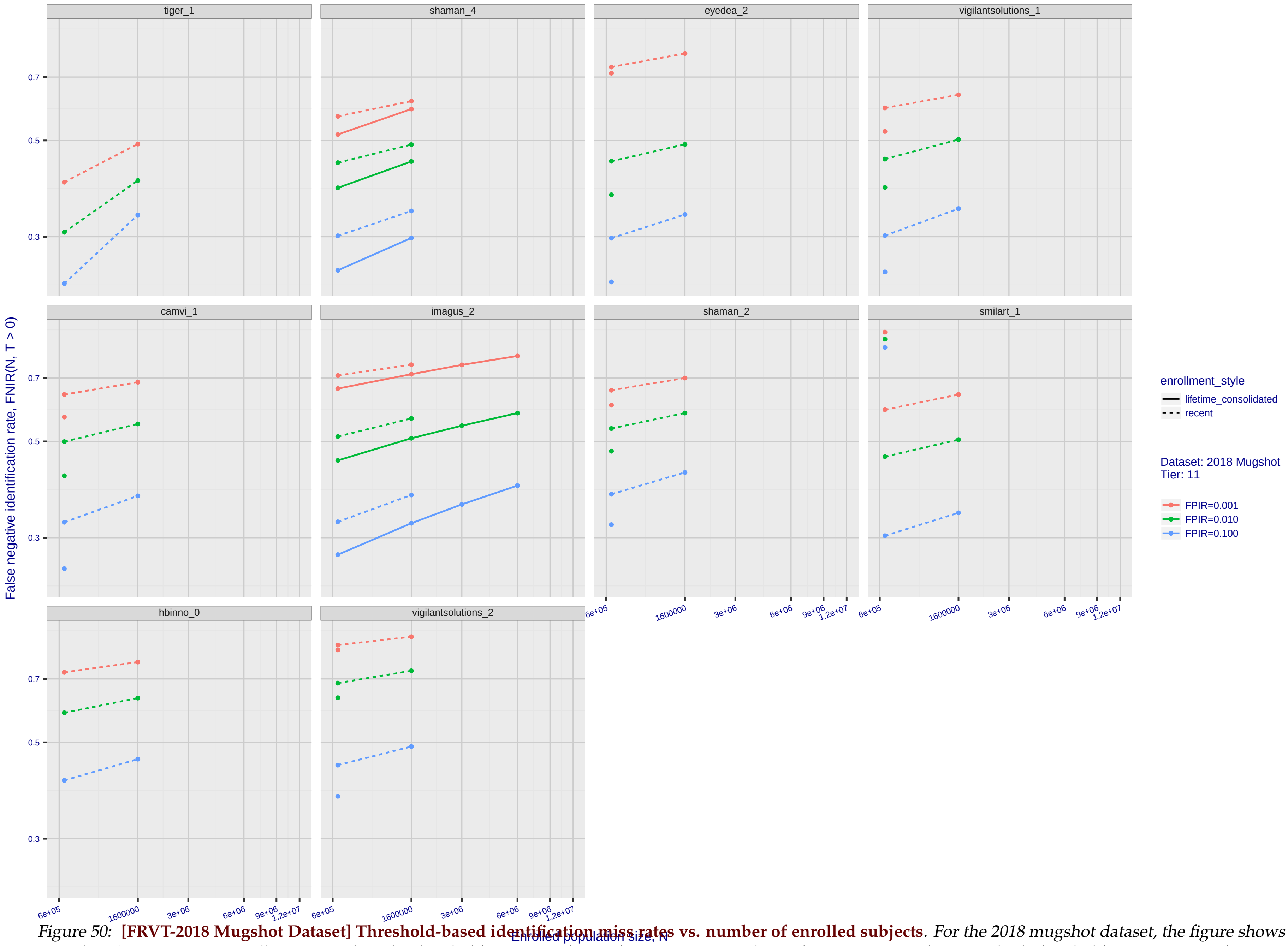

Figure 50: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

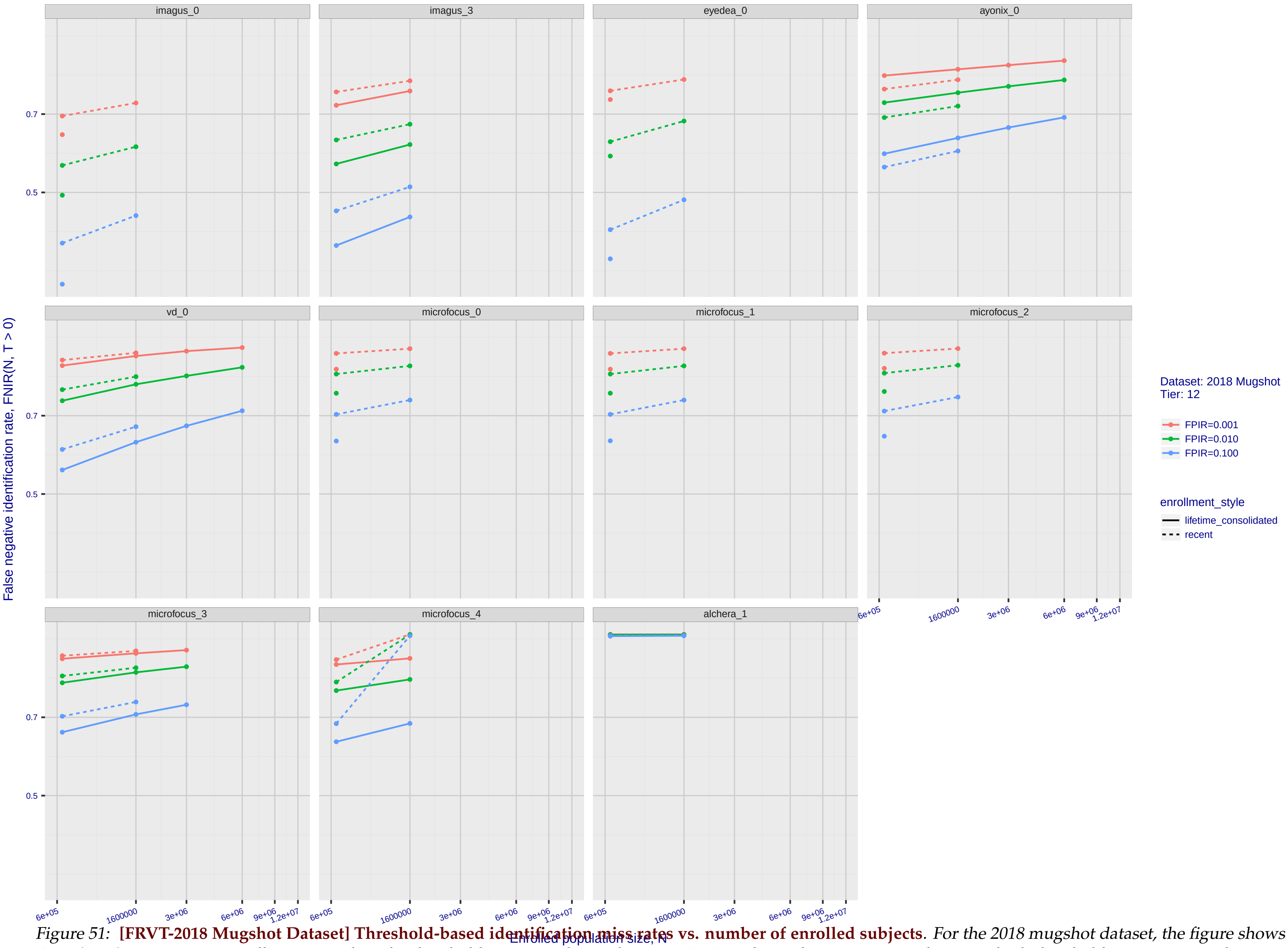

Figure 51: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects**. *For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds as mates are always at rank 1. The results are computed from the trials listed in rows 1-10 of Table 6. Less accurate algorithms were not run on large N, so results are missing. For clarity, results are sorted and reported into tiers spanning multiple pages. The tiering criteria is complicated: First paging by FNIR(N_b, 1, 0), then sorting by median FNIR(N_b, T), N_b = 640 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

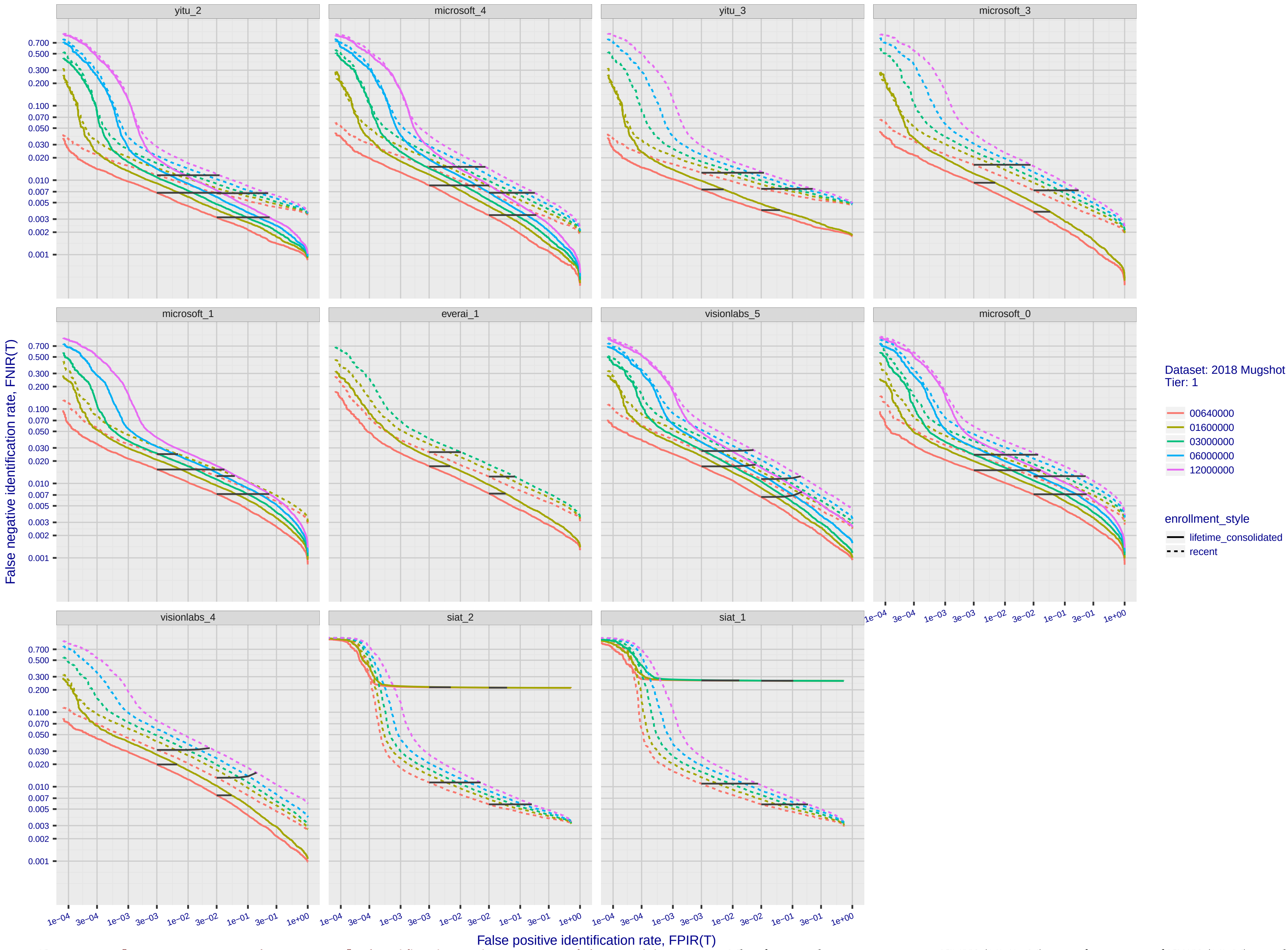

Figure 52: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8238

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

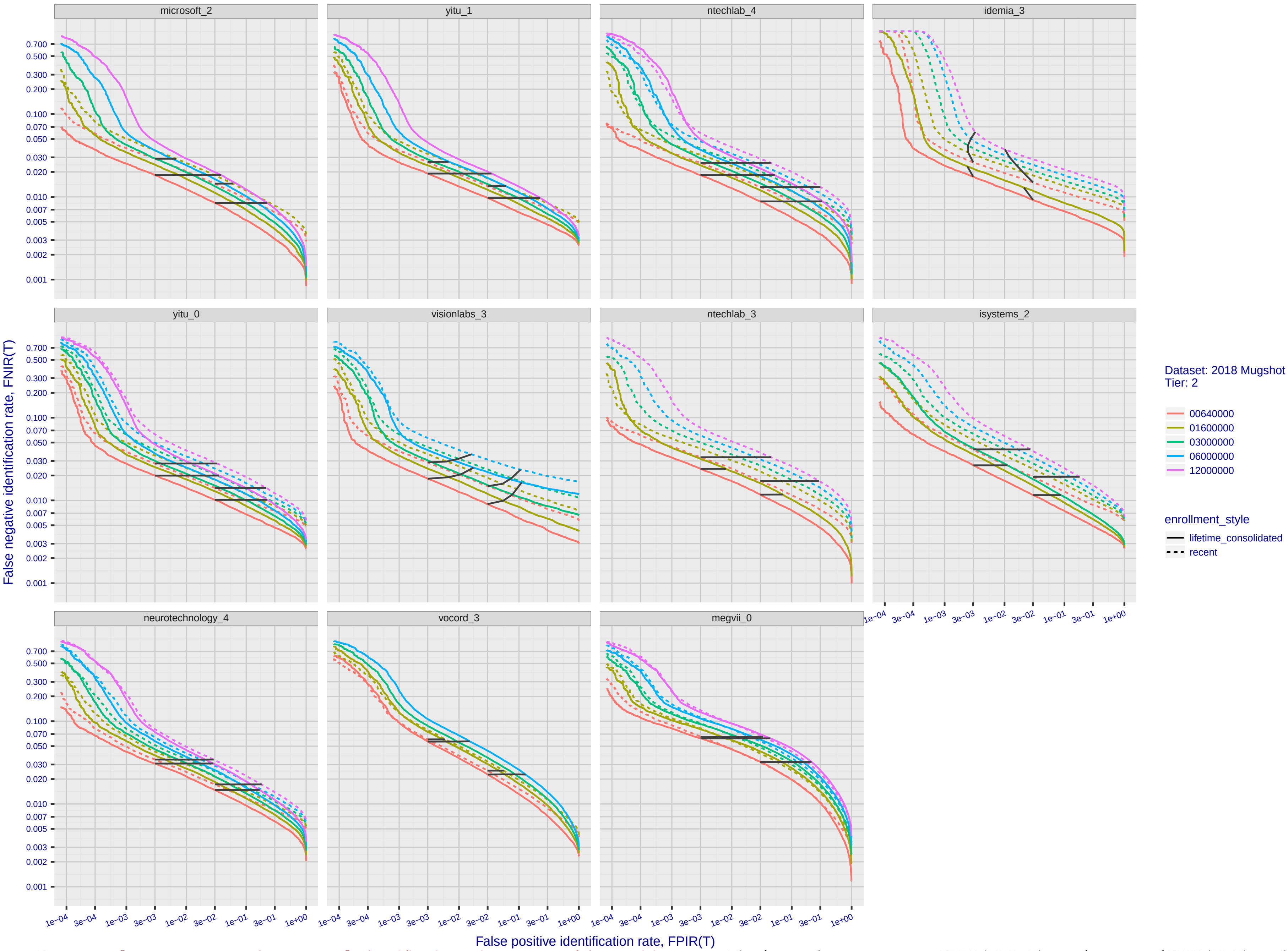

Figure 53: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

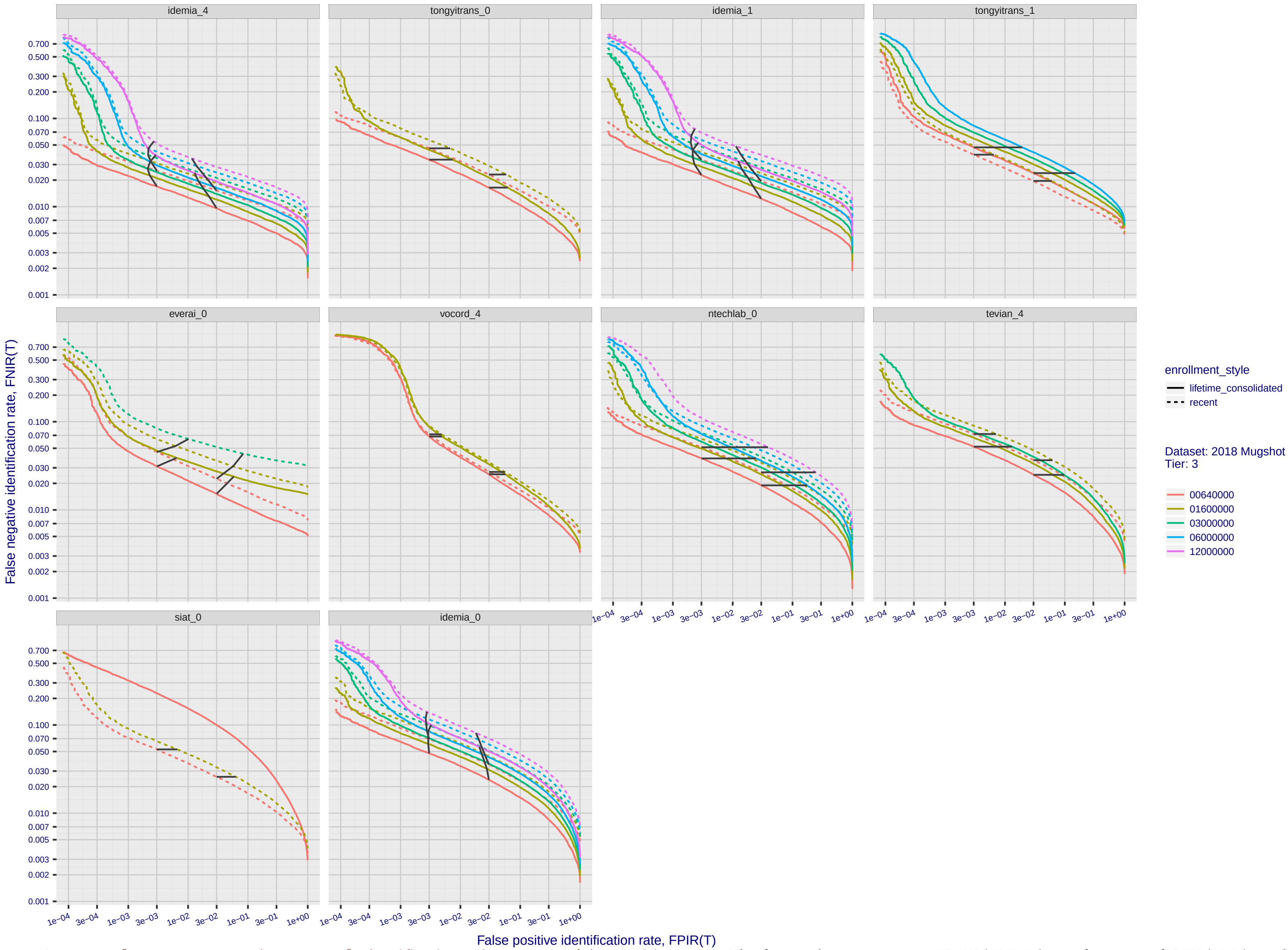

Figure 54: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

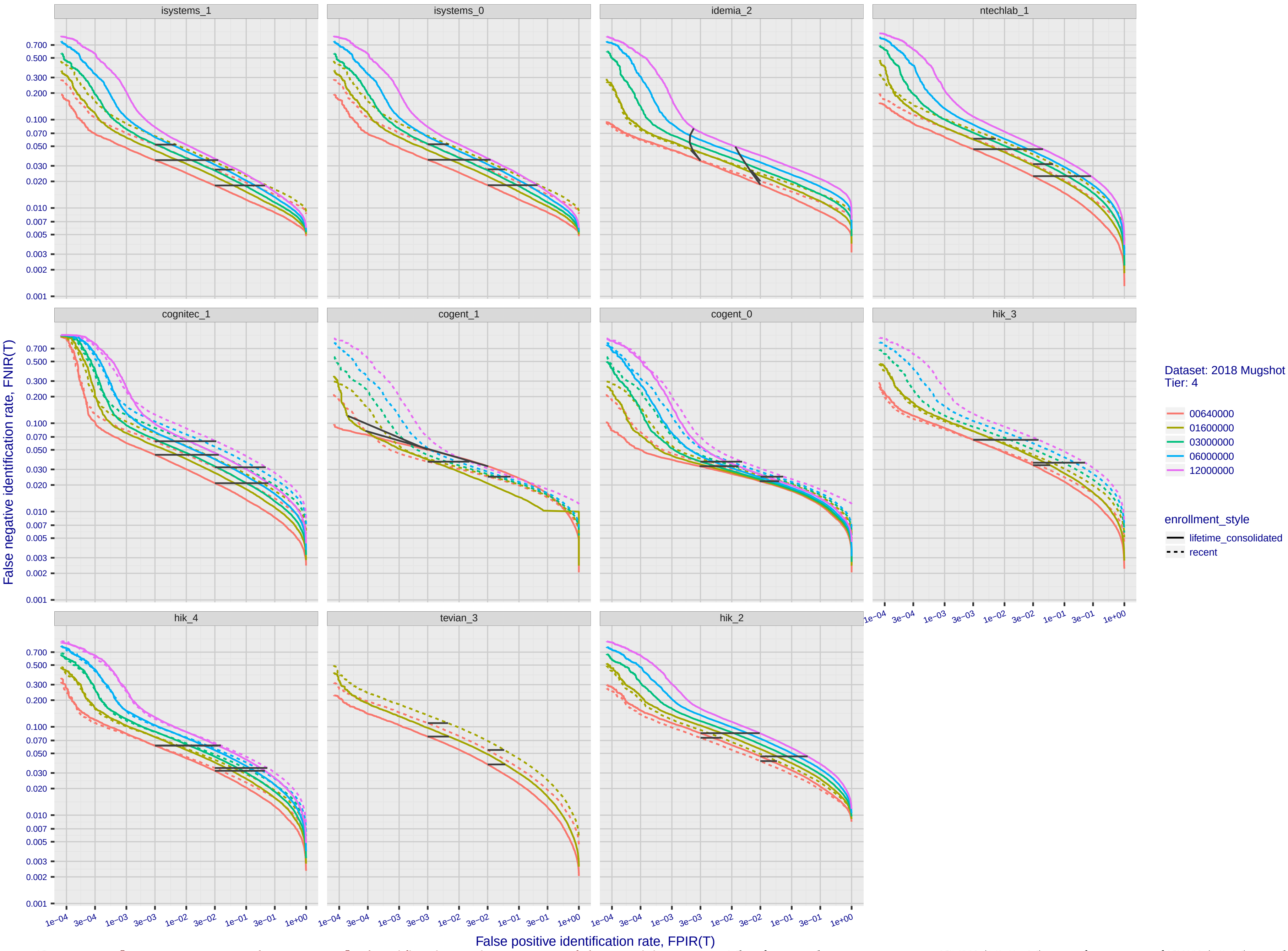

Figure 55: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

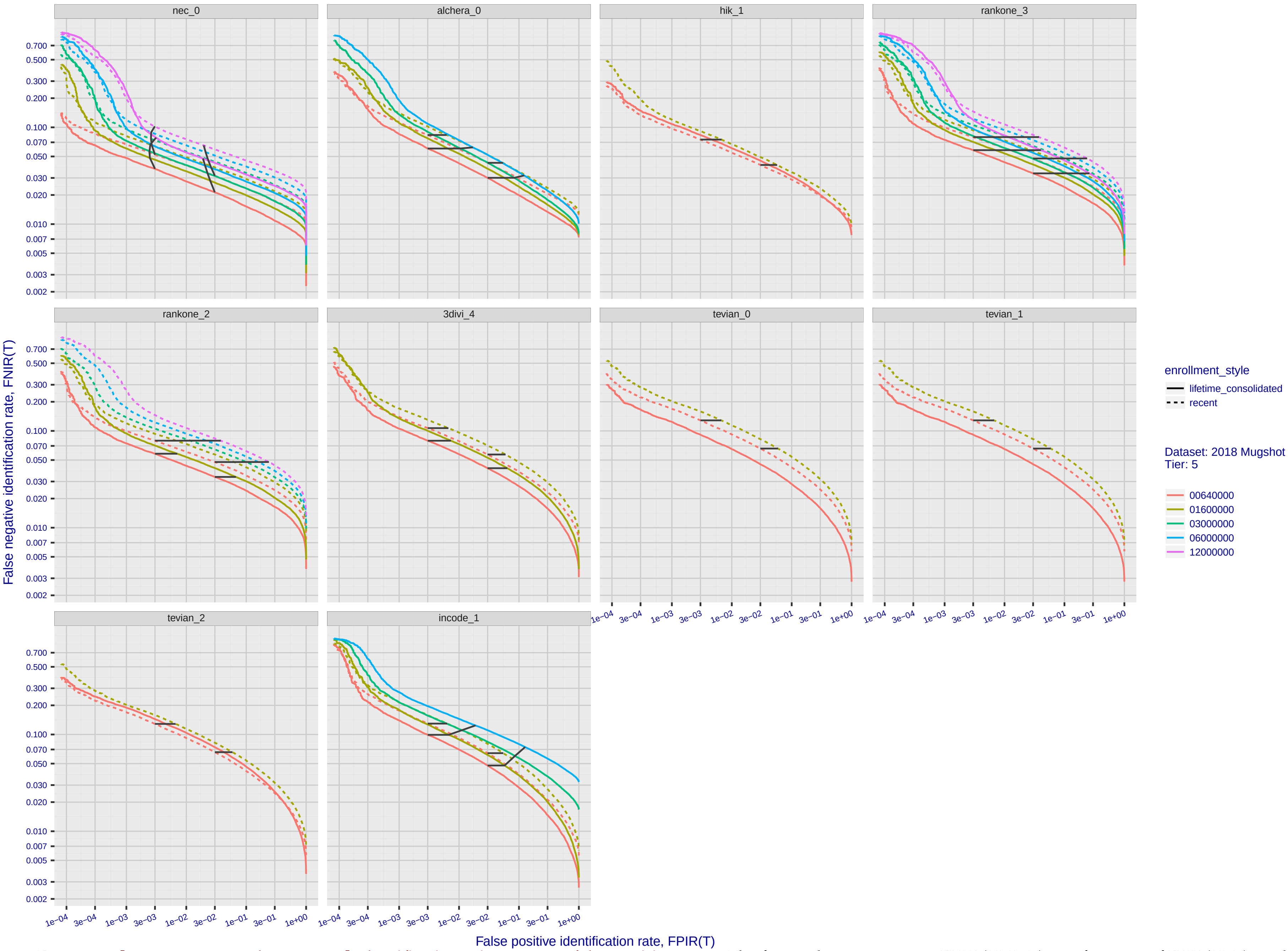

Figure 56: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

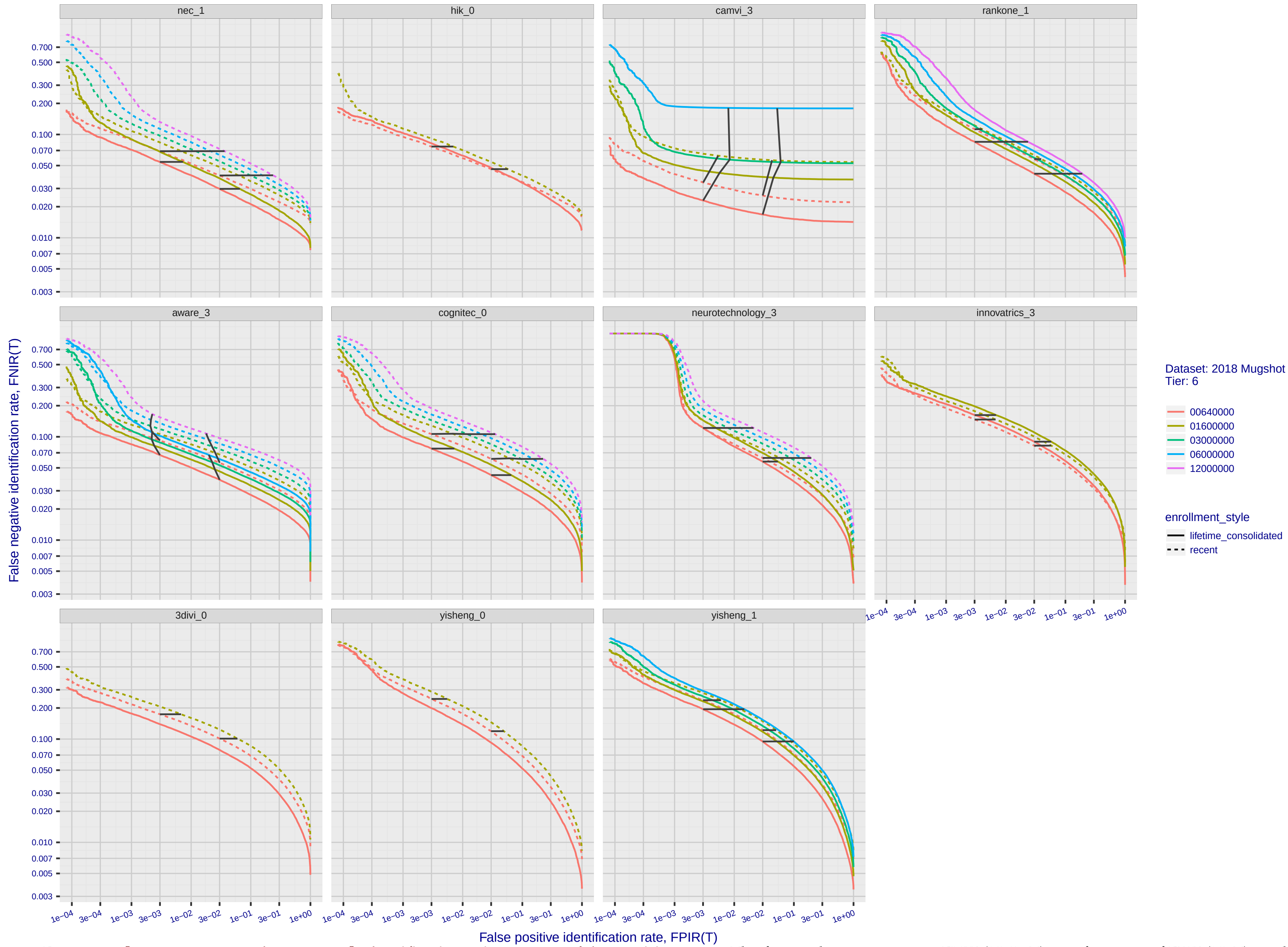

Figure 57: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

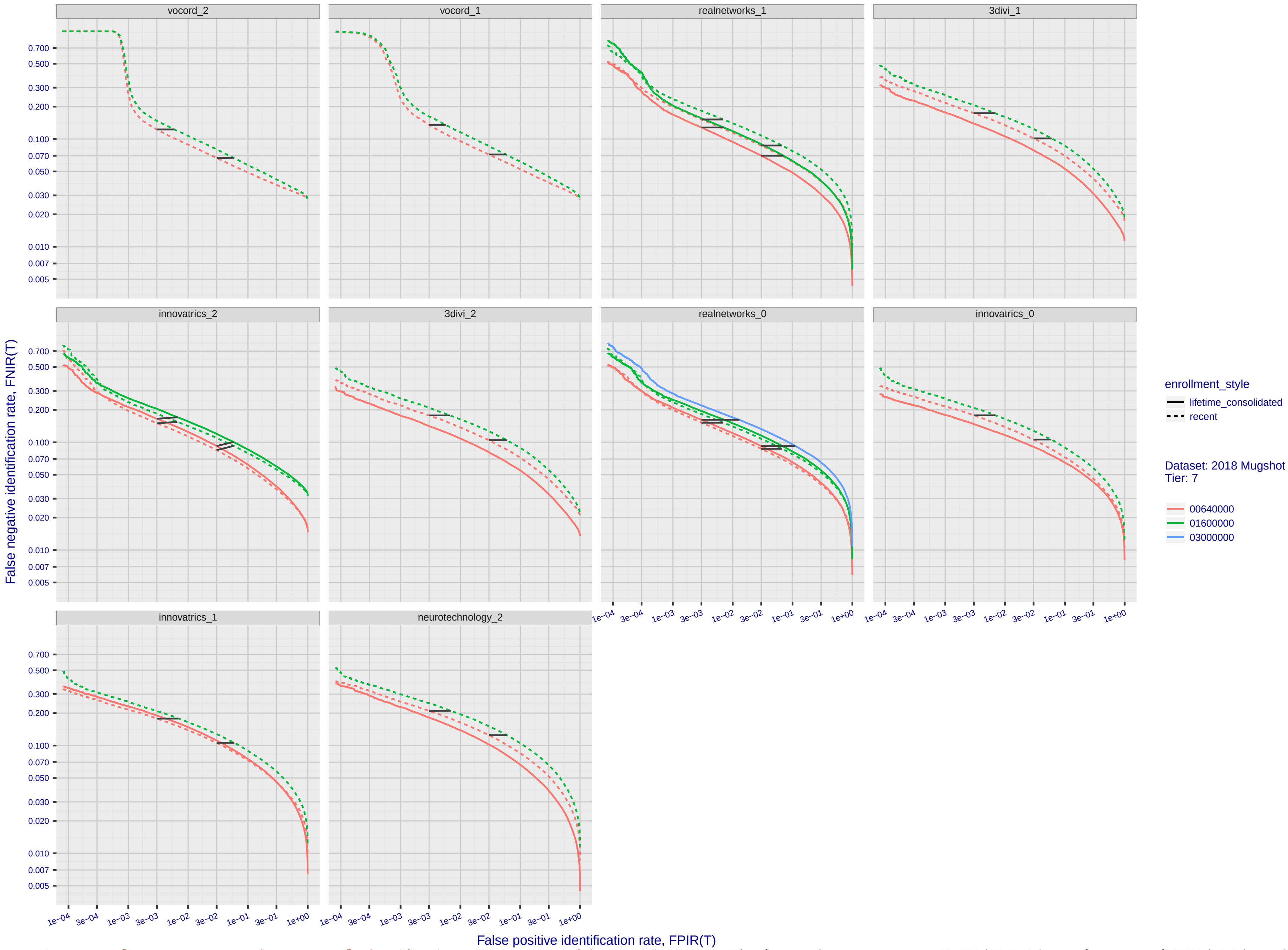

Figure 58: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

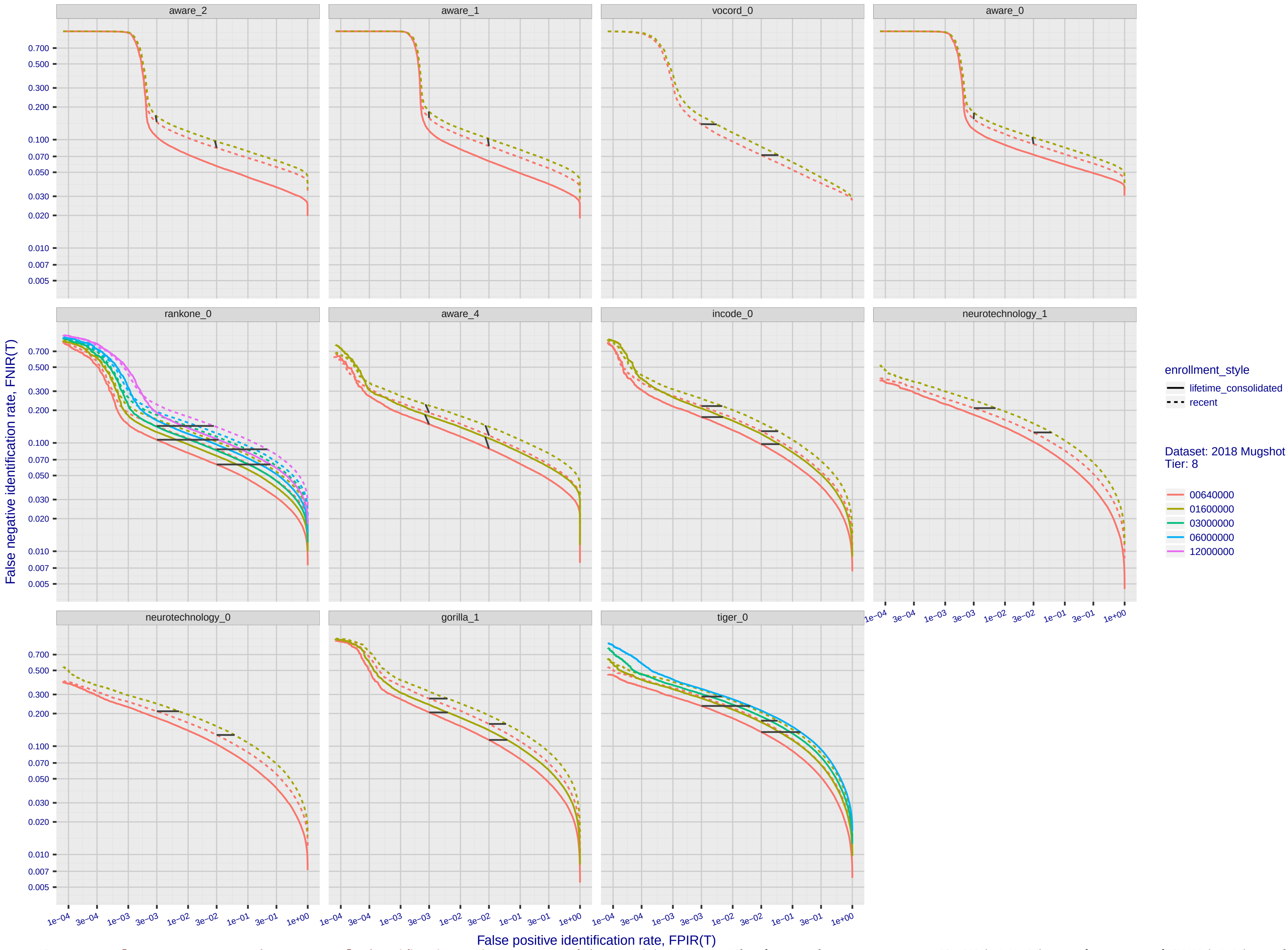

Figure 59: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate N = Num. enrolled subjects T = Threshold T = 0 → Investigation
FPIR(N, T) = False pos. identification rate R = Num. candidates examined T > 0 → Identification

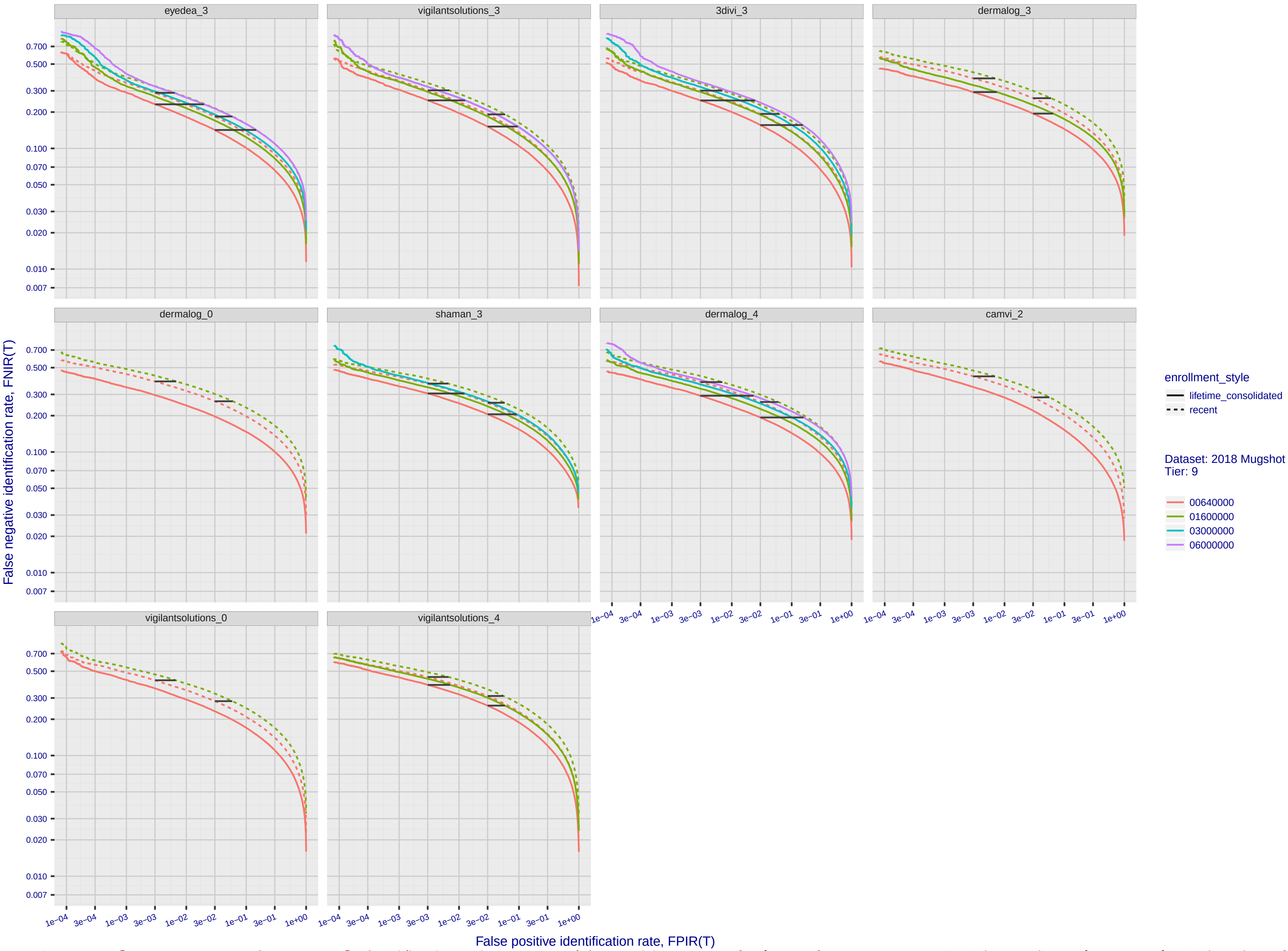

Figure 60: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

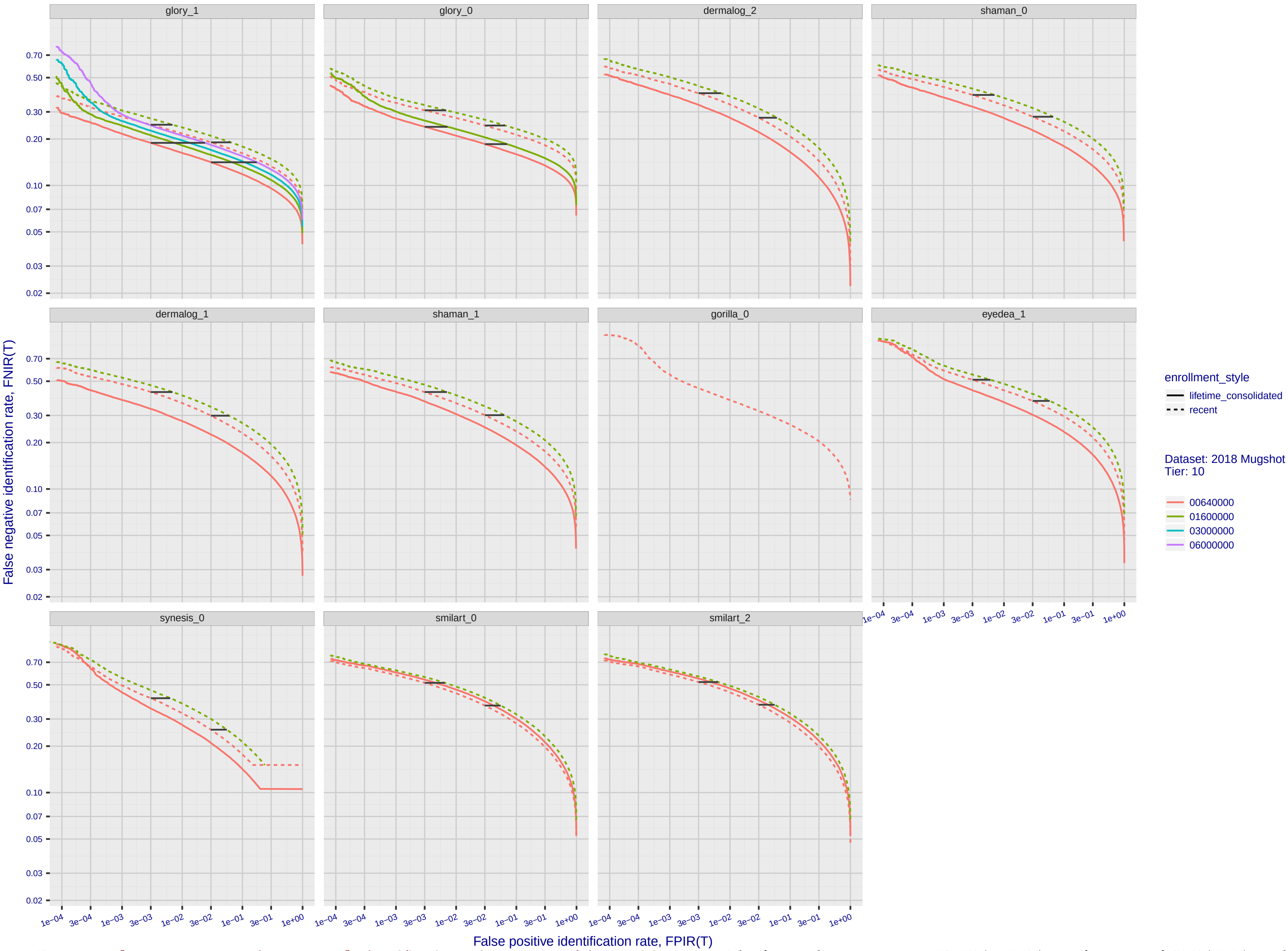

Figure 61: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

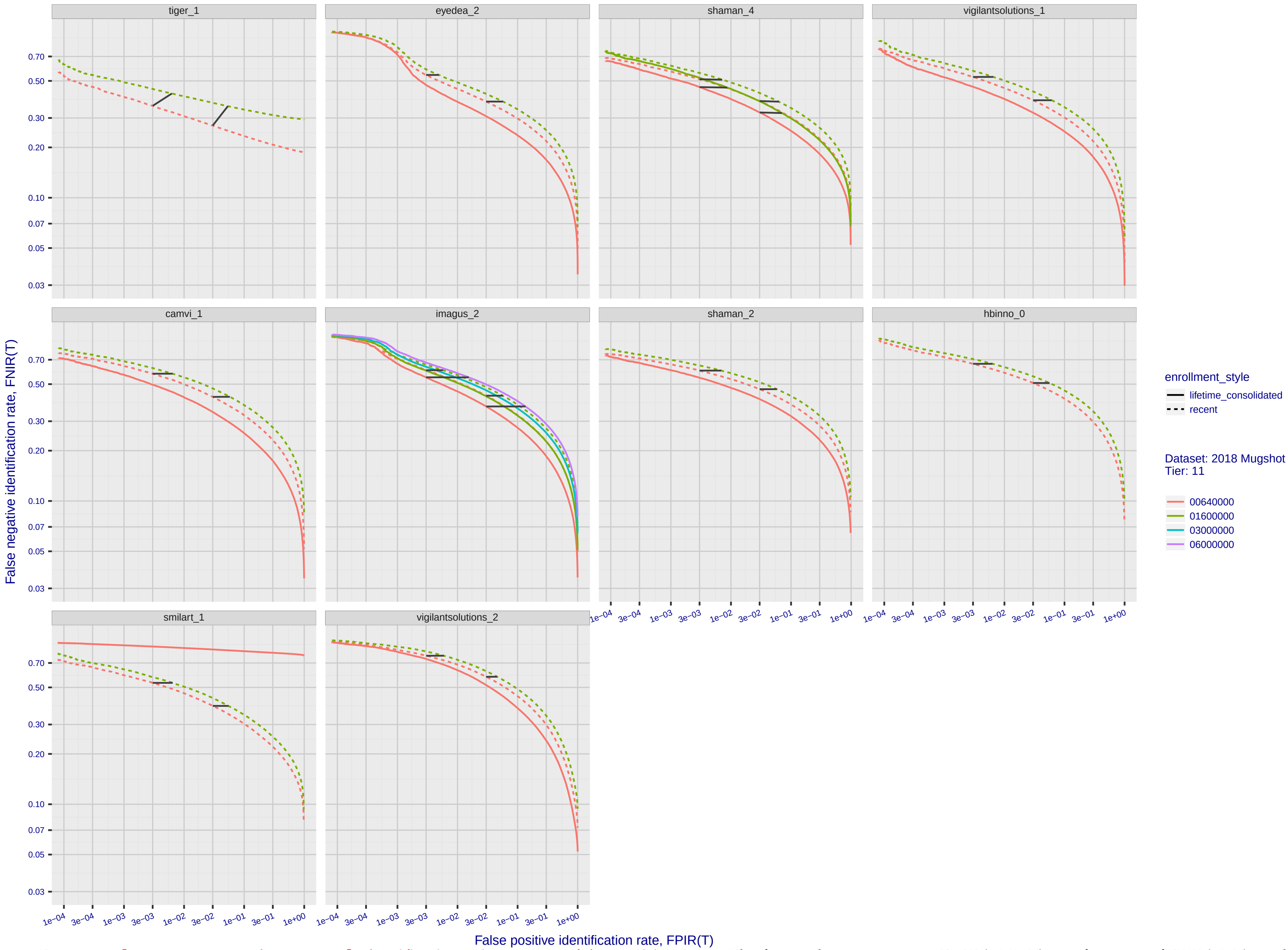

Figure 62: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8238

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate N = Num. enrolled subjects T = Threshold T = 0 → Investigation
FPIR(N, T) = False pos. identification rate R = Num. candidates examined T > 0 → Identification

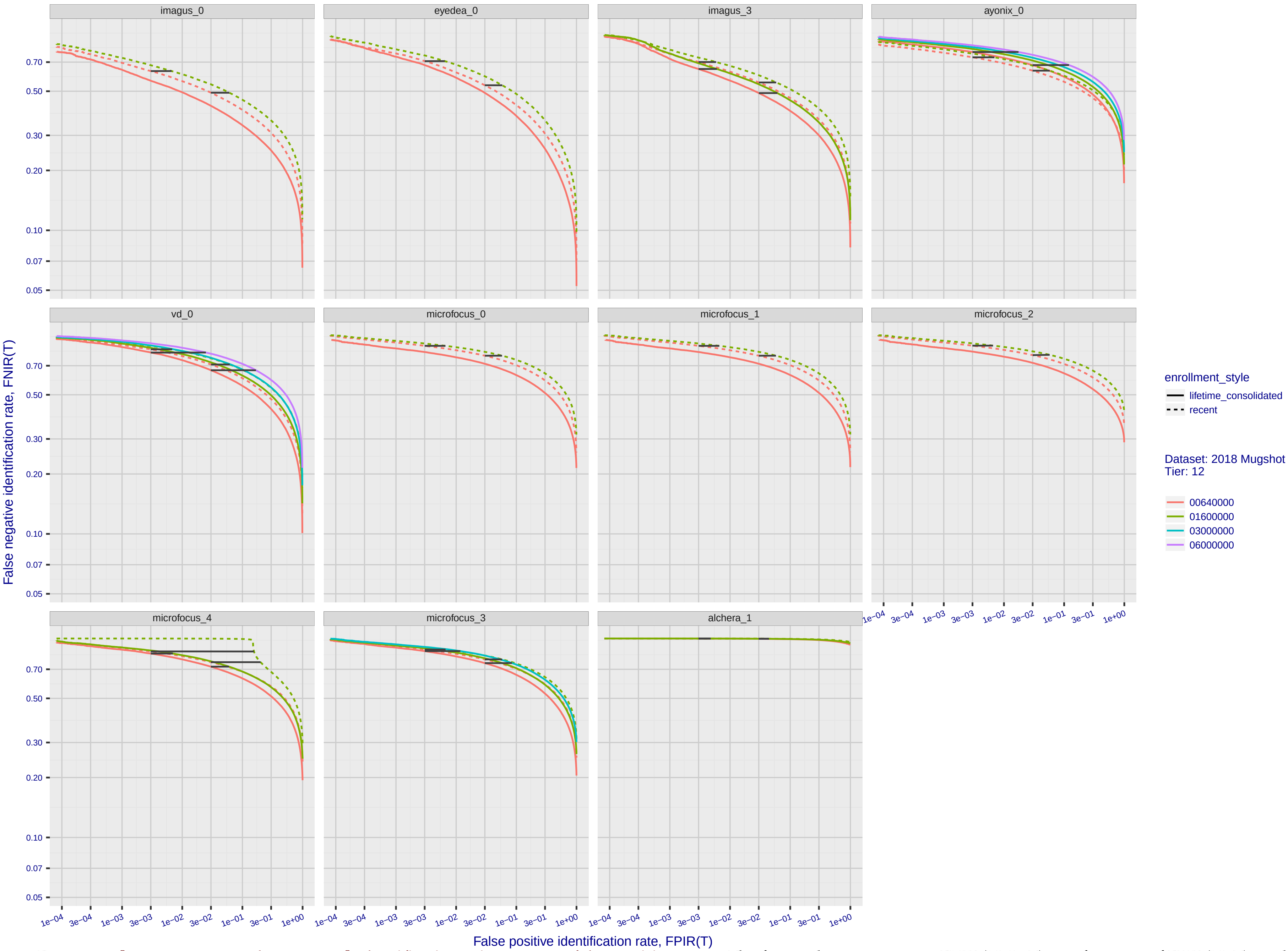

Figure 63: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates**. *The figure shows miss rates FNIR(N, L, T) as a function of FPIR(N, T), with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications where a threshold must be elevated to limit false positives, such as when human reviewer labor is not matched to the volume of searches. Dark lines join points of equal threshold: If horizontal, FPIR(T) rises with N, and mate scores are independent of N. Other algorithms adjust scores in an attempt to make FPIR independent of N.*

Appendix B Effect of time-lapse: Accuracy after face ageing

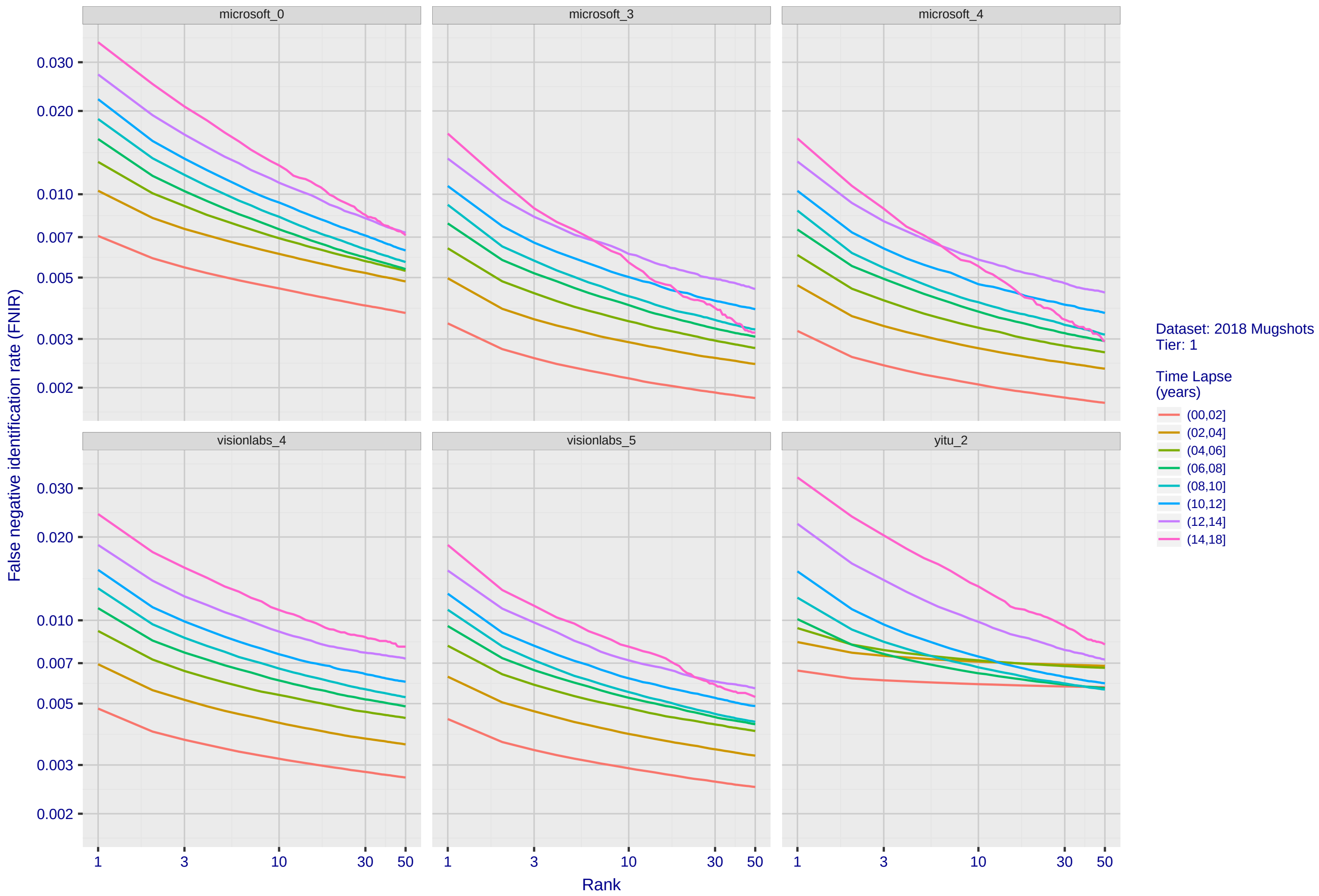

Figure 64: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. rank by time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between search and initial enrollment.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

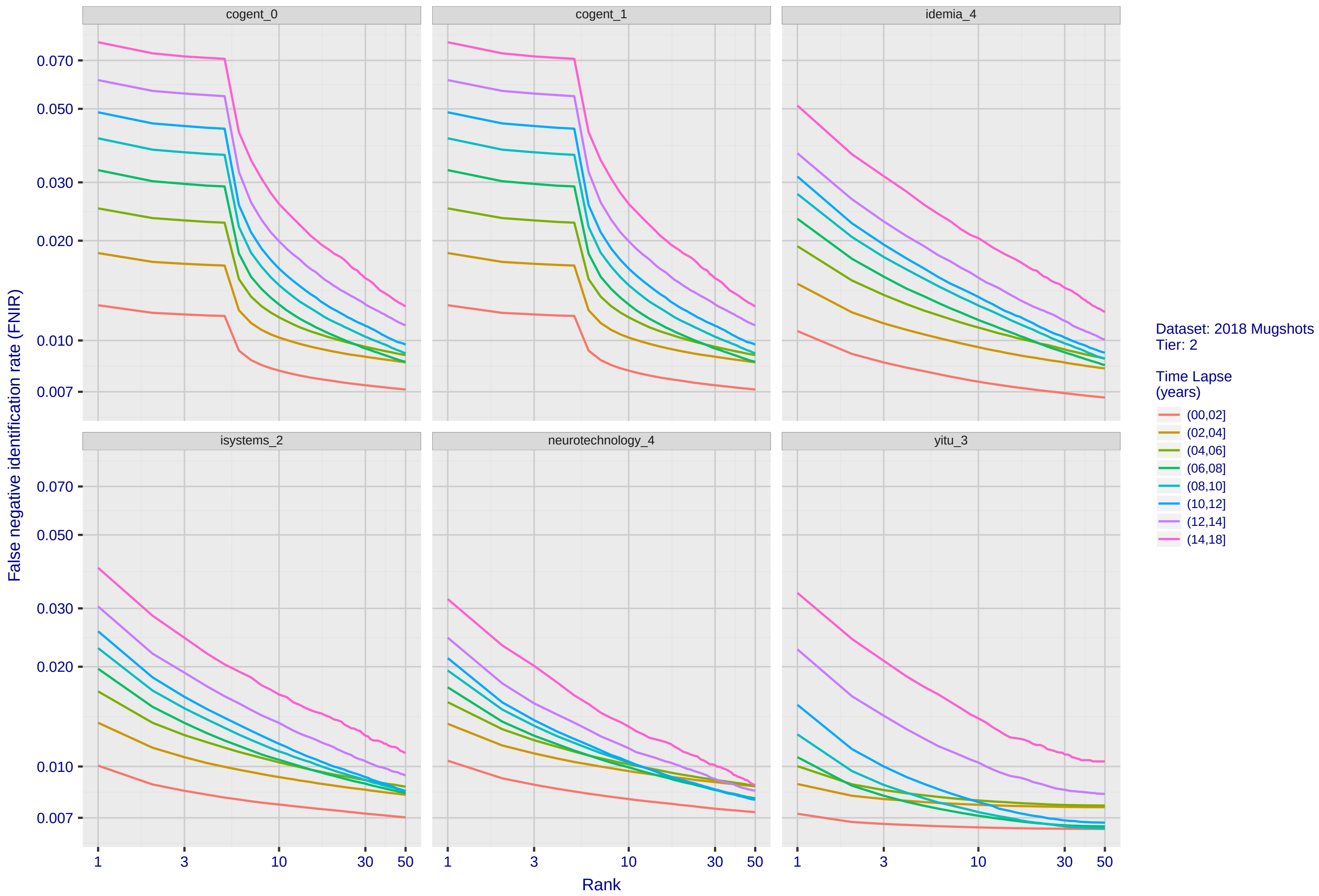

Figure 65: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. rank by time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between search and initial enrollment.*

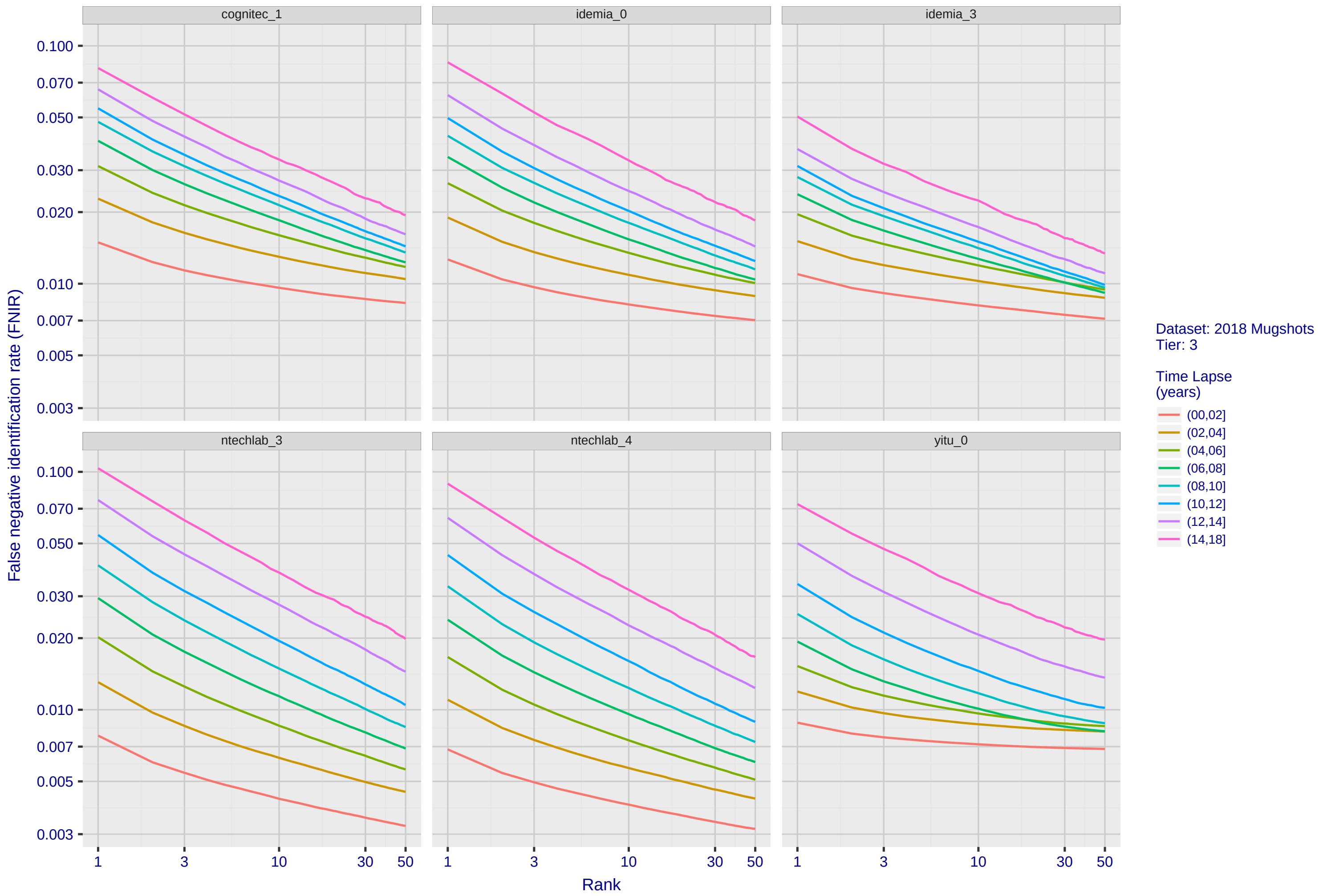

Figure 66: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. rank by time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between search and initial enrollment.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

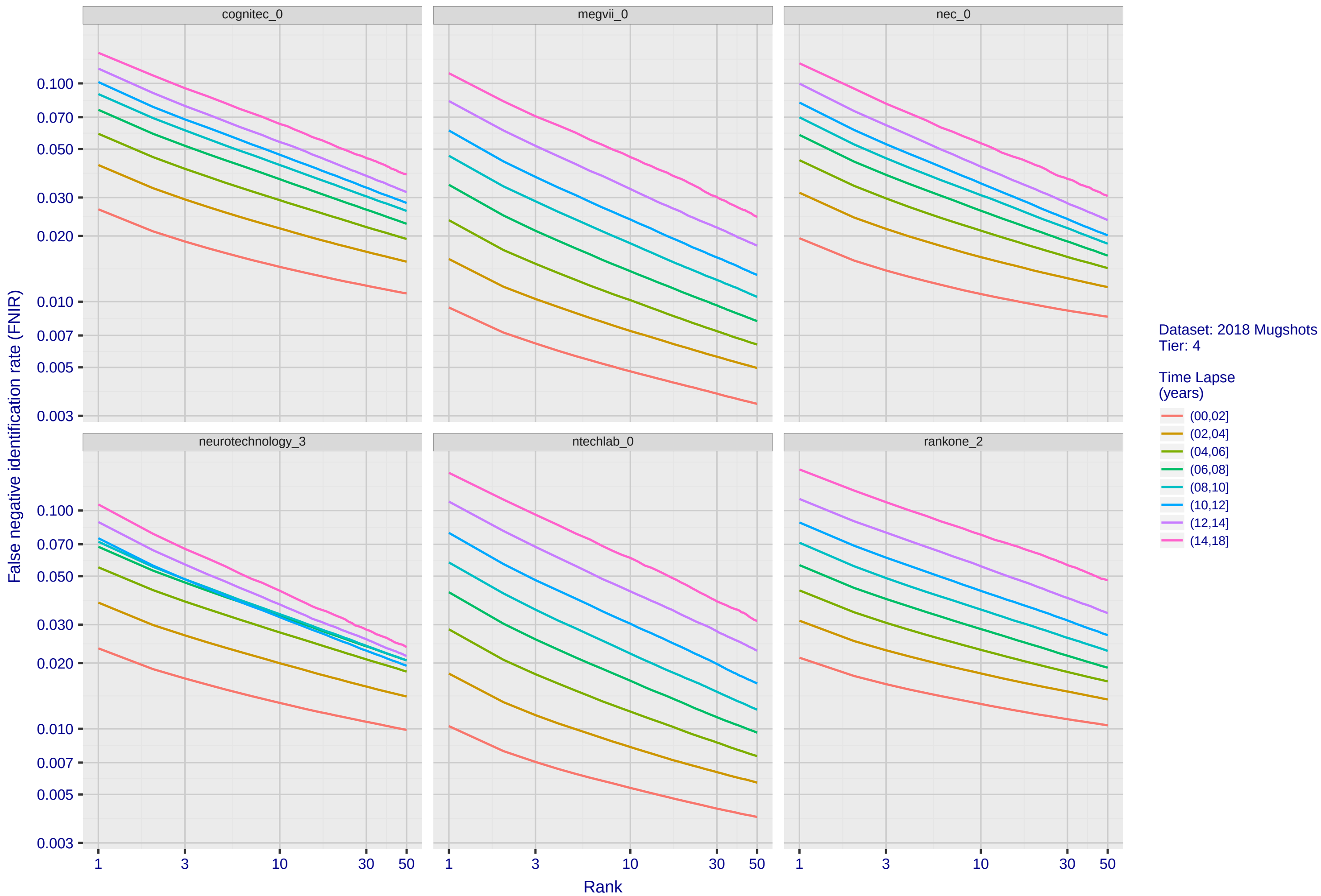

Figure 67: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. rank by time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between search and initial enrollment.*

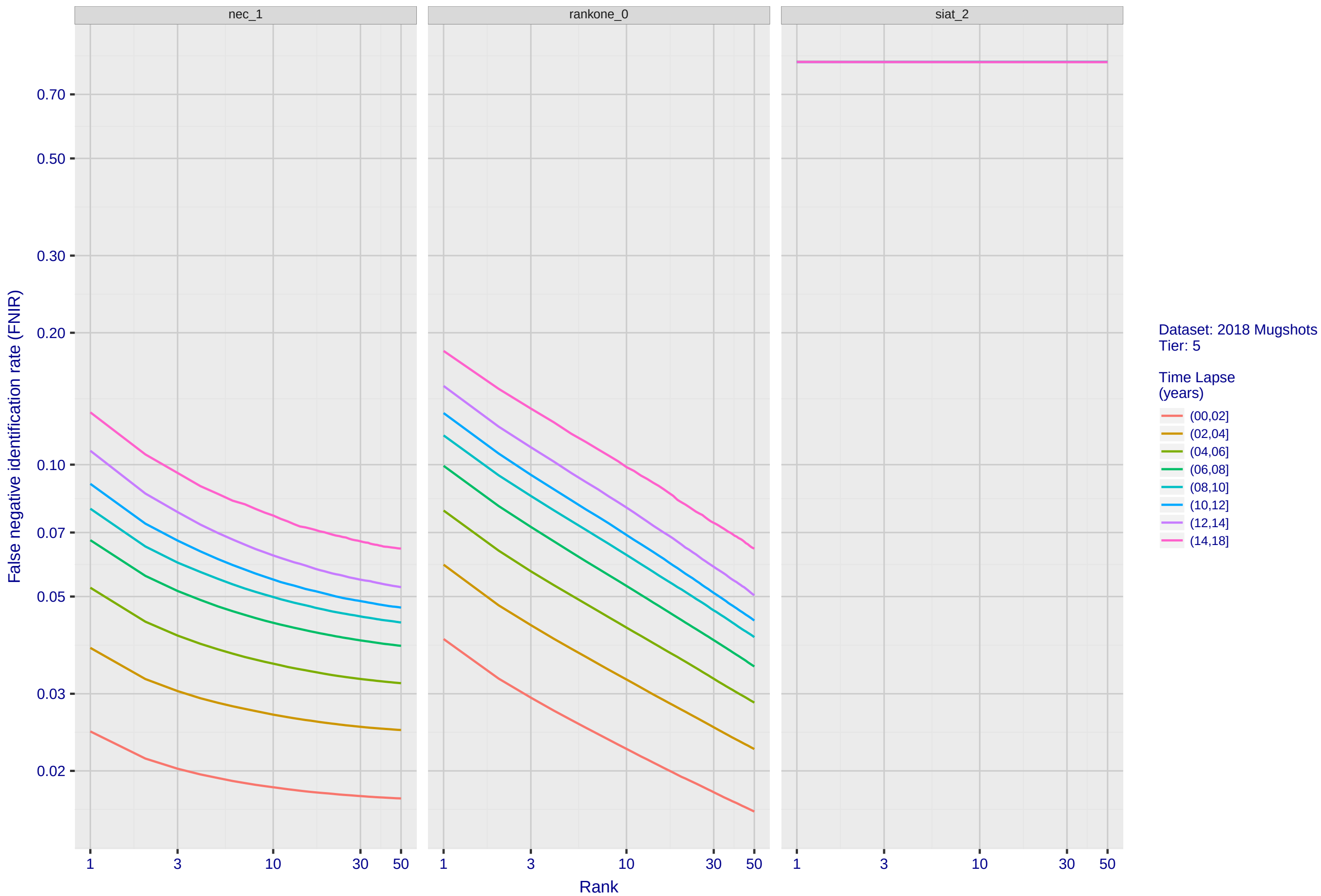

Figure 68: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. rank by time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between search and initial enrollment.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

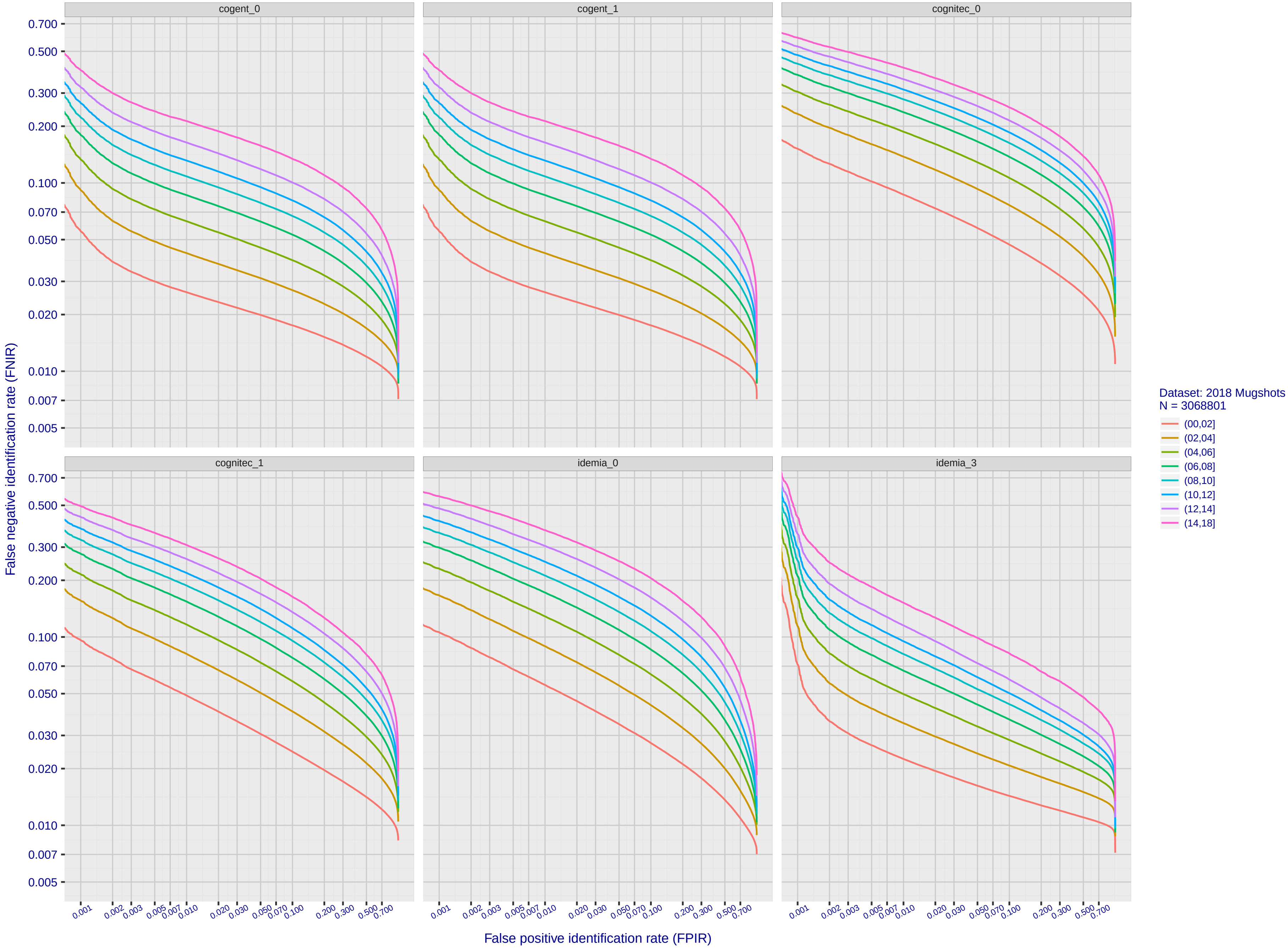

Figure 69: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. FPIR by time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between search and initial enrollment. FPIR is computed from the same FRVT 2018 non-mates noted in row 3 of Table 6 with N = 3 000 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

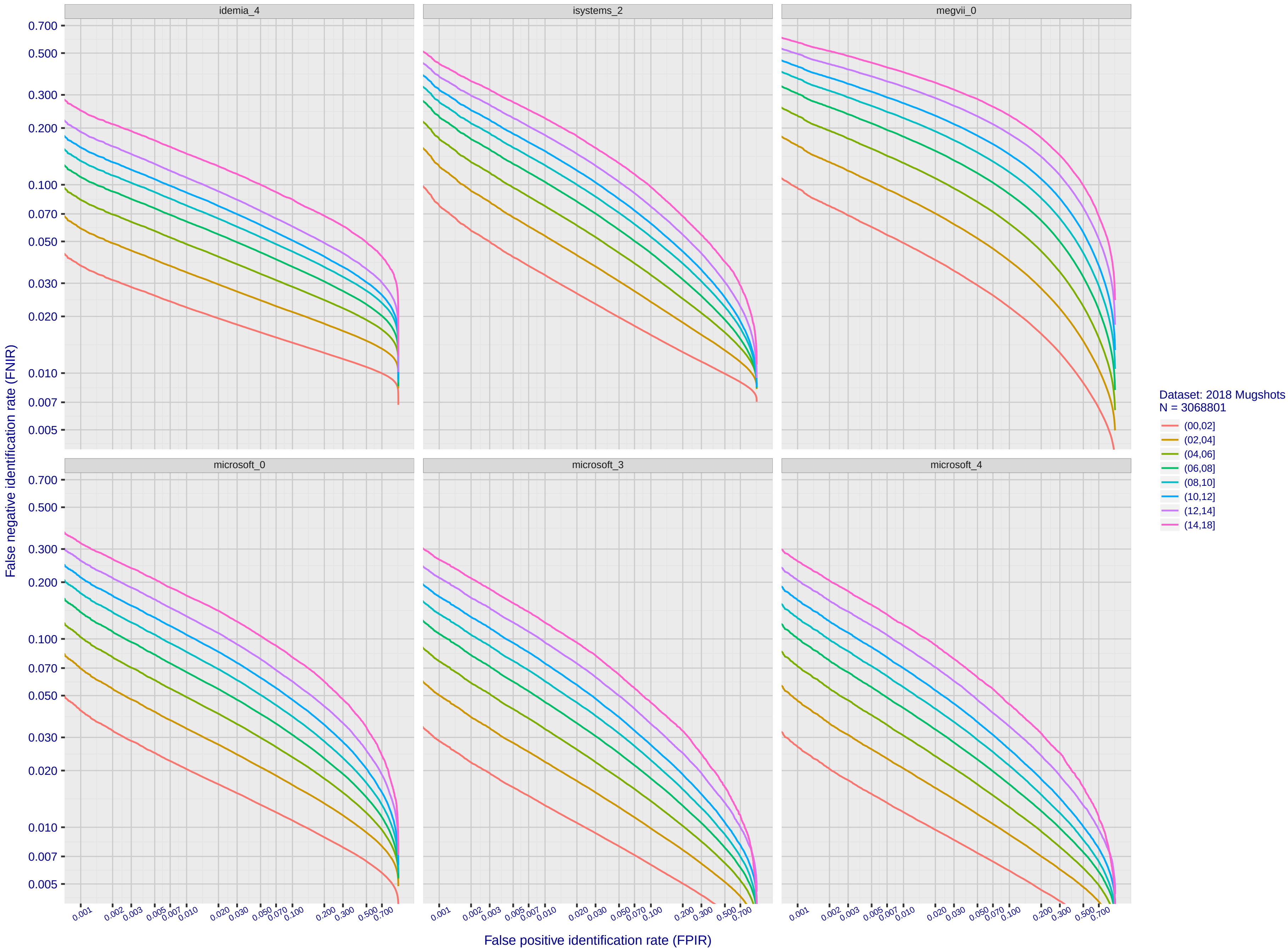

Figure 70: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. FPIR by time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between search and initial enrollment. FPIR is computed from the same FRVT 2018 non-mates noted in row 3 of Table 6 with N = 3 000 000.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

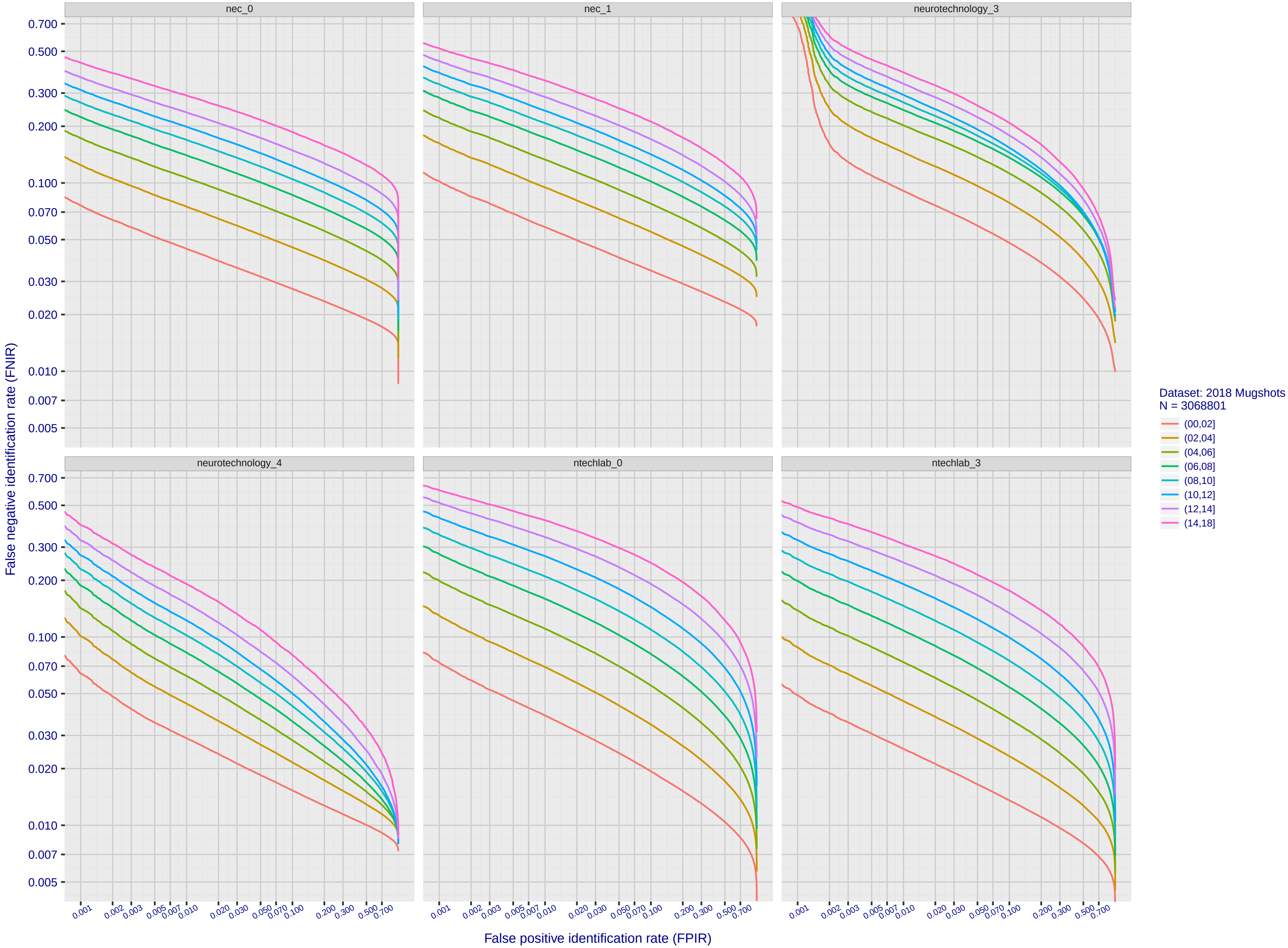

Figure 71: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. FPIR by time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between search and initial enrollment. FPIR is computed from the same FRVT 2018 non-mates noted in row 3 of Table 6 with N = 3 000 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

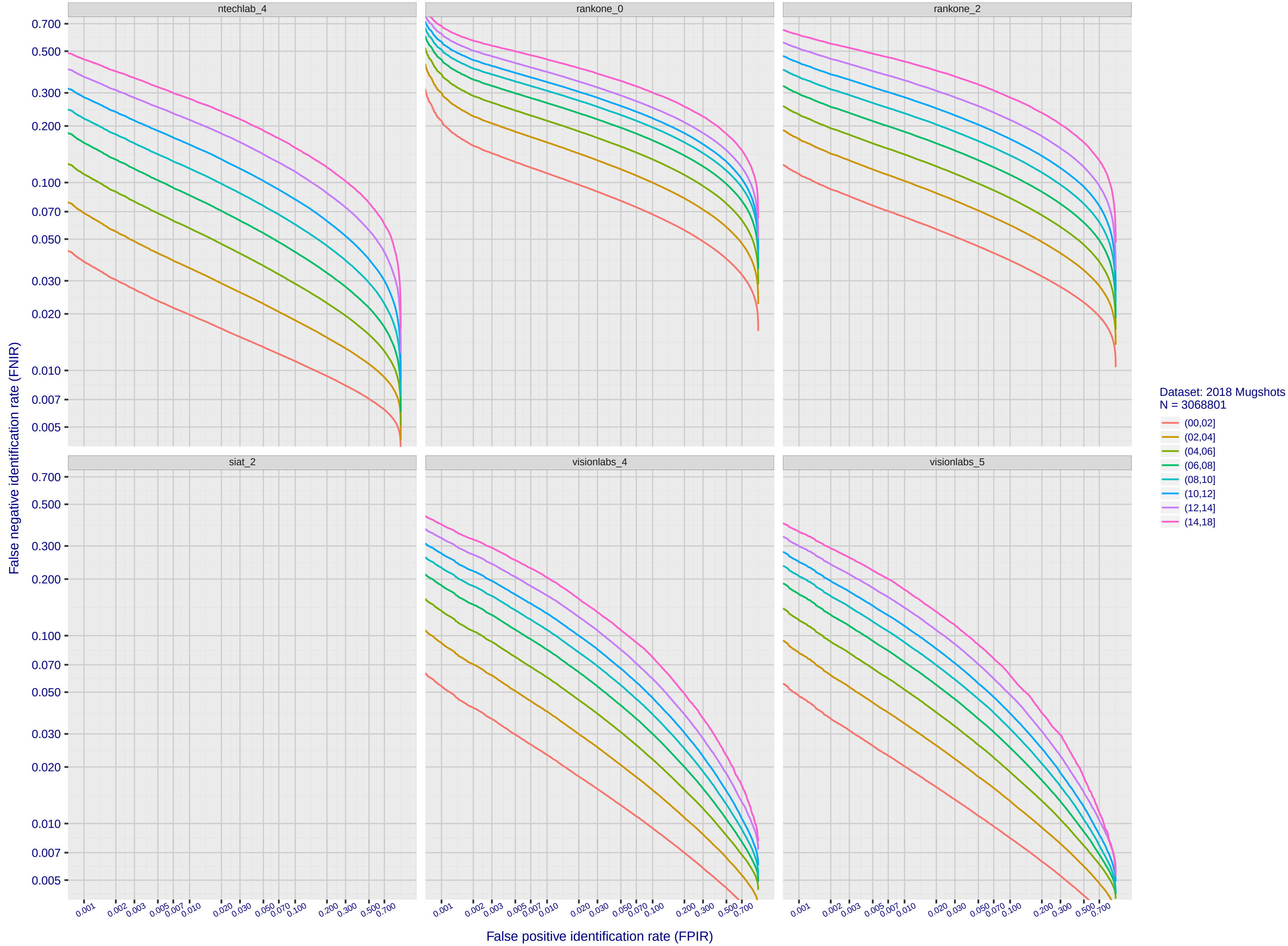

Figure 72: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. FPIR by time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between search and initial enrollment. FPIR is computed from the same FRVT 2018 non-mates noted in row 3 of Table 6 with N = 3 000 000.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

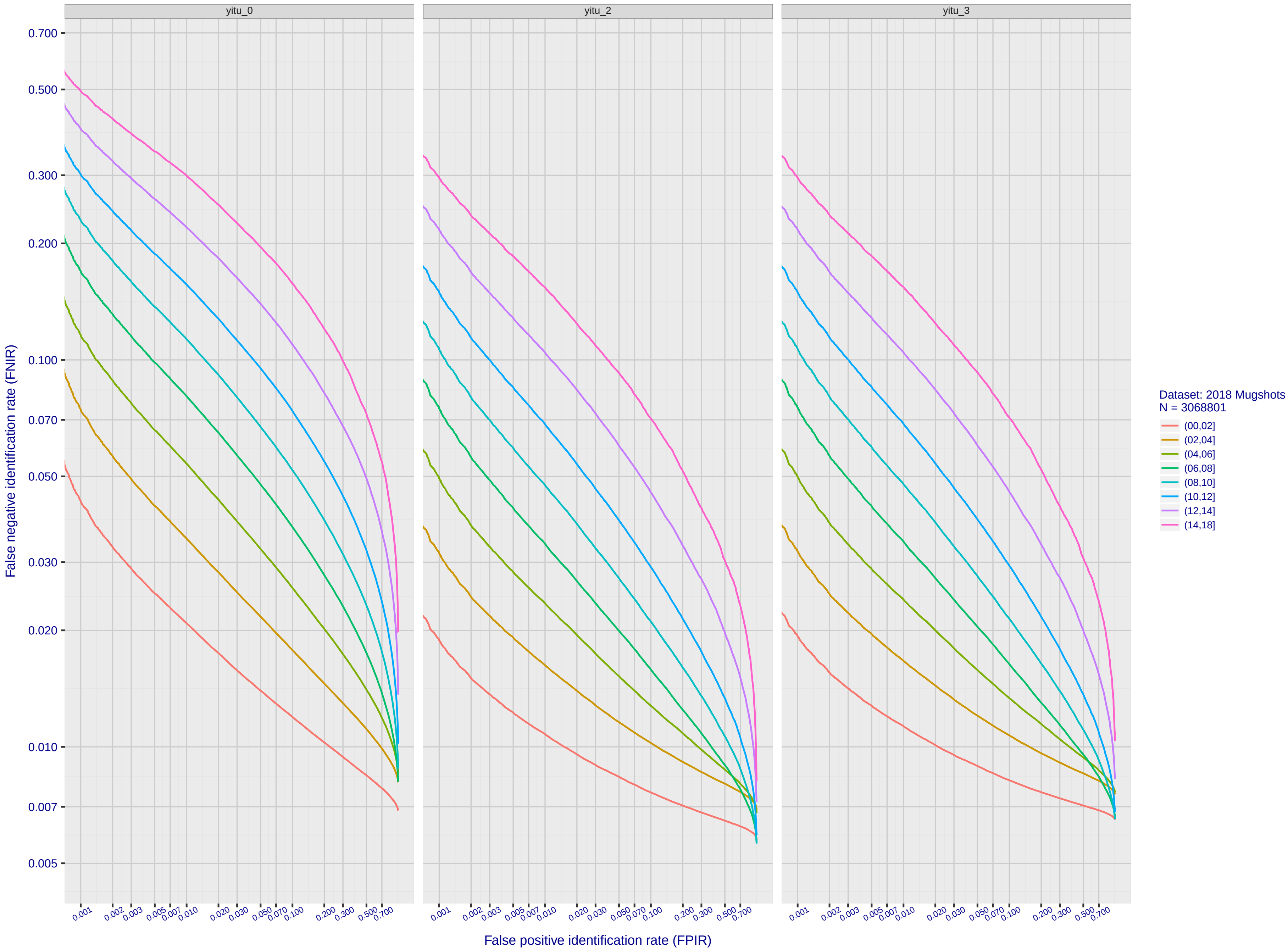

Figure 73: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. FPIR by time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between search and initial enrollment. FPIR is computed from the same FRVT 2018 non-mates noted in row 3 of Table 6 with N = 3 000 000.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

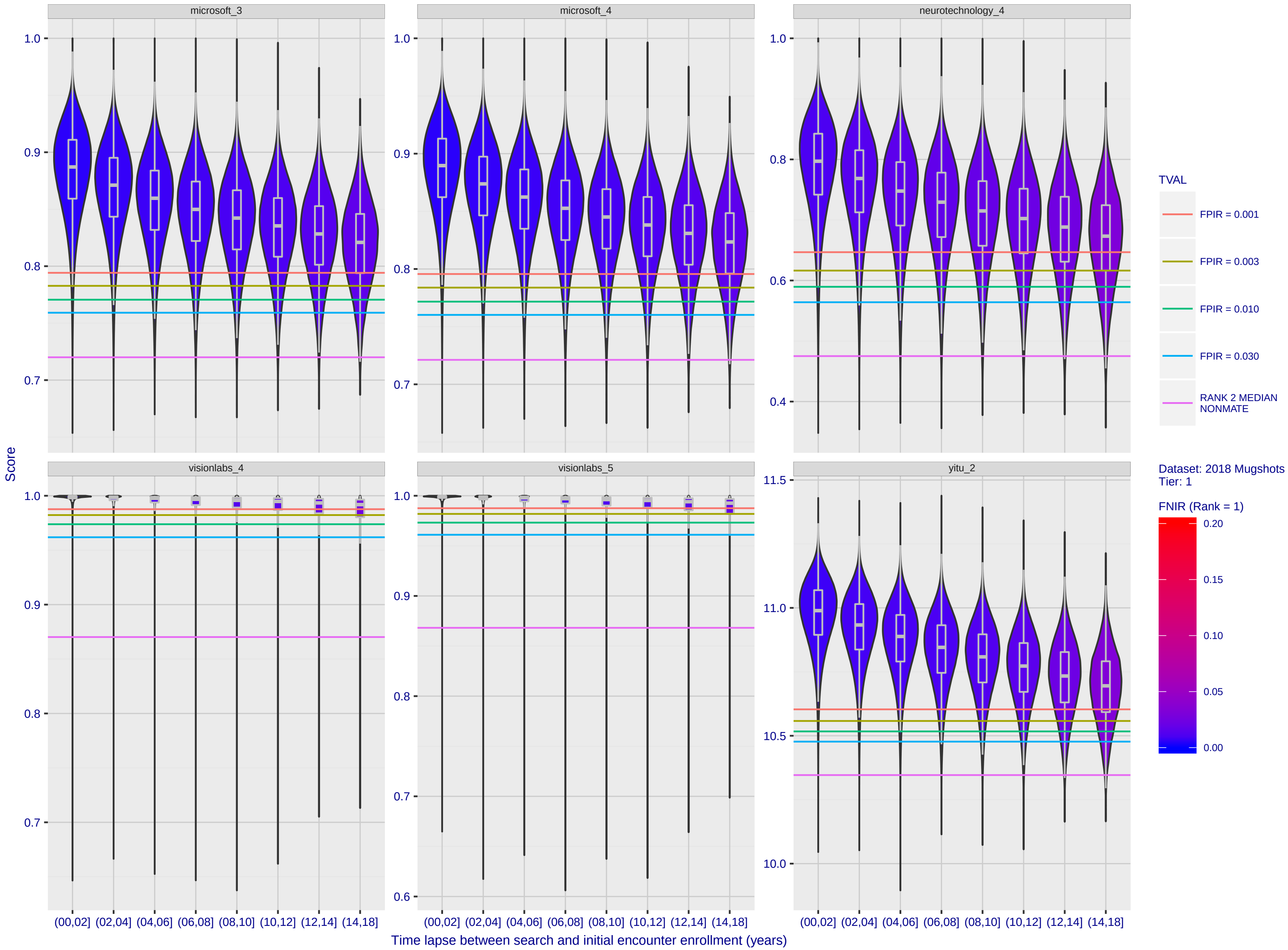

Figure 74: **[FRVT-2018 Mugshot Ageing Dataset] Native mate scores vs. time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Mated score distributions are computed over all searches noted in row 17 of Table 6 binned by number of years between search and initial enrollment.*

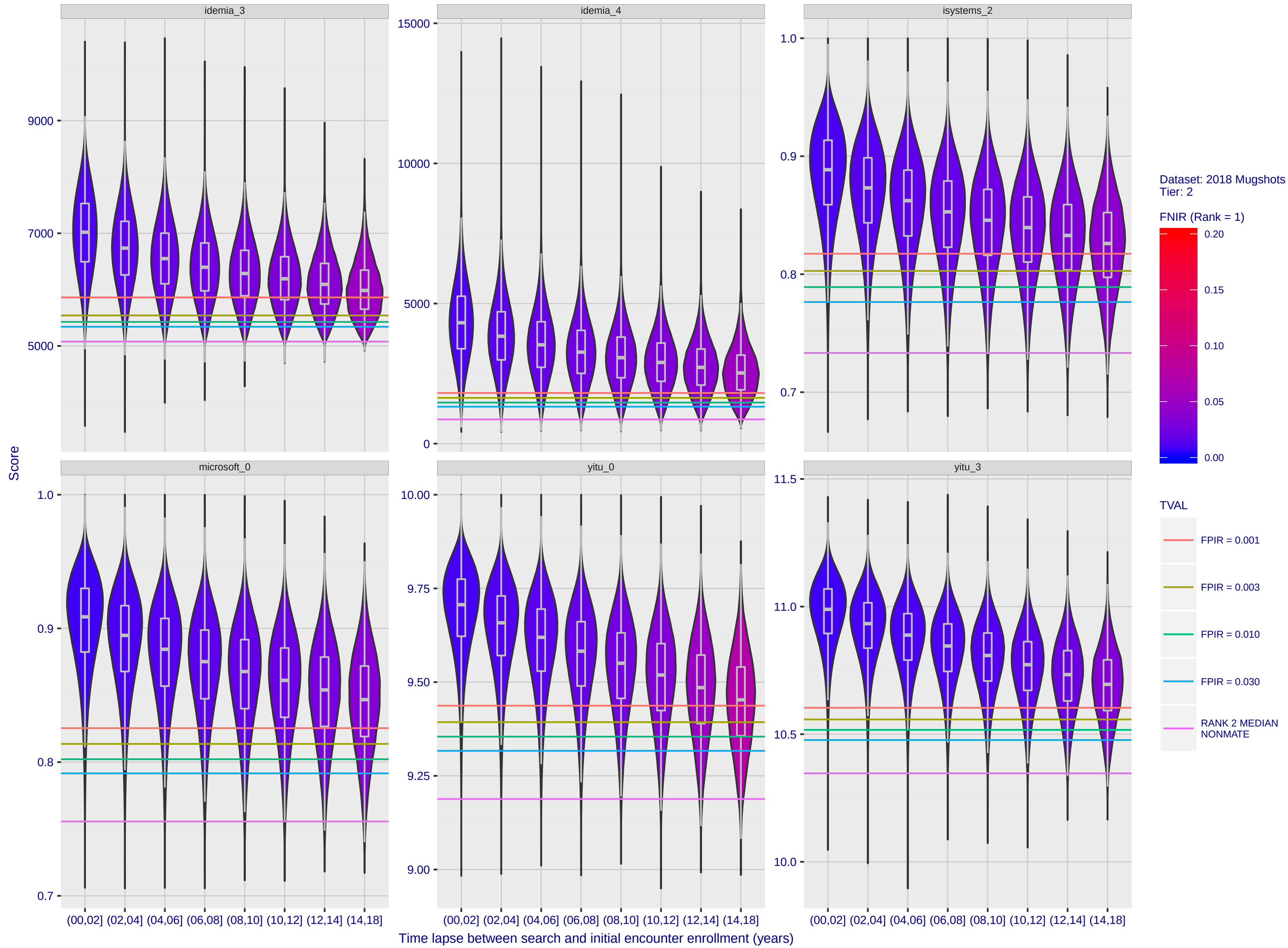

Figure 75: **[FRVT-2018 Mugshot Ageing Dataset] Native mate scores vs. time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Mated score distributions are computed over all searches noted in row 17 of Table 6 binned by number of years between search and initial enrollment.*

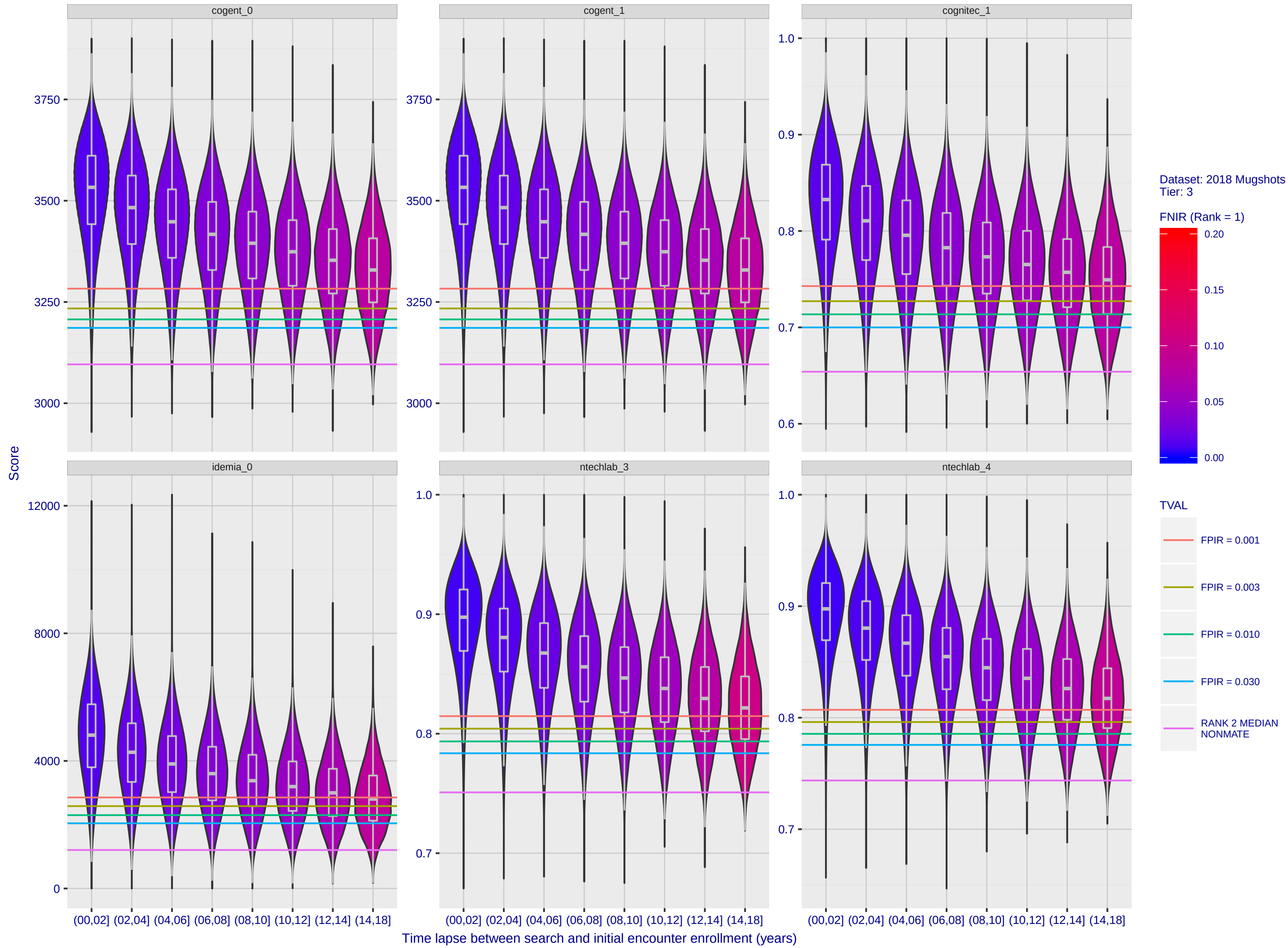

Figure 76: **[FRVT-2018 Mugshot Ageing Dataset] Native mate scores vs. time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Mated score distributions are computed over all searches noted in row 17 of Table 6 binned by number of years between search and initial enrollment.*

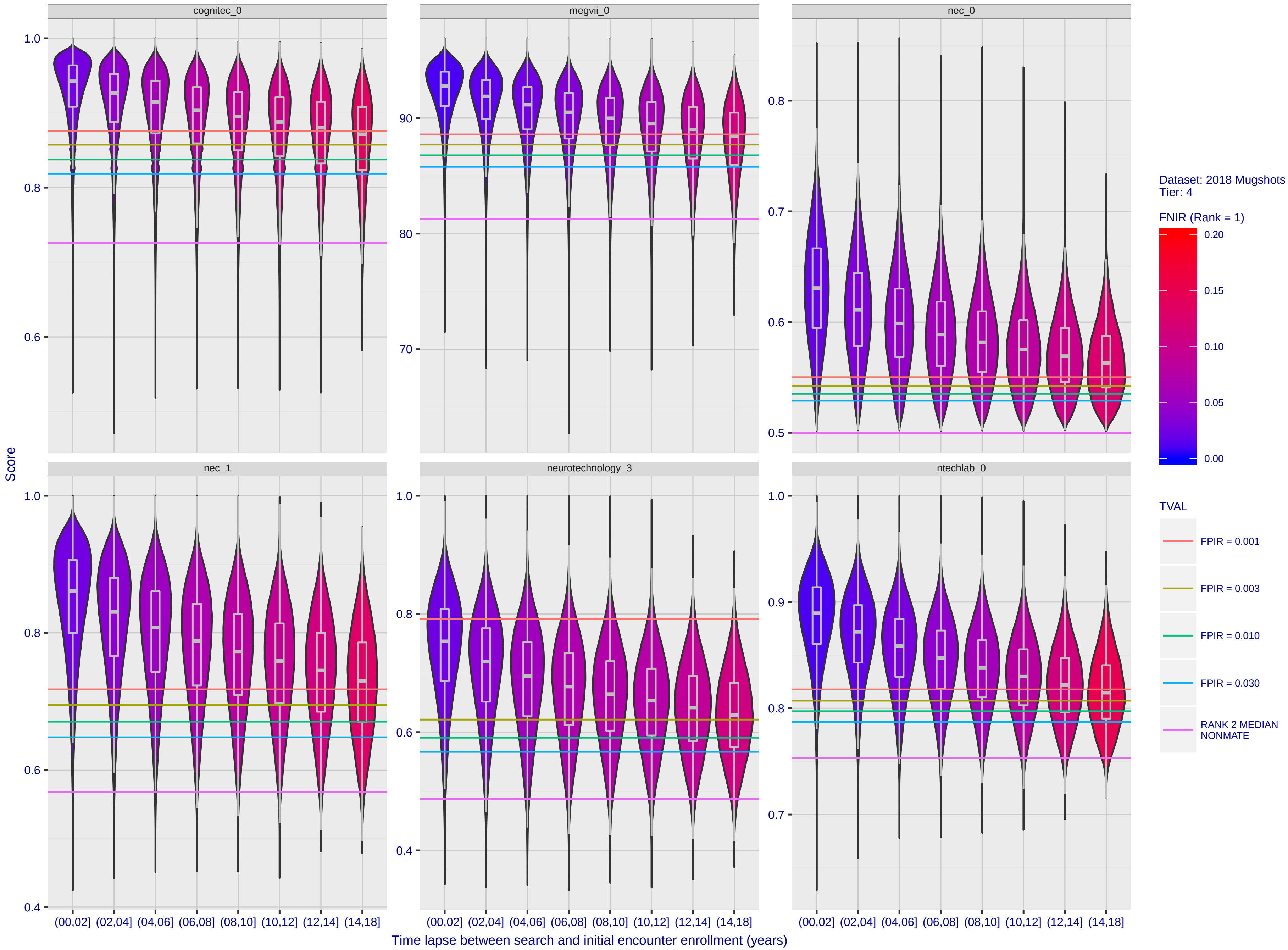

Figure 77: **[FRVT-2018 Mugshot Ageing Dataset] Native mate scores vs. time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Mated score distributions are computed over all searches noted in row 17 of Table 6 binned by number of years between search and initial enrollment.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

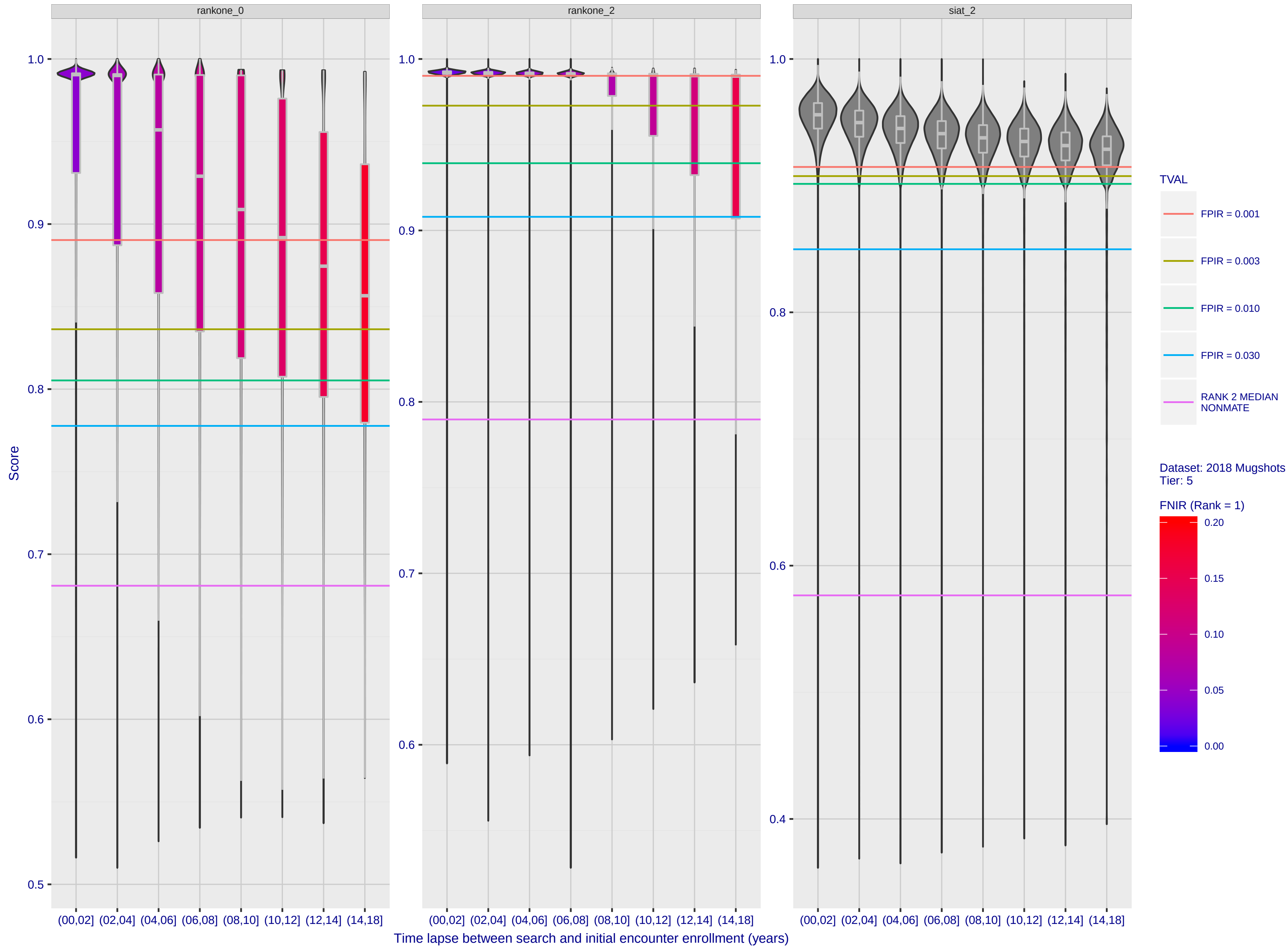

Figure 78: **[FRVT-2018 Mugshot Ageing Dataset] Native mate scores vs. time-elapsed**. *The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Mated score distributions are computed over all searches noted in row 17 of Table 6 binned by number of years between search and initial enrollment.*

Appendix C Effect of enrolling multiple images

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8238

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate

N = Num. enrolled subjects
R = Num. candidates examined

T = Threshold

T = 0 → Investigation
T > 0 → Identification

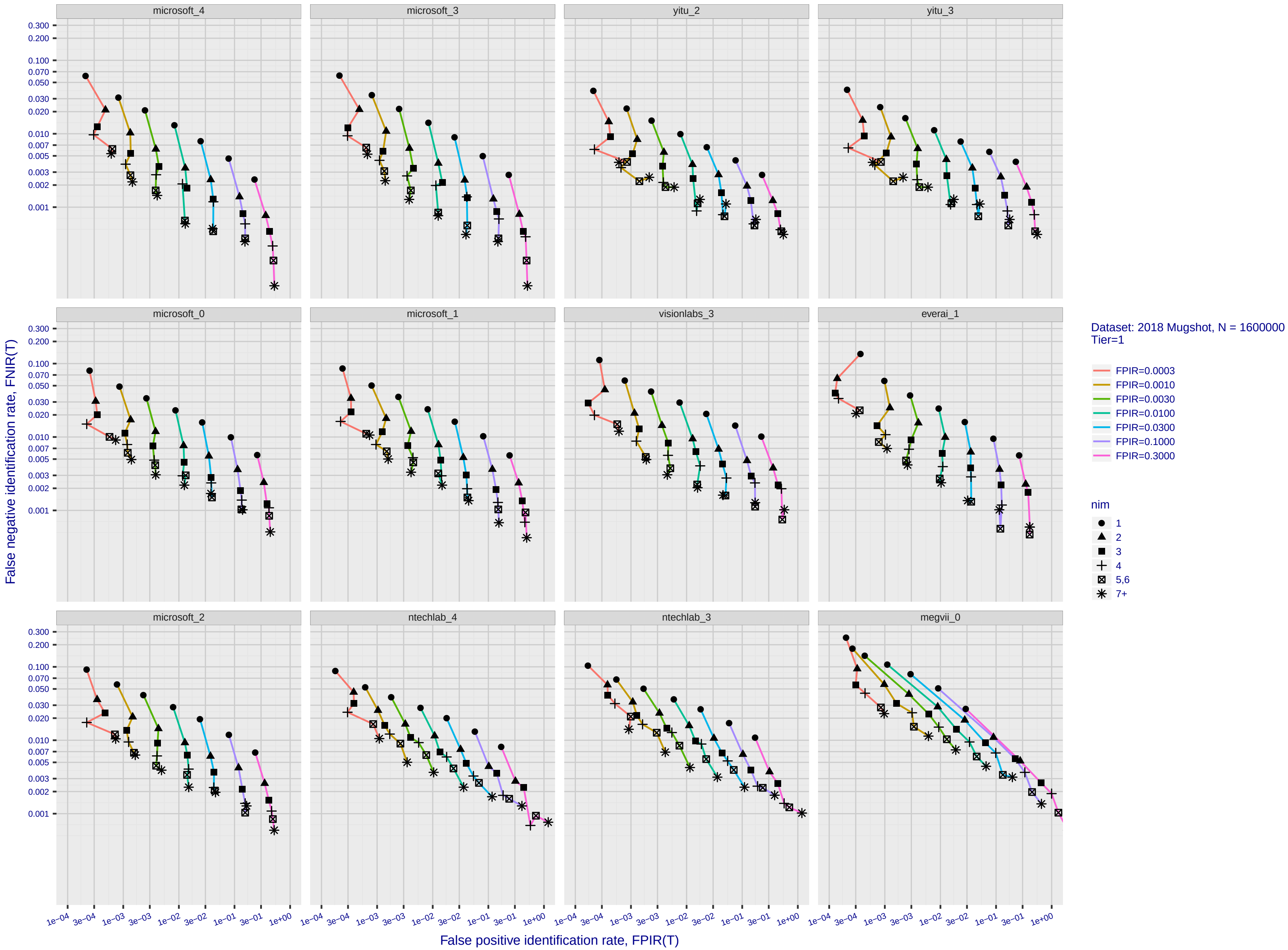

Figure 79: **[FRVT-2018 Mugshot Dataset] Effect of enrolling multiple images for each identity**. *The plot shows an identification miss rates vs. false positive rates, at seven operating thresholds. The enrolled population size is fixed. The images are enrolled with lifetime-consolidation - see section 2.3.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

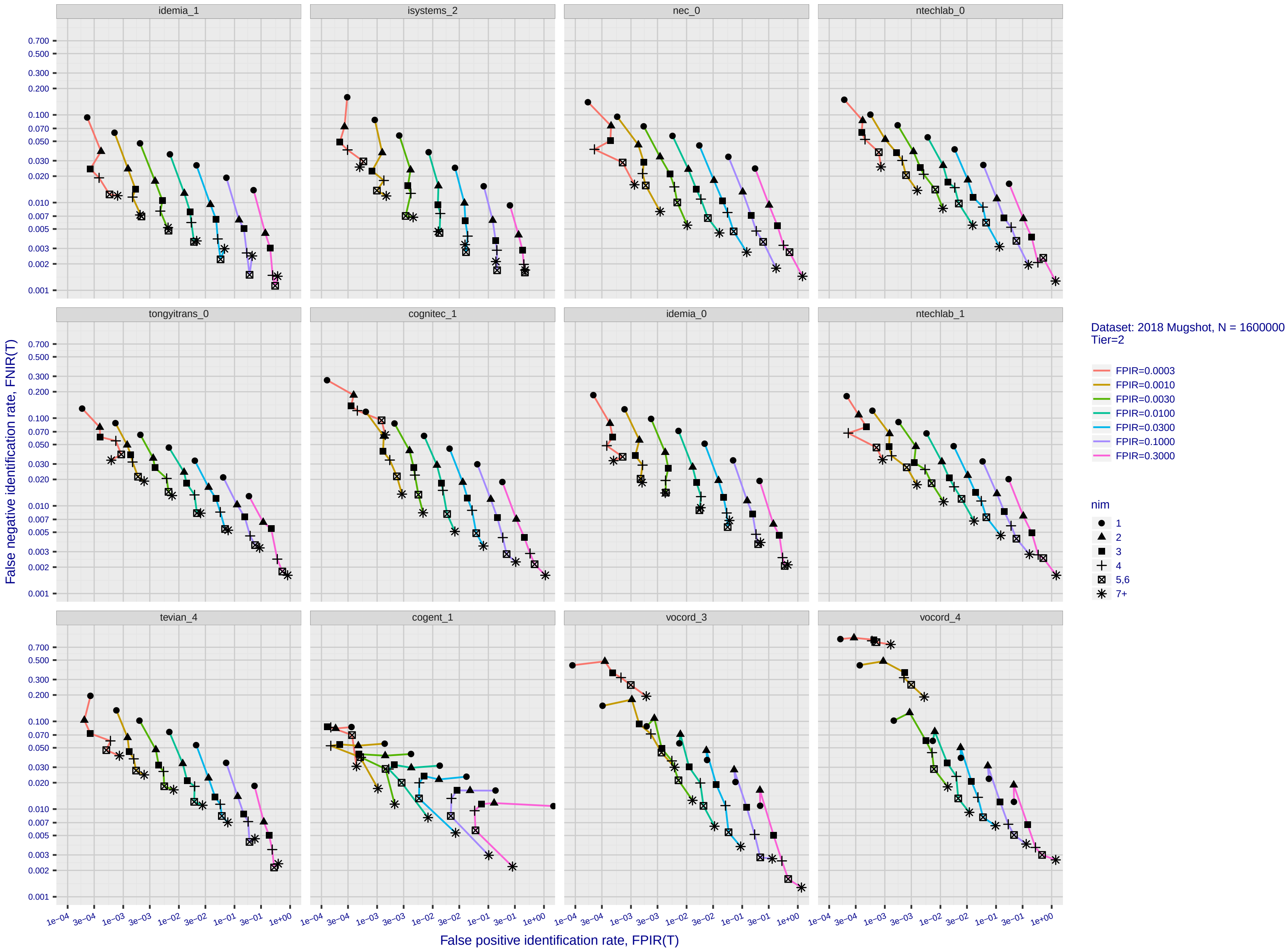

Figure 80: **[FRVT-2018 Mugshot Dataset] Effect of enrolling multiple images for each identity**. *The plot shows an identification miss rates vs. false positive rates, at seven operating thresholds. The enrolled population size is fixed. The images are enrolled with lifetime-consolidation - see section 2.3.*

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8238

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

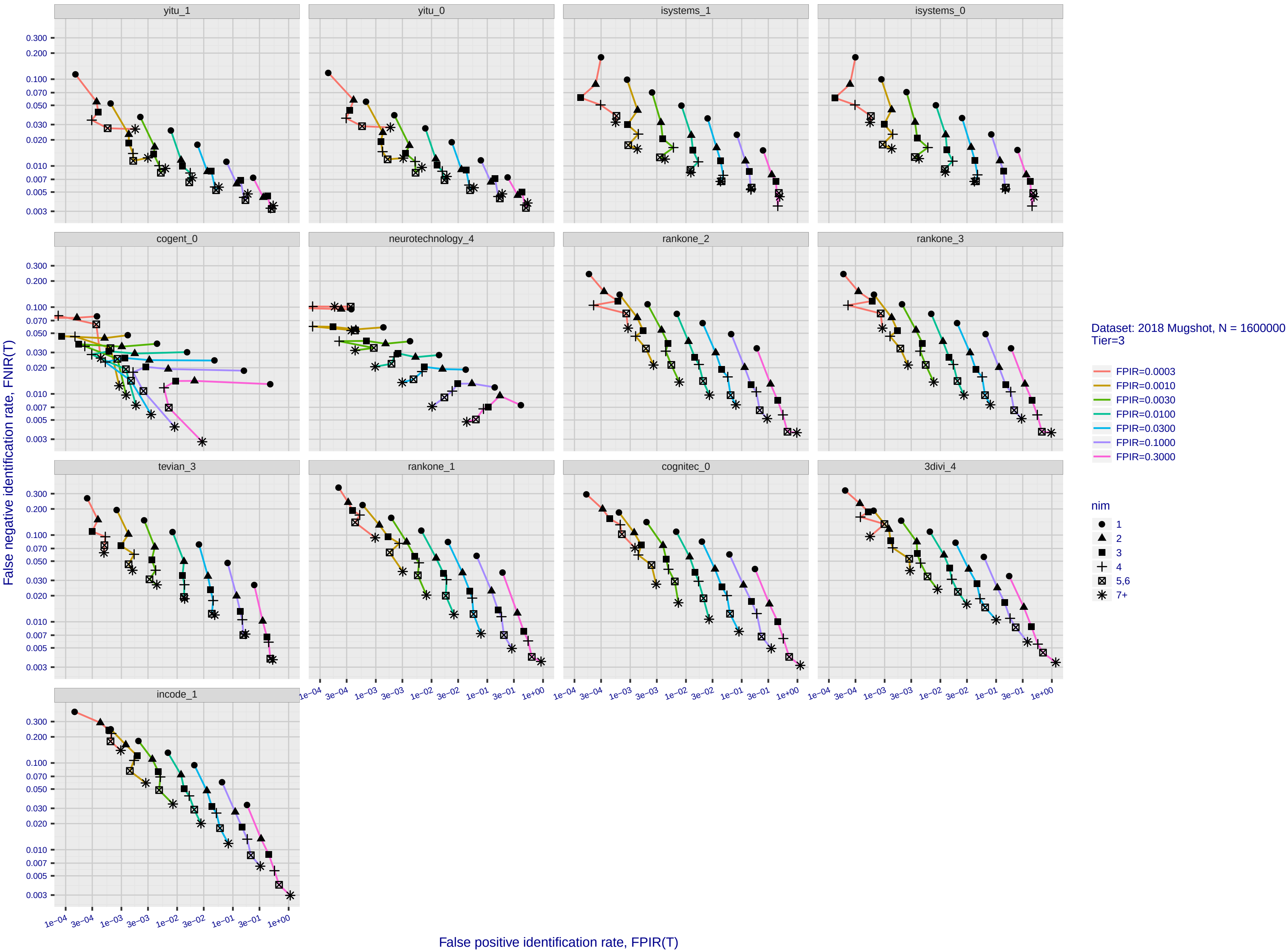

Figure 81: **[FRVT-2018 Mugshot Dataset] Effect of enrolling multiple images for each identity**. *The plot shows an identification miss rates vs. false positive rates, at seven operating thresholds. The enrolled population size is fixed. The images are enrolled with lifetime-consolidation - see section 2.3.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

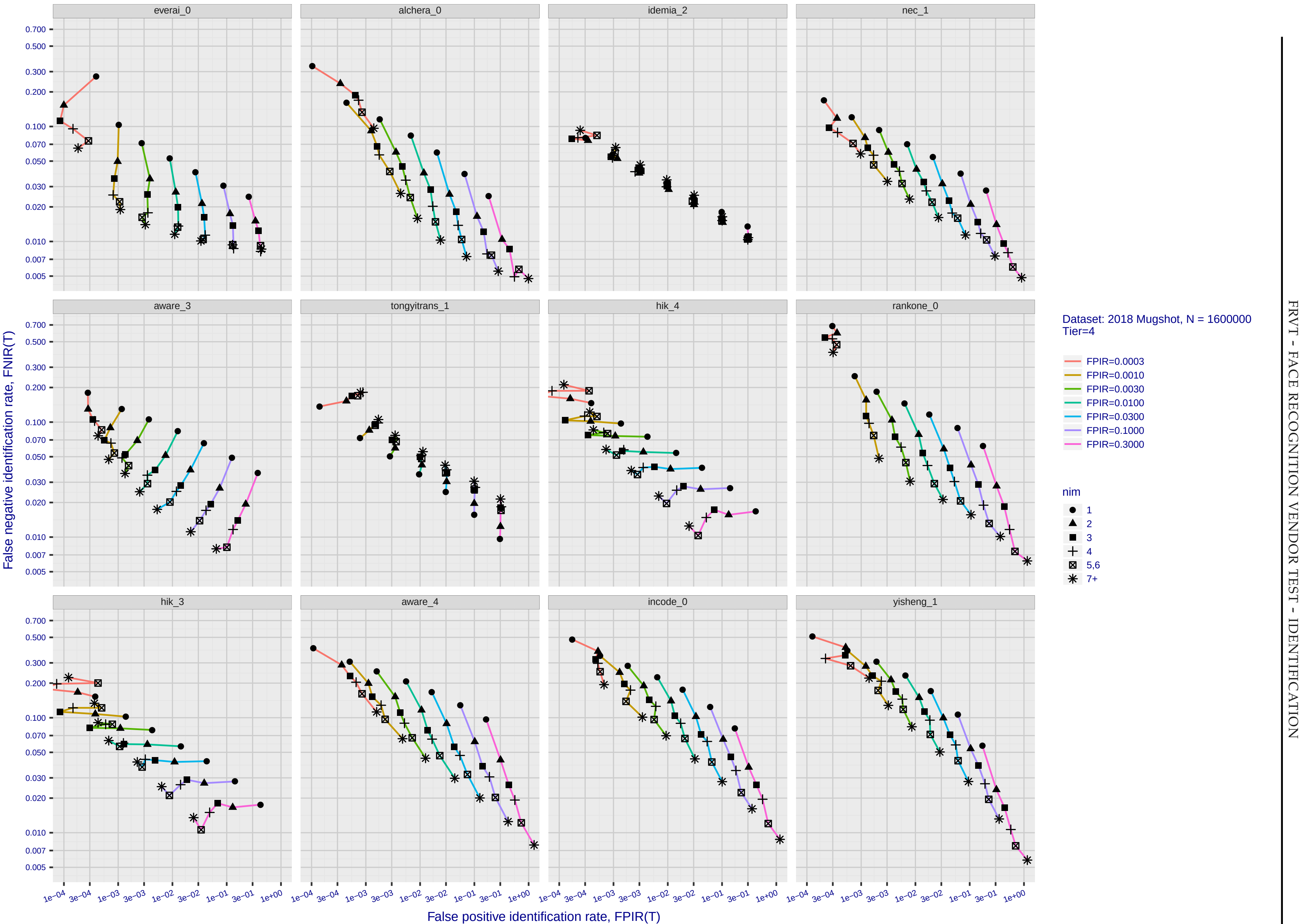

Figure 82: **[FRVT-2018 Mugshot Dataset] Effect of enrolling multiple images for each identity**. *The plot shows an identification miss rates vs. false positive rates, at seven operating thresholds. The enrolled population size is fixed. The images are enrolled with lifetime-consolidation - see section 2.3.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

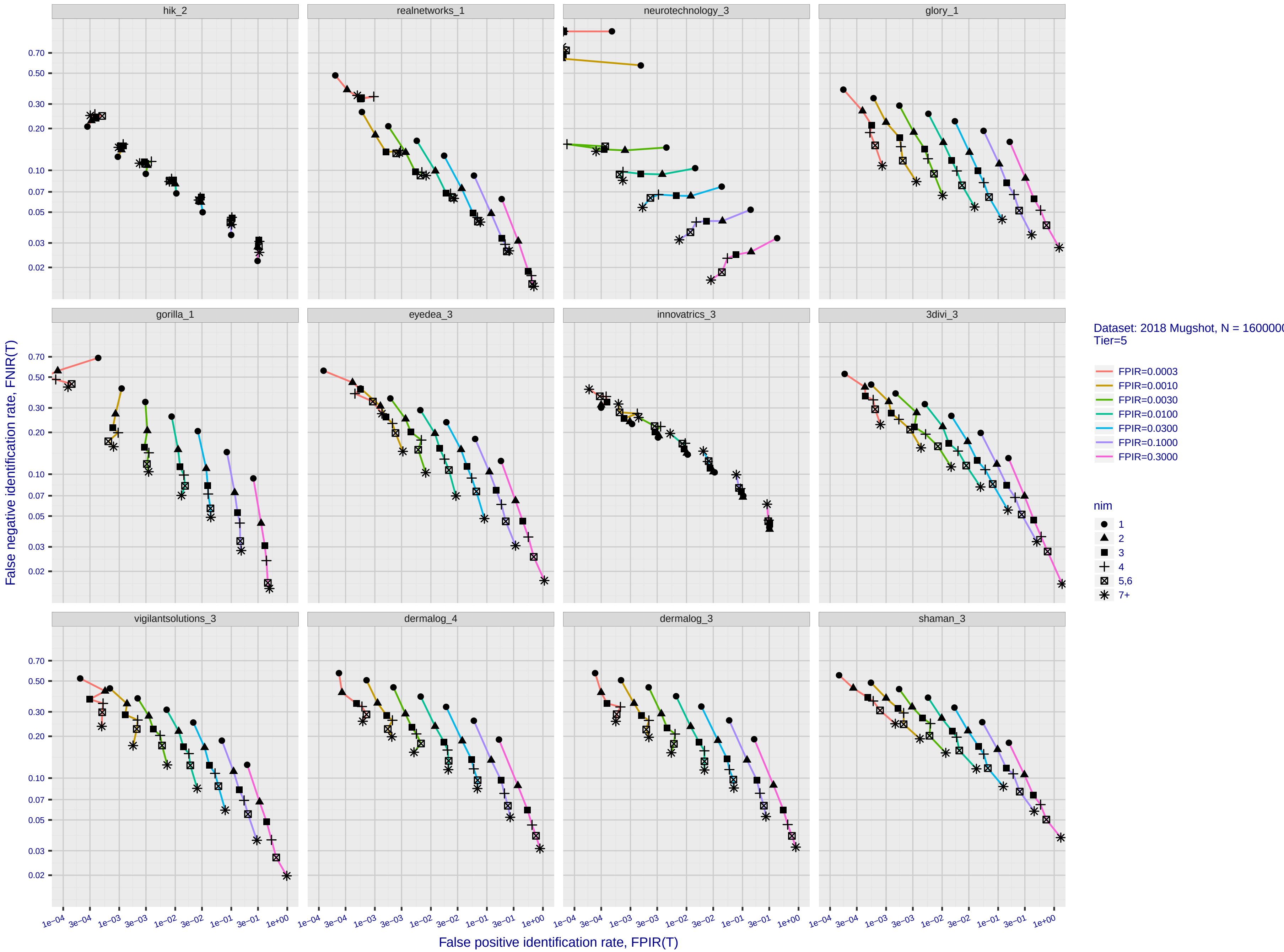

Figure 83: **[FRVT-2018 Mugshot Dataset] Effect of enrolling multiple images for each identity**. *The plot shows an identification miss rates vs. false positive rates, at seven operating thresholds. The enrolled population size is fixed. The images are enrolled with lifetime-consolidation - see section 2.3.*

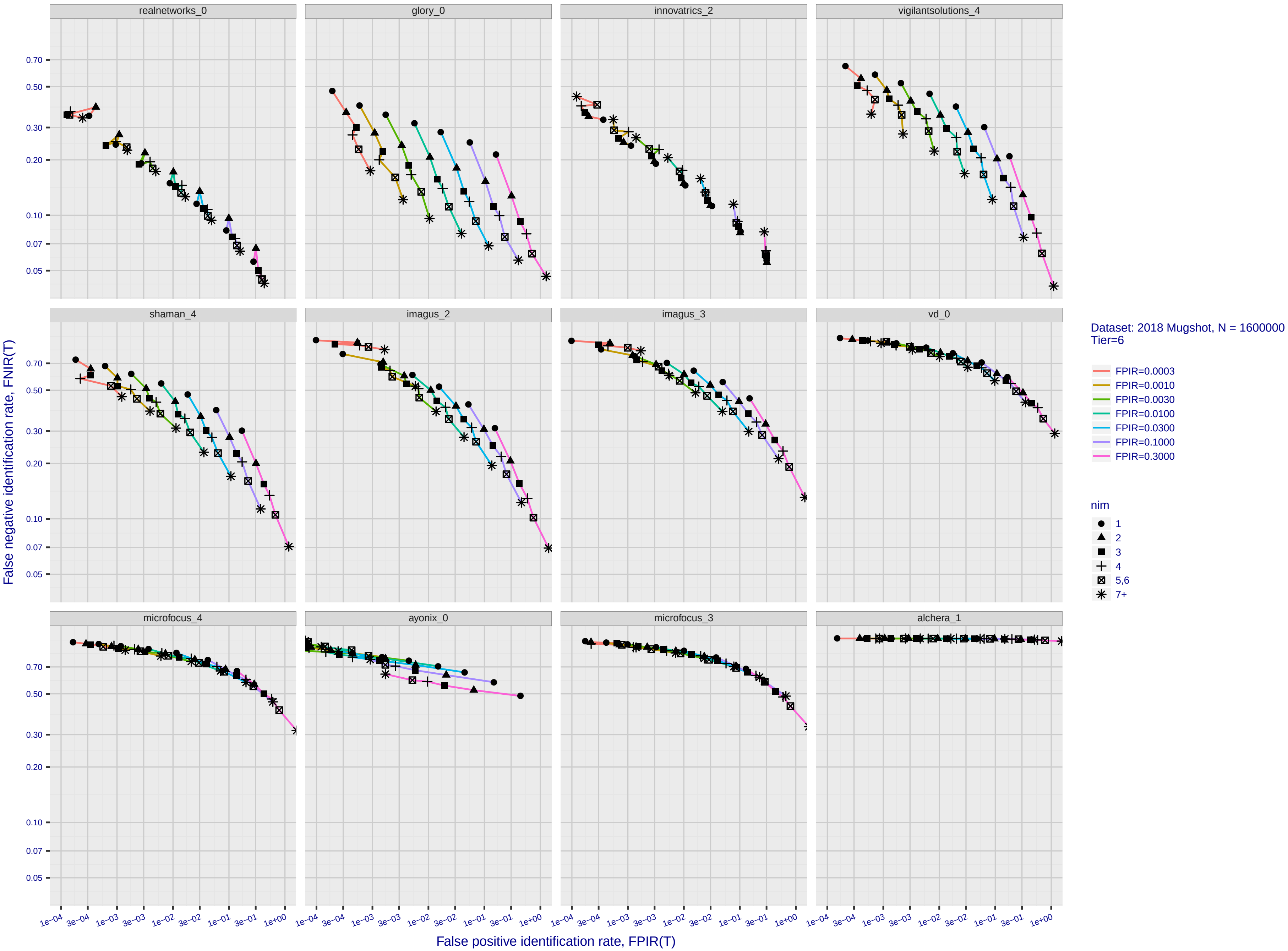

Figure 84: **[FRVT-2018 Mugshot Dataset] Effect of enrolling multiple images for each identity**. *The plot shows an identification miss rates vs. false positive rates, at seven operating thresholds. The enrolled population size is fixed. The images are enrolled with lifetime-consolidation - see section 2.3.*

Appendix D Accuracy with poor quality webcam images

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

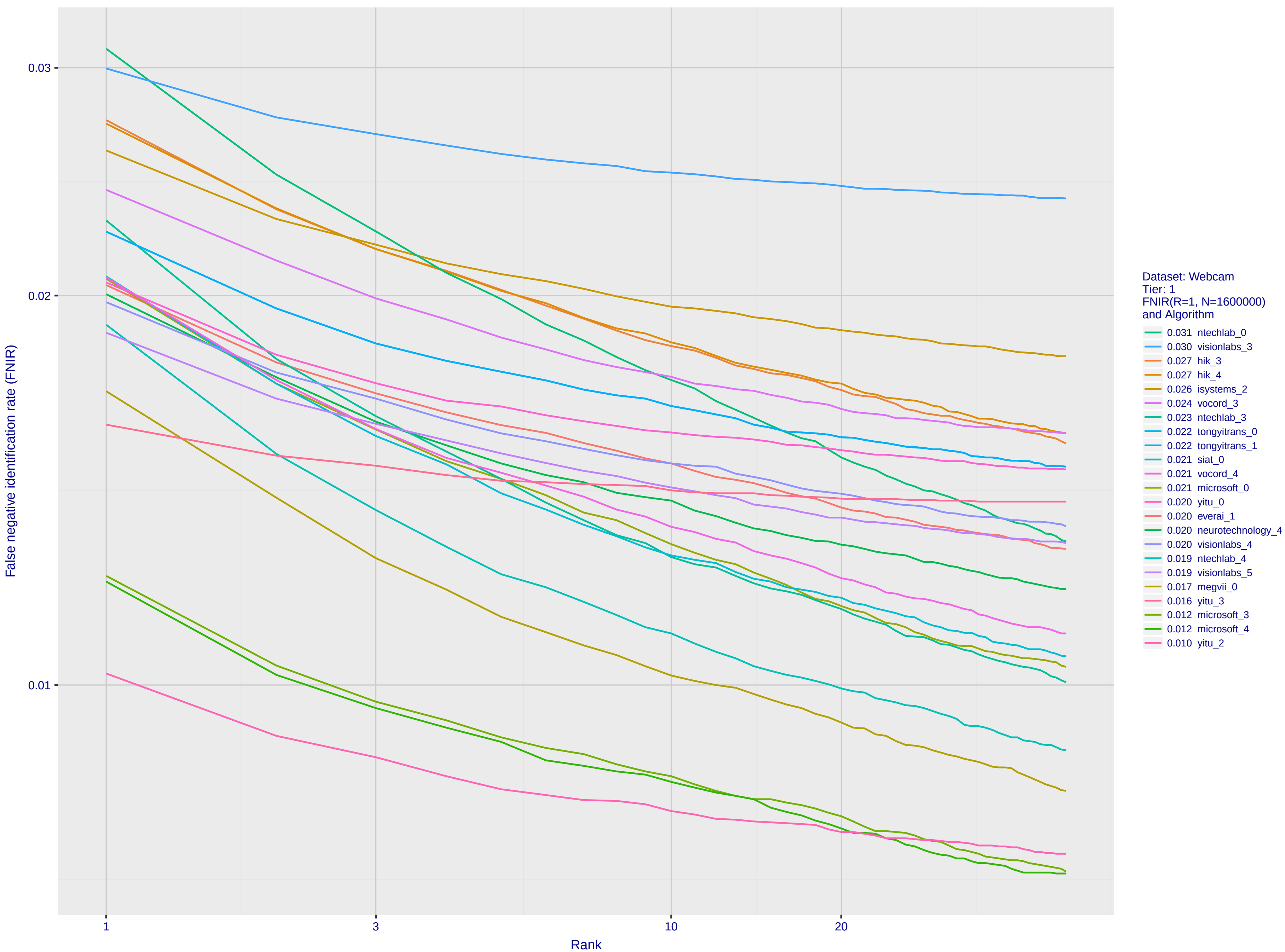

Figure 85: **[Webcam Dataset] Identification miss rates vs. rank**. *The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and less constrained subject pose in webcam images - see Figure 3.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

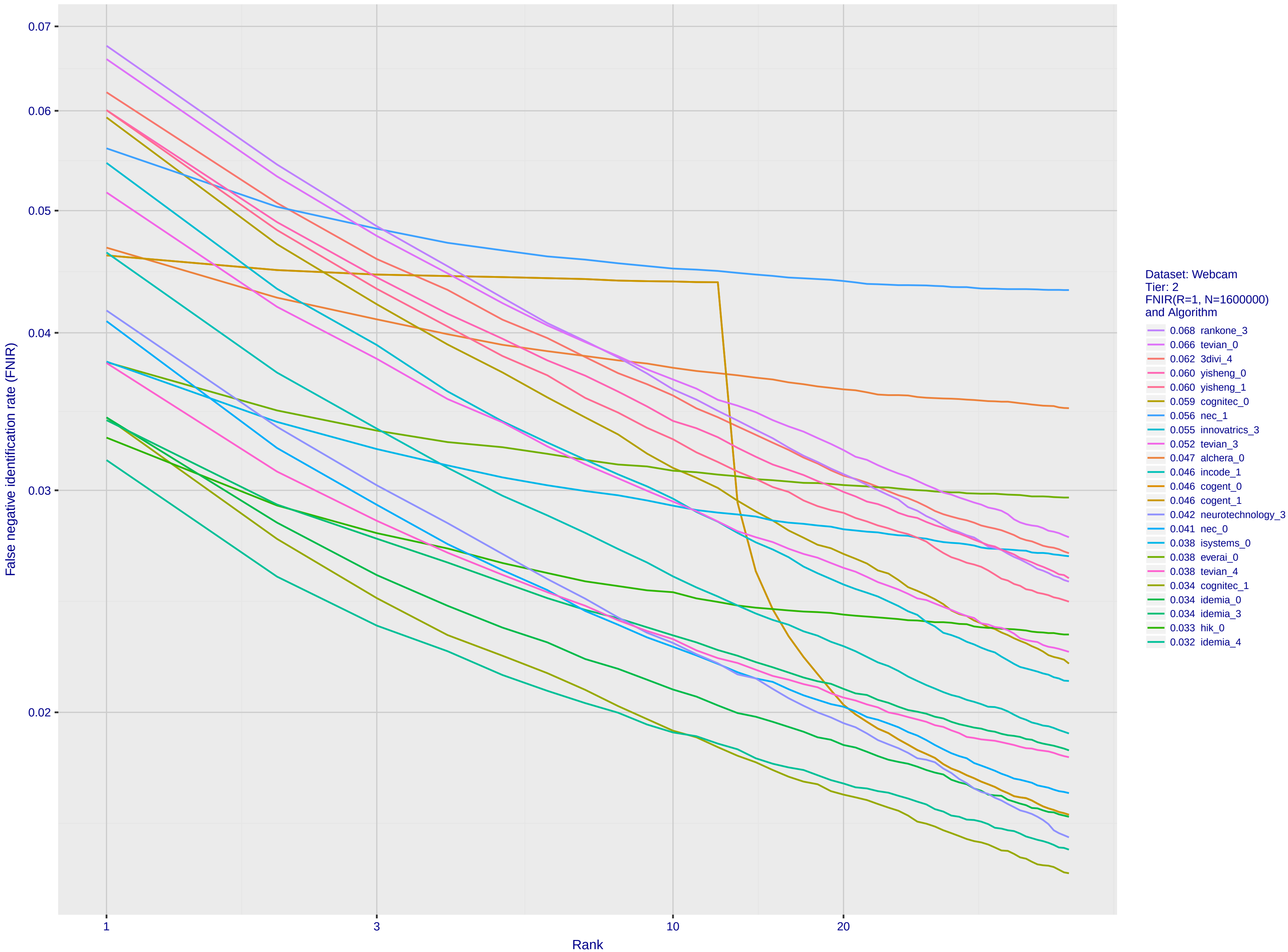

Figure 86: **[Webcam Dataset] Identification miss rates vs. rank**. *The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and less constrained subject pose in webcam images - see Figure 3.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

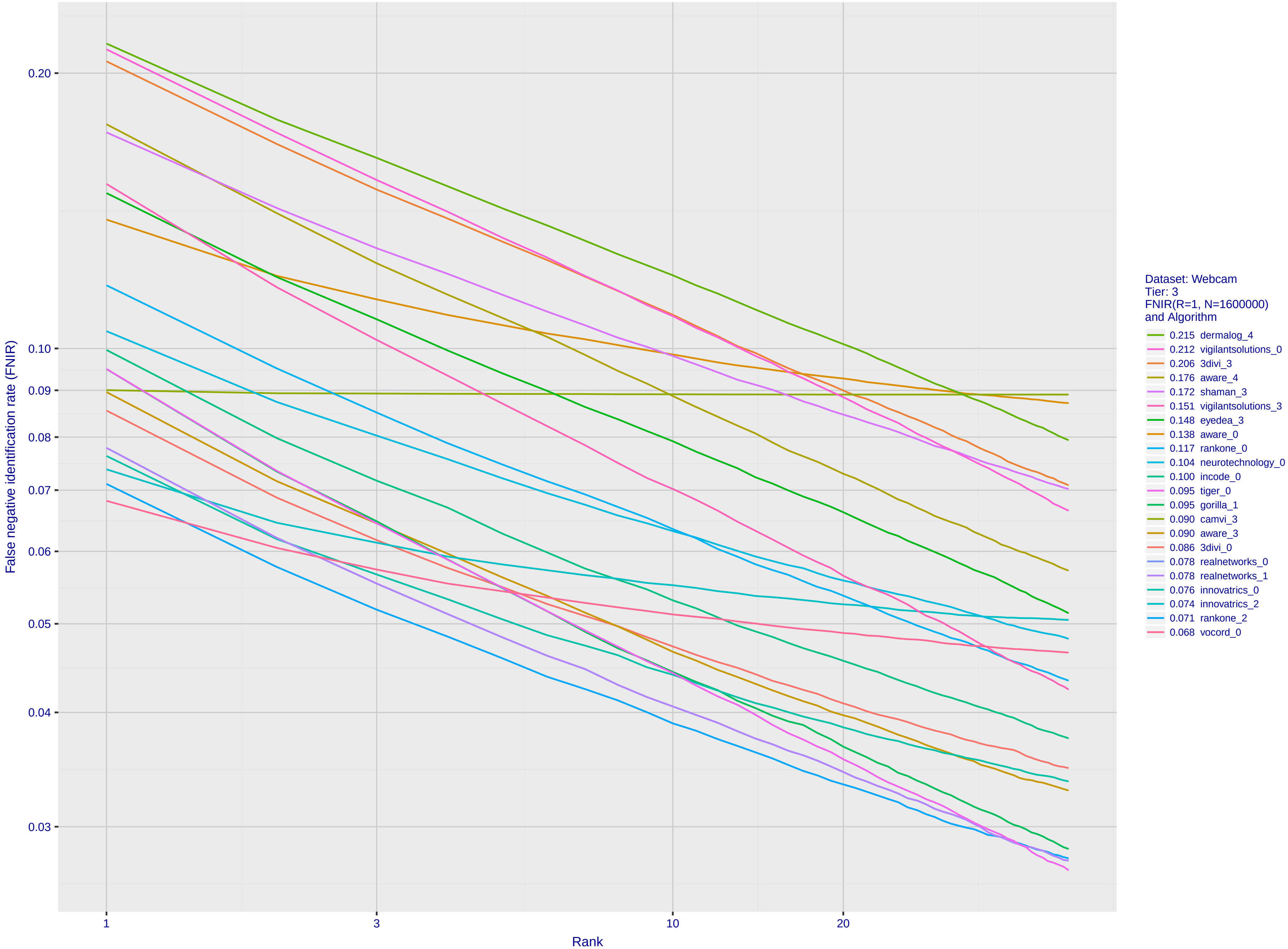

Figure 87: **[Webcam Dataset] Identification miss rates vs. rank**. *The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and less constrained subject pose in webcam images - see Figure 3.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

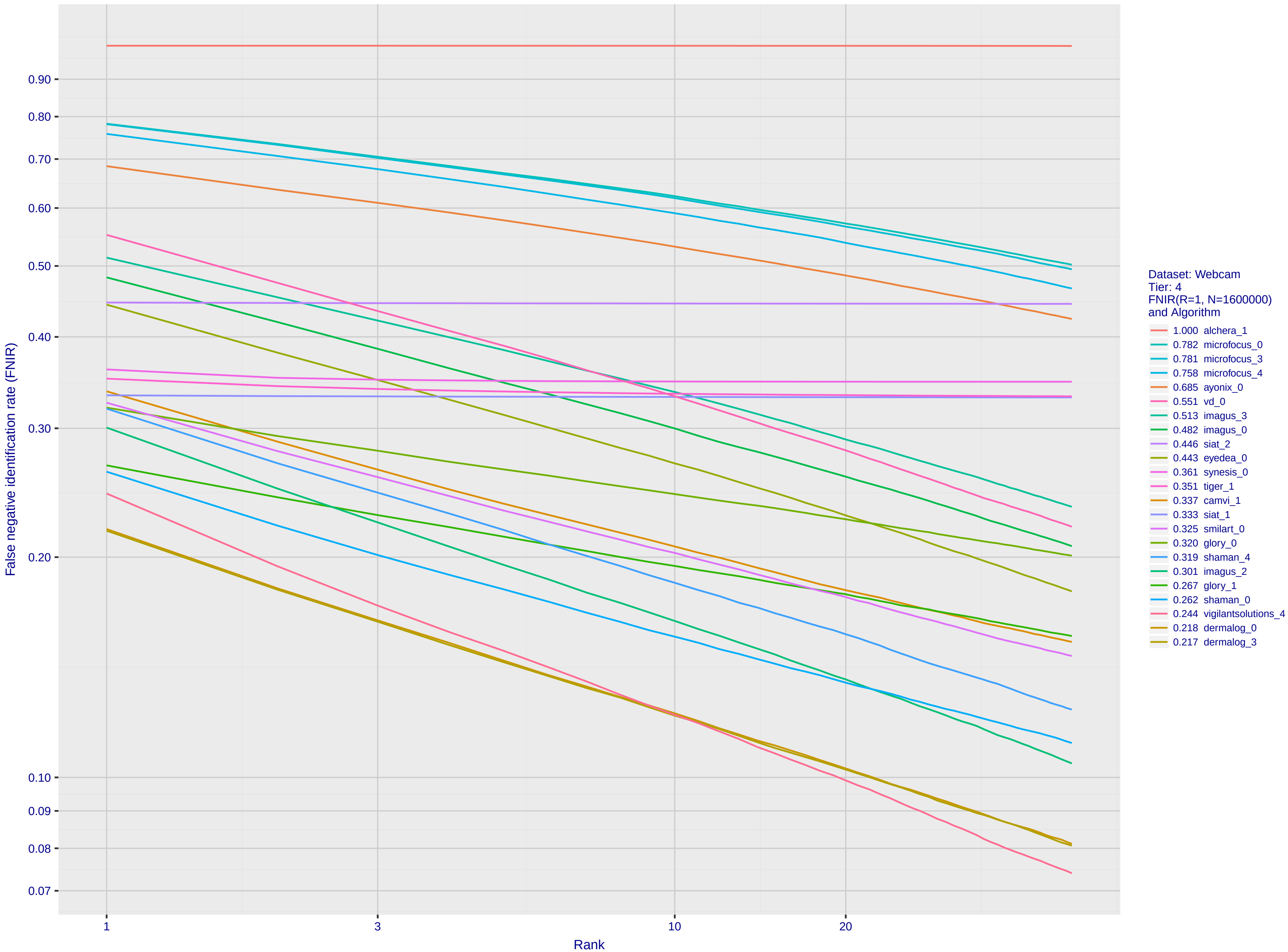

Figure 88: **[Webcam Dataset] Identification miss rates vs. rank**. *The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and less constrained subject pose in webcam images - see Figure 3.*

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

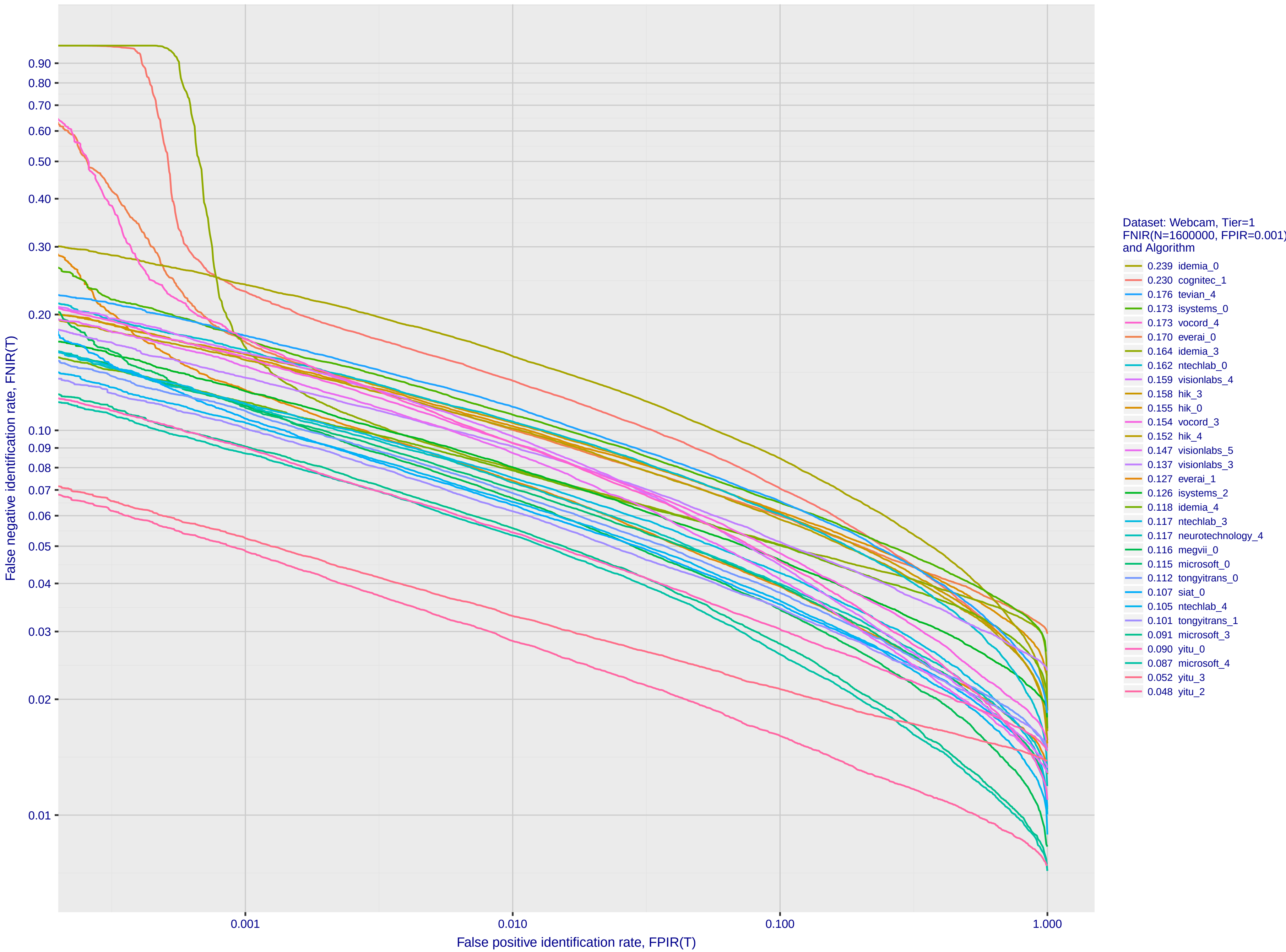

Figure 89: **[Webcam Dataset] Identification miss rates vs. false positive rates**. *The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and less constrained subject pose in webcam images - see Figure 3.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

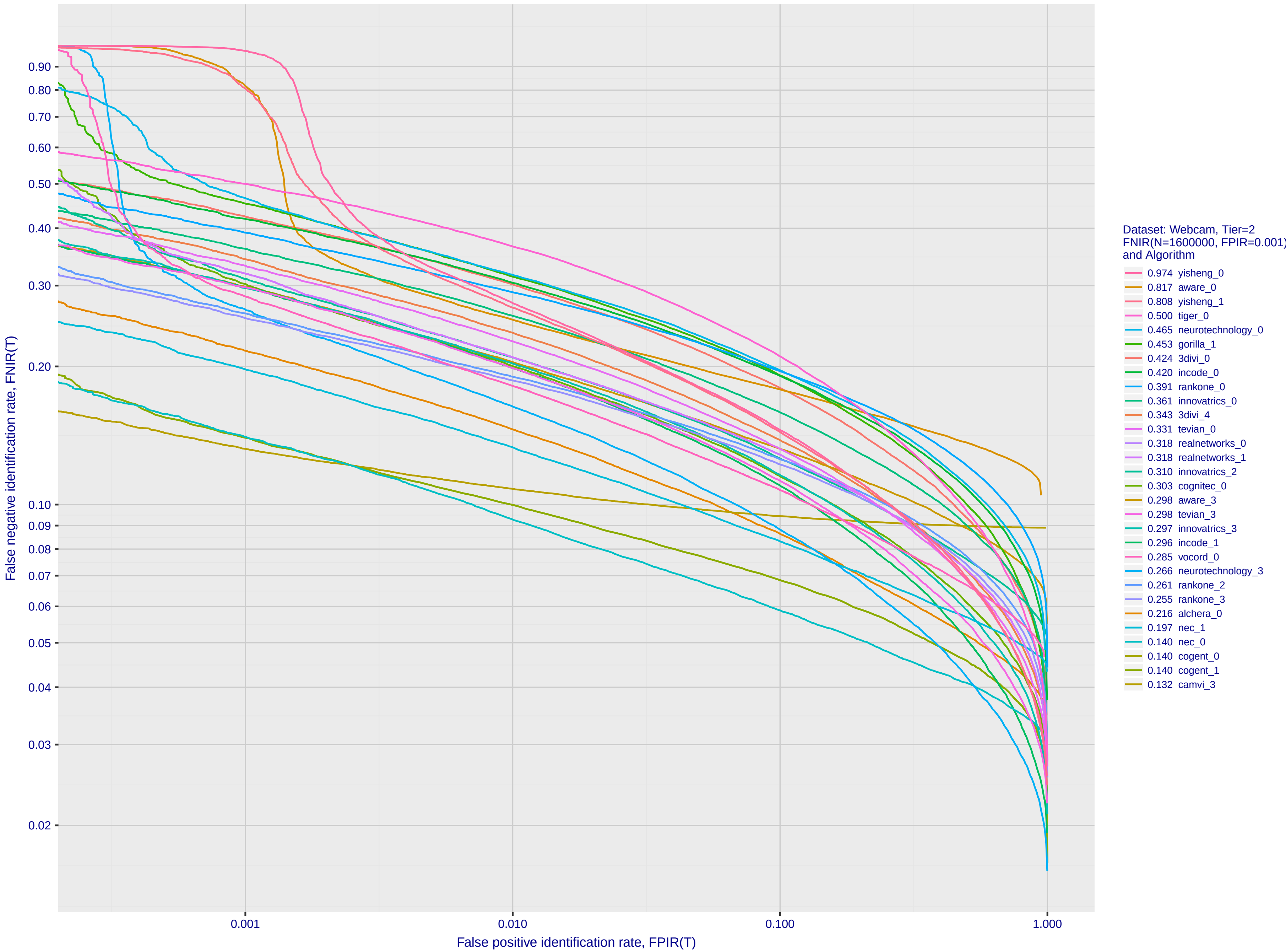

Figure 90: **[Webcam Dataset] Identification miss rates vs. false positive rates**. *The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and less constrained subject pose in webcam images - see Figure 3.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

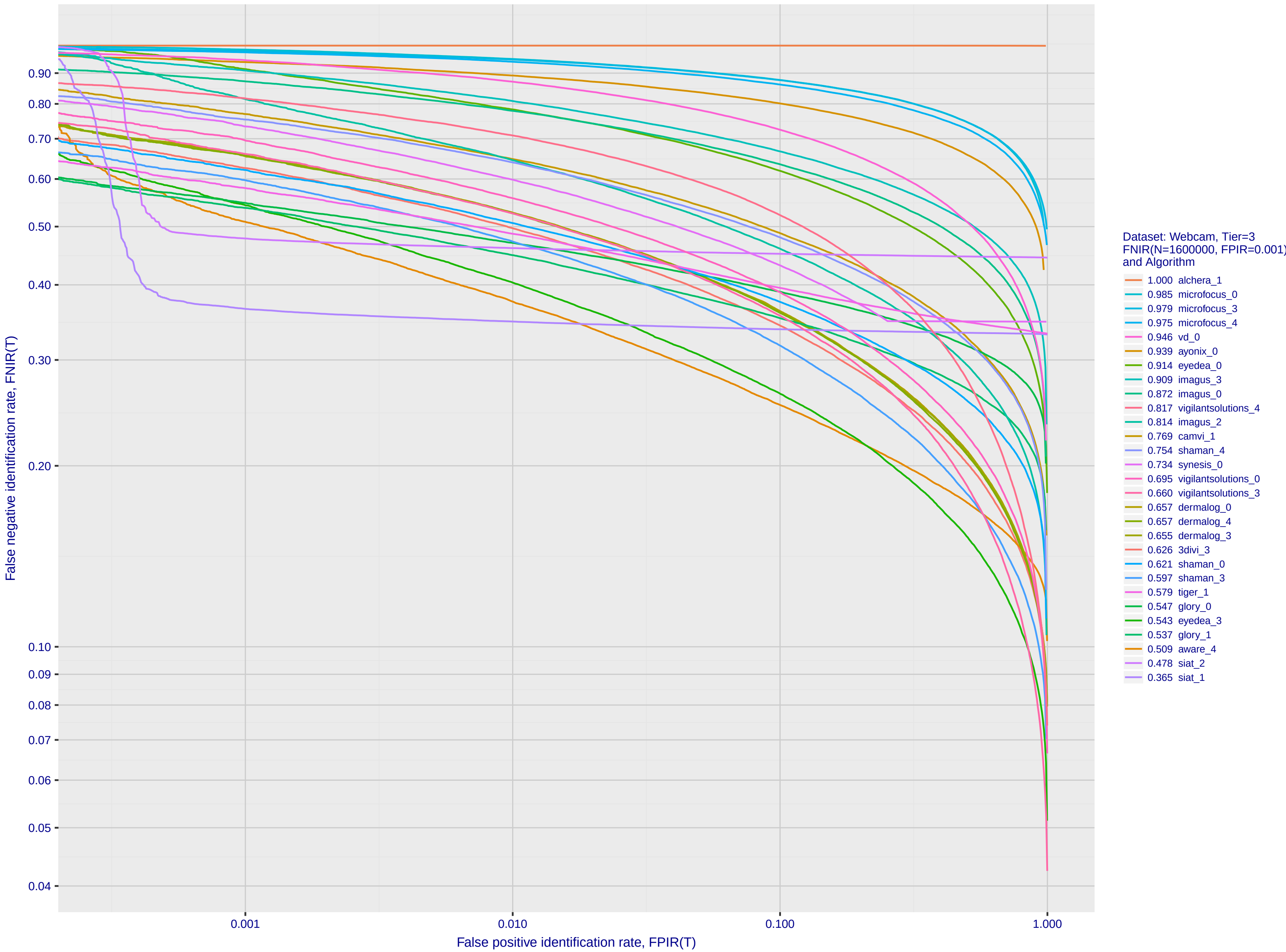

Figure 91: **[Webcam Dataset] Identification miss rates vs. false positive rates**. *The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and less constrained subject pose in webcam images - see Figure 3.*

Appendix E Accuracy with non-cooperating subjects

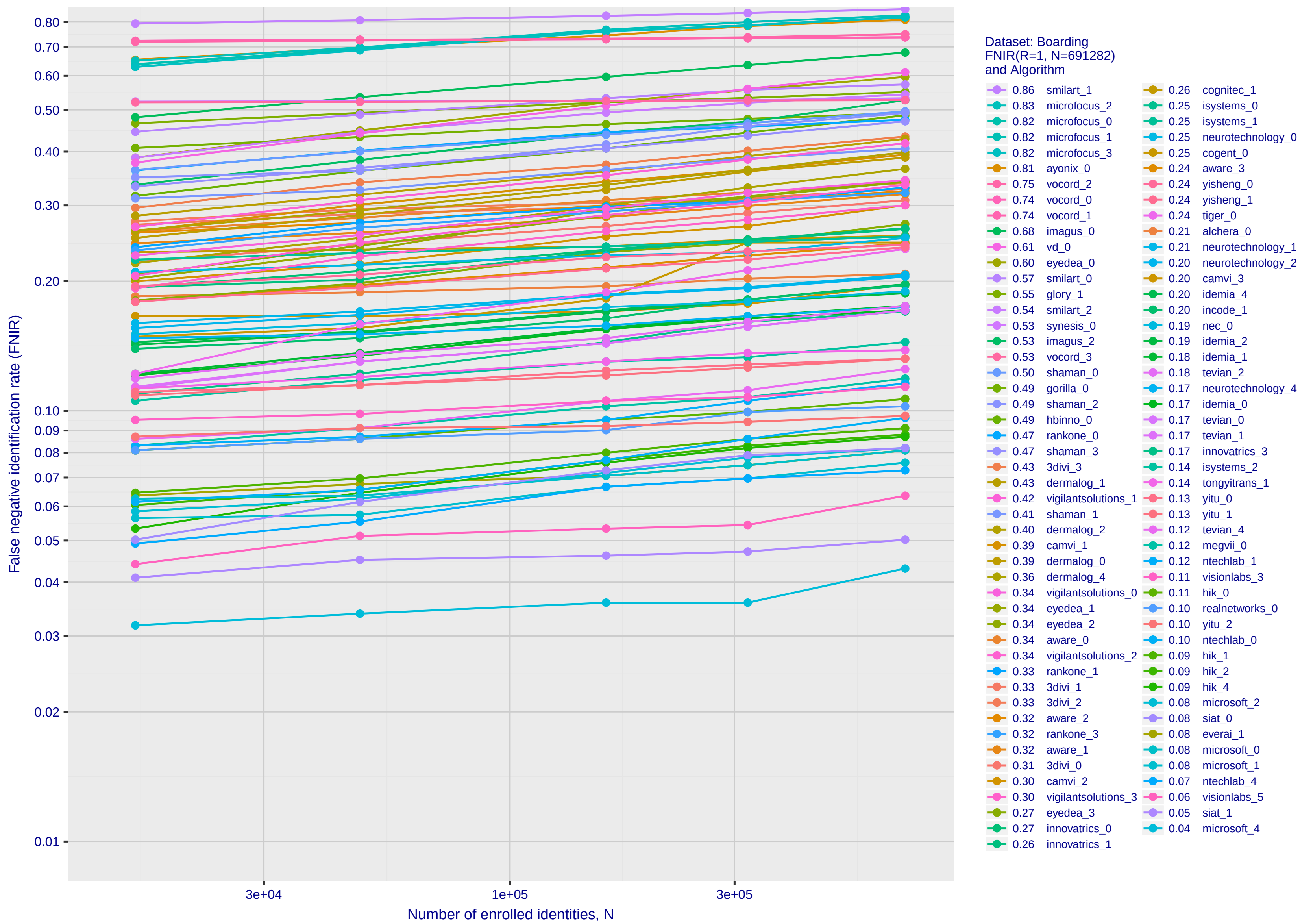

Figure 92: **[FRPC Dataset: Boarding] Miss rates vs. number of enrolled identities**. *The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of people crossing walking toward an aircraft boarding pass reader, using it, then proceeding left across the optical axis passing the camera, searched against well-controlled, portrait images of up to 691 282 individuals enrolled into a gallery. The curves show false negative identification rates at rank 1 as a function of enrolled population size, FNIR(N, 1). The threshold is set to zero. This metric is relevant to human reviewers who will traverse candidate lists in pursuit of investigations.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

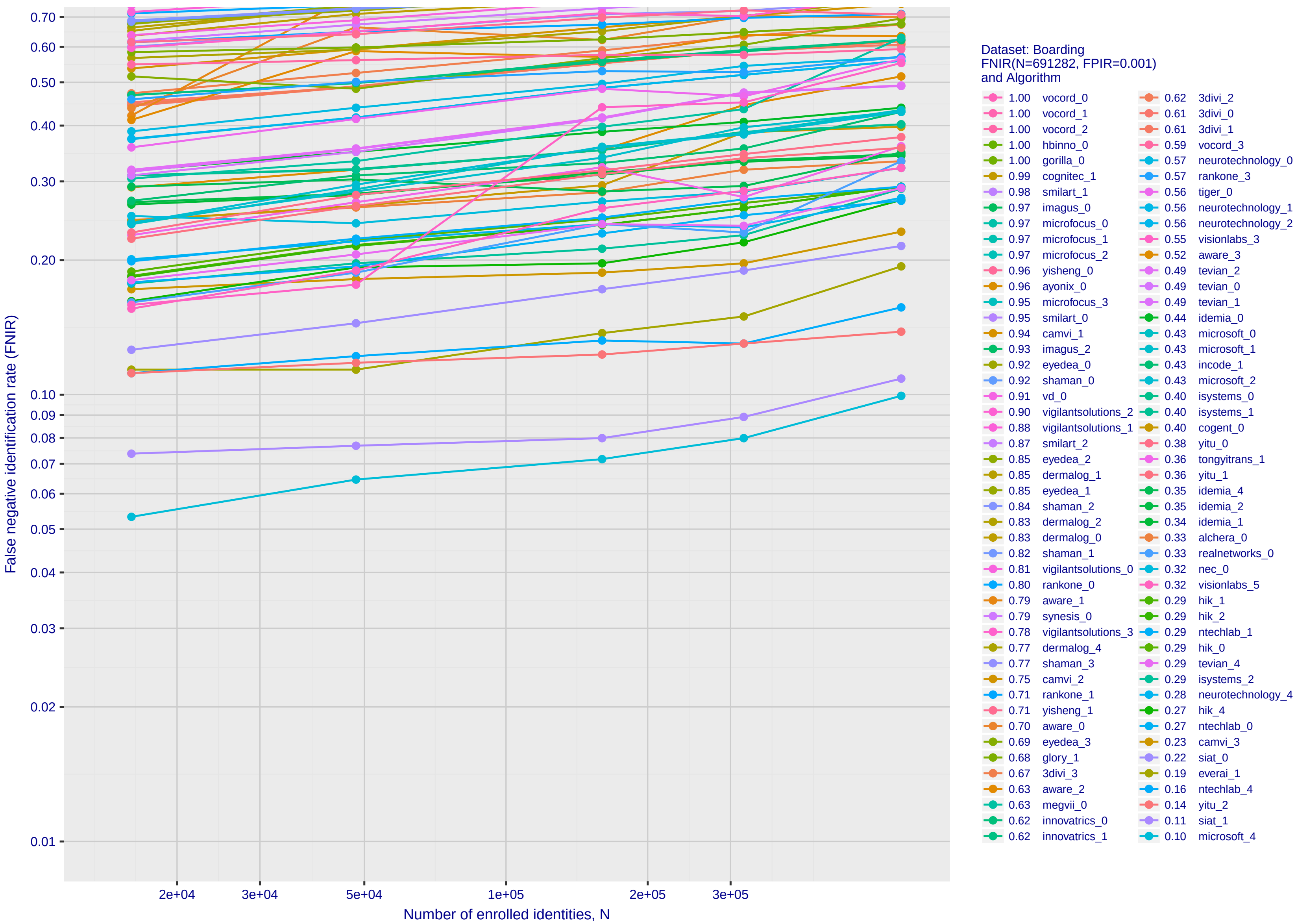

Figure 93: **[FRPC Dataset: Boarding] Miss rates vs. number of enrolled identities**. *The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of people crossing walking toward an aircraft boarding pass reader, using it, then proceeding left across the optical axis passing the camera, searched against well-controlled, portrait images of up to 691 282 individuals enrolled into a gallery. The curves show false negative identification rates vs. enrolled population size - FNIR(N, L, T) - when the threshold is set to a high value sufficient to limit false positive outcomes, FPIR = 0.001. This metric is relevant to automated watchlist applications, where most searches are from individuals who are not enrolled.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

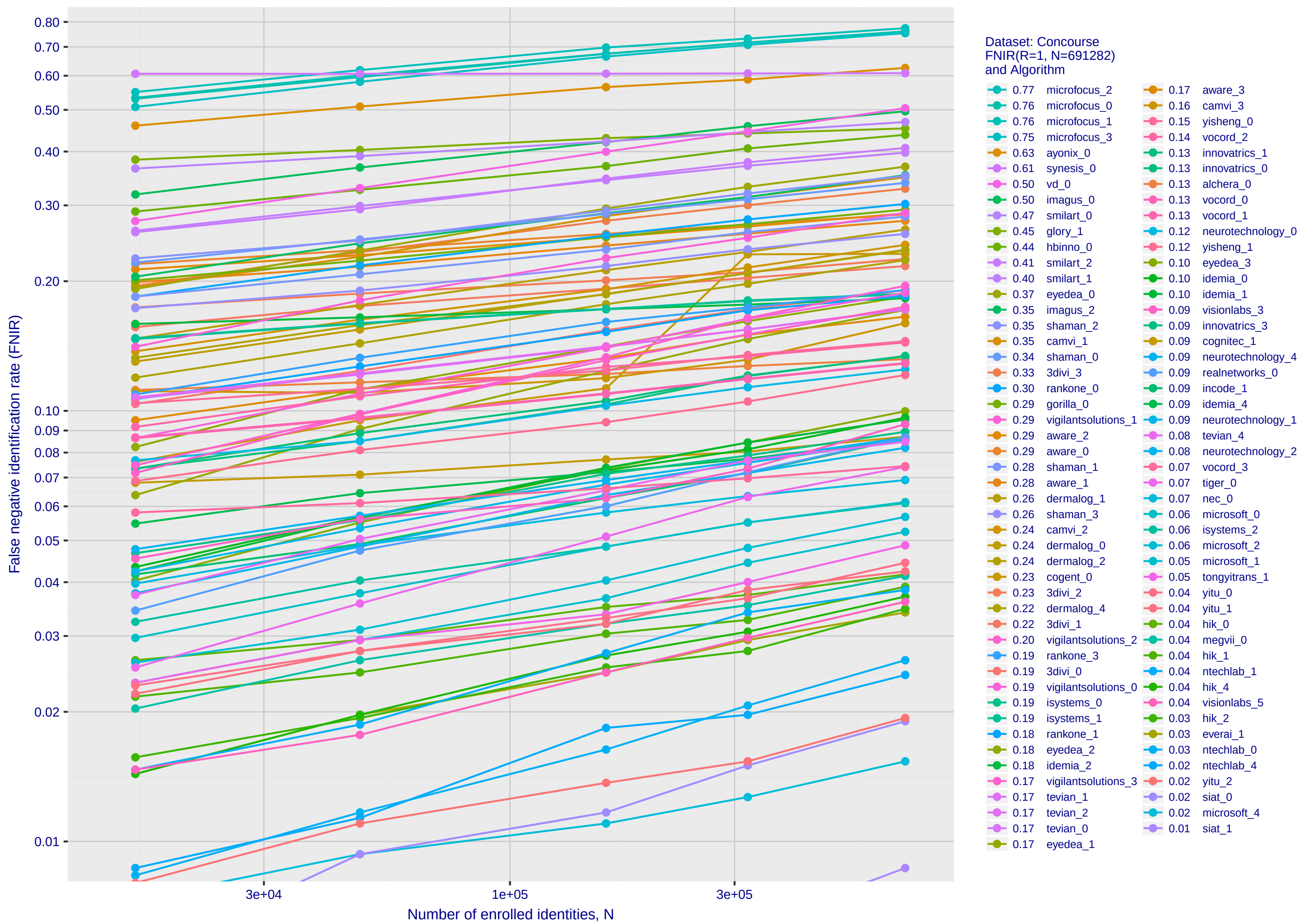

Figure 94: **[FRPC Dataset: Concourse] Miss rates vs. number of enrolled identities**. *The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of people walking down a travel concourse, searched against well-controlled, portrait images of up to 691 282 individuals enrolled into a gallery. The curves show false negative identification rates at rank 1 as a function of enrolled population size, FNIR(N, 1). The threshold is set to zero. This metric is relevant to human reviewers who will traverse candidate lists in pursuit of investigations.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

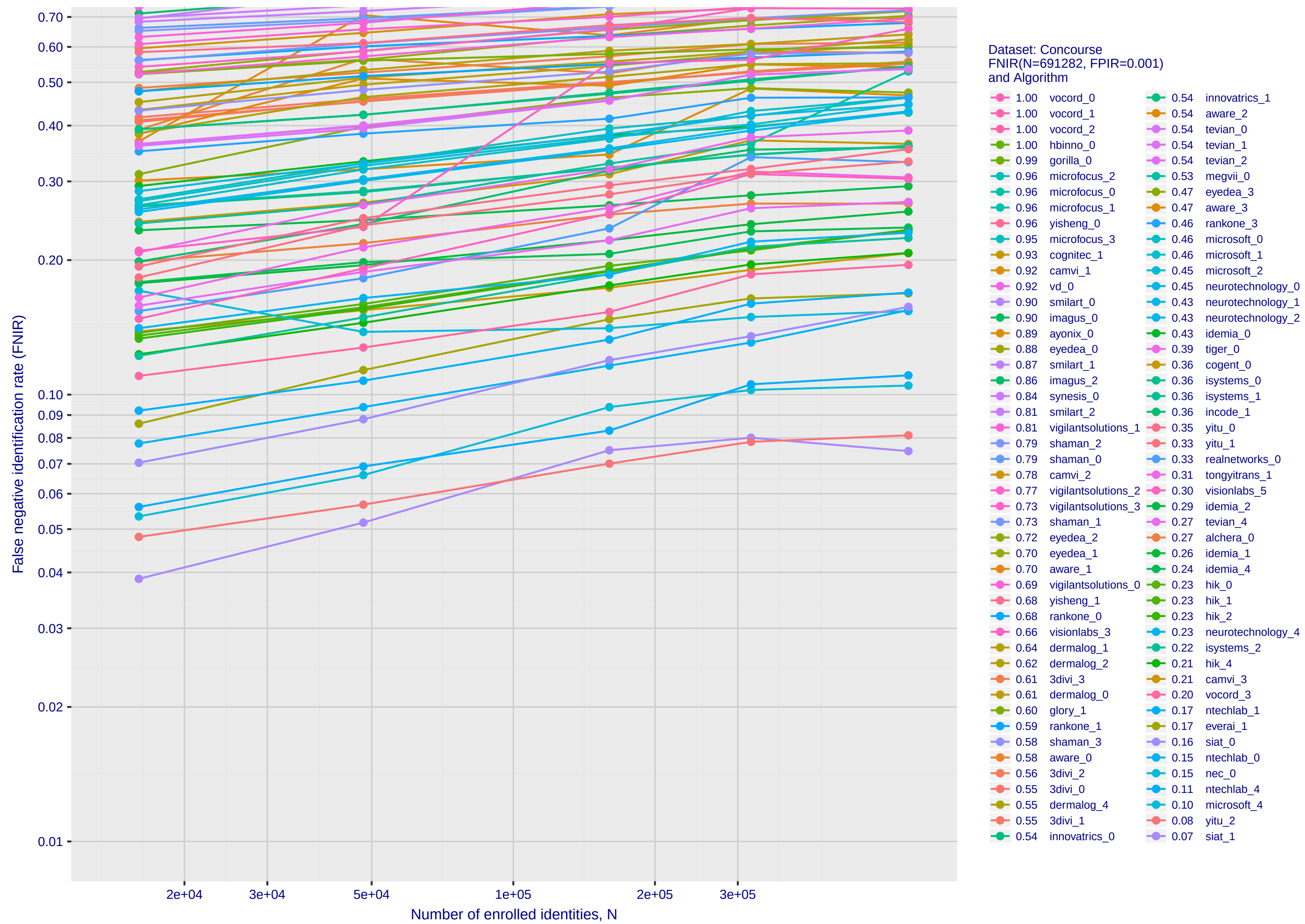

Figure 95: **[FRPC Dataset: Concourse] Miss rates vs. number of enrolled identities**. *The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of people walking down a travel concourse, searched against well-controlled, portrait images of up to 691 282 individuals enrolled into a gallery. The curves show false negative identification rates vs. enrolled population size - FNIR(N, L, T) - when the threshold is set to a high value sufficient to limit false positive outcomes, FPIR = 0.001. This metric is relevant to automated watchlist applications, where most searches are from individuals who are not enrolled.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

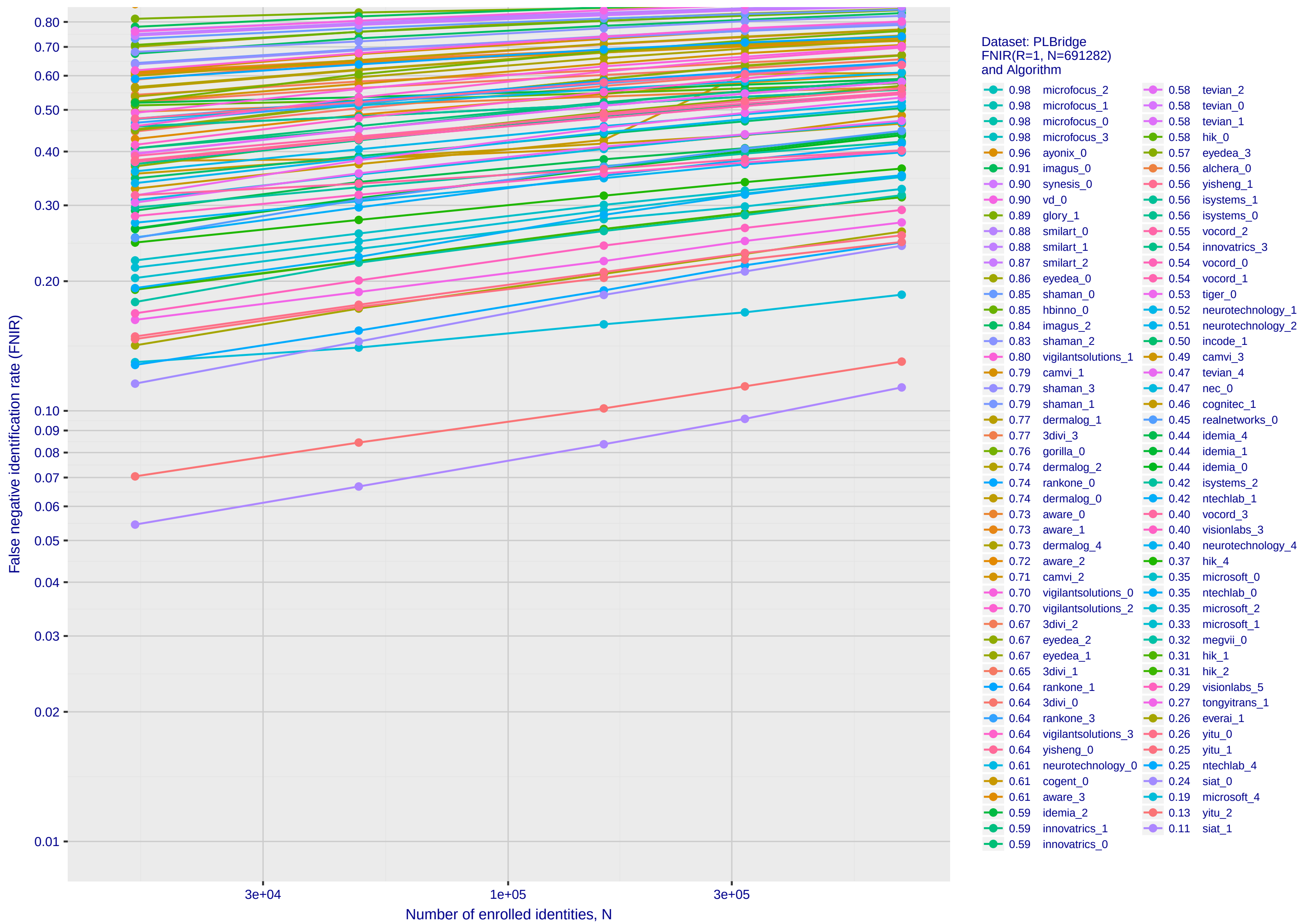

Figure 96: **[FRPC Dataset: Passenger Loading Bridge] Miss rates vs. number of enrolled identities**. *The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of subjects walking along a purpose-built simulated passenger loading bridge, searched against well-controlled, portrait images of up to 691 282 individuals enrolled into a gallery. The curves show false negative identification rates at rank 1 as a function of enrolled population size, FNIR(N, 1). The threshold is set to zero. This metric is relevant to human reviewers who will traverse candidate lists in pursuit of investigations.*

2018/11/26 07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

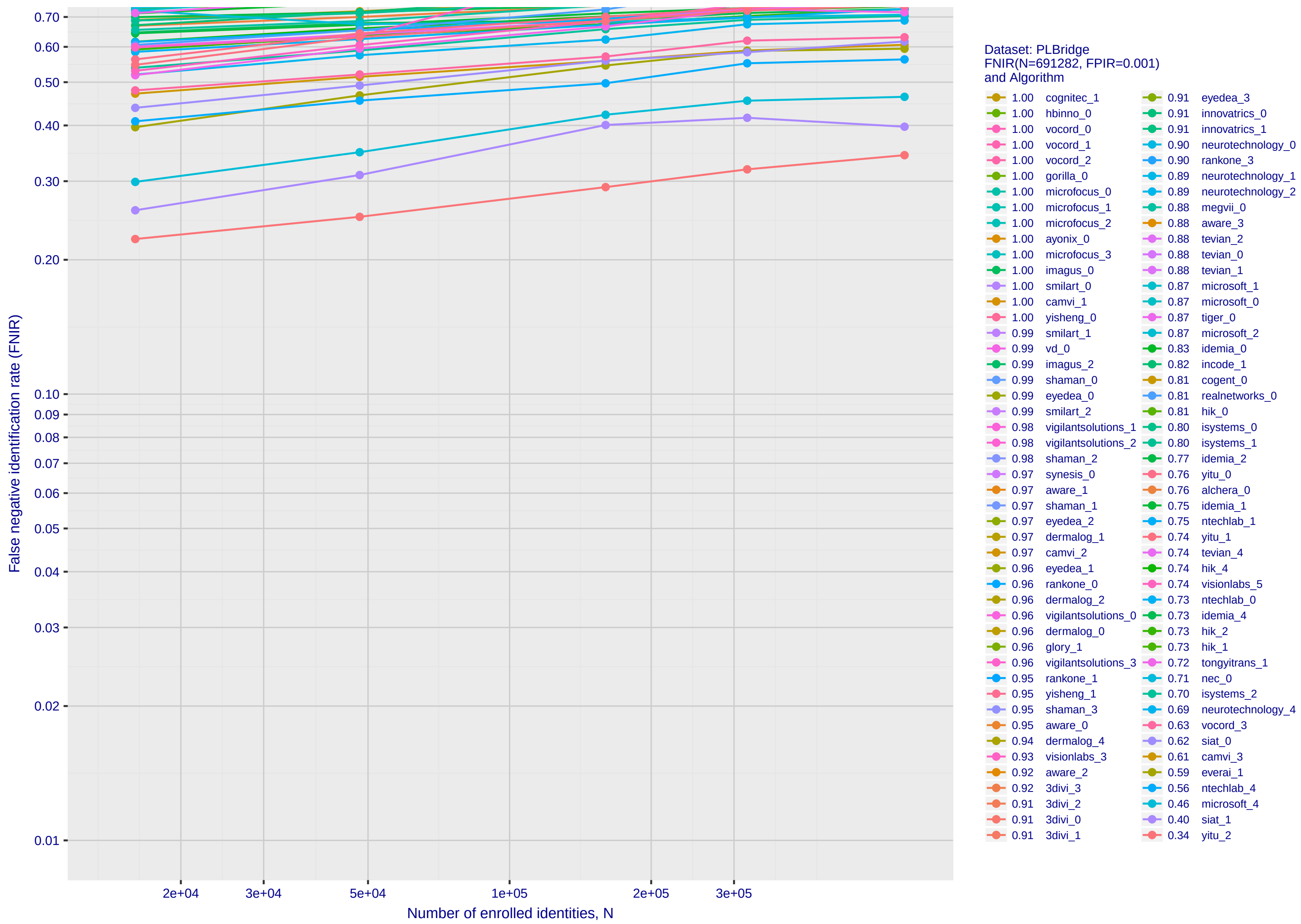

Figure 97: **[FRPC Dataset: Passenger Loading Bridge] Miss rates vs. number of enrolled identities**. *The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of subjects walking along a purpose-built simulated passenger loading bridge, searched against well-controlled, portrait images of up to 691 282 individuals enrolled into a gallery. The curves show false negative identification rates vs. enrolled population size - FNIR(N, L, T) - when the threshold is set to a high value sufficient to limit false positive outcomes, FPIR = 0.001. This metric is relevant to automated watchlist applications, where most searches are from individuals who are not enrolled.*

Appendix F Accuracy when identifying wild images

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

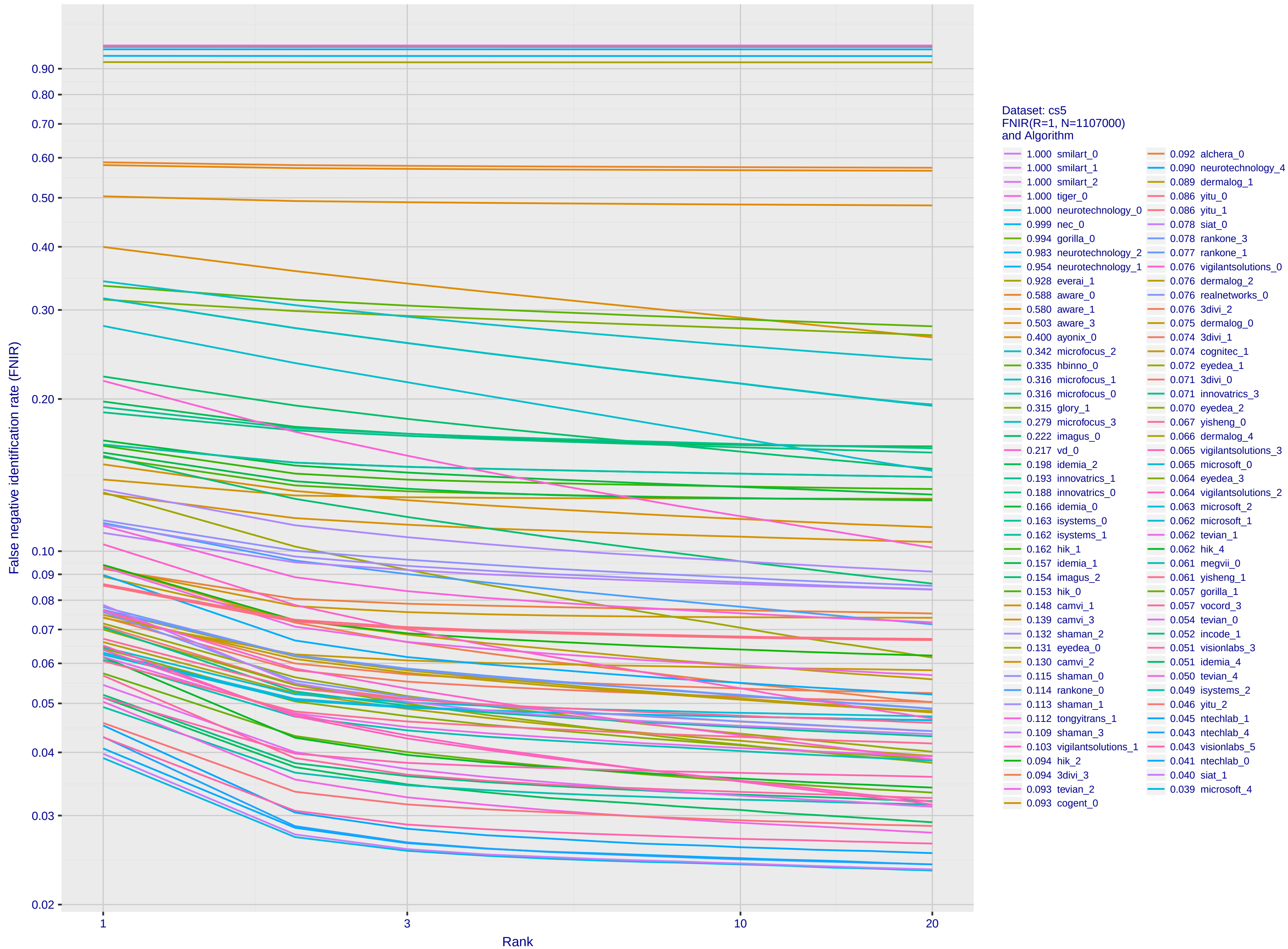

Figure 98: **[Wild Dataset] Identification miss rates vs. rank**. *For the wild dataset, the figure shows false negative identification rates (FNIR) vs. rank when the threshold is set to zero. This metric is relevant to human reviewers who will traverse candidate lists checking whether any of the returned identities match to the search imagery. Specifically, wild images were searched against 1.1 million individuals enrolled with wild images as well.*

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8238

2018/11/26
07:24:51

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

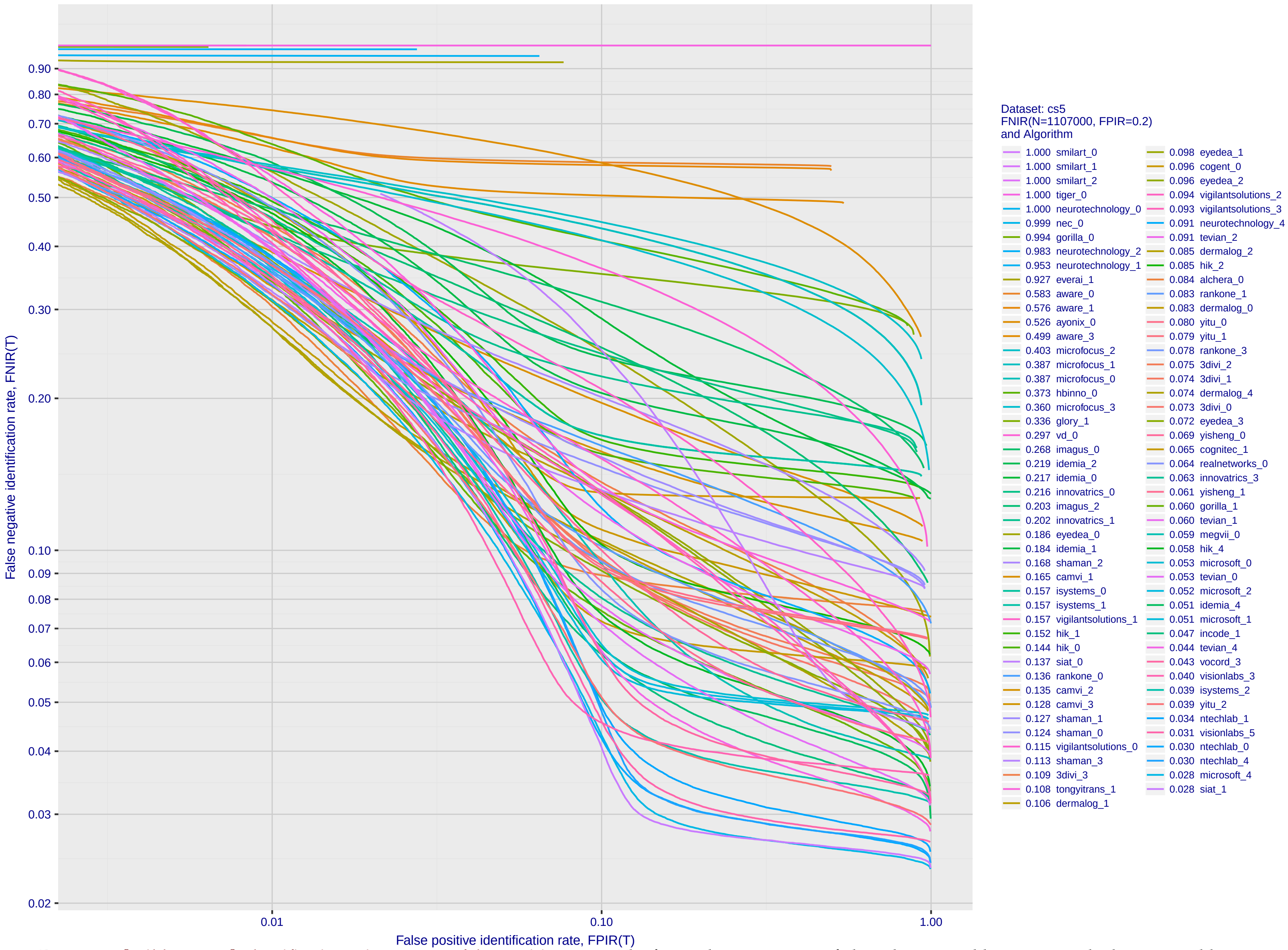

Figure 99: **[Wild Dataset] Identification miss rates vs. false positive rates**. *The figure shows accuracy of algorithms on wild images searched against wild images of 1.1 million individuals enrolled into a gallery. On the vertical axis is miss rate FNIR(N, T, L) with N = 1 107 000, as a function of false positive identification FPIR(N, T). The rapid increase in FNIR below FPIR = 0.1 suggests that some background identities in the gallery are actually present in the non-mated search sets. This issue will be addressed in the 2019 revision of this report.*

Appendix G Search duration

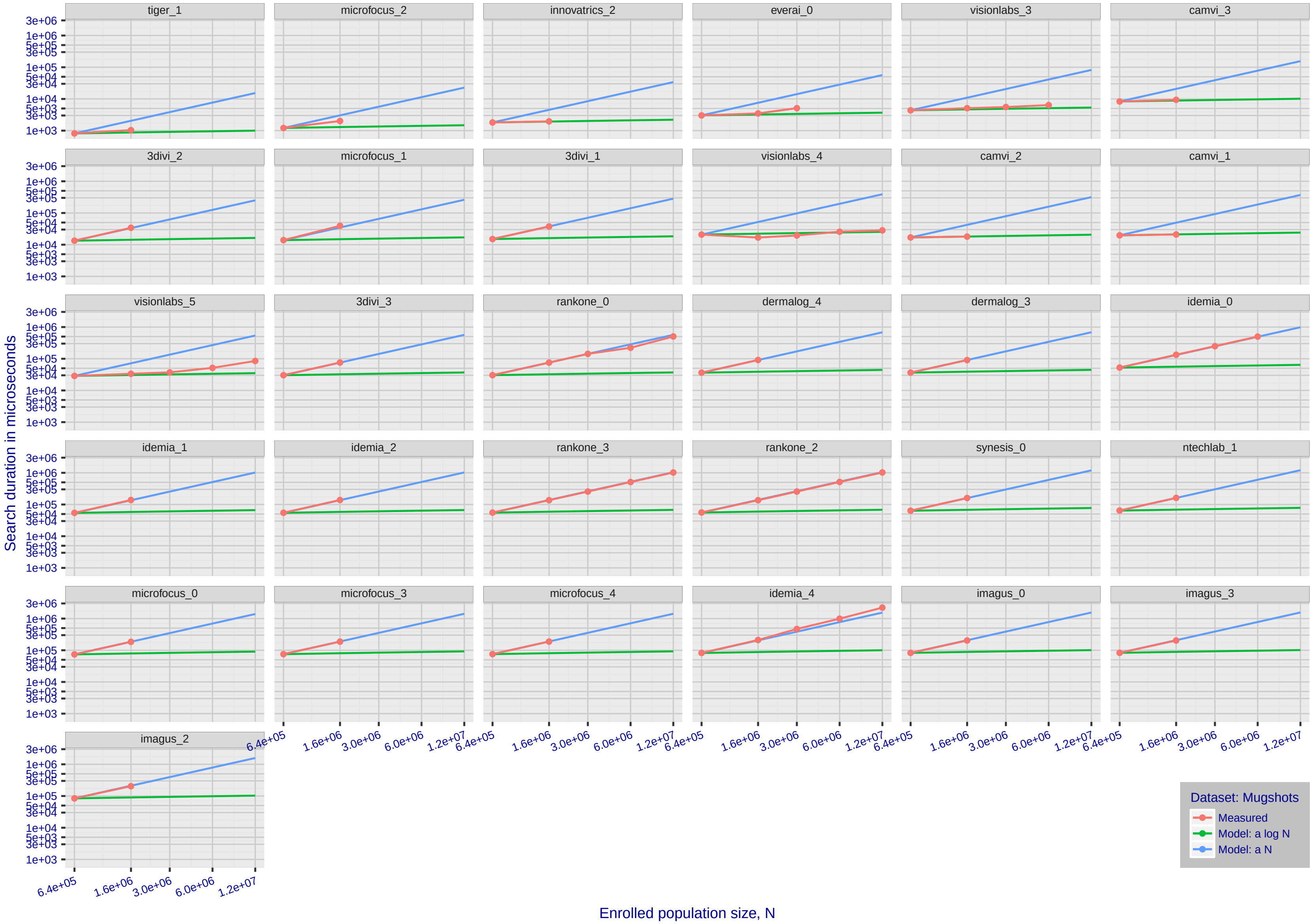

Figure 100: **[Mugshot Dataset] Search duration vs. enrolled population size**. *The red line shows actual durations measured on single c. 2016 core. The blue shows linear growth from N = 640000. The green line shows logathmic growth from that point. The red lines often covers blue. Notable sublinear growth from algorithms from Belair, Ventiane, Chongqing, and Monza. Note that search times are sometimes dominated by the template generation times shown in Table 10.*

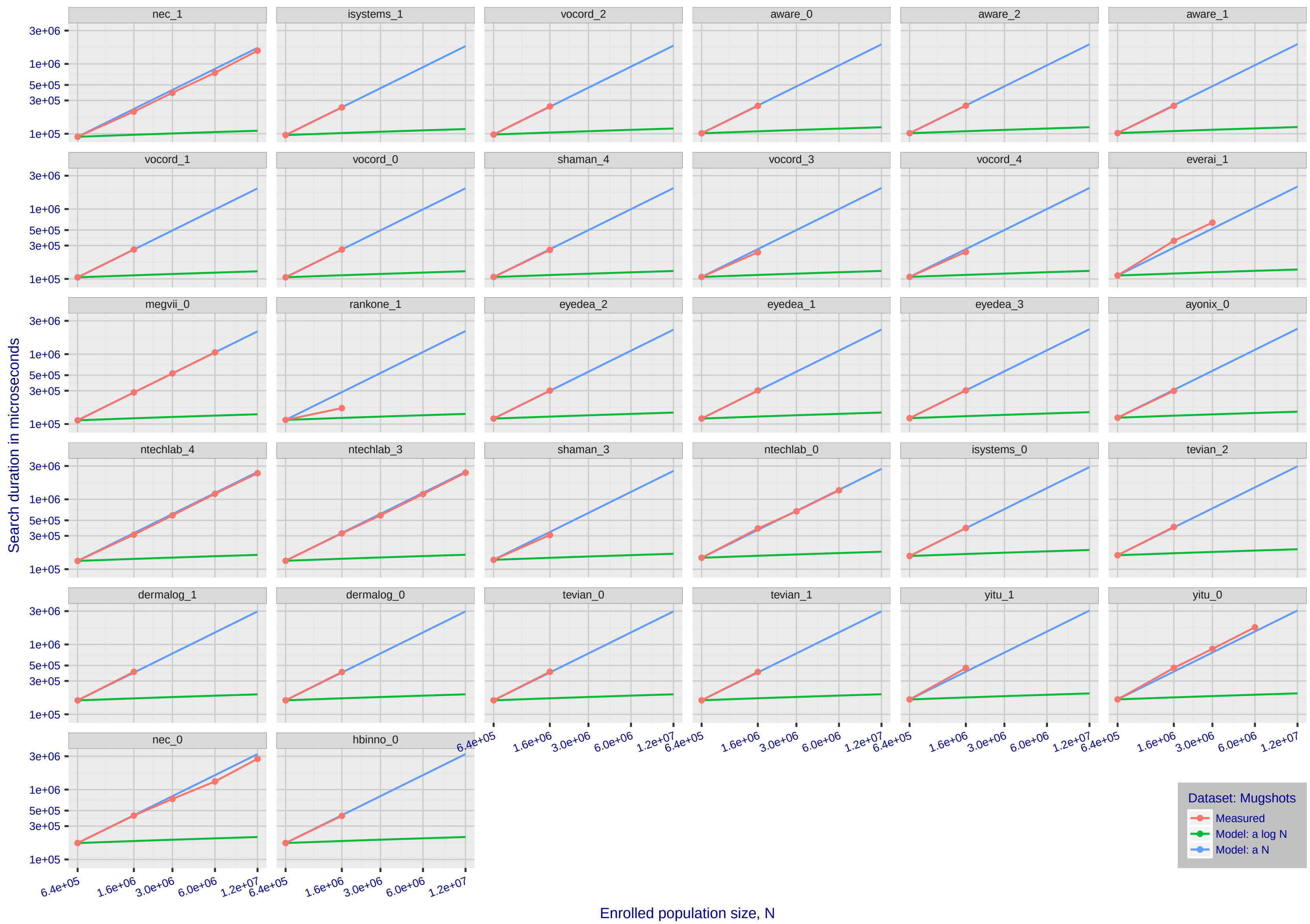

Figure 101: **[Mugshot Dataset] Search duration vs. enrolled population size**. *The red line shows actual durations measured on single c. 2016 core. The blue shows linear growth from N = 640000. The green line shows logathmic growth from that point. The red lines often covers blue. Notable sublinear growth from algorithms from Belair, Ventiane, Chongqing, and Monza. Note that search times are sometimes dominated by the template generation times shown in Table 10.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

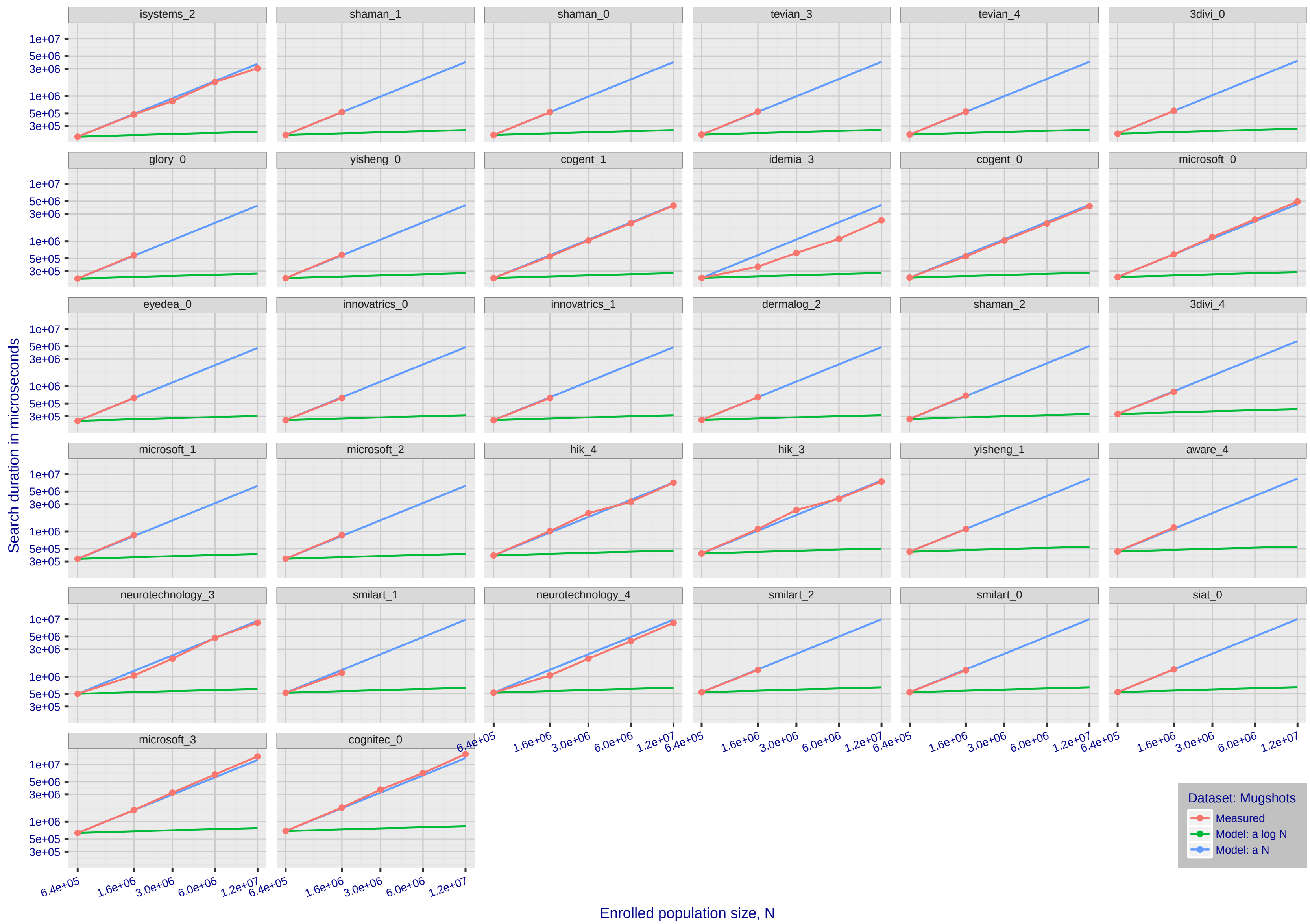

Figure 102: **[Mugshot Dataset] Search duration vs. enrolled population size**. *The red line shows actual durations measured on single c. 2016 core. The blue shows linear growth from N = 640000. The green line shows logathmic growth from that point. The red lines often covers blue. Notable sublinear growth from algorithms from Belair, Ventiane, Chongqing, and Monza. Note that search times are sometimes dominated by the template generation times shown in Table 10.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

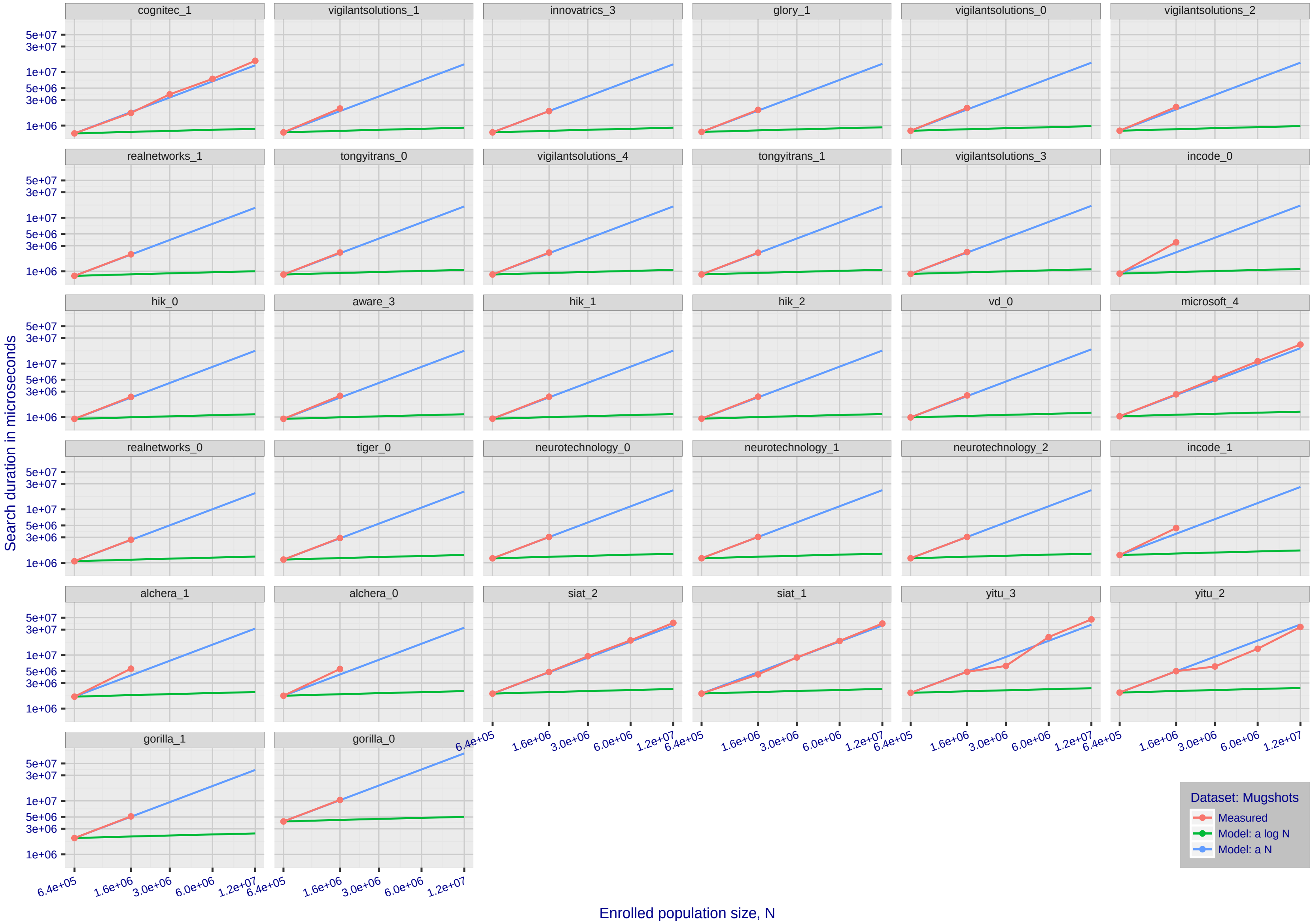

Figure 103: **[Mugshot Dataset] Search duration vs. enrolled population size**. *The red line shows actual durations measured on single c. 2016 core. The blue shows linear growth from N = 640000. The green line shows logathmic growth from that point. The red lines often covers blue. Notable sublinear growth from algorithms from Belair, Ventiane, Chongqing, and Monza. Note that search times are sometimes dominated by the template generation times shown in Table 10.*

FNIR(N, R, T) = False neg. identification rate
FPIR(N, T) = False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

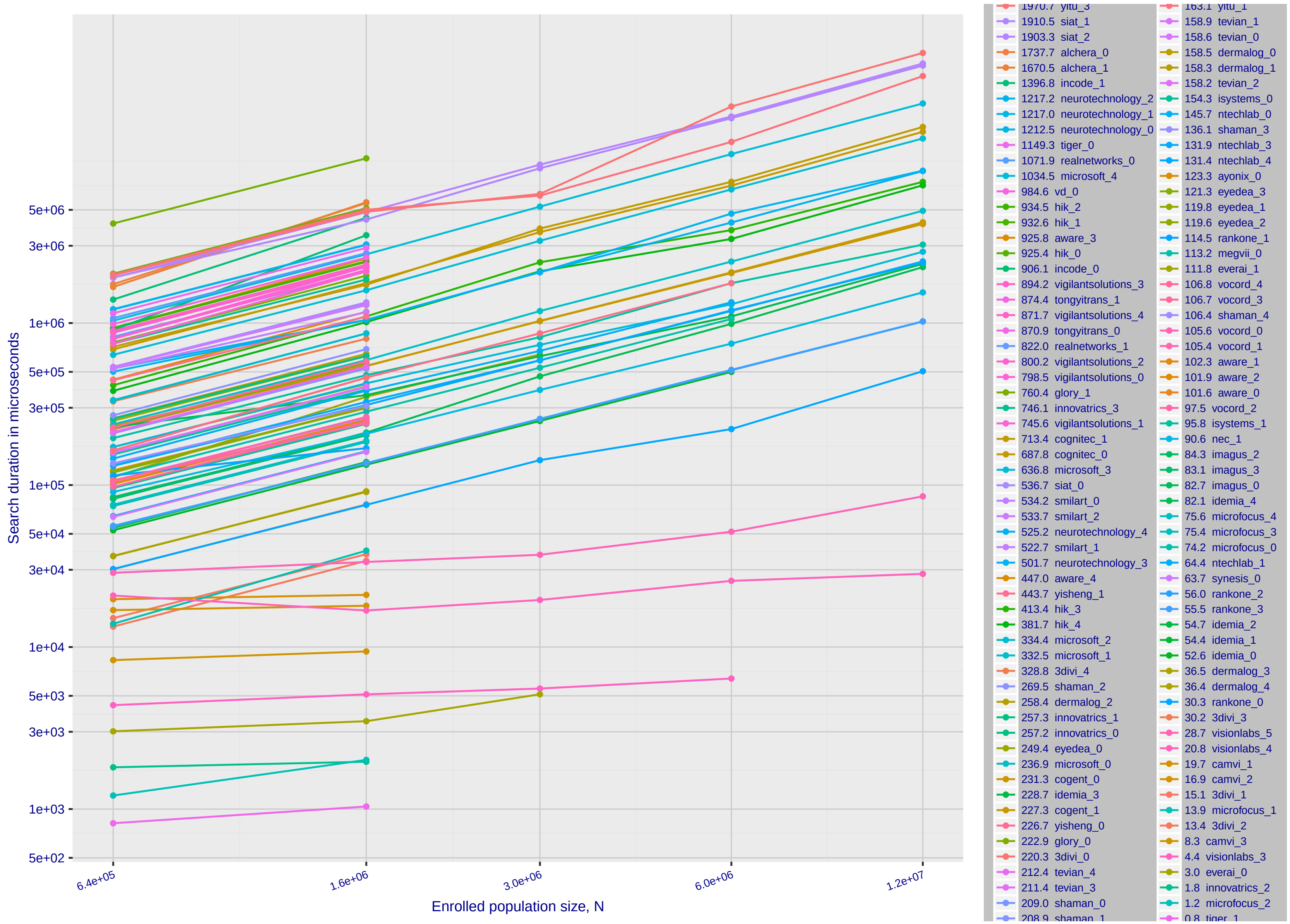

Figure 104: **[Mugshot Dataset] Search duration vs. enrolled population size**. *Alternative visualization of the same data as shown in Figure 103. Generally, only the more accurate algorithms were run on galleries with $N \geq 3\,000\,000$.*

References

[1] L. Best-Rowden and A. K. Jain. Longitudinal study of automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1):148–162, Jan 2018.

[2] Blumstein, Cohen, Roth, and Visher, editors. *Random parameter stochastic models of criminal careers*. National Academy of Sciences Press, 1986.

[3] Thomas P. Bonczar and Lauren E. Glaze. Probation and parole in the united statesm 2007, statistical tables. Technical report, Bureau of Justice Statistics, December 2008.

[4] White D., Kemp R. I., Jenkins R., Matheson M, and Burton A. M. Passport officers errors in face matching. *PLoS ONE*, 9(8), 2014. e103510. doi:10.1371/journal. pone.0103510.

[5] P. Grother, G. W. Quinn, and P. J. Phillips. Evaluation of 2d still-image face recognition algorithms. NIST Interagency Report 7709, National Institute of Standards and Technology, 8 2010. http://face.nist.gov/mbe as MBE2010 FRVT2010.

[6] P. J. Grother, R. J. Micheals, and P. J. Phillips. Performance metrics for the frvt 2002 evaluation. In *Proceedings of Audio and Video Based Person Authentication Conference (AVBPA)*, June 2003.

[7] Patrick Grother, George Quinn, and Mei Ngan. Face in video evaluation (five) face recognition of non-cooperative subjects. Interagency Report 8173, National Institute of Standards and Technology, March 2017. https://doi.org/10.6028/NIST.IR.8173.

[8] Patrick Grother, George W. Quinn, and Mei Ngan. Face recognition vendor test - still face image and video concept, evaluation plan and api. Technical report, National Institute of Standards and Technology, 7 2013. http://biometrics.nist.gov/cs_links/face/frvt/frvt2012/NIST_FRVT2012_api_Aug15.pdf.

[9] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016.

[10] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[11] Ira Kemelmacher-Shlizerman, Steven M. Seitz, Daniel Miller, and Evan Brossard. The megaface benchmark: 1 million faces for recognition at scale. *CoRR*, abs/1512.00596, 2015.

[12] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.

[13] P. Jonathon Phillips, Amy N. Yates, Ying Hu, Carina A. Hahn, Eilidh Noyes, Kelsey Jackson, Jacqueline G. Cavazos, Géraldine Jeckeln, Rajeev Ranjan, Swami Sankaranarayanan, Jun-Cheng Chen, Carlos D. Castillo, Rama Chellappa, David White, and Alice J. O'Toole. Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, 115(24):6171–6176, 2018.

[14] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015.

[15] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[16] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.

[17] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 1701–1708, Washington, DC, USA, 2014. IEEE Computer Society.

[18] Working Group 3. Ed. M. Werner. *ISO/IEC 19794-5 Information Technology - Biometric Data Interchange Formats - Part 5: Face image data*. JTC1 :: SC37, 2 edition, 2011. http://webstore.ansi.org.

[19] David White, James D. Dunn, Alexandra C. Schmid, and Richard I. Kemp. Error rates in users of automatic face recognition software. *PLoS ONE*, October 2015.

[20] Bradford Wing and R. Michael McCabe. Nist special publication 500-271: American national standard for information systems data format for the interchange of fingerprint, facial, and other biometric information part 1. Technical report, September 2015. ANSI/NIST ITL 1-2015.

[21] Andreas Wolf. Portrait quality - (reference facial images for mrtd). Technical report, ICAO, April 2018.