

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>

2018/11/26

NIST
**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 Ross, Secretary

National Institute of Standards and Technology
Walter Copan, Director

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 February and June 2018 respectively, with developers receiving technical feedback after each. A third phase commenced on October 30, 2018, results from which will be 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 2018¹.
- ▷ **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 database². 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 availability 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-

¹A 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>

²This 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	Metric	Num- subjects	Num- images	Algorithm		FNIR
Mode				Date	Name	
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

Application	Metric	Num- subjects	Enrollment type	Num- images	Algorithm	FNIR	
Mode						Raw	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.

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.

▷ **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	Metric	Num- subjects	Num- images	Algorithm	FNIR	
Mode					Raw	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.

³See 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 each other, 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 2016⁴ 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 enrolling 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

⁴This 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 core⁵, 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 indispensable 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.

⁵Intel 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 arrestees⁶. 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 referred 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.

⁶Operationally 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.

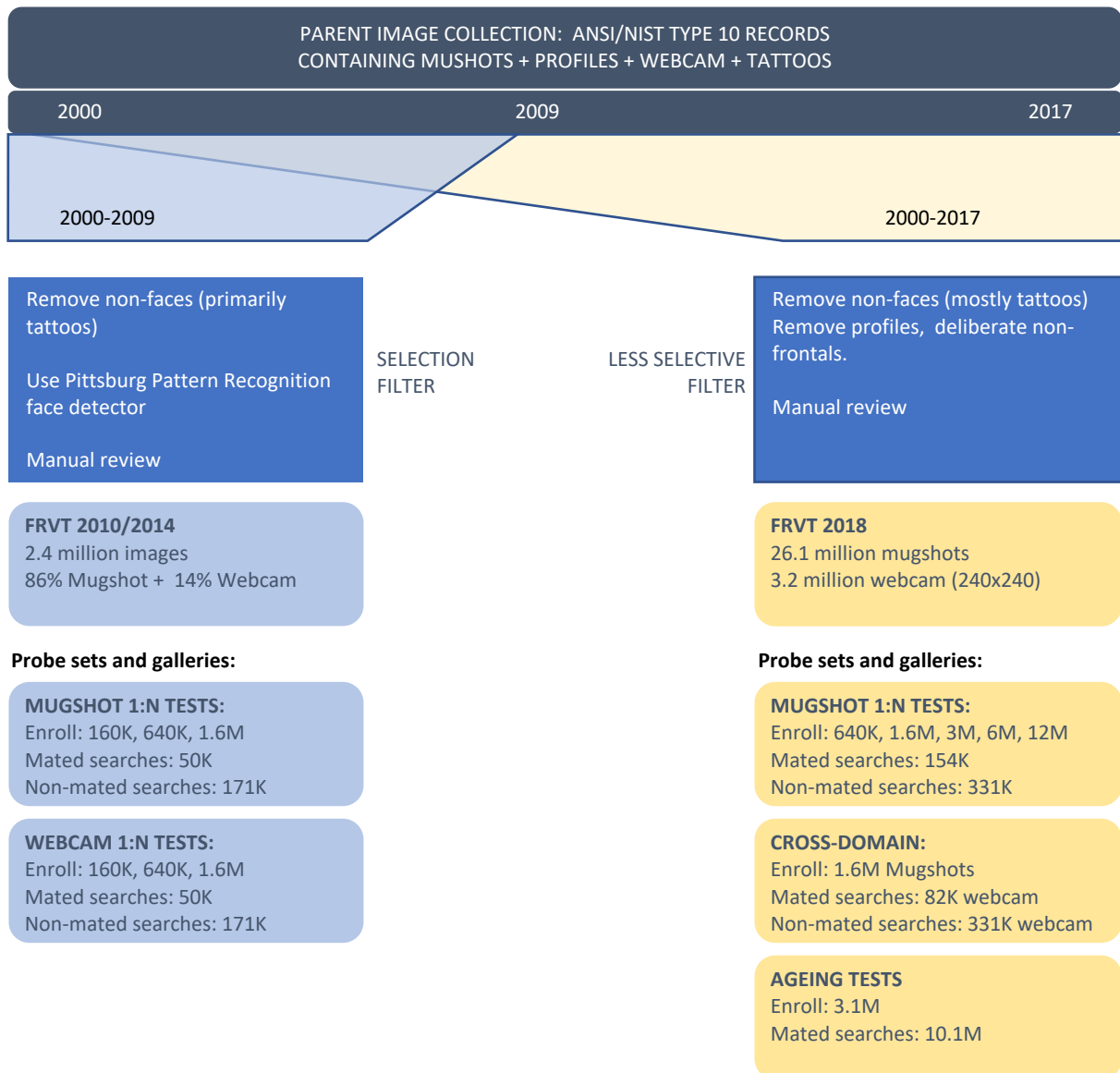


Figure 1: **Mugshot selection.** The left branch of the figure applies to the mugshots used in FRVT 2014, then termed LEO. The right hand branch shows the much larger set used in FRVT 2018. The exact details of the image selection mean that recognition of images in the FRVT 2018 dataset is more difficult 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 referred to here as LEO and FRVT2014.

Example images are shown in Figures 2 and 3.

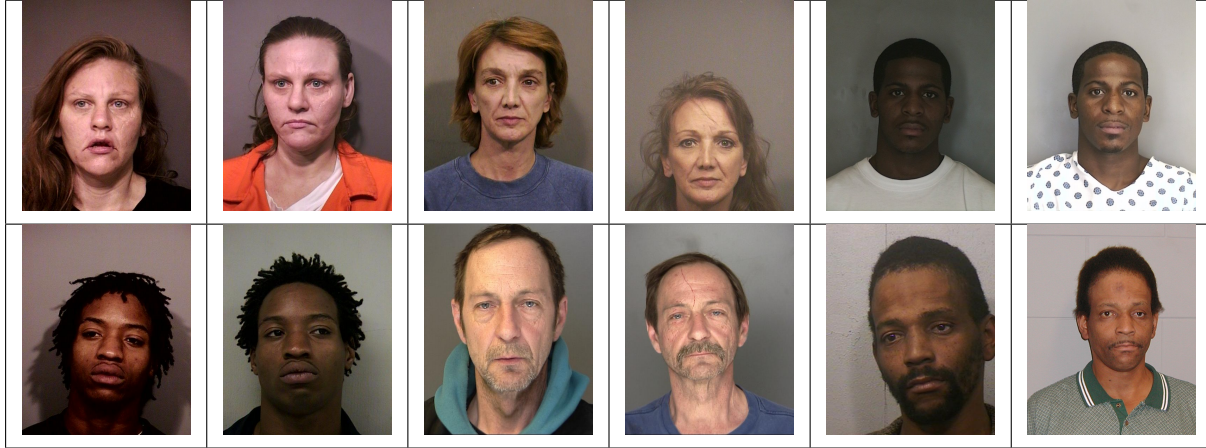


Figure 2: Six mated mugshot pairs representative of the FRVT-2014 (LEO) and FRVT-2018 datasets. The images are collected live, i.e. not scanned from paper. Image source: NIST Special Database 32



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

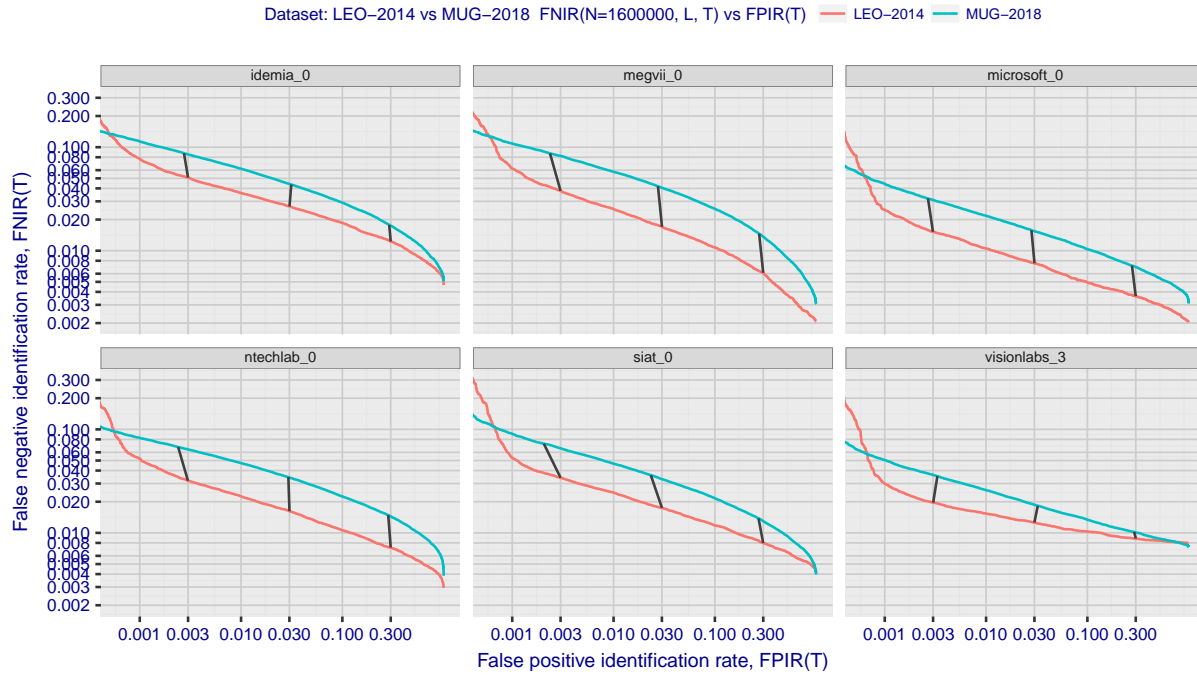


Figure 4: [Relative difficulty of 2013, 2018 datasets] The figure shows results for 2018 algorithms running on two datasets: The LEO set used in FRVT2014 and the mugshots in the FRVT2018 dataset. The axes are identification miss rates vs. false positive rates. Across most of the range the new database 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 Pittsburgh 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 con-

sisted 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.

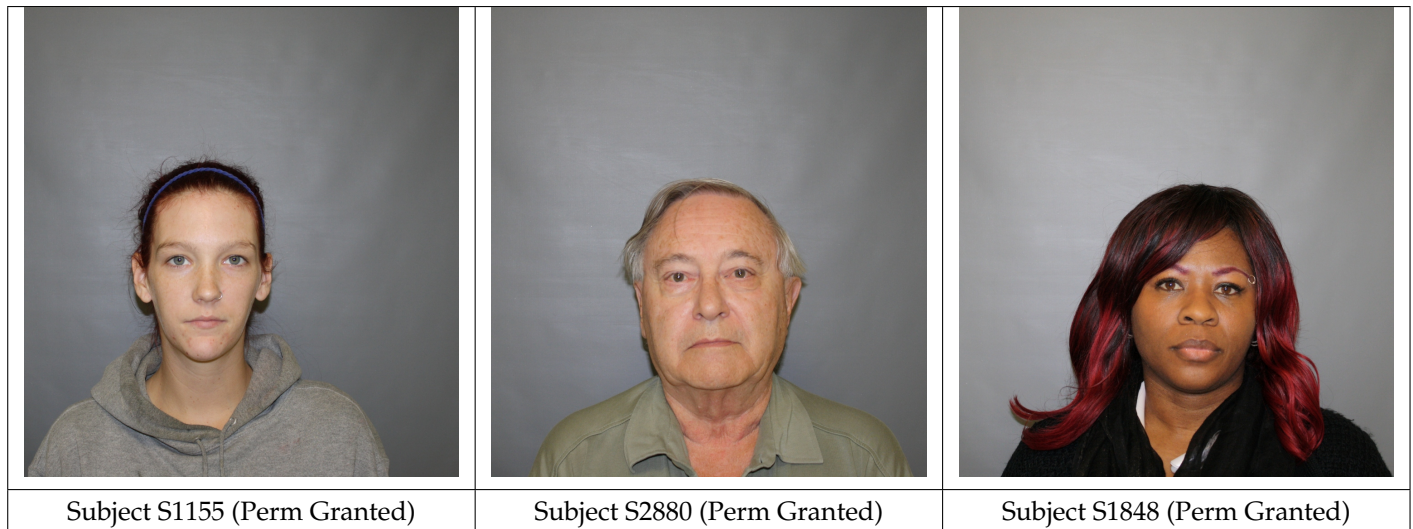


Figure 6: Examples of enrollment images collected with an SLR camera. The face images in this figure are from the DHS / S&T provided AEER dataset. The included subjects consented to release their images in public reports.

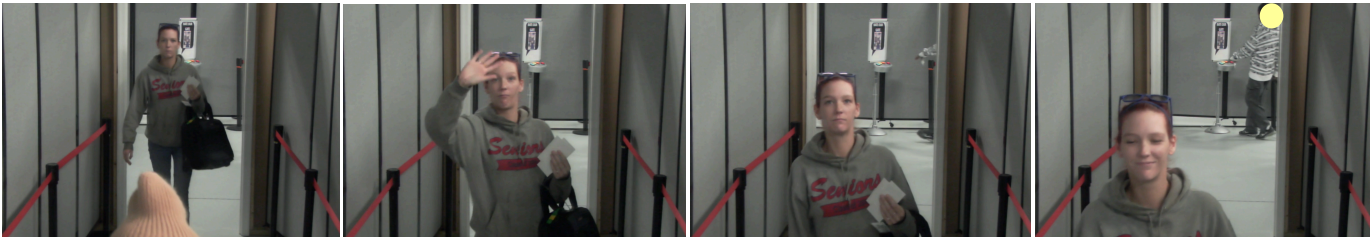


Figure 7: Example images from the ceiling mounted camera for the free movement scenarios from videos collected on an aircraft boarding ramp. The images in this table are from the subject S1115 in the DHS / S&T provided AEER dataset. The subject gave written opt-in permission to allow public release of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



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 assumption⁷. 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 complicated⁸.

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

⁷For 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.
⁸A 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” template⁹ 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 $k = 1 \dots K_i$. 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, $x_1 \dots x_{K_i-1}$. 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, $x_1 \dots x_{K_i-1}$ 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 partition 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%,

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

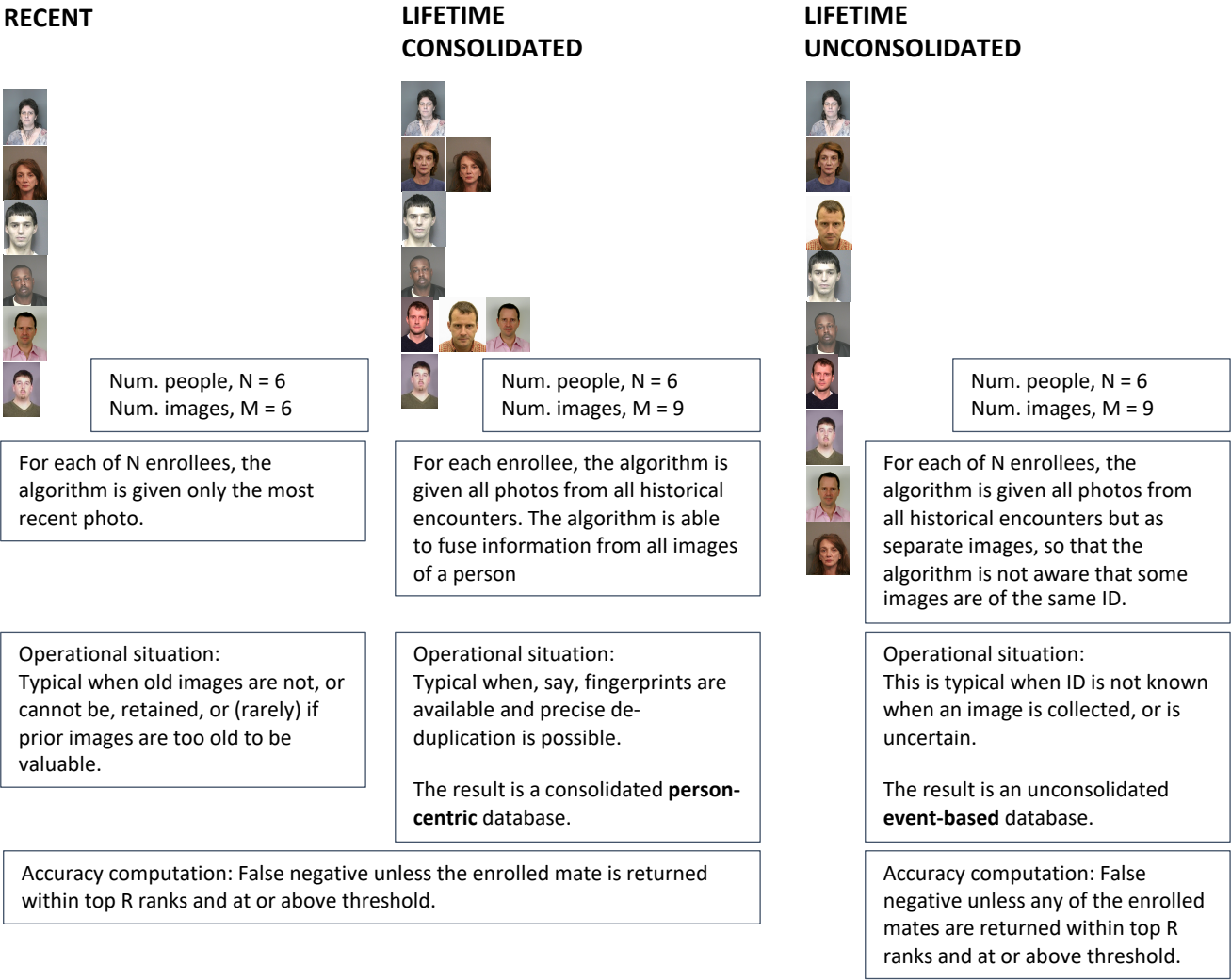


Figure 10: Enrollment database types. The figure shows the three kinds of enrollment databases examined in this report. Image source: NIST Special Database 32

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

	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:

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

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.

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

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:

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

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 subtly, 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, $\text{FNIR}(N, R, 0)$, 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. $\text{CMC}(N, 1)$. 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

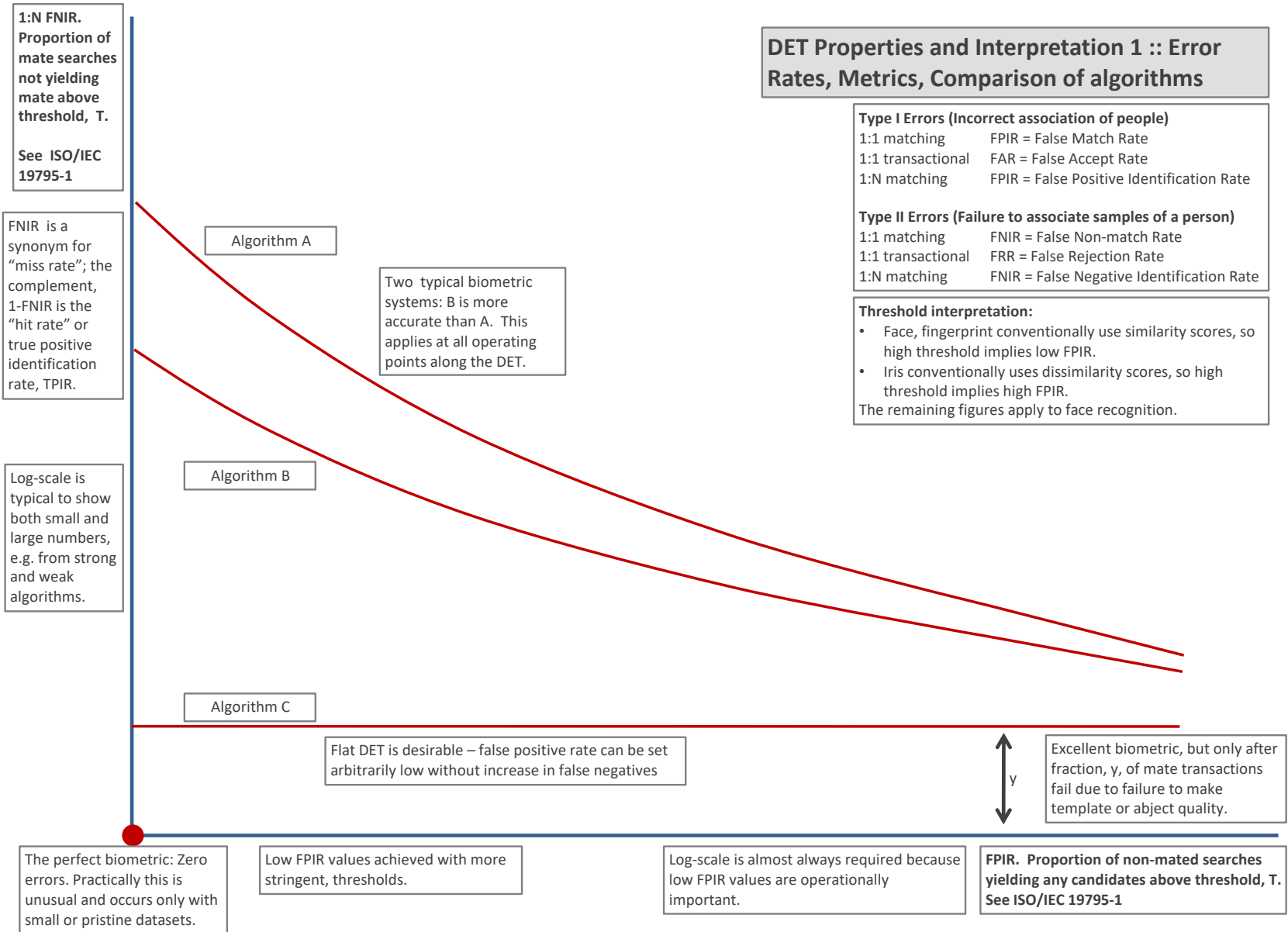


Figure 11: DET as the primary performance reporting mechanism.

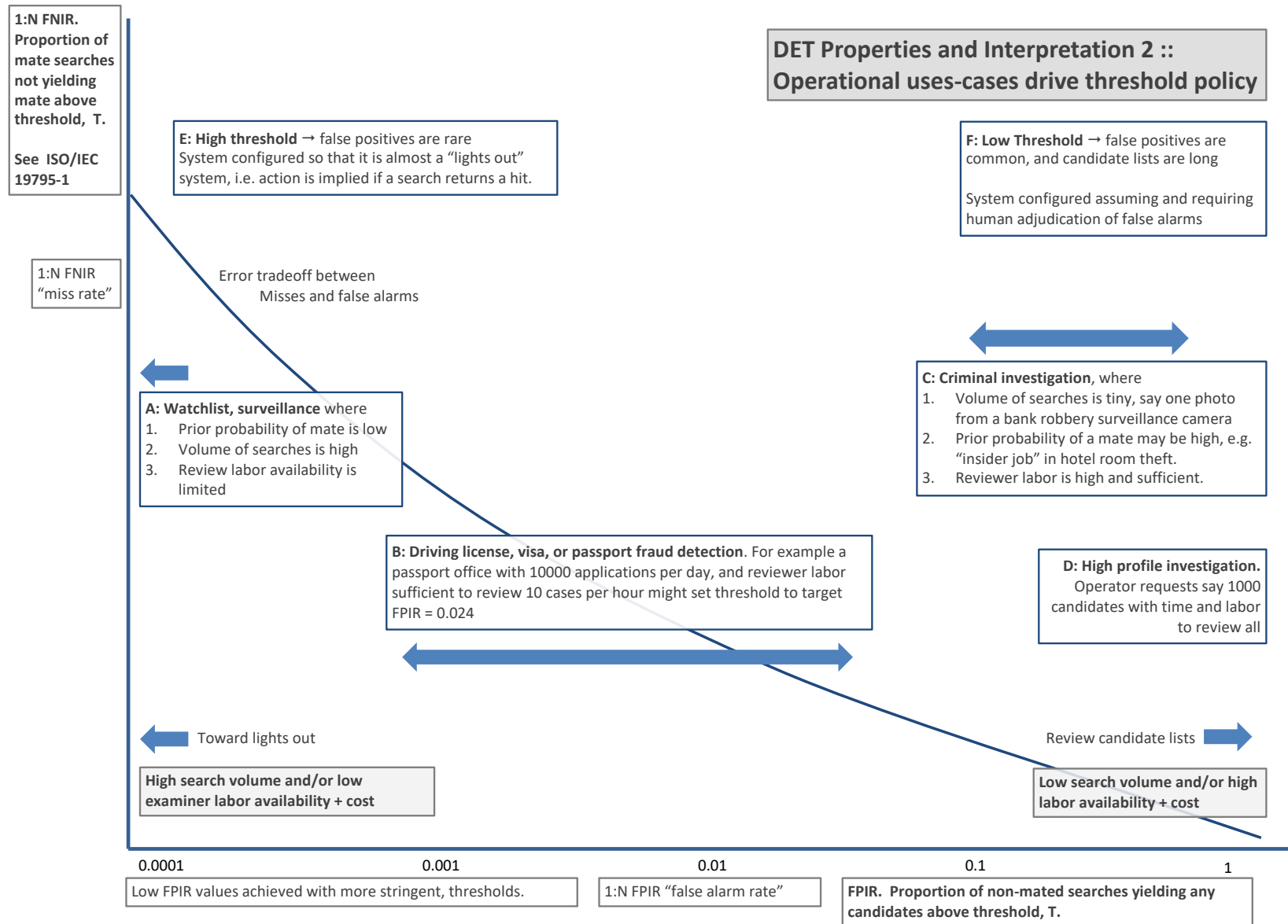


Figure 12: DET as the primary performance reporting mechanism.

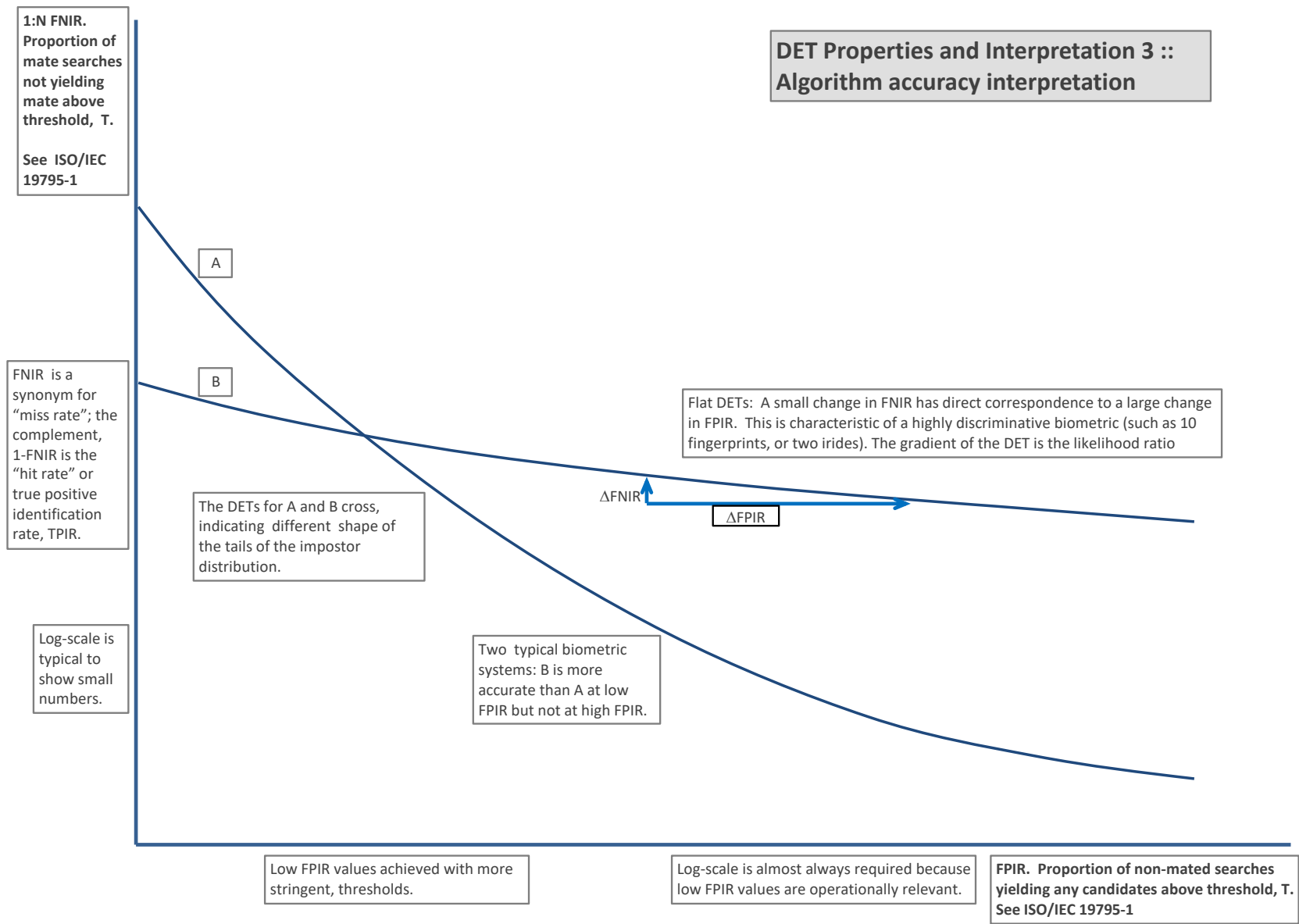


Figure 13: DET as the primary performance reporting mechanism.

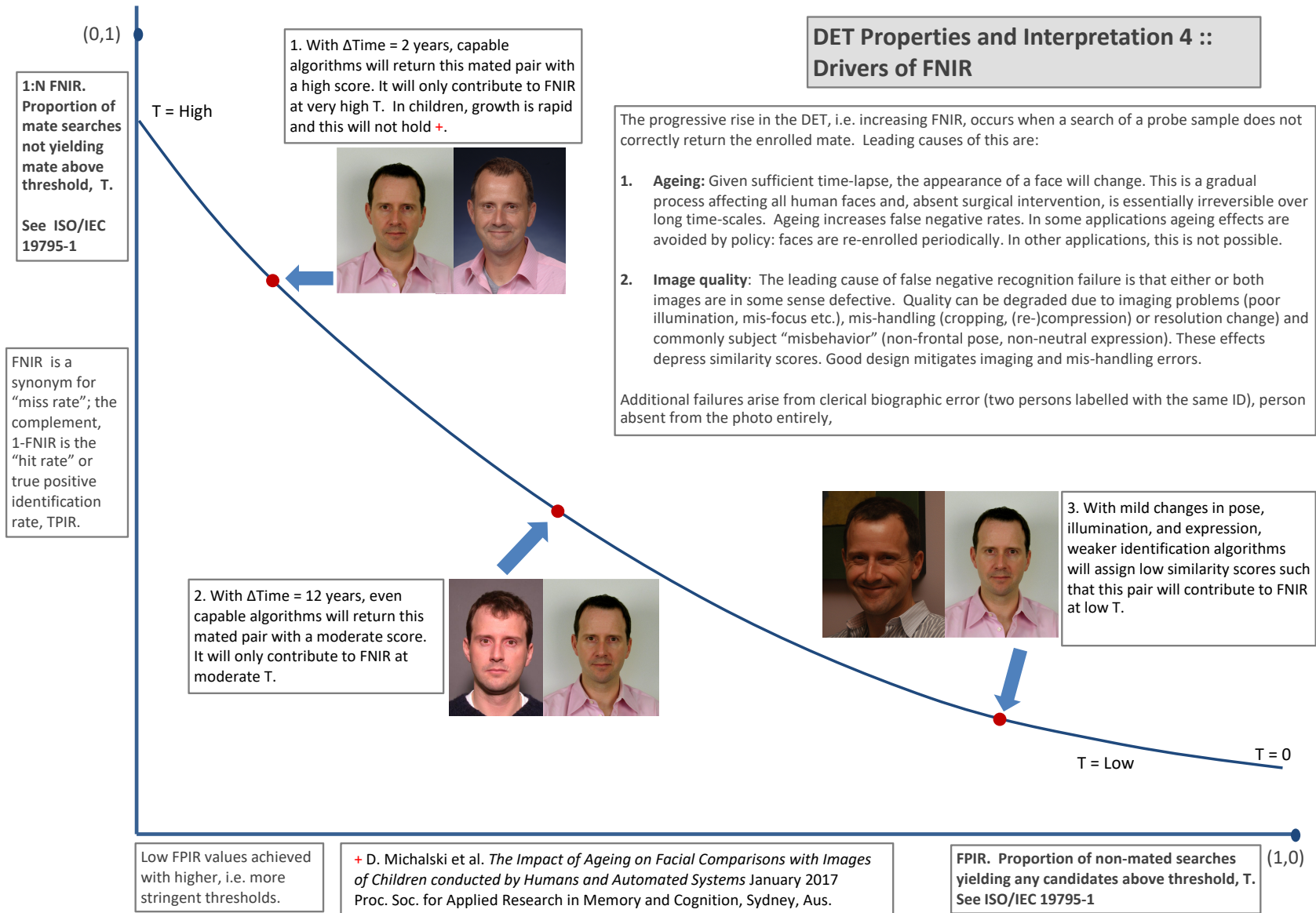


Figure 14: DET as the primary performance reporting mechanism.

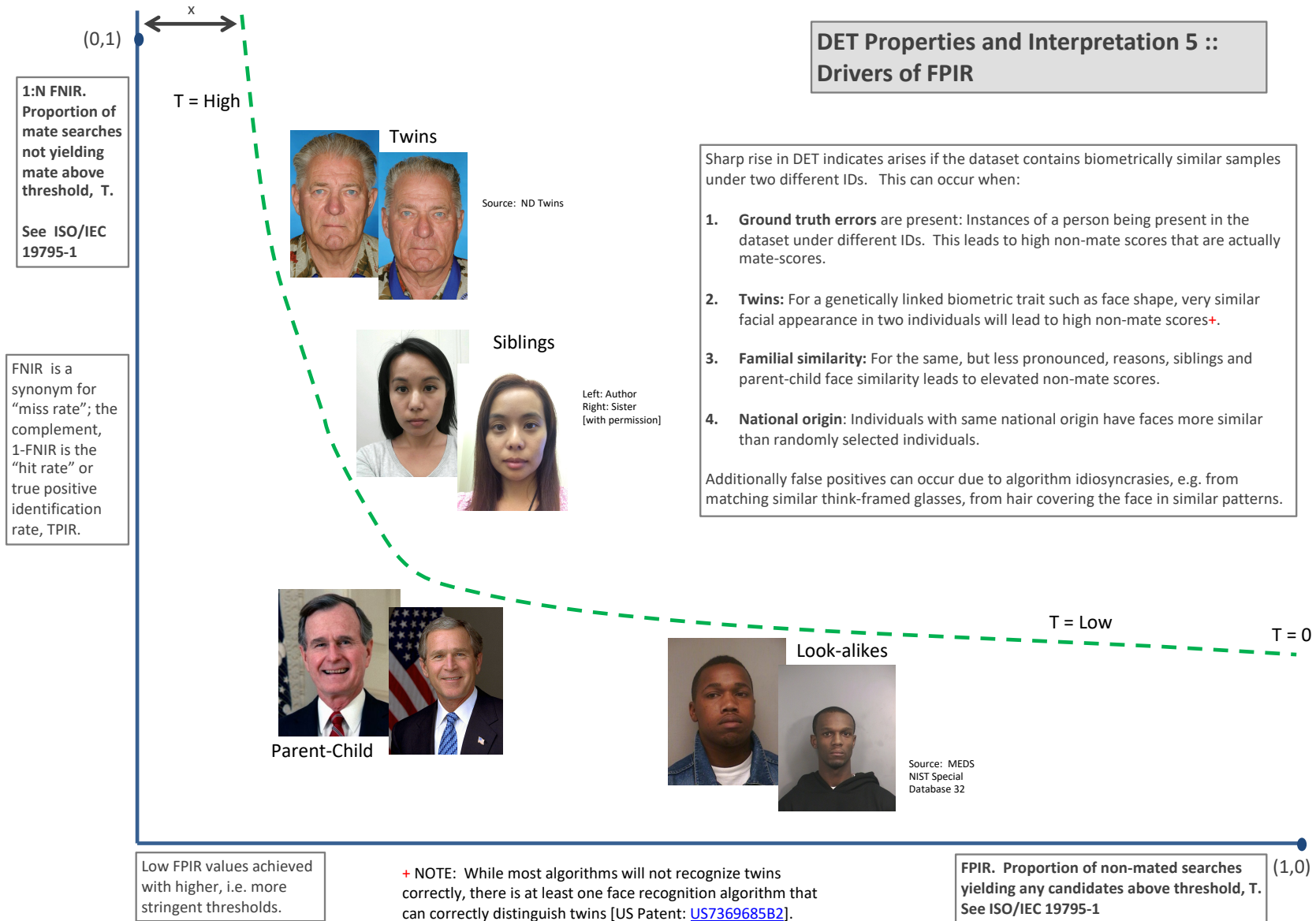
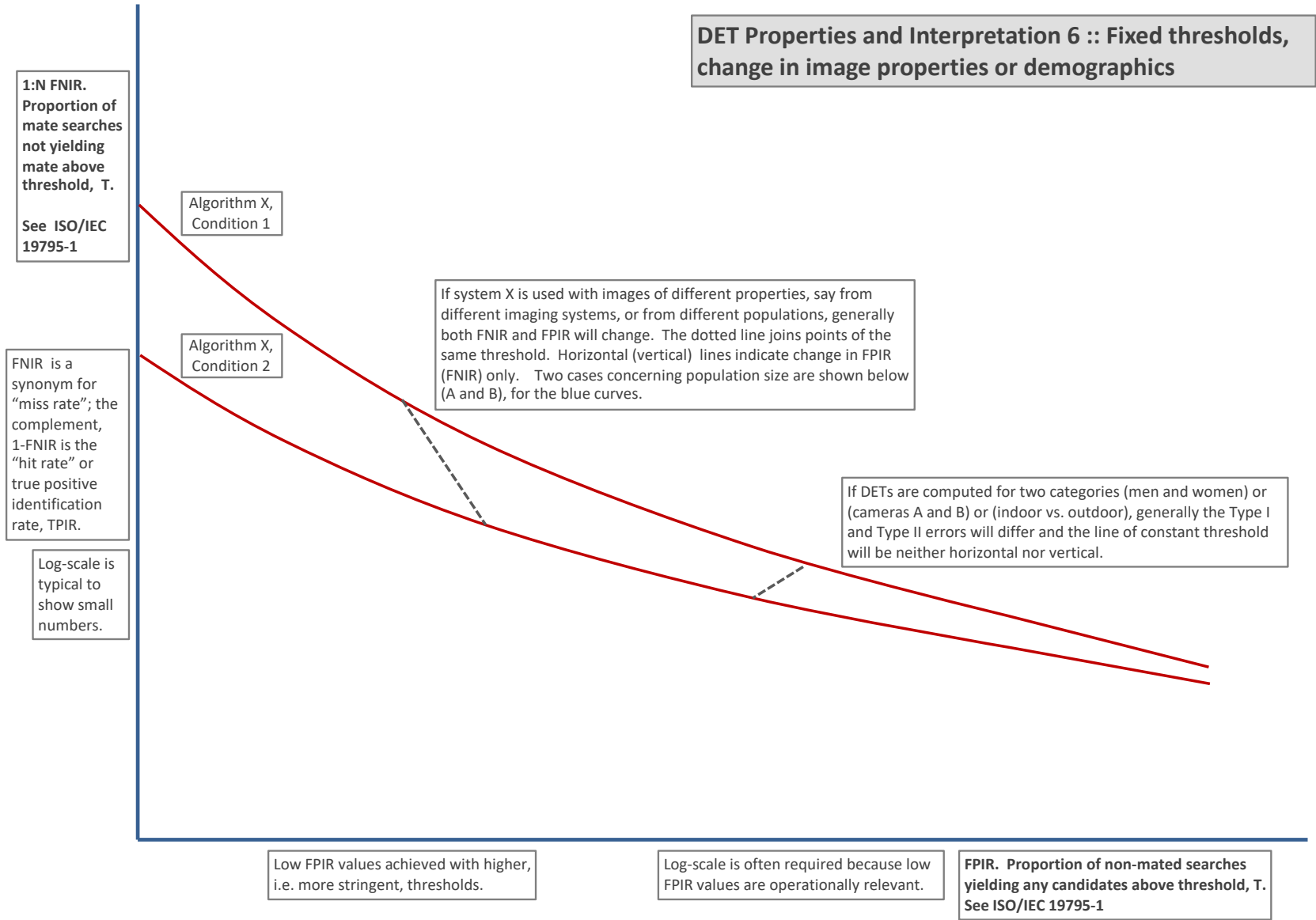


Figure 15: DET as the primary performance reporting mechanism.



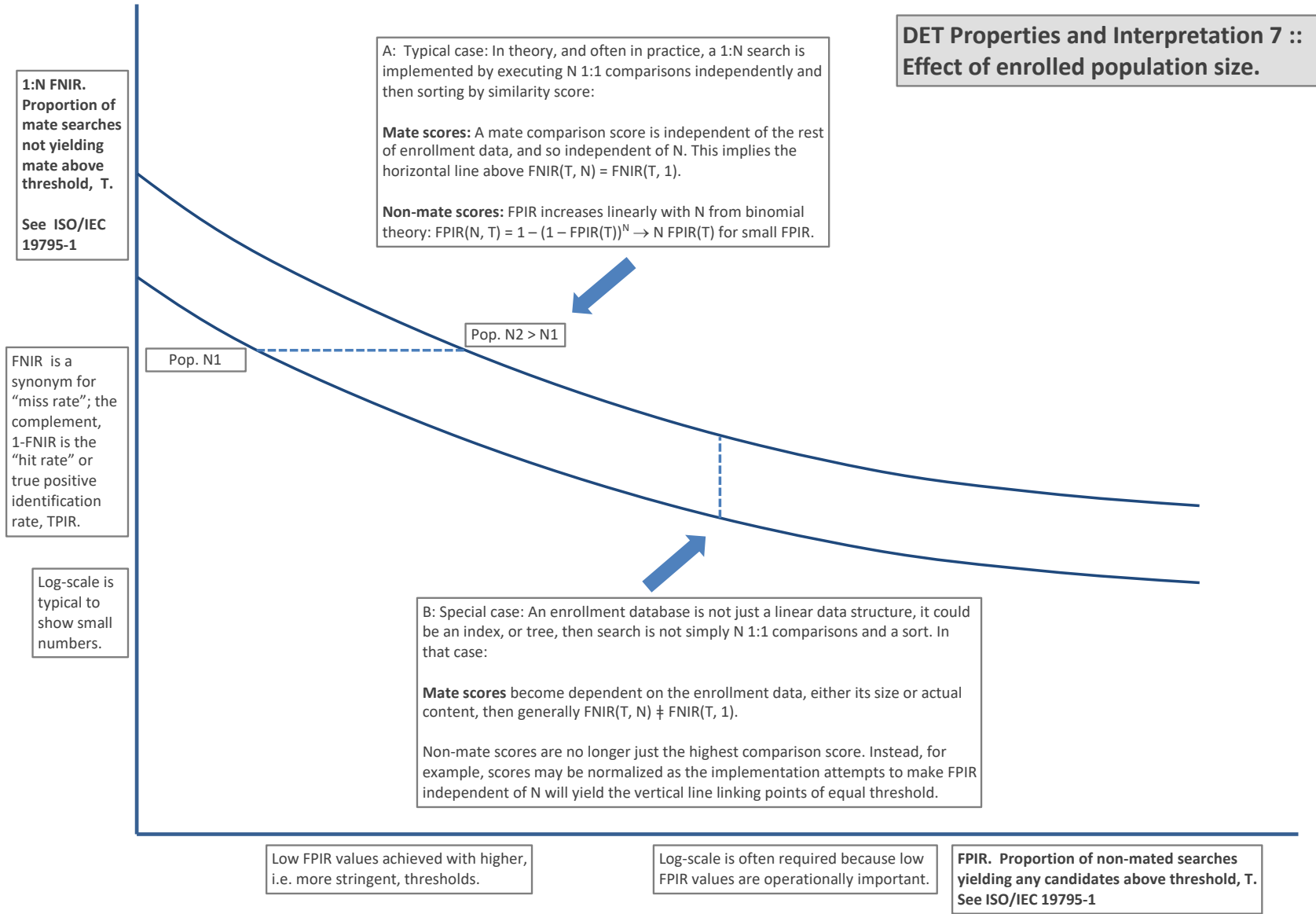


Figure 17: DET as the primary performance reporting mechanism.

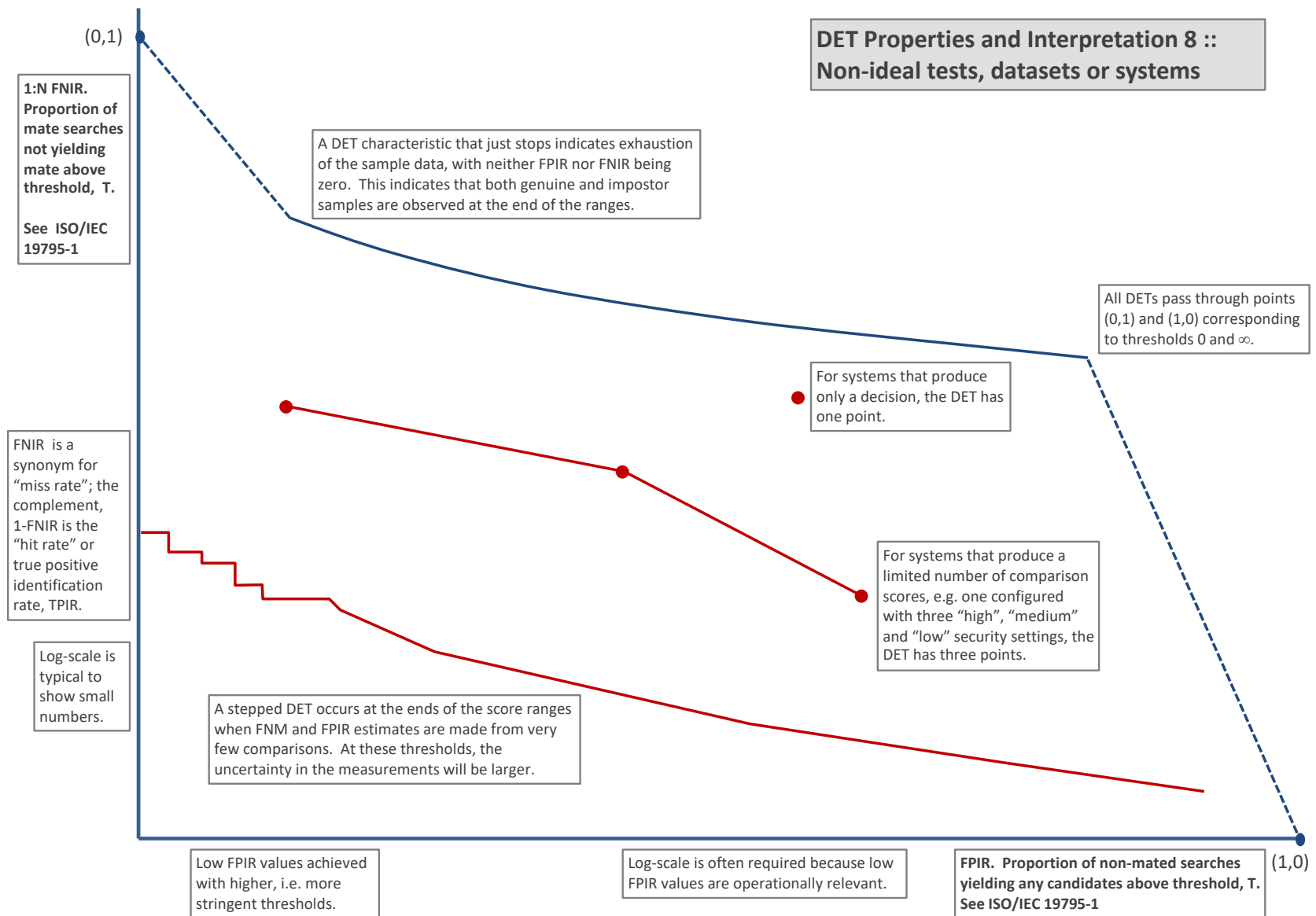


Figure 18: DET as the primary performance reporting mechanism.

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 searches¹⁰. 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 $1 - \text{FTE}$.
- ▷ **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.

$$\text{FNIR}^\dagger = \text{FTX} + (1 - \text{FTX})\text{FNIR} \quad (6)$$

$$\text{FPIR}^\dagger = (1 - \text{FTX})\text{FPIR} \quad (7)$$

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.

¹⁰For 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

$$M(R) = 1 + (1 - CMC(1)) + (1 - CMC(2)) + \dots + (1 - CMC(R-1)) \quad (8)$$

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

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:

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

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

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 $\text{FNIR} \sim 0.0039$. 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 mislabelled images can never be matched. These constitute an accuracy floor in the sample implying that FNIR cannot be below 0.0018¹¹. The profile-views and low-quality images could be successfully matched - indeed some algorithms do so. Likewise some poor quality images are matched.

¹¹This value is the sum of two partial false negative rates: $\text{FNIR}_B = 0.15 * 0.0039$ plus $\text{FNIR}_D = 0.3 * 0.0039$

For the microsoft-4 algorithm the lowest miss rate from (recent entry in Table 10) is $\text{FNIR}(640\,000, 50, 0) = 0.0018$. 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 $\text{FPIR} = 0.001$. 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: Indeterminate 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 considered¹²¹³.

- 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 February 2010 - October 2013¹⁴ 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, February to

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

¹³Note 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.

¹⁴See 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 prevalence 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

	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 (μ S)
									L=1	L=50	
1	3Divi	3divi	0	2018-02-09	186	¹¹⁶ 4096	k	⁶⁴ 426	-	⁷¹ 553	⁶⁶ 0.33 <i>N</i> 1.0
2	3Divi	3divi	1	2018-02-15	187	¹²⁴ 4224	k	⁶⁸ 428	-	¹² 37	³⁷ 0.03 <i>N</i> 1.0
3	3Divi	3divi	2	2018-02-15	187	³⁴ 528	k	⁶⁶ 428	-	¹¹ 33	⁷³ 0.02 <i>N</i> 1.0
4	3Divi	3divi	3	2018-06-19	165	²⁹ 512	k	⁹¹ 625	⁷ 76	¹⁵ 76	⁶² 0.05 <i>N</i> 1.0
5	3Divi	3divi	4	2018-06-19	186	¹¹⁴ 4096	k	⁹² 628	³⁴ 604	⁸² 801	³⁰ 0.75 <i>N</i> 1.0
6	Alchera	alchera	0	2018-06-30	168	⁹⁴ 2048	k	³⁴ 263	⁵⁷ 3296	¹²⁶ 5420	¹²⁴ 0.10 <i>N</i> 1.2
7	Alchera	alchera	1	2018-06-30	46	⁸⁰ 2048	k	⁶⁶ 66	⁵⁸ 3516	¹²⁷ 5489	¹²⁶ 0.05 <i>N</i> 1.3
8	Aware	aware	0	2018-02-16	261	⁶⁸ 1564	k	⁹⁸ 653	-	³⁸ 251	⁴⁰ 0.19 <i>N</i> 1.0
9	Aware	aware	1	2018-02-16	232	⁶⁹ 1564	k	⁹⁷ 651	-	³⁹ 251	³⁴ 0.21 <i>N</i> 1.0
10	Aware	aware	2	2018-02-16	349	¹⁰⁵ 2076	k	¹²⁶ 912	-	⁴⁰ 252	⁴² 0.19 <i>N</i> 1.0
11	Aware	aware	3	2018-06-22	350	¹⁰⁴ 2076	k	¹¹⁰ 716	⁵⁴ 2426	¹¹¹ 2508	¹¹⁷ 0.50 <i>N</i> 1.1
12	Aware	aware	4	2018-06-22	349	² 92	k	¹⁰⁸ 712	⁴⁰ 1232	⁹¹ 1187	¹¹³ 0.33 <i>N</i> 1.1
13	Ayonix	ayonix	0	2018-06-21	57	⁵⁷ 1036	k	¹ 10	²⁰ 283	⁴⁶ 298	²⁴ 0.30 <i>N</i> 1.0
14	Camvi Technologies	camvitech	1	2018-02-16	94	⁵⁰ 1024	1	¹⁹ 177	-	⁷ 23	² 7066.90 <i>N</i> 0.1
15	Camvi Technologies	camvitech	2	2018-02-16	442	⁵⁴ 1024	1	¹¹⁴ 774	-	⁸ 20	¹ 7180.65 <i>N</i> 0.1
16	Camvi Technologies	camvitech	3	2018-06-30	233	⁵⁵ 1024	1	¹⁰⁷ 707	⁴ 10	⁶ 11	⁵ 857.59 <i>N</i> 0.2
17	Gemalto Cogent	cogent	0	2018-06-20	533	³³ 525	k	⁸³ 551	³¹ 494	⁷³ 558	³⁸ 0.46 <i>N</i> 1.0
18	Gemalto Cogent	cogent	1	2018-06-20	533	²⁵ 525	k	⁸¹ 552	³² 498	⁷² 556	⁴⁶ 0.39 <i>N</i> 1.0
19	Cognitec Systems GmbH	cognitec	0	2018-06-21	364	¹⁰⁰ 2052	k	¹⁸ 176	⁴⁴ 1748	⁹⁷ 1780	¹⁰⁹ 0.57 <i>N</i> 1.0
20	Cognitec Systems GmbH	cognitec	1	2018-06-21	412	⁹⁷ 2052	k	²³ 202	⁴⁶ 1835	⁹⁶ 1735	¹¹⁵ 0.45 <i>N</i> 1.1
21	Dermalog	dermalog	0	2018-02-16	0	⁴ 128	1	⁴⁸ 344	-	⁵⁹ 404	⁸⁸ 0.19 <i>N</i> 1.0
22	Dermalog	dermalog	1	2018-02-16	0	⁶ 128	1	¹⁷ 171	-	⁶¹ 407	⁹⁹ 0.17 <i>N</i> 1.0
23	Dermalog	dermalog	2	2018-02-16	0	¹³ 256	k	⁴⁷ 344	-	⁷⁹ 640	⁶³ 0.40 <i>N</i> 1.0
24	Dermalog	dermalog	3	2018-06-21	0	⁵ 128	1	² 211	⁹ 92	¹⁶ 92	⁶⁴ 0.06 <i>N</i> 1.0
25	Dermalog	dermalog	4	2018-06-21	0	³ 128	1	²⁴ 208	⁸ 91	¹⁷ 93	⁸⁵ 0.05 <i>N</i> 1.0
26	Ever AI	everai	0	2018-06-21	142	⁹¹ 2048	1	⁷⁰ 438	³ 4	⁴ 3	⁸ 42.41 <i>N</i> 0.3
27	Ever AI	everai	1	2018-06-21	200	⁷⁵ 2048	1	⁸⁹ 590	²⁴ 336	⁵³ 356	¹²³ 0.03 <i>N</i> 1.1
28	Eyedeas Recognition	eyedeas	0	2018-02-16	644	¹²³ 4152	k	⁶³ 424	-	⁸⁰ 640	⁸⁰ 0.34 <i>N</i> 1.0
29	Eyedeas Recognition	eyedeas	1	2018-02-16	287	⁶⁸ 1036	k	⁴² 311	-	⁴⁸ 307	⁸⁷ 0.15 <i>N</i> 1.0
30	Eyedeas Recognition	eyedeas	2	2018-02-16	287	⁵⁸ 1036	k	⁶⁹ 429	-	⁴⁷ 305	⁷⁸ 0.16 <i>N</i> 1.0
31	Eyedeas Recognition	eyedeas	3	2018-06-18	284	⁵⁹ 1036	k	⁵¹ 385	²¹ 309	⁵⁰ 311	⁵³ 0.21 <i>N</i> 1.0
32	Glory Ltd	glory	0	2018-06-30	0	²² 418	k	¹¹ 160	³³ 575	⁷⁴ 575	⁹³ 0.26 <i>N</i> 1.0
33	Glory Ltd	glory	1	2018-06-30	0	⁷⁰ 1726	k	⁵⁸ 405	⁴⁷ 1864	⁹⁹ 1978	⁹¹ 0.93 <i>N</i> 1.0
34	Gorilla Technology	gorilla	0	2018-02-01	95	¹³⁰ 8300	k	⁶⁵ 427	-	¹²⁸ 10426	⁸³ 5.30 <i>N</i> 1.0
35	Gorilla Technology	gorilla	1	2018-06-19	91	¹⁰⁷ 2156	k	¹⁶ 169	⁶⁴ 5254	¹²³ 5156	⁵⁷ 3.31 <i>N</i> 1.0
36	loginface Corp	hbinno	0	2018-02-01	88	³¹ 520	-	³⁵ 265	-	⁶² 419	³⁹ 0.34 <i>N</i> 1.0
37	Hikvision Research Institute	hikvision	0	2018-02-12	378	⁷² 1808	1	¹²⁶ 875	-	¹⁰⁸ 2360	¹⁰² 0.97 <i>N</i> 1.0
38	Hikvision Research Institute	hikvision	1	2018-02-12	378	⁷⁴ 1808	1	¹¹⁸ 820	-	¹⁰⁹ 2403	⁹⁸ 1.00 <i>N</i> 1.0
39	Hikvision Research Institute	hikvision	2	2018-02-12	378	⁷³ 1808	1	¹¹⁶ 820	-	¹¹⁰ 2408	⁹⁷ 1.00 <i>N</i> 1.0
40	Hikvision Research Institute	hikvision	3	2018-06-30	408	⁶³ 1408	1	⁹⁴ 633	³⁷ 904	⁸⁹ 1108	³⁶ 0.91 <i>N</i> 1.0
41	Hikvision Research Institute	hikvision	4	2018-06-30	334	⁶² 1152	1	⁷⁵ 510	³⁶ 784	⁸⁵ 1024	³³ 0.86 <i>N</i> 1.0
42	Idemia	idemia	0	2018-02-16	371	²¹ 364	1	⁶⁰ 416	-	¹⁹ 133	⁶⁹ 0.08 <i>N</i> 1.0
43	Idemia	idemia	1	2018-02-16	371	¹⁹ 364	1	⁶¹ 417	-	²² 138	⁸⁴ 0.07 <i>N</i> 1.0
44	Idemia	idemia	2	2018-02-16	371	²⁰ 364	1	⁶² 417	-	²³ 138	⁷⁹ 0.07 <i>N</i> 1.0

Notes	
1	Configuration size does not capture static data present in libraries. Libraries are not counted because most implementations include common ancillary libraries for image processing (e.g. openCV) or numerical computation (e.g. blas).
2	This multiplier expresses the increase in template size when <i>k</i> 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.

	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 (μ s)
									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	⁵⁸ 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	¹³¹ 4497	¹²⁵ 0.06 $N^{1.3}$
52	Innovatrics	innovatrics	0	2018-02-16	0	³⁹ 530	k	⁷¹ 455	-	⁷⁶ 625	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	²²	⁵ 616.66 $N^{0.1}$
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	⁴³ 1024	1	²⁶ 222	-	³⁹ 240	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	-	⁴³ 284	⁵⁶ 0.18 $N^{1.0}$
60	Microfocus	microfocus	0	2018-02-12	101	¹³ 256	k	⁷⁴ 525	-	²³ 184	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	-	⁷⁵ 593	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	⁸³ 555	-	⁸³ 869	¹⁰⁷ 0.32 $N^{1.0}$
68	Microsoft	microsoft	3	2018-06-20	230	⁴⁷ 1024	1	⁵⁷ 404	⁴³ 1638	⁹⁶ 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	¹⁰⁸ 2592	k	⁸ 82	²² 317	⁶² 426	¹⁵ 0.73 $N^{0.9}$
71	NEC	nec	1	2018-06-29	131	¹⁰⁸ 2592	k	⁹ 88	¹⁵ 193	³¹ 208	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	¹⁰⁹ 661	-	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	-	⁵⁸ 382	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	-	¹⁵ 101	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 ancillary 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.

	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 (μ S)
									L=1	L=50	
89	Shaman Software	shaman	1	2018-02-12	0	¹¹³ 4096	k	⁸⁶ 557	-	⁶⁸ 524	⁵⁸ 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	¹⁰⁸ 2048	k	³⁵ 704	³⁵ 692	⁴⁹ 310	¹⁴ 1.04 $N^{0.9}$
92	Shaman Software	shaman	4	2018-06-30	0	⁸⁸ 2048	k	⁹⁶ 642	²⁸ 434	⁴² 267	¹⁵ 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	2018-02-30	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	¹³ 3.75 $N^{0.9}$
98	Smilart	smilart	2	2018-02-15	109	⁴⁸ 1024	k	⁸⁸ 560	-	⁹³ 1302	³³ 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	tevia	0	2018-02-16	666	⁷⁹ 2048	1	⁵³ 394	-	⁶⁰ 405	⁹⁴ 0.18 $N^{1.0}$
103	Tevian	tevia	1	2018-02-16	666	⁸⁷ 2048	1	⁵⁶ 398	-	⁵⁸ 403	⁸⁶ 0.20 $N^{1.0}$
104	Tevian	tevia	2	2018-02-16	666	⁸⁵ 2048	1	⁵⁴ 397	-	⁵⁷ 402	⁸⁹ 0.19 $N^{1.0}$
105	Tevian	tevia	3	2018-06-20	707	⁷⁷ 2048	1	⁴⁰ 300	³⁰ 473	⁷⁰ 539	⁹⁰ 0.25 $N^{1.0}$
106	Tevian	tevia	4	2018-06-20	707	⁹⁰ 2048	1	³⁹ 299	²⁸ 434	⁶⁹ 537	⁷² 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	¹⁰	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	¹⁰³ 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	¹⁰⁰ 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	-	¹⁰² 2121	¹¹³ 0.41 $N^{1.1}$
115	Vigilant Solutions	vigilant	3	2018-06-21	335	⁶⁵ 1544	k	¹²² 832	⁵⁵ 2453	¹⁰⁷ 2307	¹⁰³ 0.93 $N^{1.0}$
116	Vigilant Solutions	vigilant	4	2018-06-21	337	⁶⁴ 1544	k	¹²⁸ 830	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 ancillary 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	UNCONSOLE	
#	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	⁷⁰ 0.020	⁷² 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	⁶⁵ 0.040					
4	3DIVI-3	²⁵ 512	⁸⁸ 625	⁷³ 0.053	⁶⁹ 0.024	⁶⁷ 0.020	⁸⁸ 0.086	⁸⁸ 0.037	⁸³ 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	⁴⁸ 0.021	⁵⁸ 0.018	⁵⁹ 0.018	⁴⁴ 0.019	⁵⁹ 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	⁸² 0.042	⁸² 0.039	⁸² 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	⁷⁵ 0.025	⁴⁸ 0.014	⁴⁹ 0.012	⁶² 0.033	⁵⁹ 0.015	⁵⁹ 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	⁹² 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	¹¹⁷ 774	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	⁶² 0.035	⁷⁴ 0.035	⁷⁷ 0.035	⁷⁹ 0.054	⁸⁸ 0.054	⁹⁰ 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	²⁹ 0.006	⁴² 1.111	³² 0.013	³⁵ 0.011	¹⁰ 0.007			
19	COGNITEC-0	⁹² 2052	¹⁶ 76	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	³ 128	²⁵ 211				⁹³ 0.128	⁹² 0.063	⁹¹ 0.050	⁹² 1.752	⁹³ 0.128	⁶⁹ 0.097				
25	DERMALOG-4	³ 128	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	⁶⁹ 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	⁷ 1.038	⁹ 0.006	⁹ 0.003				
28	EYEDea-0	¹²⁰ 4152	⁶¹ 424	⁹² 0.201	⁹⁷ 0.100	⁹⁶ 0.081	¹¹⁵ 0.300	¹¹³ 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	⁶⁹ 429	⁸⁷ 0.110	⁸¹ 0.054	⁸³ 0.044	¹⁰⁶ 0.200	¹⁰⁵ 0.107	¹⁰⁵ 0.089	¹⁰⁶ 2.246	¹⁰⁶ 0.200					
31	EYEDea-3	⁵⁵ 1036	⁴⁹ 385	⁷¹ 0.044	⁶⁶ 0.021	⁶⁸ 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 1.133				
33	GLORY-1	⁶⁸ 1726	⁵⁷ 405	⁸⁵ 0.109	⁹² 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	⁷⁵ 1.331	⁸² 0.063	⁵⁸ 0.041				
36	HBBINO-0	²⁹ 520	³³ 265	⁹⁸ 0.191	⁹⁸ 0.102	⁹⁷ 0.086	¹¹⁴ 0.275	¹¹² 0.152	¹¹² 0.126	¹¹⁴ 2.743	¹¹⁴ 0.275					
37	HIK-0	⁷⁰ 1808	¹²³ 875	⁸⁵ 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	⁷⁸ 0.073	⁸⁰ 0.071	⁸⁰ 0.070	⁴³ 0.017	⁴⁷ 0.011	⁵¹ 0.010	⁴⁶ 1.116	⁴³ 0.017					
39	HIK-2	⁷² 1808	¹¹⁵ 820	⁷⁵ 0.013	⁴² 0.010	⁴² 0.010	⁴² 0.017	⁴⁶ 0.011	⁴⁸ 0.010	⁴⁴ 1.115	⁴² 0.017	⁴⁷ 0.019				
40	HIK-3	⁶¹ 1408	⁹⁰ 633				³⁹ 0.014	³² 0.007	³⁹ 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	²⁹ 0.006	²⁹ 0.006	²⁸ 1.070	²⁸ 0.011	²¹ 0.006				
43	IDEMIA-1	²⁰ 364	⁶⁰ 417	²³ 0.008	²¹ 0.005	²² 0.005	³⁰ 0.012	³¹ 0.007	³¹ 0.006	³⁰ 1.072	³⁰ 0.012	²⁵ 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	³⁶ 0.011	³⁹ 0.008	³⁹ 0.007	²⁴ 0.010	²⁹ 0.006	³⁰ 0.006	²⁶ 1.066	²⁴ 0.010	¹⁸ 0.005	⁹ 0.005			
46	IDEMIA-4	³³ 528	⁹⁹ 669	²⁵ 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	¹¹⁶ 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	¹⁰⁸ 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	⁷¹ 1.285	⁷⁸ 0.051	⁵⁷ 0.038				
51	INCODE-1	⁷³ 2048	¹⁰¹ 690	³¹ 0.012	²⁶ 0.008	²⁸ 0.007	⁴⁵ 0.019	³⁸ 0.009	⁴⁰ 0.008	⁴⁰ 1.106	⁴⁵ 0.019	⁴¹ 0.013				
52	INNOVATRIS-0	³⁶ 530	⁶⁹ 455	⁶² 0.029	⁵³ 0.017	⁵⁴ 0.016	⁷⁰ 0.042	⁶⁷ 0.019	⁶⁶ 0.							

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				WORK-10			RECENT	LIFETIME	UNCONSOLE
#	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	² 0.002	² 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	² 0.002	² 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	⁵³ 404	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	³⁶ 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	¹⁰⁵ 2592	⁸ 88	⁵² 0.025	⁶² 0.021	⁶⁰ 0.021	⁵⁰ 0.024	⁶⁰ 0.015	⁶³ 0.014	⁵⁸ 1.158	⁵⁰ 0.024	⁴³ 0.016				
72	NEUROTECHNOLOGY-0	¹²³ 5214	¹⁰² 702	⁶⁴ 0.031	⁵⁴ 0.018	⁵⁸ 0.016	⁷⁰ 0.050	⁷² 0.023	⁷¹ 0.019	⁷³ 1.278	⁷⁰ 0.050					
73	NEUROTECHNOLOGY-1	¹²³ 5214	⁹⁷ 661	⁵⁹ 0.028	⁴⁹ 0.014	⁴⁷ 0.012	⁷⁰ 0.047	⁶⁹ 0.020	⁶⁷ 0.016	⁷⁰ 1.250	⁷⁰ 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	⁹⁶ 2048	⁷⁸ 543	³⁵ 0.014	⁴³ 0.012	⁴⁵ 0.011	¹⁶ 0.008	²² 0.006	²⁶ 0.006	¹⁹ 1.058	¹⁷ 0.008	¹⁹ 0.006			⁶ 0.004	
77	NTECHLAB-0	¹²² 4442	¹⁰⁸ 730	¹⁵ 0.006	¹² 0.003	¹² 0.003	²⁰ 0.012	²⁰ 0.005	¹⁵ 0.005	²⁵ 1.064	²⁰ 0.012	²³ 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	¹⁵ 0.007	⁹ 0.004	⁸ 0.003	¹¹ 1.041	¹⁷ 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	⁵³ 0.032				
82	RANKONE-1	¹⁷ 324	¹¹ 136	⁶⁵ 0.032	⁶² 0.021	⁶⁰ 0.019	⁵⁶ 0.025	⁵⁰ 0.012	⁵² 0.010	⁵³ 1.145	⁵⁶ 0.025	⁴⁰ 0.019				
83	RANKONE-2	⁷ 133	¹¹³ 505	⁵⁵ 0.025	⁵⁷ 0.018	⁵⁵ 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	⁴⁰ 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	⁶¹ 1.222	⁷² 0.043	⁵⁹ 0.044				
86	REALNETWORKS-1	¹¹⁵ 4104	²⁹ 243				⁷¹ 0.043	⁶² 0.017	⁶⁰ 0.013	⁶⁴ 1.222	⁷¹ 0.043	⁵³ 0.033				
87	SHAMAN-0	¹¹¹ 4096	⁷⁶ 538	⁸⁹ 0.119	⁹⁶ 0.076	⁸⁸ 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	⁸³ 557	⁹⁷ 0.180	⁹⁷ 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	⁹⁷ 0.127	⁶⁸ 0.097				
91	SHAMAN-4	⁷⁶ 2048	⁹³ 642				¹¹⁰ 0.224	¹⁰⁸ 0.126	¹⁰⁸ 0.107	¹¹⁰ 2.431	¹¹⁰ 0.224	⁷² 0.187				
92	SIAT-0	⁵⁰ 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	³ 0.003	³ 0.003	³ 1.031	³ 0.004	⁷⁵ 0.264			² 0.001	
94	SIAT-2	²⁰ 52	¹²⁴ 906	⁸¹ 0.081	⁹¹ 0.080	⁹¹ 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	¹² 168	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	⁸³ 560	⁹¹ 0.132	⁸⁵ 0.069	⁸⁸ 0.058	¹⁰⁴ 0.195	¹⁰² 0.102	¹⁰¹ 0.084	¹⁰¹ 2.196	¹⁰¹ 0.195					
98	SYNOPSIS-0	²³ 512	²² 237	⁸⁴ 0.108	⁹⁶ 0.100	⁹⁷ 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	³² 0.022	⁴³ 0.010	⁴³ 0.008	⁴⁸ 1.122	³² 0.022					
100	TEVIAN-1	⁹¹ 2048	⁵⁴ 398	⁴⁰ 0.017	³⁶ 0.010	³⁸ 0.009	³² 0.022	⁴⁴ 0.010	⁴⁴ 0.008	⁴⁹ 1.122	³² 0.022					
101	TEVIAN-2	⁸¹ 2048	⁵² 397	⁴² 0.017	⁴⁰ 0.010	³⁹ 0.009	³¹ 0.022	⁴² 0.010	⁴² 0.008	⁴⁷ 1.121	³¹ 0.022					
102	TEVIAN-3	⁷⁷ 2048	³⁹ 300				⁴¹ 0.017	³⁹ 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	⁹² 2052	⁶³ 428	⁶⁶ 0.033	⁴⁶ 0.014	⁴³ 0.011	⁸⁵ 0.064	⁷⁶ 0.026	⁷³ 0.020	⁷⁹ 1.334	⁸⁵ 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	²⁴ 0.006	²⁴ 0.005	²³ 0.010	²³ 0.006	²³ 0.005	²³ 1.062	²³ 0.010	³⁰ 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	¹¹⁸ 823	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	⁹⁵ 2056	¹⁰⁵ 739	⁹⁰ 0.120	⁸² 0.054	⁸¹ 0.043	¹⁰⁴ 0.204	¹⁰⁰ 0.100	⁹⁹ 0.080	¹⁰³ 2.210	¹⁰⁴ 0.204					
111	VIGILANTSOLUTIONS-2	⁶² 1544	¹¹² 820	⁹⁶ 0.159	⁹⁵ 0.090	⁹⁰ 0.077	¹¹² 0.239	¹¹¹ 0.139	¹¹¹ 0.118	¹¹² 2.555	¹¹² 0.239					
112	VIGILANTSOLUTIONS-3	⁶⁴ 1544	¹¹² 832	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	³⁰ 0.008	²⁰ 0.009	³⁵ 0.008	³⁹ 0.007	²⁹ 1.072	²⁰ 0.009	¹⁵ 0.005				
115																

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	FPFR=0.001	FPFR=0.01	FPFR=0.1	FPFR=0.001	FPFR=0.01	FPFR=0.1	FPFR=0.001	FPFR=0.01	FPFR=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	⁶⁵ 0.058	⁶⁷ 0.255	⁷⁴ 0.164	⁷⁵ 0.089			
4	3DIVI-3	⁷⁴ 0.287	⁷⁷ 0.183	⁷⁶ 0.105	⁸⁰ 0.402	⁸⁹ 0.284	⁸⁸ 0.168	⁶⁸ 0.626	⁷⁰ 0.497	⁶⁶ 0.343
5	3DIVI-4				⁵⁸ 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	⁴⁸ 0.035	³⁵ 0.216	³⁶ 0.146	³⁶ 0.087
7	ALCHERA-1				¹²⁵ 0.999	¹²⁵ 0.999	¹²⁵ 0.995	¹⁰⁸ 1.000	¹⁰⁹ 1.000	⁹⁸ 1.000
8	AWARE-0	⁹⁷ 0.775	⁶¹ 0.092	⁶⁹ 0.065	¹²⁴ 0.983	⁶⁶ 0.128	⁷¹ 0.085	⁷⁰ 0.817	⁵² 0.253	⁵⁸ 0.178
9	AWARE-1	¹⁰³ 0.863	⁵⁷ 0.084	⁵⁹ 0.055	¹²⁴ 0.996	⁶³ 0.127	⁷⁰ 0.081			
10	AWARE-2	⁵⁸ 0.757	⁶⁰ 0.090	⁷⁰ 0.067	¹²² 0.977	⁶⁴ 0.120	⁶⁸ 0.078			
11	AWARE-3	⁴⁵ 0.096	⁴³ 0.056	⁵⁰ 0.035	⁴⁹ 0.131	⁵¹ 0.085	⁵¹ 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	⁶¹ 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	⁹⁵ 0.557	⁹⁴ 0.409	⁹⁵ 0.255	¹⁰⁷ 0.684	¹¹¹ 0.549	¹¹¹ 0.375	⁷⁸ 0.770	⁸⁰ 0.648	⁸⁰ 0.488
15	CAMVI-2	⁸¹ 0.408	⁸⁵ 0.265	⁸¹ 0.147	⁹⁰ 0.537	⁹⁹ 0.402	⁹⁶ 0.242			
16	CAMVI-3	²⁴ 0.046	³⁶ 0.038	³⁰ 0.036	²⁶ 0.074	⁴¹ 0.060	⁶⁰ 0.055	¹⁷ 0.132	³¹ 0.108	³⁸ 0.094
17	COGENT-0	¹³ 0.033	²¹ 0.021	²¹ 0.015	²¹ 0.056	²³ 0.032	²⁸ 0.020	²⁸ 0.140	²⁶ 0.100	³⁴ 0.069
18	COGENT-1	¹³ 0.033	²⁰ 0.021	²⁰ 0.015	²² 0.056	²² 0.032	²⁷ 0.020	¹⁸ 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	⁴⁶ 0.115
20	COGNITEC-1	³⁵ 0.063	³² 0.031	³¹ 0.018	³⁵ 0.105	³⁶ 0.055	³⁵ 0.027	³⁵ 0.230	³⁵ 0.135	³¹ 0.071
21	DERMALOG-0	⁷¹ 0.348	⁷⁹ 0.233	⁷⁵ 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	⁸⁰ 0.172	⁹⁰ 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	⁶⁹ 0.657	⁷² 0.526	⁶⁹ 0.359
26	EVERAI-0				³¹ 0.092	³² 0.047	³⁷ 0.028	³⁰ 0.170	²⁵ 0.100	²⁶ 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	EYDEA-0	⁹⁵ 0.724	⁹⁹ 0.549	⁹⁵ 0.357	¹¹³ 0.812	¹¹⁷ 0.679	¹¹⁸ 0.484	⁸⁵ 0.914	⁸³ 0.783	⁸⁵ 0.619
29	EYDEA-1	⁸⁵ 0.459	⁸⁸ 0.324	⁸⁵ 0.207	¹⁰⁰ 0.632	¹⁰⁴ 0.480	¹⁰⁵ 0.335			
30	EYDEA-2	⁹³ 0.570	⁸⁹ 0.327	⁸⁹ 0.208	¹¹² 0.794	¹⁰⁷ 0.490	¹⁰⁷ 0.338			
31	EYDEA-3	⁷⁴ 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				⁷¹ 0.369	⁹⁰ 0.297	⁹² 0.233	⁶¹ 0.547	⁶⁷ 0.470	⁷³ 0.390
33	GLORY-1	⁶⁹ 0.240	⁷⁶ 0.182	⁷⁶ 0.140	⁷⁴ 0.307	⁸⁴ 0.238	⁸⁹ 0.179	⁶² 0.537	⁶⁵ 0.448	⁶⁹ 0.352
34	GORILLA-0				⁸⁵ 0.408	⁸⁵ 0.248	⁸⁴ 0.136	⁵⁷ 0.453	⁵⁹ 0.314	⁵⁸ 0.191
35	GORILLA-1				¹¹⁵ 0.766	¹¹⁵ 0.632	¹¹⁵ 0.458			
36	HBINNO-0	⁹⁸ 0.632	⁹⁸ 0.498	⁹⁸ 0.336	¹¹⁵ 0.766	¹¹⁵ 0.632	¹¹⁵ 0.458			
37	HIK-0	⁴⁰ 0.078	⁴¹ 0.049	³¹ 0.035	⁴² 0.114	⁴⁷ 0.070	⁴⁷ 0.040	²⁴ 0.155	²⁸ 0.103	²⁸ 0.061
38	HIK-1	³⁵ 0.131	⁶² 0.095	⁷⁴ 0.081	⁴⁵ 0.120	⁴⁵ 0.067	⁴⁴ 0.034			
39	HIK-2	³⁰ 0.076	³⁵ 0.037	³⁰ 0.022	⁴⁵ 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	³⁴ 0.077	³⁴ 0.036	³⁴ 0.019	⁴¹ 0.114	⁴² 0.062	³⁸ 0.029	³⁴ 0.240	³⁷ 0.156	³⁸ 0.085
43	IDEMIA-1	¹⁵ 0.041	¹⁹ 0.021	¹⁹ 0.013	¹⁷ 0.054	²¹ 0.031	²¹ 0.018			
44	IDEMIA-2	²¹ 0.043	²⁵ 0.023	²⁰ 0.016	¹⁷ 0.054	²⁴ 0.032	²⁴ 0.019			
45	IDEMIA-3	¹⁸ 0.041	²² 0.021	²⁰ 0.015	¹⁴ 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	²¹ 0.013	⁷ 0.040	¹³ 0.024	¹⁸ 0.014	¹⁴ 0.118	¹⁵ 0.079	²⁴ 0.050
47	IMAGUS-0	⁵⁹ 0.592	⁹⁷ 0.468	⁹⁷ 0.329	¹⁰⁷ 0.734	¹¹⁴ 0.608	¹¹⁴ 0.453	⁸¹ 0.872	⁸² 0.779	⁸¹ 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				¹¹⁵ 0.808	¹¹⁶ 0.670	¹¹⁸ 0.512	⁸⁴ 0.909	⁸⁴ 0.809	⁸⁴ 0.667
50	INCODE-0				⁷⁵ 0.313	⁸¹ 0.201	⁸¹ 0.107	⁵⁸ 0.420	⁵⁸ 0.304	⁵⁷ 0.191
51	INCODE-1	⁴⁷ 0.127	⁴⁶ 0.061	³⁰ 0.027	⁵⁸ 0.214	⁵⁸ 0.114	⁵¹ 0.050	⁴⁷ 0.296	⁴² 0.198	⁴⁶ 0.110
52	INNOVATRIS-0	⁵⁹ 0.171	⁶⁴ 0.100	⁶² 0.055	⁶⁶ 0.255	⁷⁶ 0.165	⁷⁶ 0.089	⁵³ 0.361	⁵³ 0.258	⁵⁴ 0.159
53	INNOVATRIS-1	⁵⁸ 0.171	⁶³ 0.100	⁶¹ 0.055	⁶⁵ 0.255	⁷⁵ 0.165	⁷⁶ 0.089			
54	INNOVATRIS-2				⁶⁵ 0.237	⁷¹ 0.142	⁶⁹ 0.079	⁴⁷ 0.310	⁴⁷ 0.209	⁴⁶ 0.126
55	INNOVATRIS-3	⁵⁰ 0.133	⁵⁴ 0.068	⁴⁰ 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	³⁵ 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	¹⁸ 0.011	²⁵ 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	¹⁰³ 0.877	¹⁰⁵ 0.793	¹⁰³ 0.641	¹²¹ 0.933	¹²³ 0.867	¹²⁴ 0.749	⁸⁶ 0.985	⁸⁹ 0.950	⁸⁹ 0.877
61	MICROFOCUS-1	¹⁰³ 0.877	¹⁰⁴ 0.793	¹⁰³ 0.641	¹¹⁹ 0.933	¹²² 0.867	¹²⁴ 0.749			
62	MICROFOCUS-2	¹⁰⁸ 0.878	¹⁰⁶ 0.796	¹⁰⁸ 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 FPFR = 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	²³ 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	⁸⁹ 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	⁸² 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	⁵⁹ 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	⁸⁵ 0.514	⁹¹ 0.375	⁹¹ 0.233	¹⁰² 0.629	¹⁰⁸ 0.492	¹⁰⁴ 0.325			
98	SYNOPSIS-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	⁴⁵ 0.198	⁴¹ 0.113
103	TEVIAN-4	³⁶ 0.074	³² 0.035	³¹ 0.018	⁴⁵ 0.120	⁴⁴ 0.066	⁴² 0.031	⁴³ 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	¹⁰⁵ 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.

MISSSES NOT AT RANK 1 FNIR(N, T = 0, R > 1)		ENROL LIFETIME DATASET: FRVT 2018					ENROL MOST RECENT DATASET: FRVT 2018						
#	ALGORITHM	N=0.64M	N=1.6M	N=3.0M	N=6.0M	N=12.0M	$a N^b$	N=0.64M	N=1.6M	N=3.0M	N=6.0M	N=12.0M	$a N^b$
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	⁴⁶ 0.0106	³⁹ 0.0121	²⁸ 0.0135	²⁶ 0.0170		⁴⁹ 0.0006 N ^{0.209 20}	⁴⁸ 0.0167	⁴⁴ 0.0186				⁵⁴ 0.0035 N ^{0.117 11}
3	AWARE-3	³⁷ 0.0165	³² 0.0209	³³ 0.0247	³¹ 0.0297		³⁹ 0.0005 N ^{0.283 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	¹⁸ 0.4198	⁷⁸ 0.4649	⁴⁸ 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 ^{0.106 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	³⁵ 0.0047 N ^{0.089 2}	⁴¹ 0.0127	³³ 0.0131	²³ 0.0136	¹⁹ 0.0141	¹⁸ 0.0151	⁵⁹ 0.0058 N ^{0.058 2}
7	COGENT-1	⁴² 0.0103	³⁵ 0.0106				³⁷ 0.0074 N ^{0.085 1}	⁴⁰ 0.0127	³² 0.0131	²⁴ 0.0136	¹⁸ 0.0141	¹⁷ 0.0151	⁵⁶ 0.0058 N ^{0.058 1}
8	COGNITEC-0	³¹ 0.0146	³⁰ 0.0205				¹⁵ 0.0001 N ^{0.376 58}	³⁸ 0.0221	³⁹ 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	²⁹ 0.0165	²⁵ 0.0192	²⁴ 0.0225	²⁵ 0.0006 N ^{0.226 41}
10	DERMALOG-4	⁸⁸ 0.0759	⁶⁷ 0.0961	⁴⁸ 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	²⁸ 0.0065	⁴⁶ 0.0166				¹ 0.0000 N ^{0.1029 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 ^{0.272 24}	⁸ 0.0047	⁹ 0.0056	⁹ 0.0061			²² 0.0005 N ^{0.166 20}
13	EYEDAE-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	⁹ 0.0818	⁶⁶ 0.0932	⁴⁸ 0.1007	³⁹ 0.1091		³⁹ 0.0147 N ^{0.129 6}	⁹⁷ 0.1154	⁹⁴ 0.1291				⁶¹ 0.0223 N ^{0.125 14}
15	HIK-2	³⁸ 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	³⁶ 0.0083	³⁴ 0.0104	²⁵ 0.0121	²² 0.0146	¹⁹ 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	¹⁸ 0.0048	²¹ 0.0063	¹² 0.0076	¹² 0.0095	¹² 0.0116	¹⁷ 0.0001 N ^{0.301 48}	²⁵ 0.0093	²⁵ 0.0113	²⁵ 0.0131	²⁰ 0.0153	²⁰ 0.0182	¹⁶ 0.0004 N ^{0.227 42}
19	IDEMIA-1	²⁴ 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	²⁴ 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	³ 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	²³ 0.0001 N ^{0.252 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	INCOCODE-1	³⁸ 0.0098	⁴¹ 0.0131	³⁶ 0.0286	³⁴ 0.0466		⁷ 0.0000 N ^{0.729 60}	⁴³ 0.0151	⁴³ 0.0190				²⁴ 0.0005 N ^{0.290 50}
25	ISYSTEMS-0	³⁷ 0.0074	²⁸ 0.0085	¹⁷ 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	³⁸ 0.0074	²⁷ 0.0085	¹⁹ 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	³⁸ 0.0072	³² 0.0099	²⁸ 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	¹¹⁸ 0.4791	⁸⁰ 0.5389	⁵⁸ 0.5771			⁶¹ 0.0951 N ^{0.121 5}	¹²² 0.5417	¹²² 0.5953				⁶⁴ 0.1370 N ^{0.108 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	⁹ 0.0020	⁷ 0.0026	⁶ 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	⁹ 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	⁷ 0.0009	² 0.0011				⁷ 0.0000 N ^{0.255 30}	⁷ 0.0028	⁷ 0.0032	⁷ 0.0035	⁷ 0.0039	⁷ 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	²⁸ 0.0154	²⁷ 0.0185	²² 0.0223	²⁵ 0.0002 N ^{0.284 42}	⁴⁶ 0.0157	⁴⁶ 0.0196	²⁴ 0.0229	²⁷ 0.0270	²⁶ 0.0320	²⁷ 0.0006 N ^{0.243 48}
36	NEC-1	⁴⁸ 0.0136	⁴⁵ 0.0164				⁴⁹ 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	³⁶ 0.0161	⁴⁹ 0.0199				⁴³ 0.0007 N ^{0.234 27}	⁵⁴ 0.0204	⁵⁰ 0.0250	³⁴ 0.0288	³¹ 0.0331	³⁰ 0.0386	⁴⁰ 0.0011 N ^{0.216 37}
38	NEUROTECHNOLOGY-4	²⁸ 0.0049	¹⁹ 0.0058	¹² 0.0065	¹¹ 0.0075	¹⁰ 0.0087	³⁴ 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	²⁸ 0.0056	²⁴ 0.0077	¹⁷ 0.0094	¹⁸ 0.0114		¹⁵ 0.0001 N ^{0.323 54}	²⁸ 0.0092	²⁹ 0.0115	²³ 0.0137	²² 0.0164	²² 0.0196	¹⁰ 0.0003 N ^{0.261 53}
40	NTECHLAB-1	²⁰ 0.0070	³⁰ 0.0097	²³ 0.0119	²¹ 0.0146	¹⁹ 0.0179	²⁹ 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	⁷ 0.0030	¹¹ 0.0040	⁷ 0.0049	⁸ 0.0060	⁹ 0.0075	¹⁰ 0.0000 N ^{0.315 50}	¹² 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	⁵⁹ 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	³³ 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	⁴⁸ 0.0117	⁴³ 0.0149	³⁰ 0.0172	²⁸ 0.0200	²³ 0.0236	³⁵ 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	²⁶ 0.0337	³⁹ 0.0443	²⁷ 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	¹⁷ 0.2638	⁷⁵ 0.2639	⁴ 0.2640			-	⁴ 0.0037	³ 0.0039	³ 0.0041	³ 0.0044	⁴ 0.0049	³⁸ 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	TONGVITRANS-1	³⁷ 0.0096	³⁸ 0.0114	²⁷ 0.0127	²³ 0								

MISSES NOT AT RANK 50 FNIR(N, T=0, R>50)		ENROL LIFETIME DATASET: FRVT 2018					ENROL MOST RECENT 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	⁶ 0.0073	⁵¹ 0.0076	²⁵ 0.0079	³⁰ 0.0101		⁵⁷ 0.0012 N ^{0.133 14}	⁷² 0.0125	⁶⁶ 0.0129				⁶⁷ 0.0079 N ^{0.084 5}
3	AWARE-3	⁵⁰ 0.0039	⁴³ 0.0050	²⁹ 0.0061	²⁸ 0.0077		²⁴ 0.0001 N ^{0.299 40}	⁶⁰ 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		⁶⁴ 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		⁷ 0.0000 N ^{0.080 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	²⁴ 0.0021	²² 0.0024				⁴⁰ 0.0002 N ^{0.189 25}	²⁶ 0.0047	²⁴ 0.0050	¹⁵ 0.0054	²⁰ 0.0062	²³ 0.0122	⁷ 0.0001 N ^{0.288 52}
8	COGNITEC-0	⁴⁸ 0.0039	⁵⁰ 0.0067				⁷ 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}	³⁷ 0.0056	³⁷ 0.0060	²⁶ 0.0066	²⁴ 0.0072	²² 0.0081	⁴⁰ 0.0010 N ^{0.128 26}
10	DERMLOG-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				⁷ 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	⁹ 0.0415	⁷⁰ 0.0490	⁴⁸ 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	⁷⁹ 0.0084	⁵⁶ 0.0090	³⁰ 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				³¹ 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	²⁵ 0.0023	³² 0.0028	²² 0.0033	²⁰ 0.0039	¹⁹ 0.0048	²⁵ 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	¹⁸ 0.0016	¹⁷ 0.0019	¹⁷ 0.0023	¹⁰ 0.0026	⁹ 0.0031	²⁷ 0.0001 N ^{0.228 31}	²⁵ 0.0045	²⁶ 0.0051	²⁵ 0.0055	¹⁸ 0.0060	¹⁹ 0.0067	³⁸ 0.0008 N ^{0.134 27}
19	IDEMIA-1	¹⁹ 0.0019	²¹ 0.0024	²⁰ 0.0029	¹⁸ 0.0036	¹⁷ 0.0046	¹³ 0.0000 N ^{0.300 42}	³⁴ 0.0049	³⁶ 0.0058	²⁹ 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}	⁴³ 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	¹⁹ 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	⁹ 0.0348	¹³ 0.0510	⁴⁶ 0.0641	⁴¹ 0.0804		⁴⁰ 0.0002 N ^{0.375 53}	¹⁰⁰ 0.0468	¹⁰⁰ 0.0657				¹⁹ 0.0003 N ^{0.371 59}
24	INCOCODE-1	²⁴ 0.0026	³⁴ 0.0033	³⁶ 0.0167	³⁶ 0.0323		¹ 0.0000 N ^{1.217 62}	³² 0.0055	³⁰ 0.0063				³⁴ 0.0007 N ^{0.153 31}
25	ISYSTEMS-0	⁵⁸ 0.0048	⁴² 0.0050	²⁹ 0.0053	²³ 0.0056	²¹ 0.0060	²⁴ 0.0017 N ^{0.076 6}	⁵⁹ 0.0086	⁵⁴ 0.0089				⁵⁸ 0.0048 N ^{0.044 7}
26	ISYSTEMS-1	⁵⁸ 0.0048	⁴⁴ 0.0050	²⁹ 0.0053	²² 0.0056	²⁰ 0.0060	³⁵ 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			⁶² 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	¹⁹ 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}	¹²² 0.2518	¹²² 0.3113				⁶⁴ 0.0114 N ^{0.292 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	³ 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	⁷ 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	²⁷ 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				²⁹ 0.0038 N ^{0.051 2}	⁷⁵ 0.0135	⁶⁷ 0.0138	³⁴ 0.0142	³² 0.0147	³¹ 0.0154	⁶⁸ 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	²⁴ 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	¹² 0.0013	¹² 0.0016	⁷ 0.0021	⁹ 0.0026		¹⁹ 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	¹² 0.0013	¹⁵ 0.0018	¹⁰ 0.0022	¹¹ 0.0029	¹³ 0.0038	⁹ 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	⁶⁹ 0.0074	⁵⁸ 0.0100	³⁶ 0.0120	³⁴ 0.0146	²⁶ 0.0176	³⁷ 0.0001 N ^{0.299 39}	⁷⁴ 0.0127	⁷² 0.0159	³⁷ 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	⁴ 0.0037	⁴⁰ 0.0047				³⁶ 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	⁴⁸ 0.0037	³⁹ 0.0047	²⁷ 0.0055	²⁵ 0.0067	²³ 0.0079	³⁷ 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	⁶⁴ 0.0059	⁵⁴ 0.0083	²⁹ 0.0108			¹⁵ 0.0000 N ^{0.393 55}	⁵⁸ 0.0077	⁵⁸ 0.0098				¹³ 0.0002 N ^{0.269 51}
48	SHAMAN-3	⁹⁴ 0.0344	⁸⁹ 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	¹¹⁴ 0.2635	⁸⁰ 0.2635	⁴⁹ 0.2636			-	¹¹ 0.0029	⁶ 0.0030	⁵ 0.0031	⁵ 0.0032	⁵ 0.0033	⁴⁸ 0.0016 N ^{0.046 9}
50	SIAT-2	¹¹⁴ 0.2124	⁷⁶ 0.2124				-	¹² 0.0031	¹¹ 0.0032	⁸ 0.0032	⁴ 0.0033	⁴ 0.0034	⁴⁸ 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	TONGVITRANS-1	⁶⁰ 0.0057	⁴⁸ 0.0060	³⁰ 0.0062	²⁶ 0.0067								

MISSSES BELOW THRESHOLD, T FNIR(N, T > 0, R > L)		ENROL LIFETIME DATASET: FRVT 2018					ENROL MOST RECENT 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	⁴⁴ 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	⁷ 0.8490	⁴⁵ 0.8640	⁴⁵ 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	²⁴ 0.0513				²³ 0.0455	²⁰ 0.0557	¹⁵ 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	²⁶ 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		⁸⁰ 0.4380	⁸⁹ 0.4813			
11	EVERAI-0	²⁴ 0.0460	²⁸ 0.0676				³¹ 0.0681	²⁴ 0.0921	²⁵ 0.1223		
12	EVERAI-1	¹² 0.0255	¹⁵ 0.0360				¹³ 0.0383	¹⁶ 0.0518	¹⁴ 0.0686		
13	EYDEA-3	²⁹ 0.2911	⁶⁴ 0.3283	⁴² 0.3673	³⁹ 0.4154		⁸³ 0.3498	⁸¹ 0.3893			
14	GLORY-1	⁶⁶ 0.2160	²⁸ 0.2447	³⁷ 0.2618	²⁸ 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	²⁷ 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	³⁶ 0.1013	²⁴ 0.1173	²⁴ 0.1498	²⁶ 0.2503
18	IDEMIA-0	³³ 0.0645	²⁸ 0.0802	²³ 0.0986	²⁴ 0.1237	¹⁷ 0.1872	⁴¹ 0.0920	⁴¹ 0.1135	³⁰ 0.1332	²⁷ 0.1628	²⁶ 0.2208
19	IDEMIA-1	¹⁸ 0.0304	¹⁶ 0.0377	¹¹ 0.0465	⁷ 0.0623	⁹ 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	¹⁴ 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	²⁴ 0.1400	¹⁸ 0.1796	³⁰ 0.2159	²⁸ 0.2741		⁵⁶ 0.1763	⁵⁶ 0.2143			
25	ISYSTEMS-0	²⁵ 0.0485	²⁶ 0.0633	²⁰ 0.0795	¹⁹ 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	²⁶ 0.1489	¹⁶ 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	⁵ 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	²⁶ 0.0815	²¹ 0.0961	¹⁷ 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	²⁰ 0.0575	¹⁷ 0.0711	¹⁵ 0.0954	¹⁴ 0.1845	²⁵ 0.0493	²⁴ 0.0656	¹⁹ 0.0810	¹⁶ 0.1167	¹⁵ 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	⁸ 0.1528
43	RANKONE-0	⁵⁵ 0.1485	³⁸ 0.1788	³⁶ 0.2210	²⁶ 0.3260	²⁰ 0.4758	⁶⁰ 0.1899	⁵⁹ 0.2192	³⁵ 0.2635	³⁰ 0.2992	³¹ 0.4301
44	RANKONE-1	⁵¹ 0.1211	⁴⁰ 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	²¹ 0.2946	⁴⁷ 0.0998	⁴⁴ 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	⁶⁷ 0.2727	³⁸ 0.2758			² 0.0160	² 0.0201	² 0.0260	¹ 0.0380	¹ 0.1069
50	SIAT-2	⁶⁸ 0.2198	⁵⁸ 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	²⁷ 0.1328		²⁶ 0.0545	²⁶ 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	¹⁰ 0.0353	⁹ 0.0441	⁹ 0.0628	¹² 0.1727	¹⁷ 0.0396	¹⁶ 0.0531	¹³ 0.0654	¹⁶ 0.0878	¹⁶ 0.1894
59	VOCORD-3	⁴⁴ 0.0969	⁴⁵ 0.1295	³³ 0.1627	³⁰ 0.2361		⁴⁵ 0.0973	⁴⁶ 0.1258			
60	YISHENG-1	⁷³ 0.2539	⁶¹ 0.3002	⁴⁰ 0.3366	²⁷ 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	¹ 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	⁵⁸ 0.026	⁶³ 0.034	⁵³ 0.086	⁶⁰ 0.191	³⁰ 0.071	⁶⁸ 0.103	⁷² 0.160	⁵⁷ 0.302	⁵⁴ 0.435	²⁸ 0.095	0.004	0.003		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	⁶⁵ 0.030	⁶⁵ 0.040		⁶⁵ 0.225	³⁵ 0.076	⁶⁷ 0.105	⁷⁴ 0.164		⁵⁶ 0.439	³⁵ 0.096	0.004	0.003		0.011	0.013
4	3DIVI-3	⁷³ 0.053	⁸⁸ 0.086	⁶⁶ 0.206	⁷⁵ 0.328	⁴⁸ 0.094	⁷⁴ 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	²⁸ 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	⁷⁴ 0.053	⁸⁴ 0.064	⁶¹ 0.138	⁷⁴ 0.286	⁸⁸ 0.588	⁶¹ 0.092	⁶⁰ 0.128	⁵² 0.253	⁵² 0.421	⁸¹ 0.587	0.013	0.006	0.054	0.129	0.143
9	AWARE-1	⁶⁹ 0.043	⁸¹ 0.059		⁷² 0.276	⁸² 0.580	⁵⁰ 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	³⁵ 0.025	⁶² 0.033	⁵⁴ 0.090	⁴⁵ 0.165	⁸¹ 0.503	⁴⁵ 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	⁶² 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	⁵⁹ 0.143	¹¹¹ 0.227	⁷⁹ 0.337	⁸¹ 0.349	⁶³ 0.148	⁵⁹ 0.409	¹¹¹ 0.449	⁸⁰ 0.648	⁸⁵ 0.771	⁶¹ 0.196	0.006	0.005	0.009	0.050	0.058
15	CAMVI-2	⁷⁰ 0.076	⁹⁵ 0.129		⁶⁹ 0.243	⁵⁰ 0.130	⁸⁰ 0.265	⁹⁹ 0.402		⁷² 0.608	⁵⁰ 0.157	0.006	0.005		0.050	0.058
16	CAMVI-3	⁶⁵ 0.035	⁷⁹ 0.054	⁵⁵ 0.090	⁴⁷ 0.160	⁶⁹ 0.139	³⁰ 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	³⁰ 0.011	³² 0.013	³⁵ 0.046	⁶⁶ 0.232	⁴⁹ 0.093	²⁴ 0.021	²² 0.032	²⁵ 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			²⁴ 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			⁴⁴ 0.054	⁵⁵ 0.098	⁴⁴ 0.200			0.002	0.003	0.002		
20	COGNITEC-1	³⁴ 0.013	⁴⁰ 0.014	²² 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	⁷⁶ 0.075	⁹⁶ 0.131	⁷⁰ 0.218	⁶⁸ 0.237	³⁸ 0.075	⁷⁶ 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	⁸⁰ 0.096	⁹⁸ 0.156		⁷¹ 0.264	⁴⁰ 0.089	⁸⁰ 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	³⁰ 0.076	⁸¹ 0.248	⁹⁶ 0.378		⁶² 0.507	³⁰ 0.105	0.004	0.003		0.011	0.020
24	DERMALOG-3		⁹³ 0.128	⁶⁹ 0.217				⁹² 0.362	⁷⁴ 0.526				0.002	0.002		
25	DERMALOG-4	⁷⁸ 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	EYDEA-0	⁵⁹ 0.201	¹¹⁵ 0.300	⁸² 0.443	⁸⁴ 0.369	³⁰ 0.131	⁵⁹ 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	EYDEA-1	⁸⁰ 0.109	¹⁰⁵ 0.198		⁴⁹ 0.172	³⁰ 0.072	⁸⁰ 0.324	¹⁰⁸ 0.480		⁶⁴ 0.534	⁴⁸ 0.131	0.001	0.001		0.008	0.008
30	EYDEA-2	⁸⁷ 0.110	¹⁰⁶ 0.200		⁵⁵ 0.184	²⁹ 0.070	⁸⁹ 0.327	¹⁰⁹ 0.490		⁶⁷ 0.548	⁴⁸ 0.130	0.001	0.000		0.007	0.005
31	EYDEA-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	⁸⁸ 0.109	⁹⁴ 0.129	⁷³ 0.267	⁸⁸ 0.453	⁷⁶ 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	⁸⁰ 0.994				⁸⁰ 0.708	⁸⁰ 0.994	⁸⁰ 0.994	0.001	0.001		0.004	0.008
35	GORILLA-1		⁸² 0.063	⁵⁶ 0.095		¹⁴ 0.057		⁸⁵ 0.248	⁵⁹ 0.314		⁸⁰ 0.994		0.001	0.001		0.007
36	HBINNO-0	⁵⁸ 0.191	¹¹⁴ 0.275		⁸⁷ 0.437	⁷³ 0.335	⁵⁸ 0.498	¹¹⁵ 0.632		⁹³ 0.975	⁷³ 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	⁶⁰ 0.153	⁴⁴ 0.049	⁴⁷ 0.070	²⁸ 0.103	¹⁸ 0.160	⁵⁸ 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	³⁵ 0.013	⁴² 0.017		⁸ 0.035	⁵⁹ 0.094	³⁵ 0.037	⁴⁶ 0.067		¹⁵ 0.158	³⁵ 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	²⁵ 0.008	³⁷ 0.014	²⁶ 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	²⁵ 0.011	²⁹ 0.034	³⁶ 0.096	⁶⁶ 0.166	³² 0.036	⁴² 0.062	³⁴ 0.156	³⁹ 0.302	⁷¹ 0.288	0.003	0.003	0.000	0.003	0.002
43	IDEMIA-1	²³ 0.008	³⁰ 0.012		³⁵ 0.095	⁶⁰ 0.157	¹⁰ 0.021	²¹ 0.031		²⁴ 0.191	⁶³ 0.205	0.003	0.003	0.003	0.003	0.002
44	IDEMIA-2	³⁰ 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	²³ 0.010	²⁹ 0.034			²³ 0.021	¹³ 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	³⁶ 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	¹⁰⁸ 0.216	¹¹⁶ 0.305	⁸⁴ 0.482	⁹⁰ 0.496	⁷⁹ 0.222	⁹⁹ 0.468	¹¹⁹ 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	⁶⁵ 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	⁵⁷ 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	⁶⁹ 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	0.06

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				
#	ALGORITHM	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	²³ 0.065	⁷ 0.010	⁷ 0.022	¹¹ 0.071	²⁶ 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	² 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	³⁸ 0.125	⁸⁹ 1.000	⁷⁶ 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	⁵⁸ 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	⁴ 0.026	⁵ 0.041	²³ 0.023	³¹ 0.047	³⁰ 0.105	² 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	⁴ 0.024	⁵ 0.043	⁵ 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	²⁸ 0.302	⁵⁵ 0.114	⁵⁹ 0.090	⁶⁷ 0.129	⁵⁶ 0.291	⁷⁷ 0.584	⁵⁷ 0.161	0.000	0.000	0.000	0.002	0.000
82	RANKONE-1	⁶⁸ 0.032	⁵⁶ 0.025		³⁶ 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	⁴⁵ 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	⁷⁷ 0.628	⁵² 0.146	0.020	0.020	0.011	0.098	0.043
88	SHAMAN-1	⁸⁸ 0.118	¹⁰¹ 0.172		⁷⁹ 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	⁹² 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	⁸⁸ 0.244	⁹¹ 0.348	⁶⁸ 0.472	⁷⁰ 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	¹⁹ 0.007	²⁶ 0.010	¹⁴ 0.021	⁷ 0.019	⁴¹ 0.078	²⁷ 0.025	²⁹ 0.047	⁷ 0.064	³ 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	⁰ 0.009	² 0.040	⁷ 0.007	¹ 0.009	⁶¹ 0.348	⁰ 0.033	⁰ 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	¹⁰⁵ 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	³² 0.022	⁴⁵ 0.066	³⁰ 0.172	¹⁵ 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	³³ 0.022		³² 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		³¹ 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	²⁴ 0.074	¹⁰⁸ 1.000	⁷⁴ 0.151	⁸⁶ 0.263	⁶² 0.366	³⁴ 0.256	¹⁰⁸ 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	²³ 0.008	²³ 0.010	¹⁵ 0.022	¹⁷ 0.049	⁵³ 0.112	¹⁷ 0.020	²⁵ 0.035	⁶ 0.062	¹⁴ 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	³⁷ 0.188	³⁸ 0.076	⁸² 0.260	⁹⁸ 0.394	⁷⁶ 0.557	⁶⁸ 0.552	⁵⁵ 0.152	0.001	0.000	0.001	0.005	0.003
110	VIGILANTSOLUTIONS-1	⁹⁰ 0.120	¹⁰⁷ 0.204		²⁶ 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		⁶⁶ 0.195	²² 0.064	¹⁰⁸ 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	⁵⁸ 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	²⁹ 0.009	²⁰ 0.009	²² 0.030	³⁴ 0.093	¹¹ 0.051	¹⁴ 0.015	¹⁷ 0.026	¹⁹ 0.091	⁴³ 0.246	⁴ 0.046	0.004	0.002	0.003	0.014	0.014
115	VISIONLABS-4	⁴ 0.003	⁵ 0.004	⁸ 0.020			¹⁰ 0.012	¹⁸ 0.026	²³ 0.097			0.001	0.001	0.001		
116	VISIONLABS-5	² 0.003	⁵ 0.004	⁶ 0.019	⁷ 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									

2018/11/26
07:24:51

FNIR(N, R, T) =
FPIR(N, T) =

False neg. identification rate
False pos. identification rate

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

T = Threshold

T = 0 → Investigation
T > 0 → Identification

False negative identification rate (FNIR)

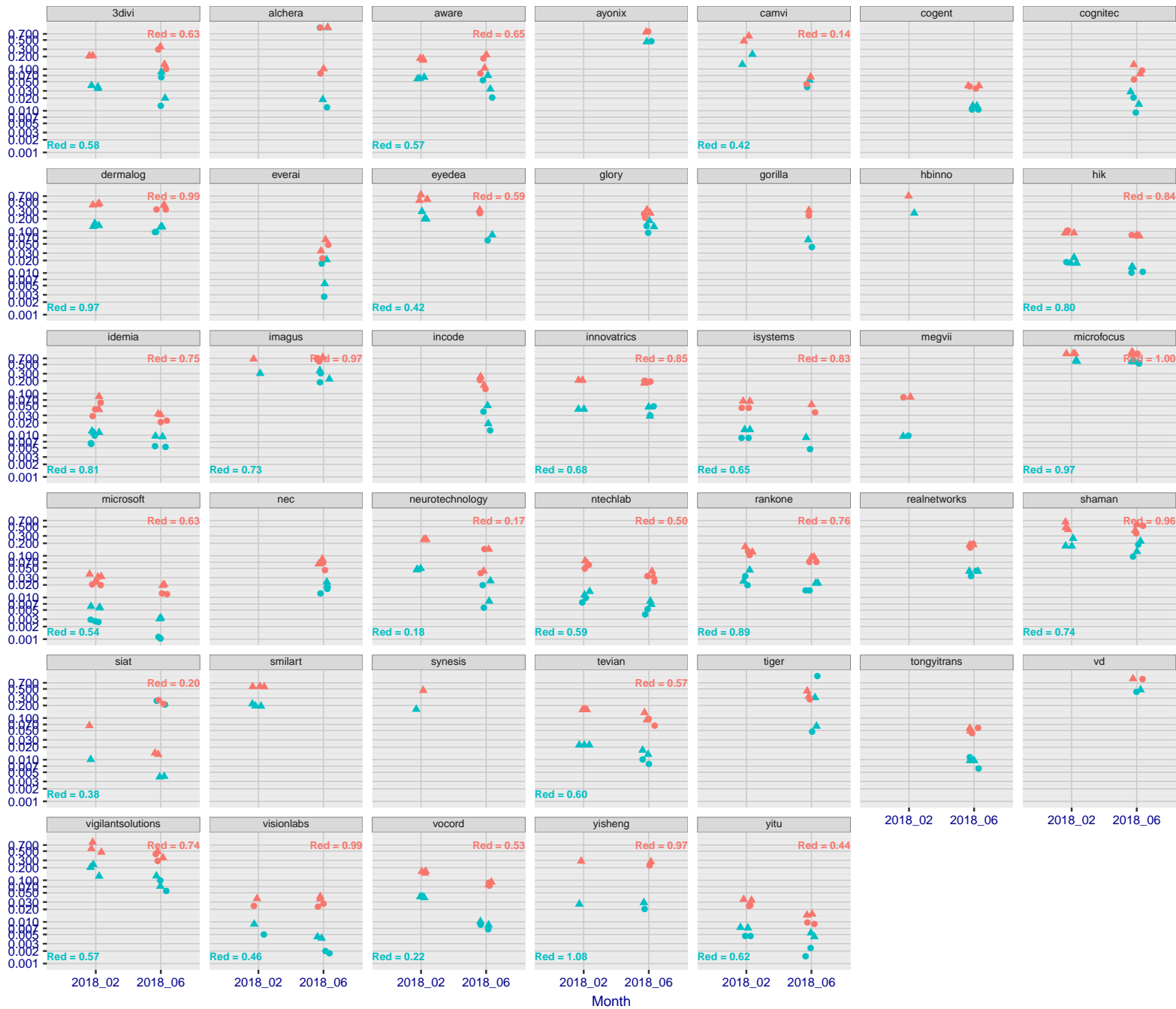


Figure 19: . [Mugshot Dataset] Error rate reductions in 2018. For each FRVT2018 participant, the plot shows accuracy gains between Phase 1 (Feb 2018) and Phase 2 (Jun 2018) according to two metrics: rank one miss rate, $FNIR(N, 1, 0)$, and high threshold, $FNIR(N, N, T)$, set to achieve $FPIR = 0.003$.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

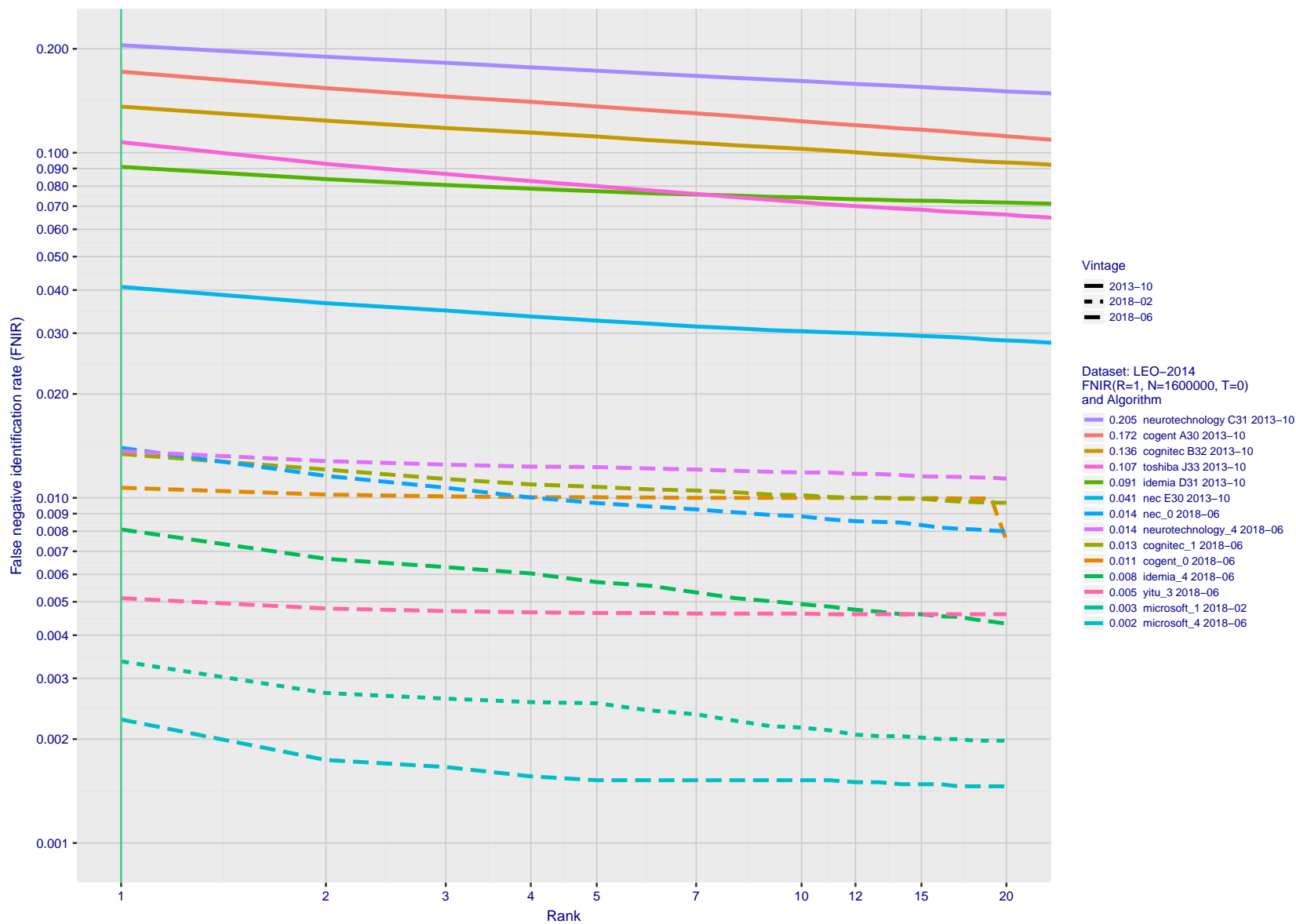


Figure 20: **[Gains 2013-2018]** On the LEO set used in FRVT2014, the figure shows investigational miss rates vs. rank for the most accurate algorithms submitted to NIST in October 2013 and in February/June 2018. The reduction in error rates is an order of magnitude. For the most accurate algorithms, miss rates fell approximately twelvefold from 4.1% to 0.34%.

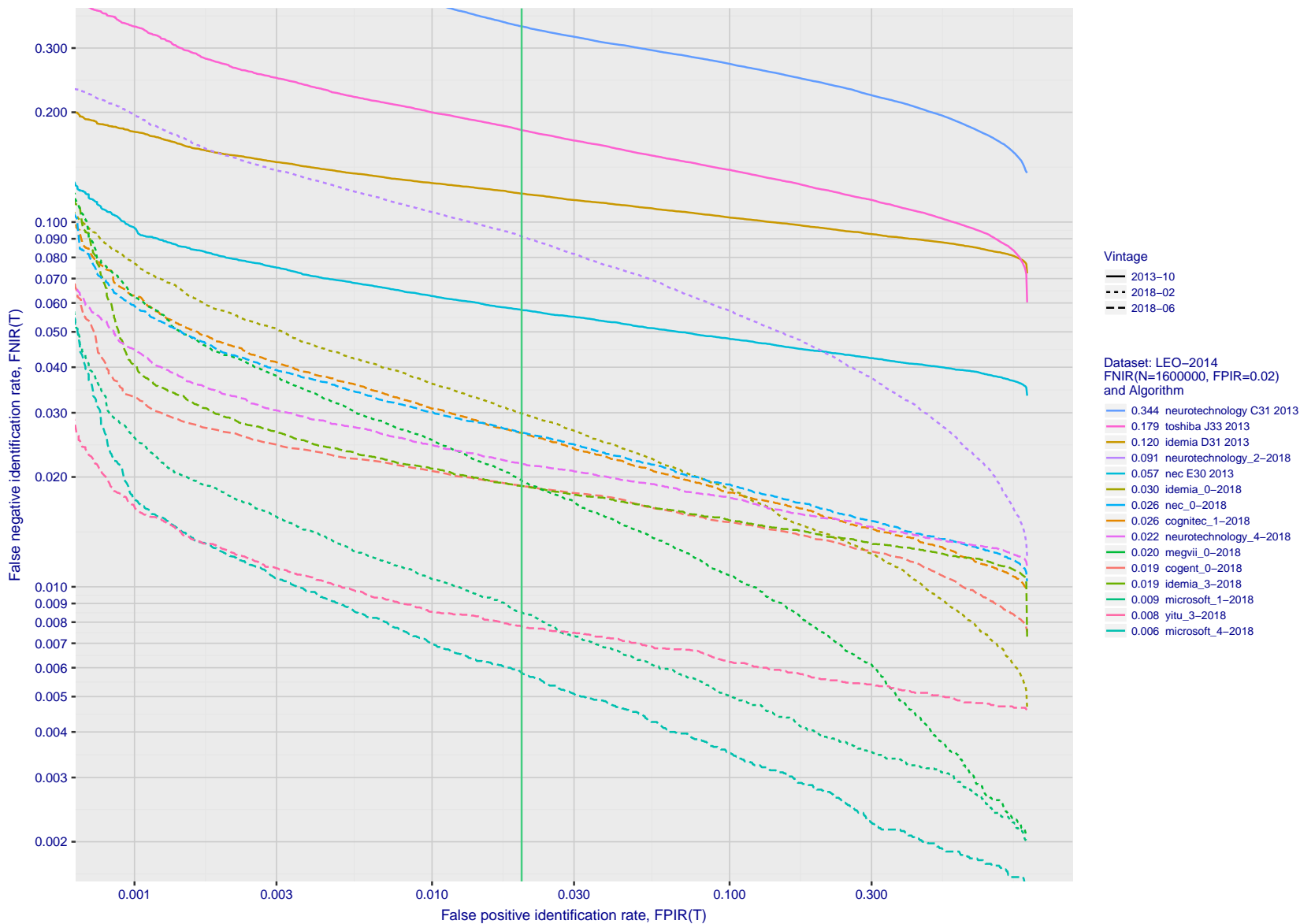


Figure 21: **[Gains 2013-2018]** On the LEO set used in FRVT2014, the figure shows identification miss rates vs. false positive rates for the most accurate algorithms submitted to NIST in October 2013 and February/June 2018. The reduction in error rates is not as large as for rank-based miss rates but, 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.

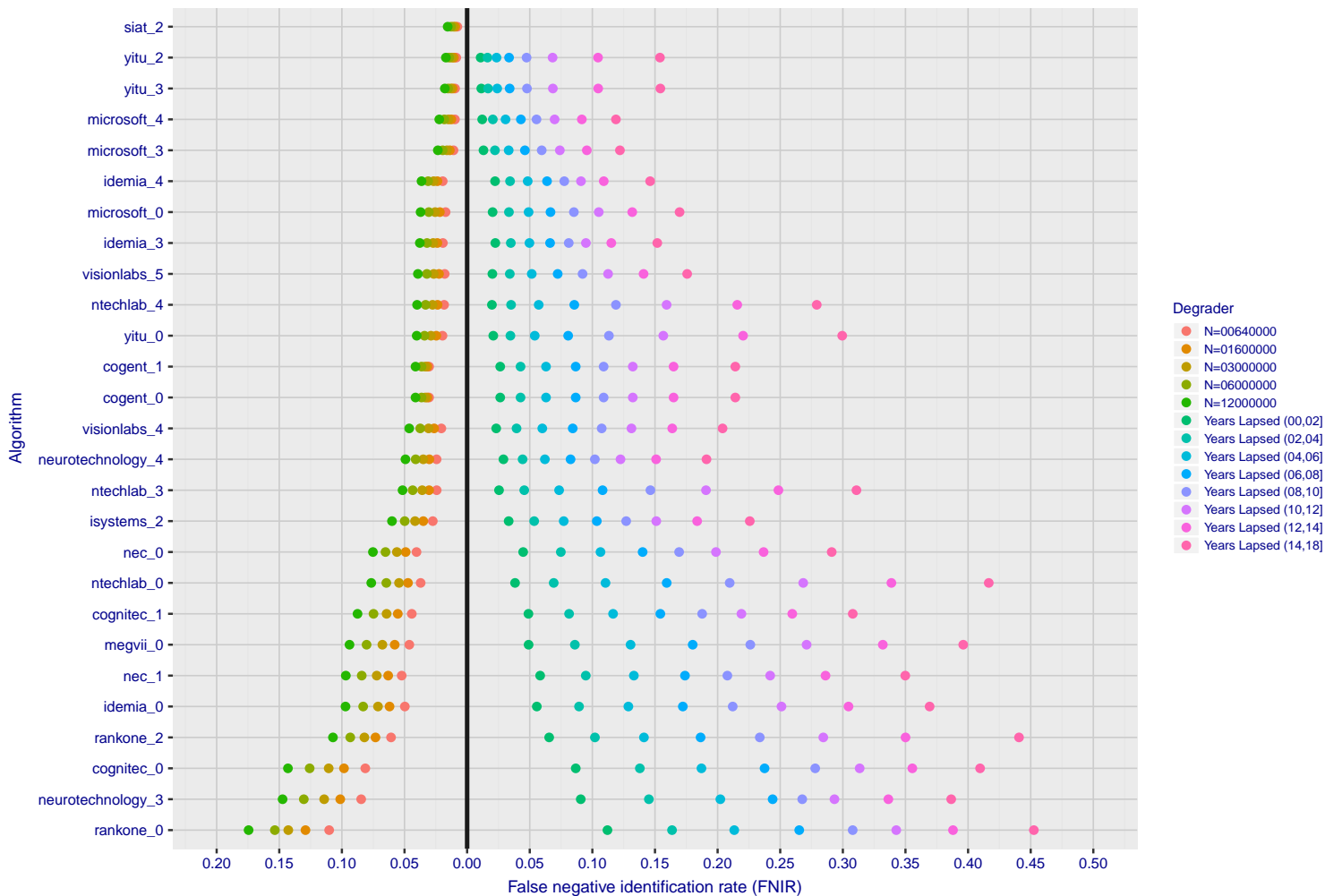


Figure 22: **[FRVT-2018 Mugshot Ageing Dataset] Contrast of ageing and population size dependency..** The Figure shows, at left, the dependence $FNIR(N)$ for the FRVT-2018, as tabulated in Table 14. At right, is $FNIR(N = 3\,000\,000, \Delta T)$ from Figure:68. Ageing miss rates are computed over all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

2018/11/26
07:24:51FNIR(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 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

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

2018/11/26
07:24:51FNIR(N, R, T) =
FPIR(N, T) =False neg. identification rate
False pos. identification rateN = Num. enrolled subjects
R = Num. candidates examined

T = Threshold

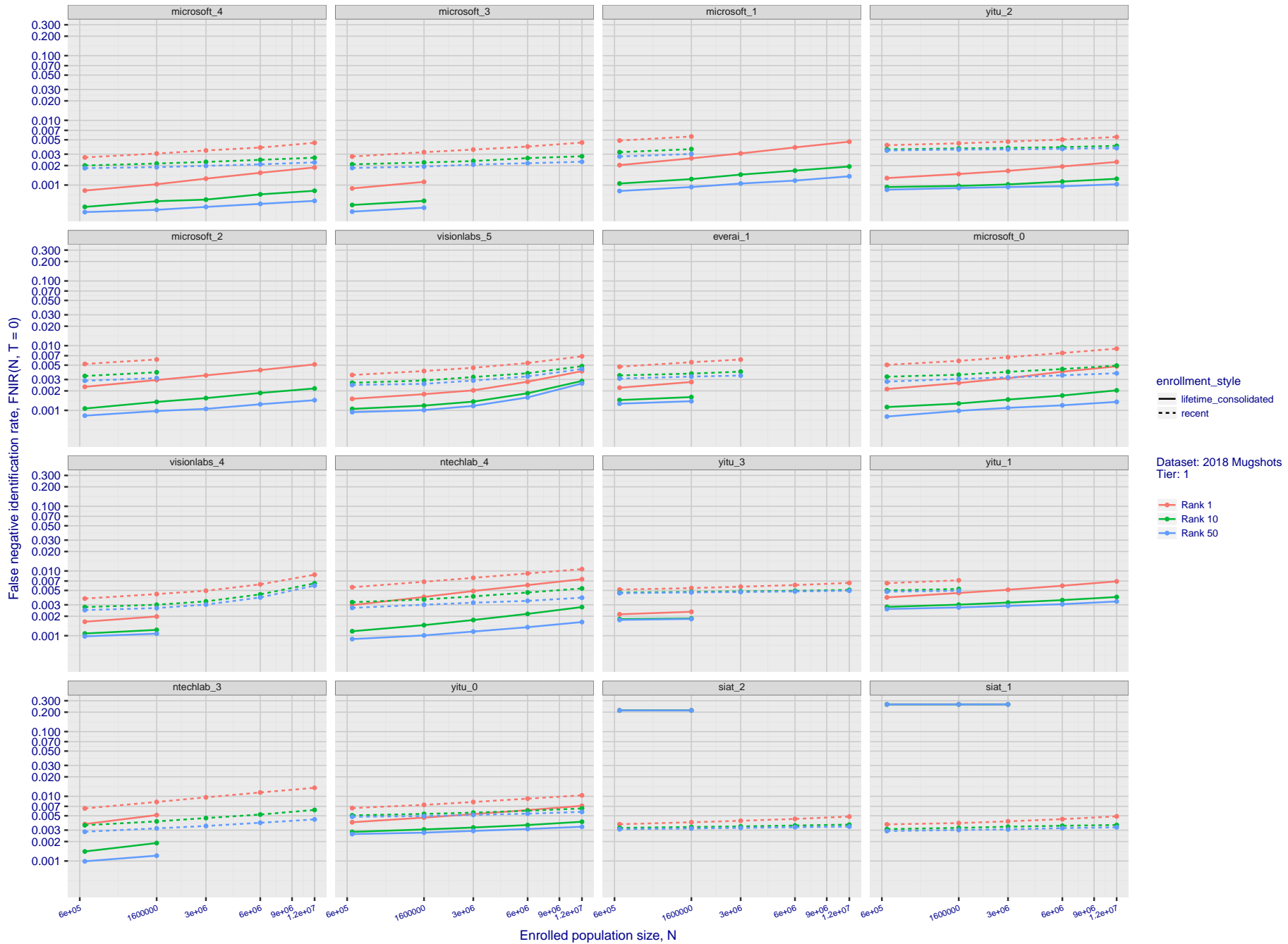
T = 0 → Investigation
T > 0 → Identification

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.

2018/11/26
07:24:51
FNIR(N, R, T) =
FPNR(N, T) =
False neg. identification rate
False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

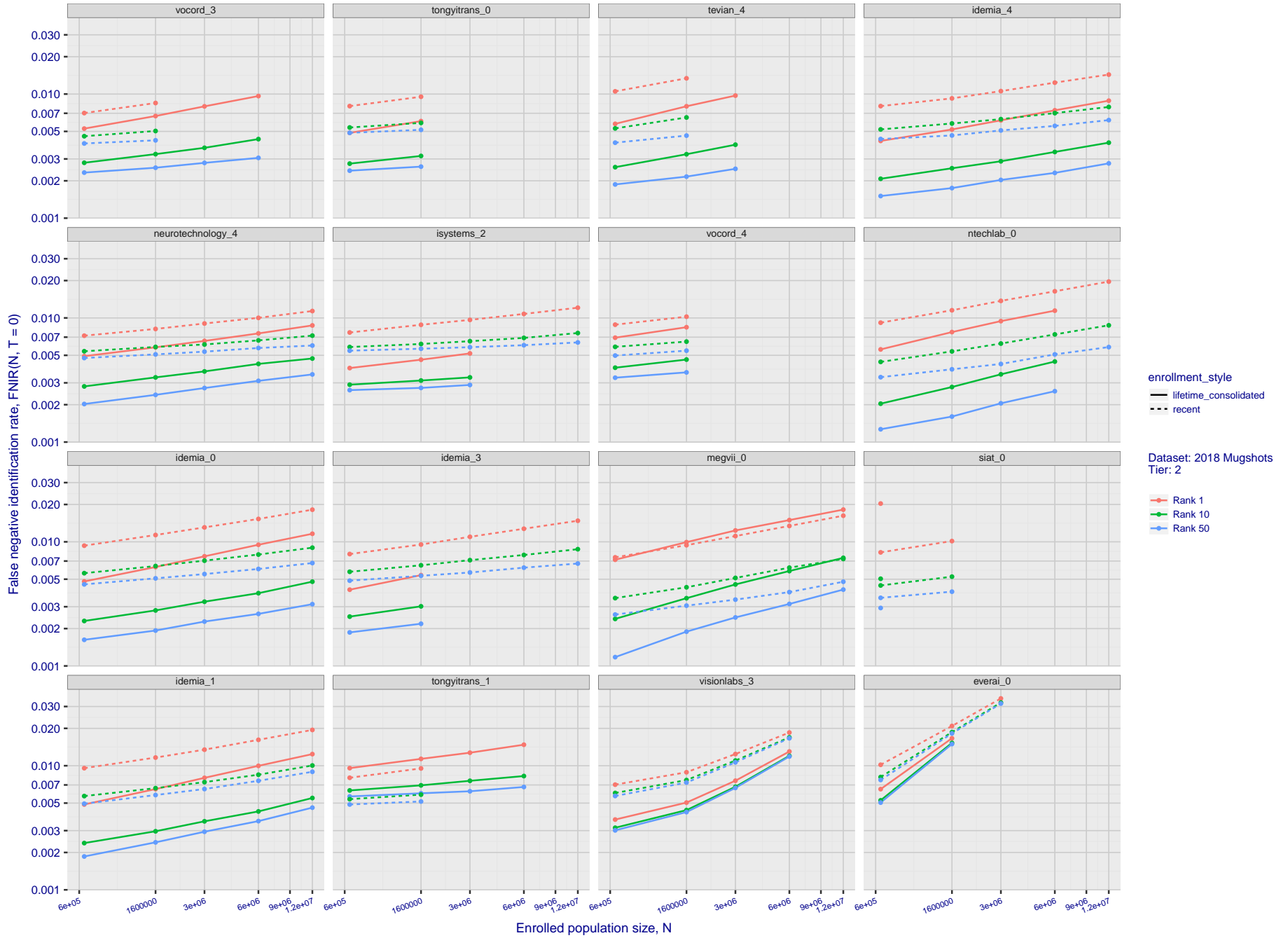


Figure 25: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects.** For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this 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) =
FPR(N, T) =

False neg. identification rate
False pos. identification rate

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

T = Threshold

T = 0 → Investigation
T > 0 → Identification

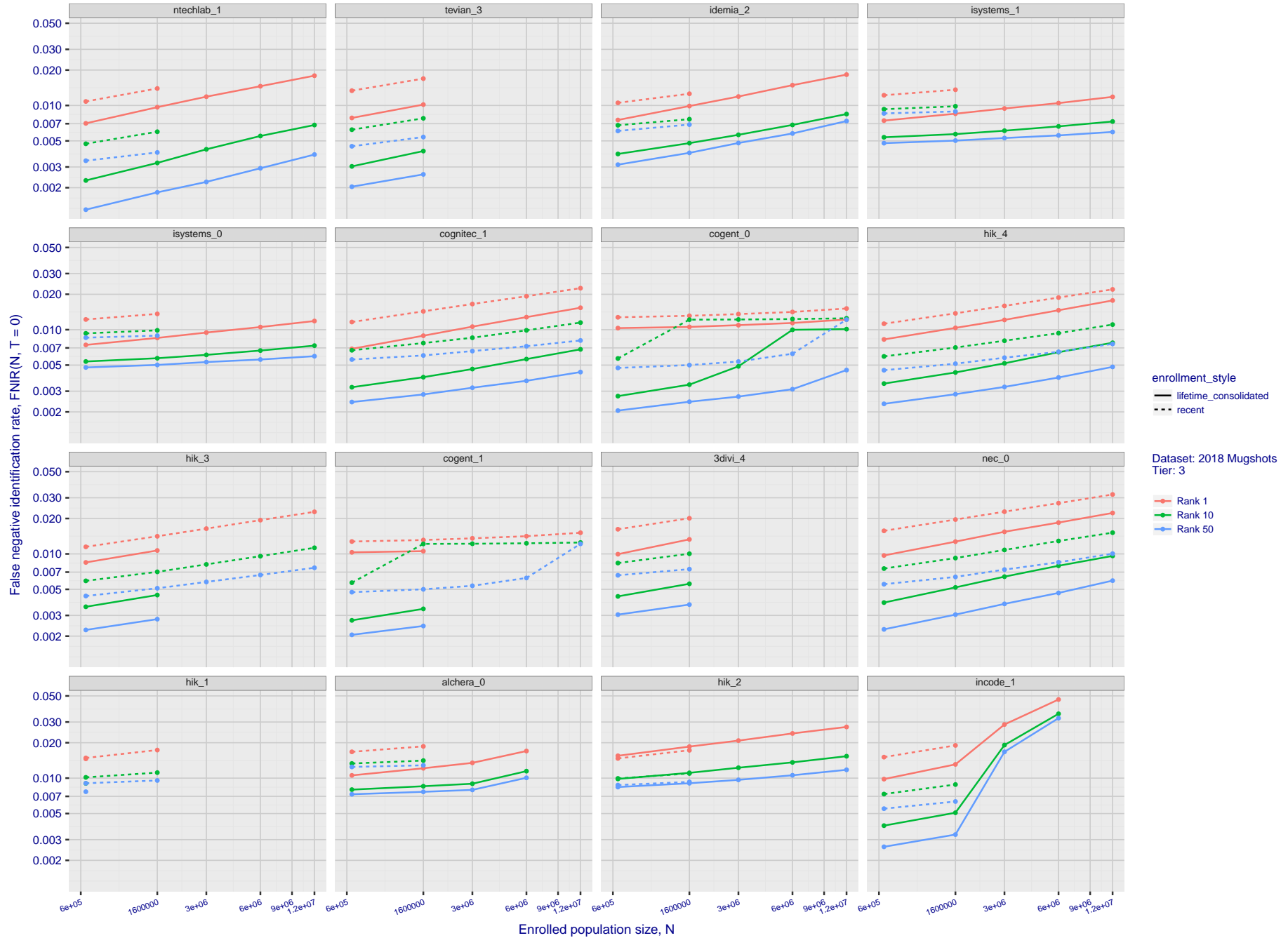


Figure 26: [FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects. For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric 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) =
FPR(N, T) =
False neg. identification rate
False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

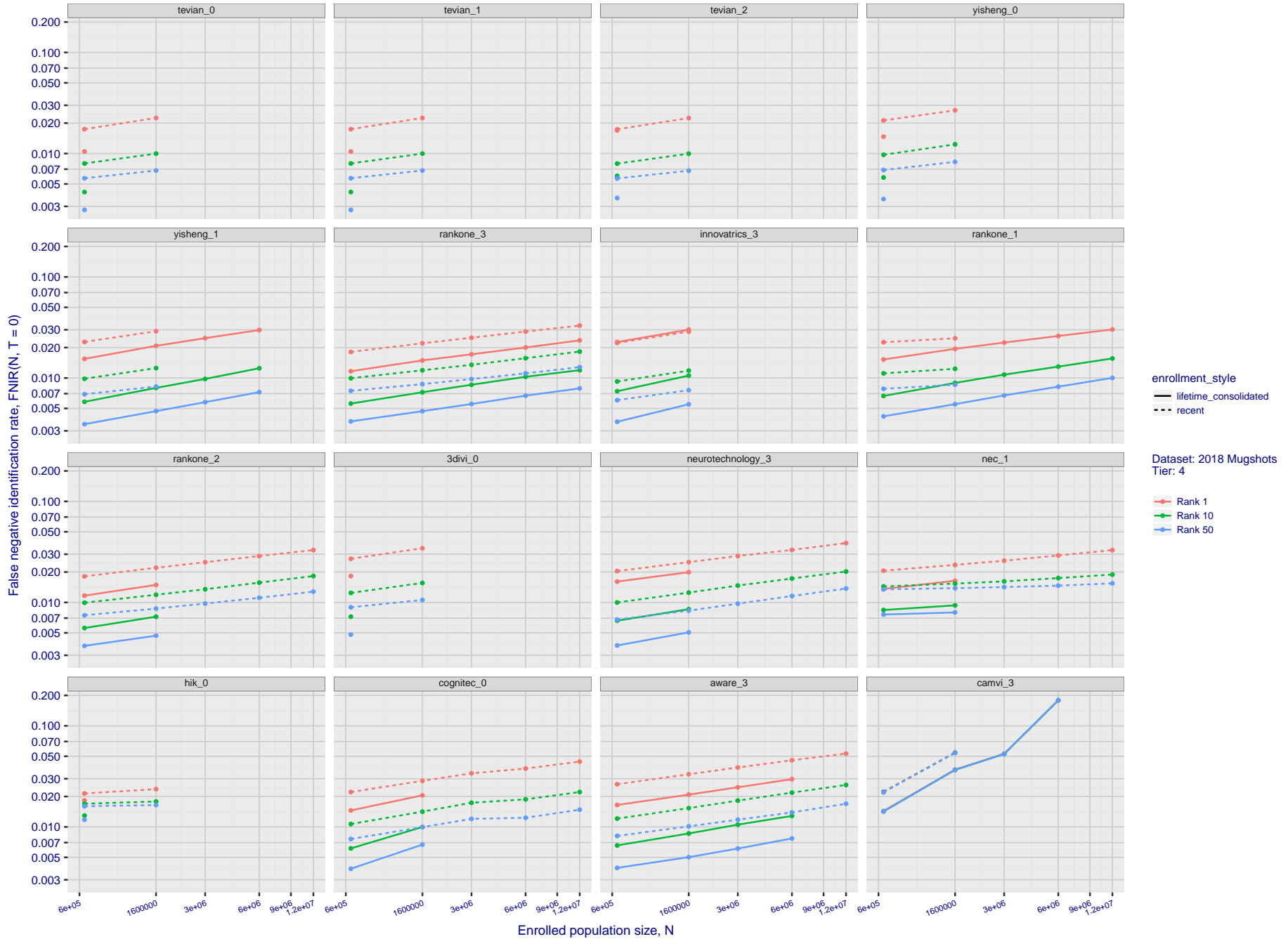


Figure 27: [FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects. For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric 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.

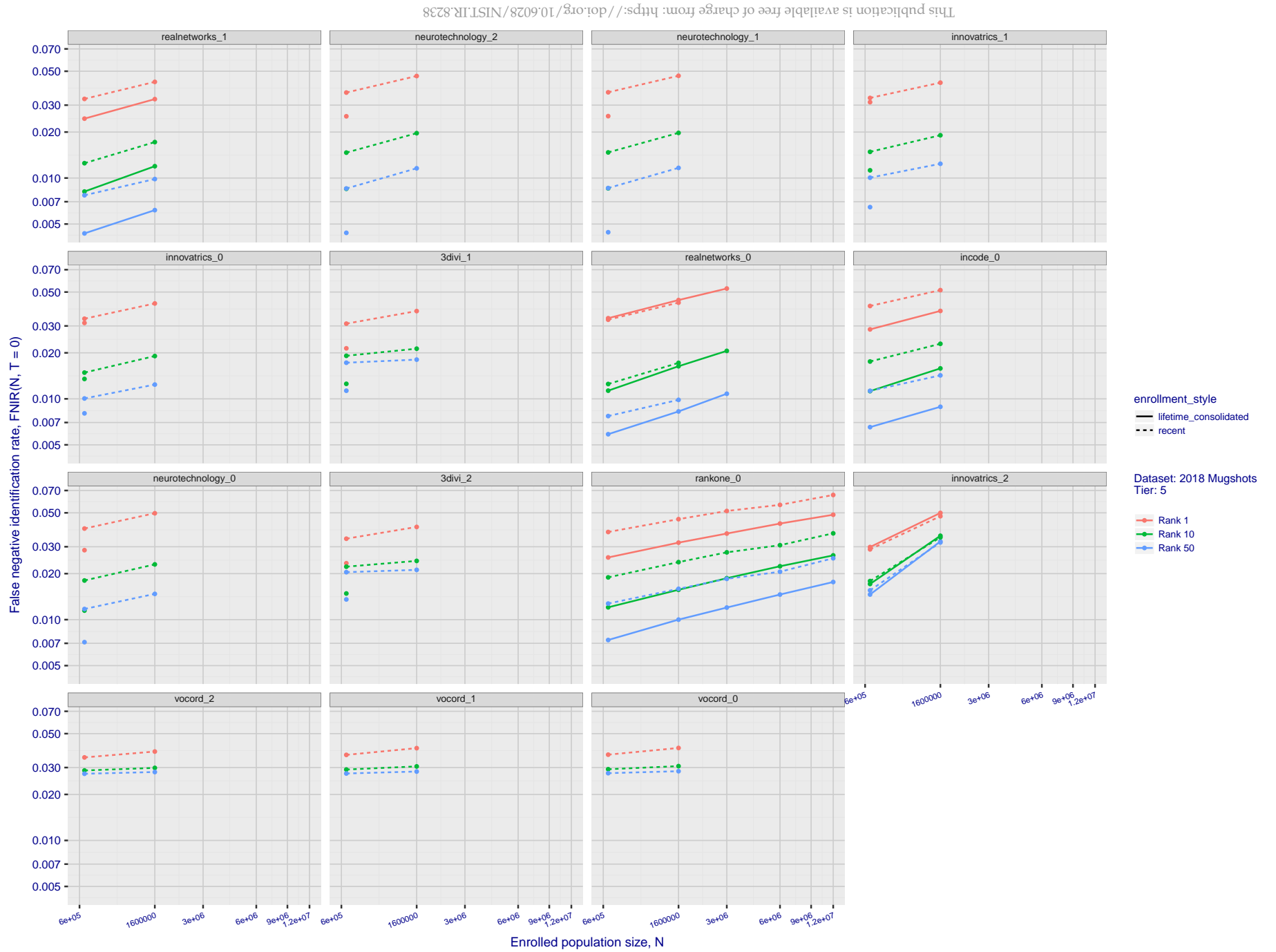


Figure 28: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects.** For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this 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.

2018/11/26
07:24:51

FNIR(N, R, T) =
FPR(N, T) =

False neg. identification rate
False pos. identification rate

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

T = Threshold

T = 0 → Investigation
T > 0 → Identification

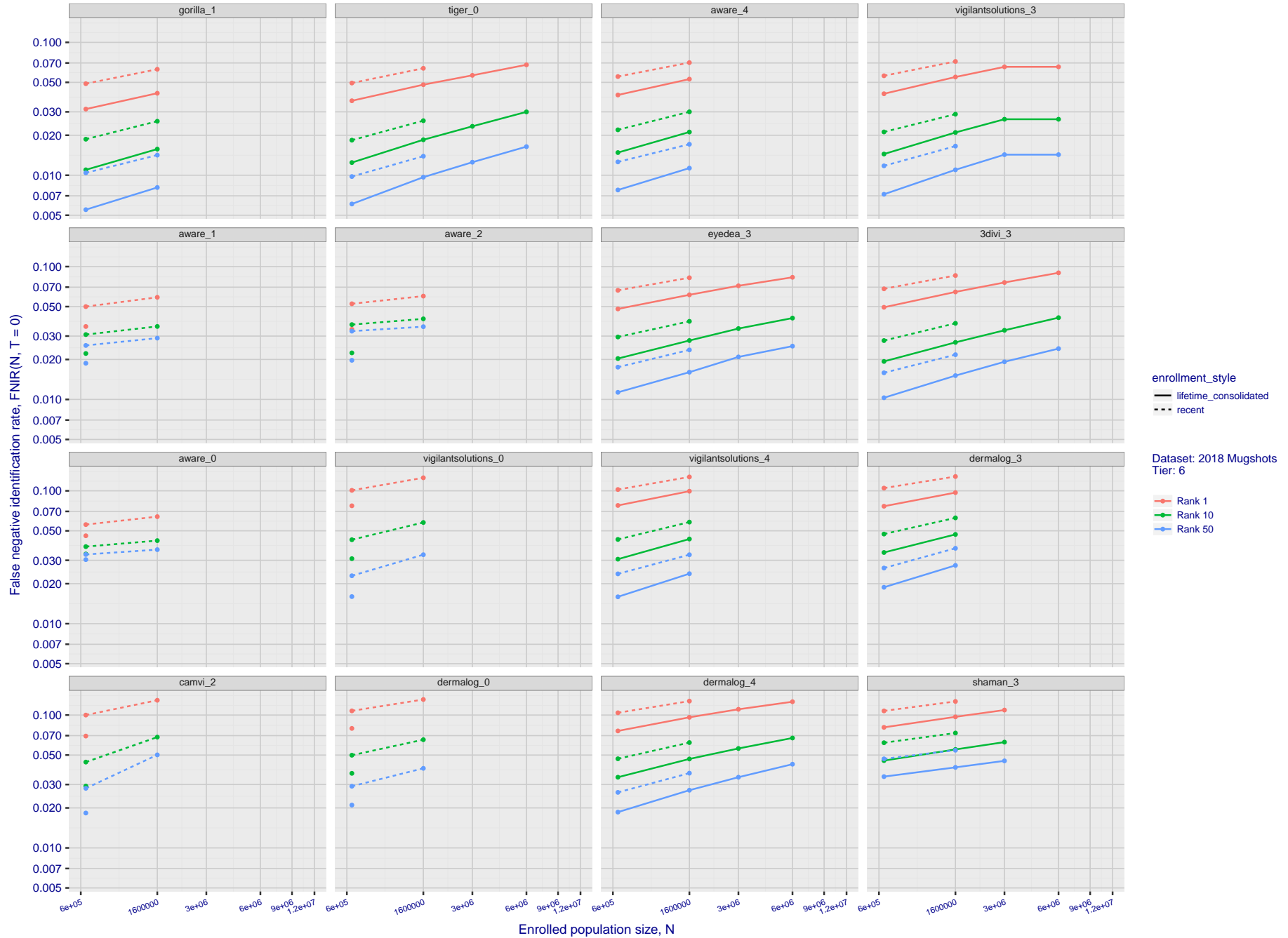


Figure 29: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects.** For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this 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.

False negative identification rate, FNIR(N, T = 0)

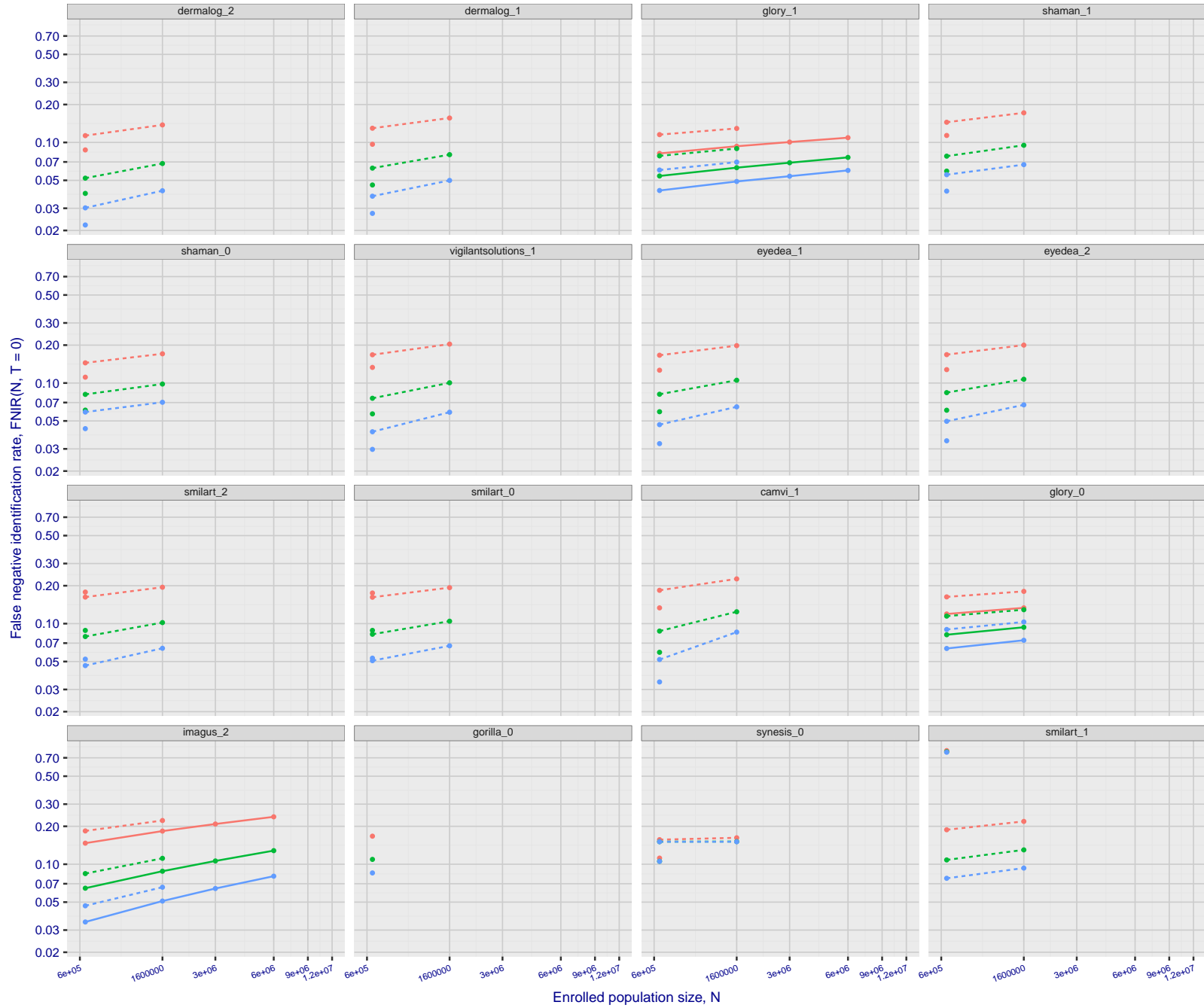


Figure 30: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects.** For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this 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) =
FPR(N, T) =
False neg. identification rate
False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

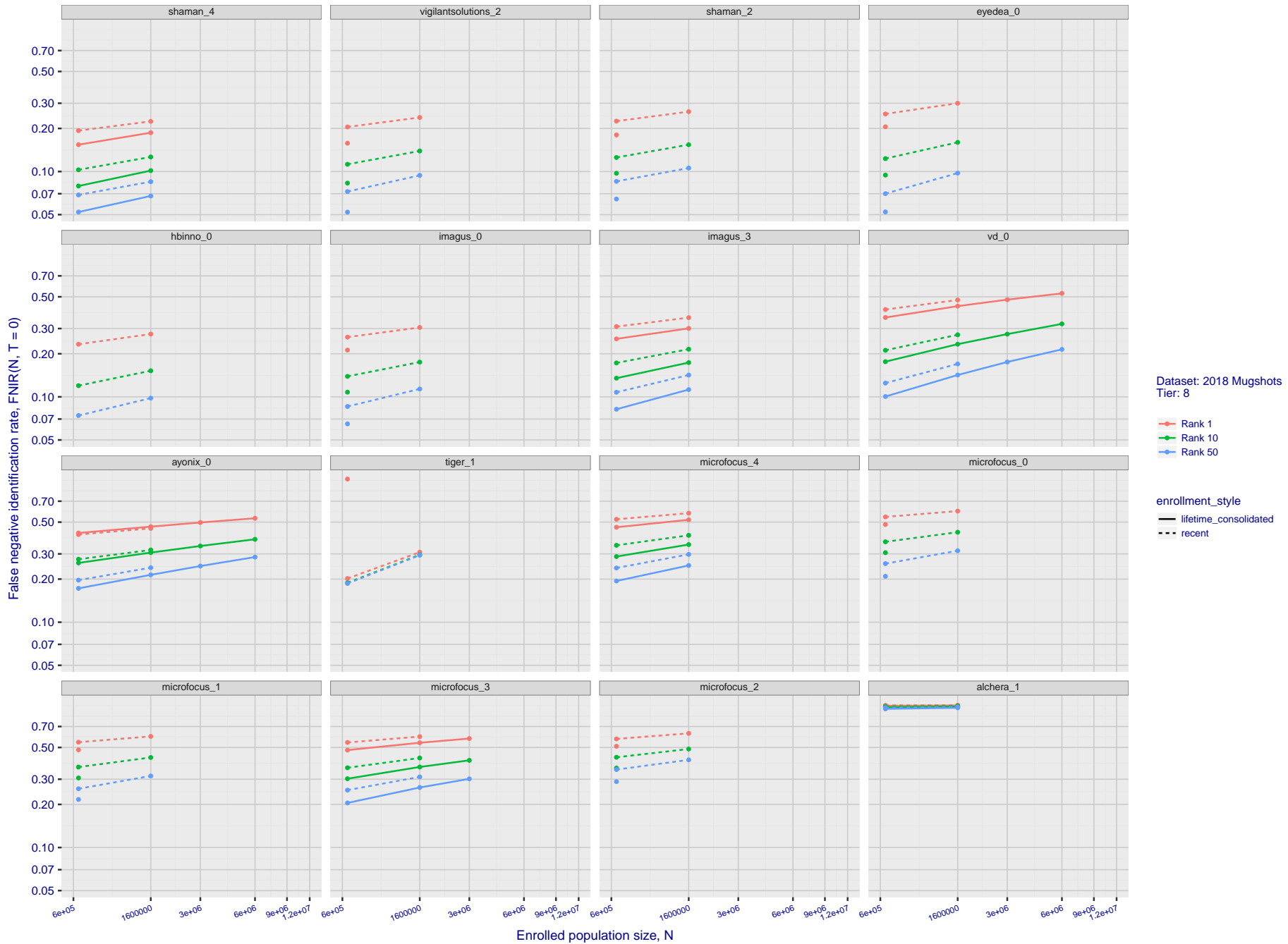


Figure 31: [FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. number of enrolled subjects. For the 2018 mugshots dataset, the figure shows false negative identification rates, FNIR(N, R), across various gallery sizes and ranks 1, 10 and 50. The threshold is set to zero, so this metric 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.

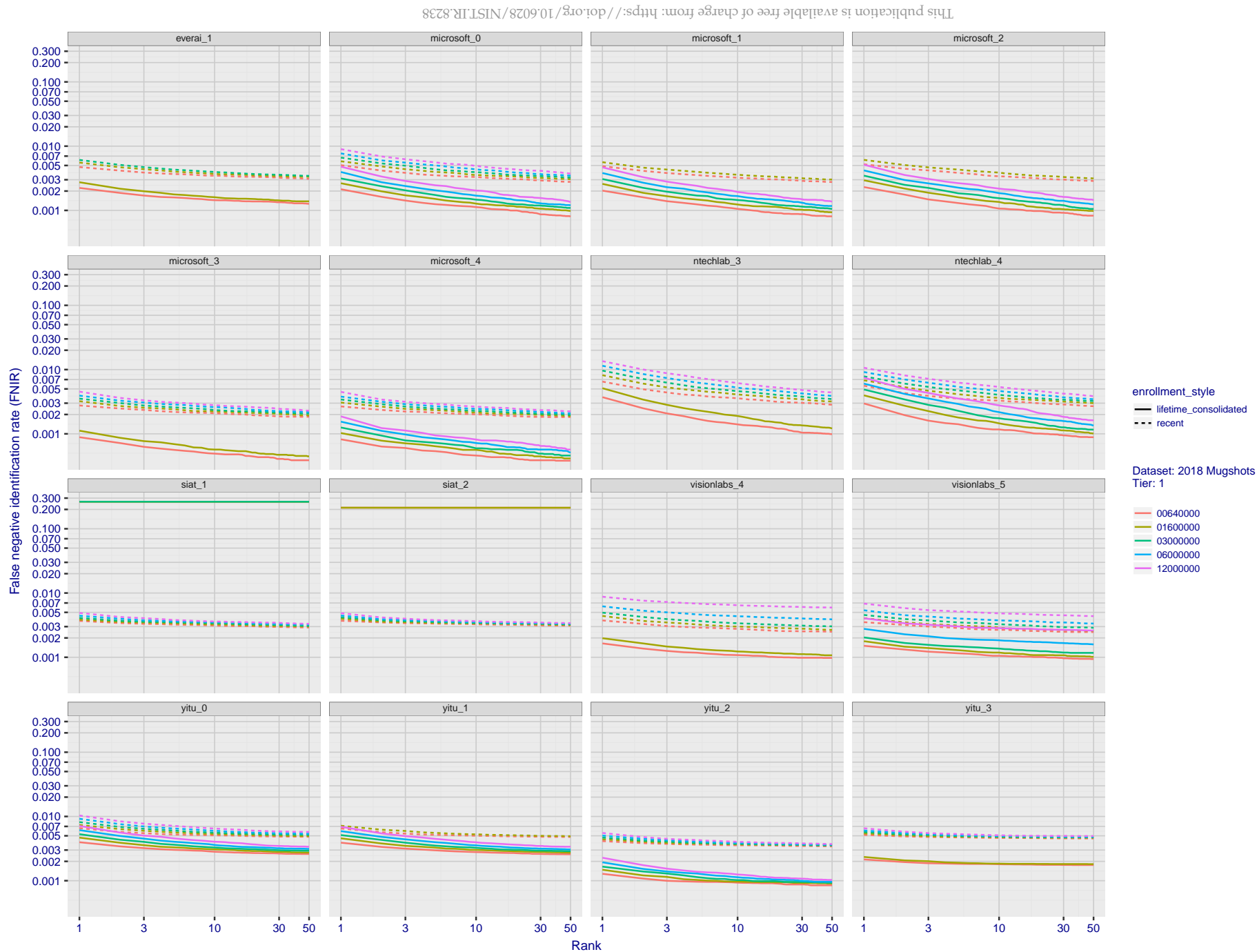


Figure 32: [FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank. For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted 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.

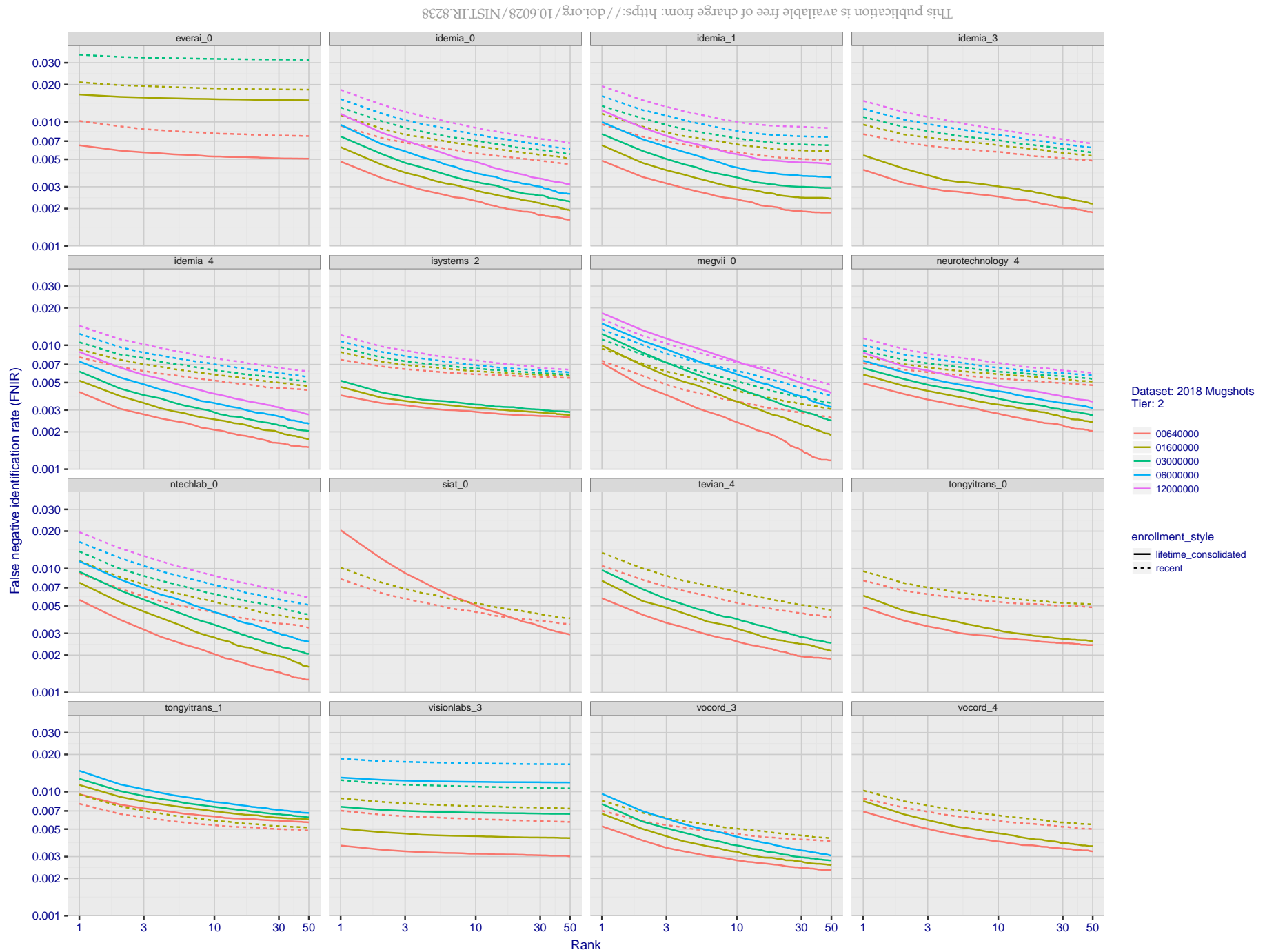


Figure 33: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank.** For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted 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.

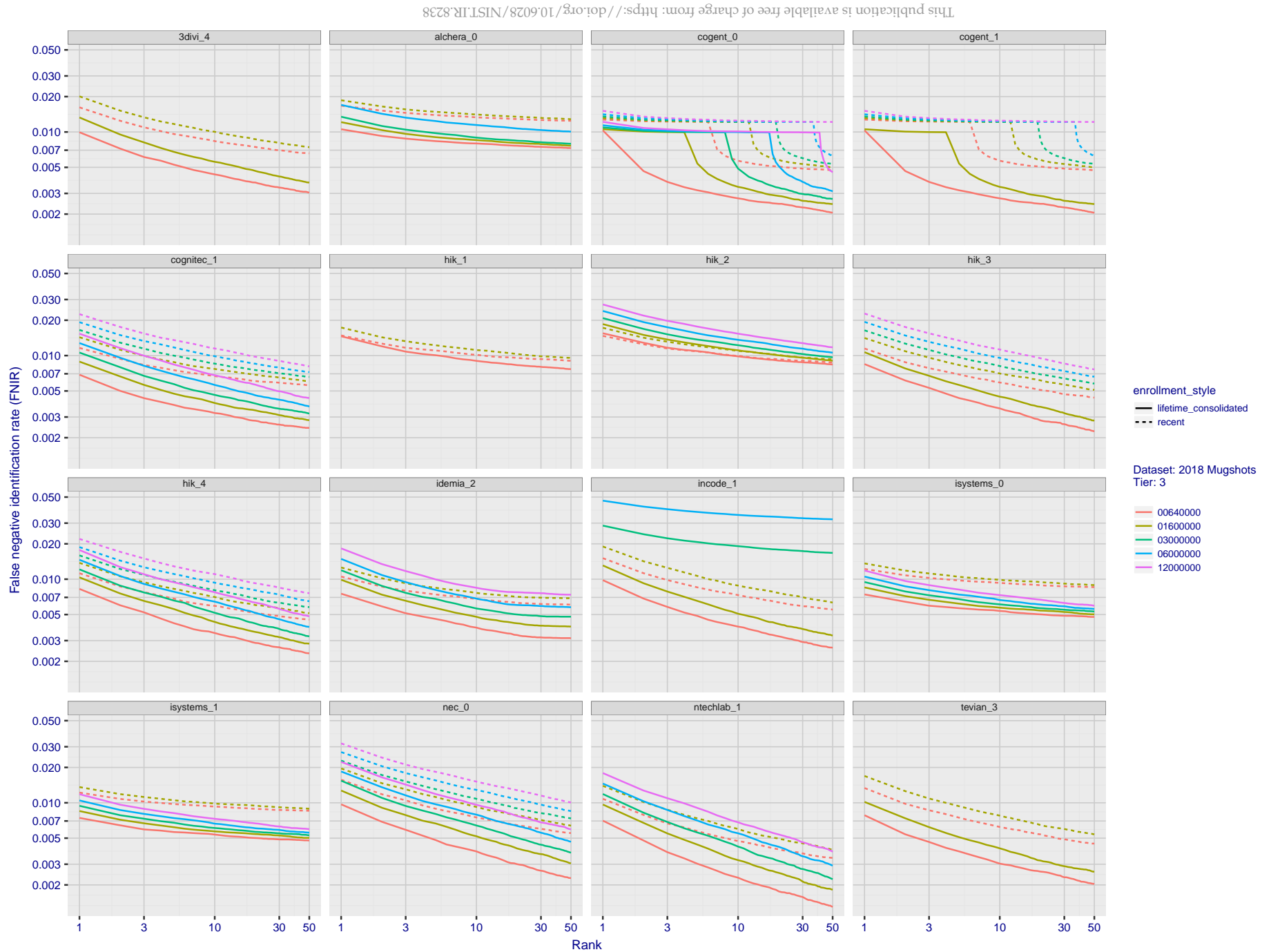


Figure 34: [FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank. For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted 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.

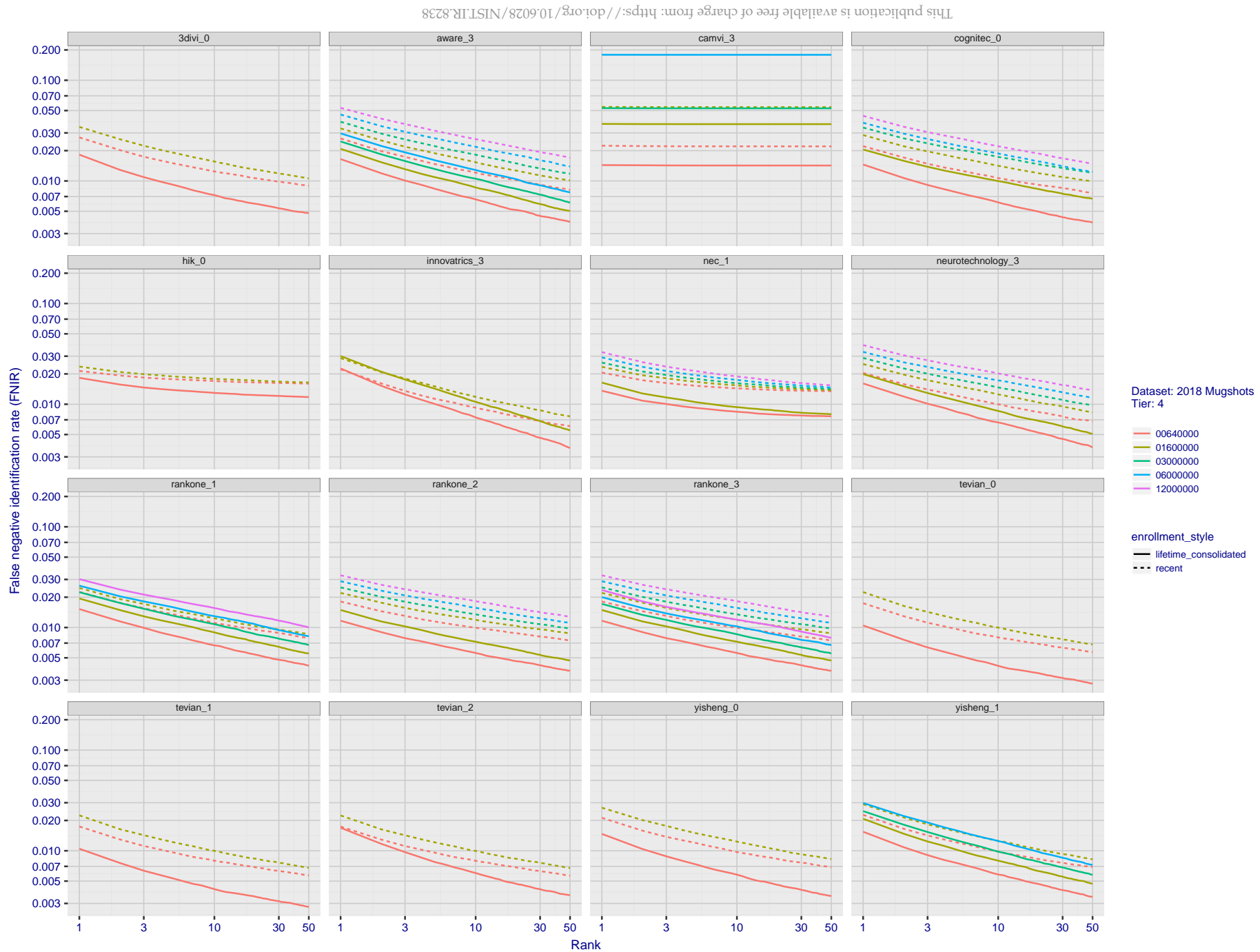


Figure 35: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank.** For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted 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.

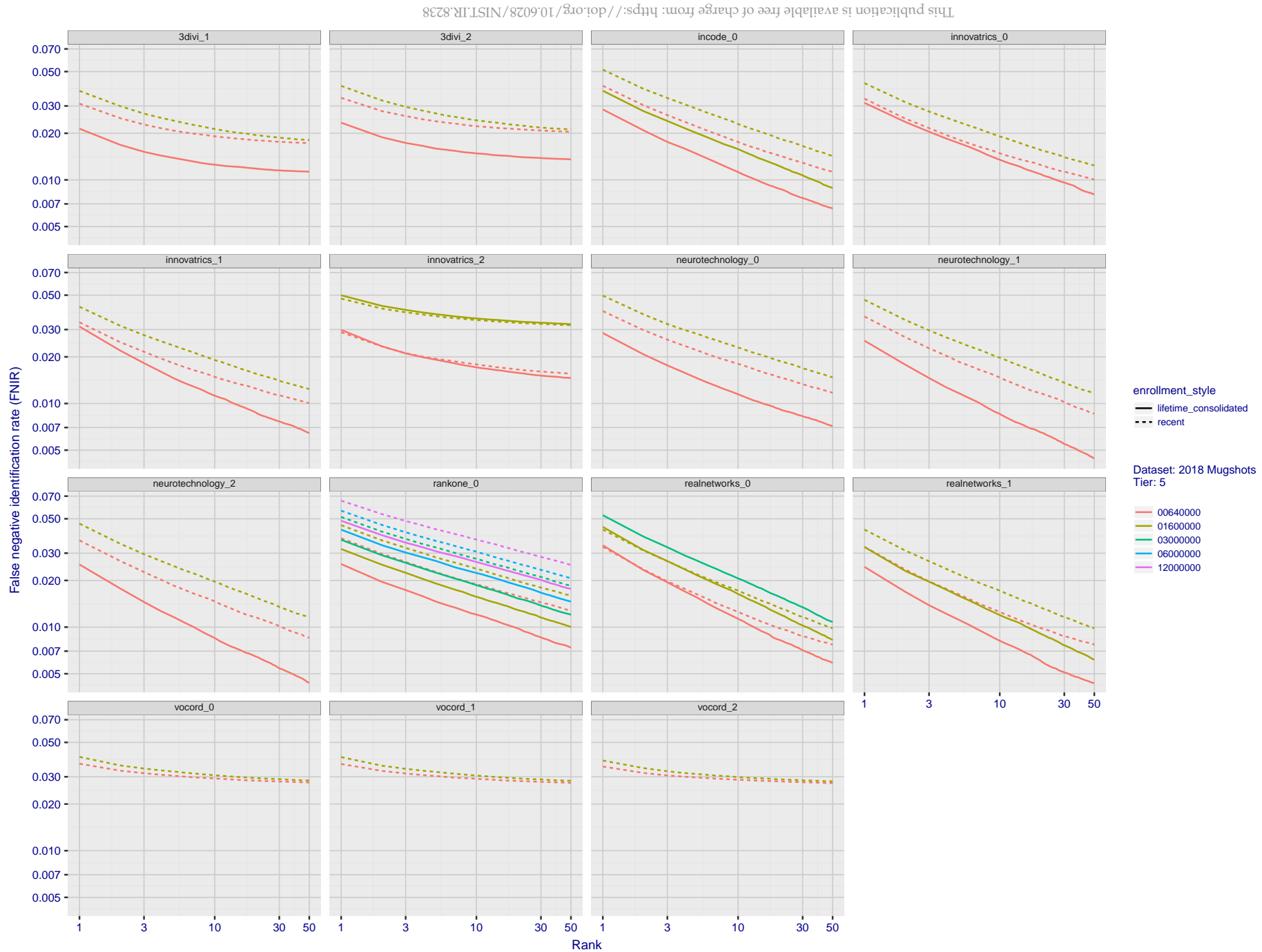


Figure 36: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank.** For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted 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.

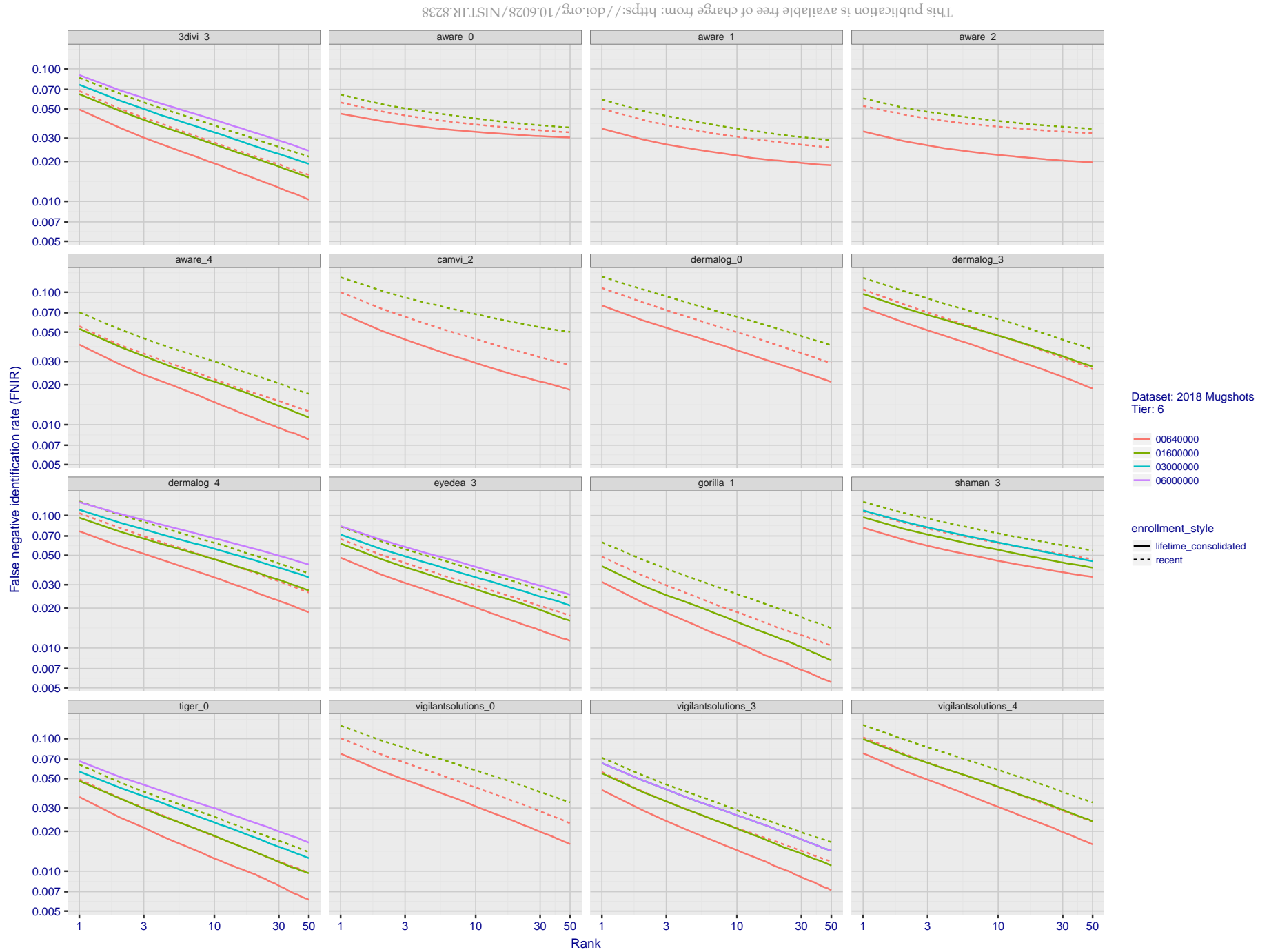


Figure 37: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank.** For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted 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.

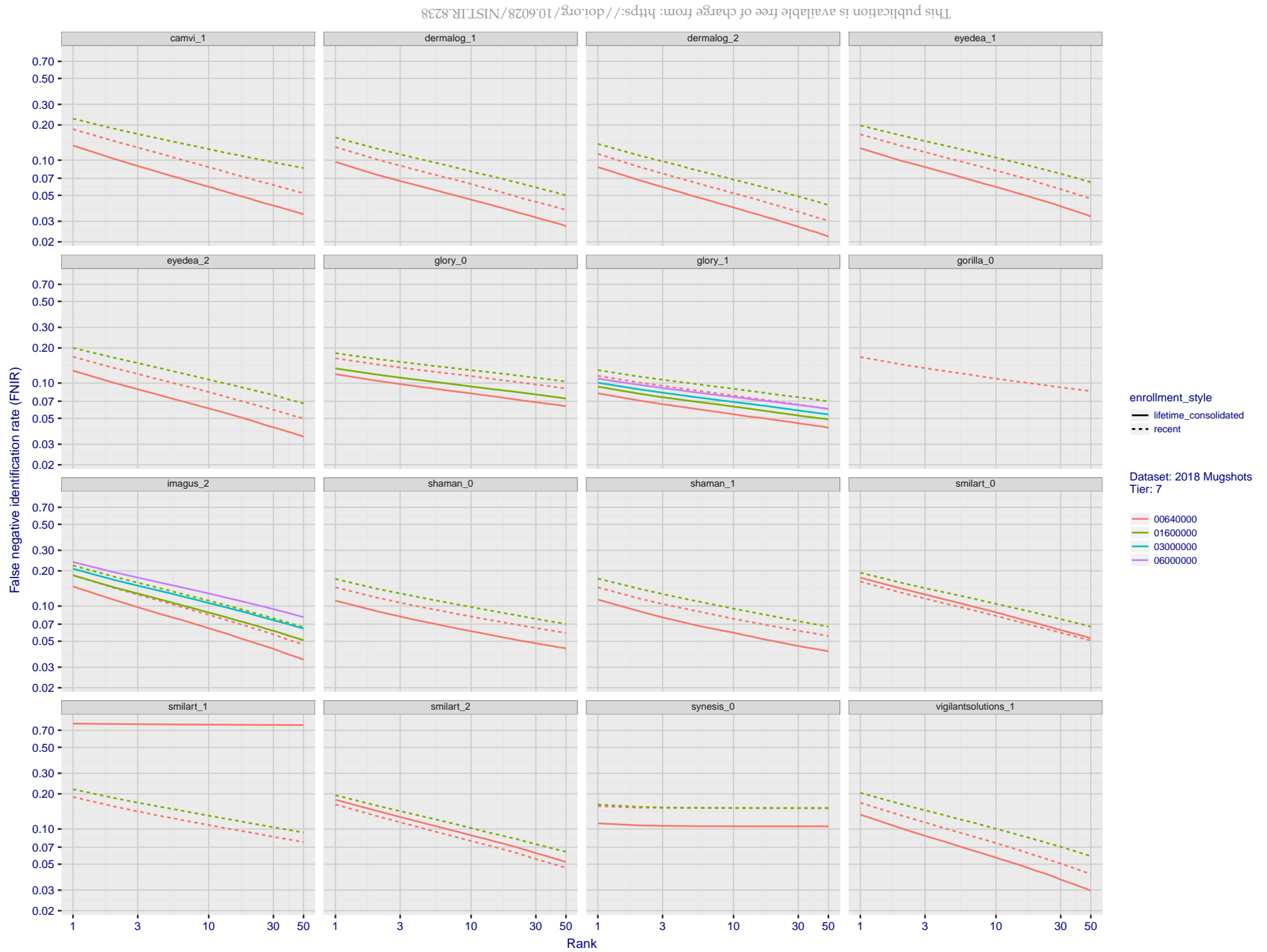


Figure 38: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank.** For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted 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.

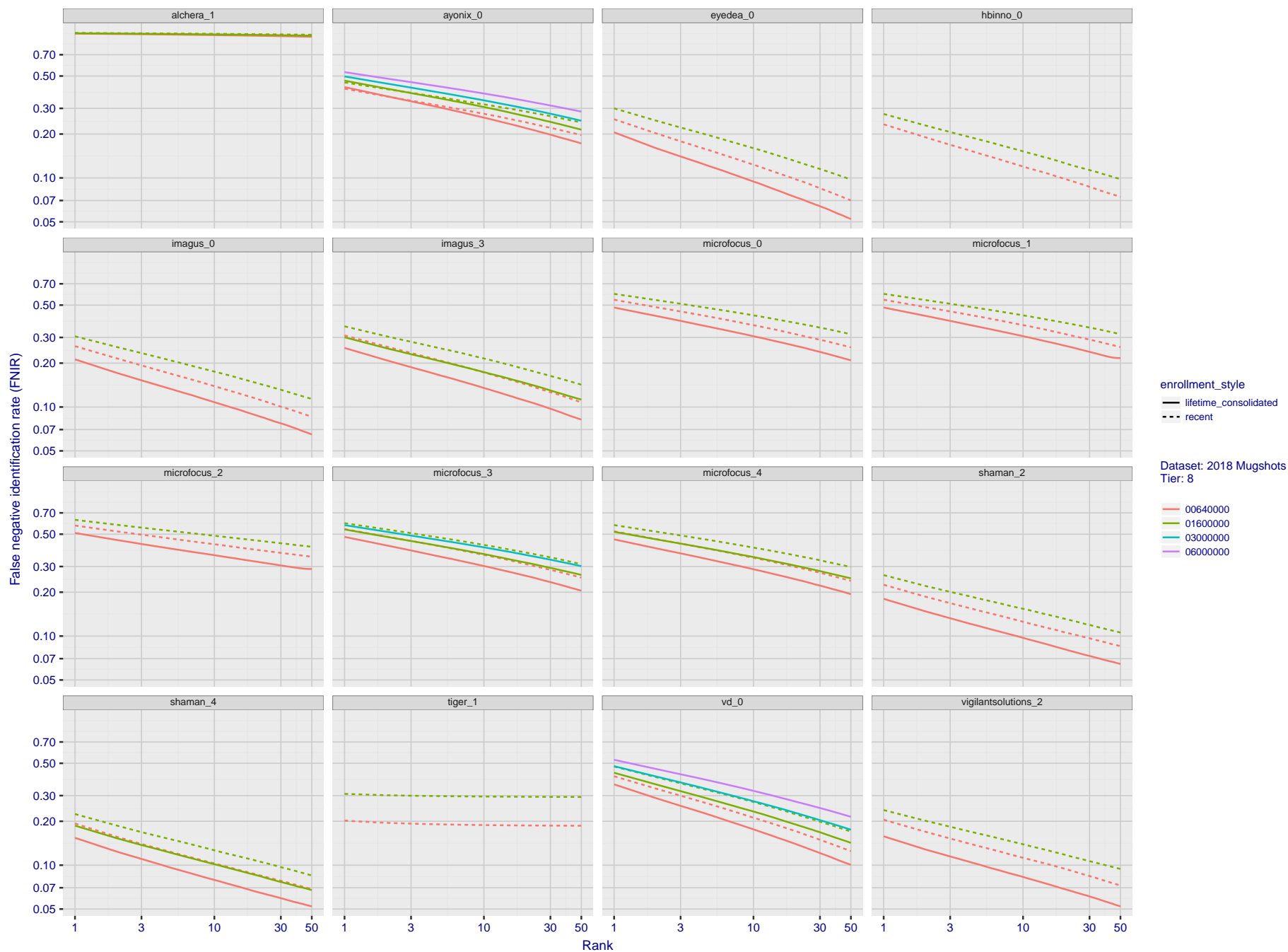


Figure 39: **[FRVT-2018 Mugshot Dataset] Rank-based identification miss rates vs. rank.** For the 2018 mugshots dataset, the figure shows false negative identification rates (FNIR) for ranks up to 50. This metric is appropriate to investigational applications where human reviewers will adjudicate sorted 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
07:24:51FNIR(N, R, T) =
FPIR(N, T) =False neg. identification rate
False pos. identification rateN = Num. enrolled subjects
R = Num. candidates examined

T = Threshold

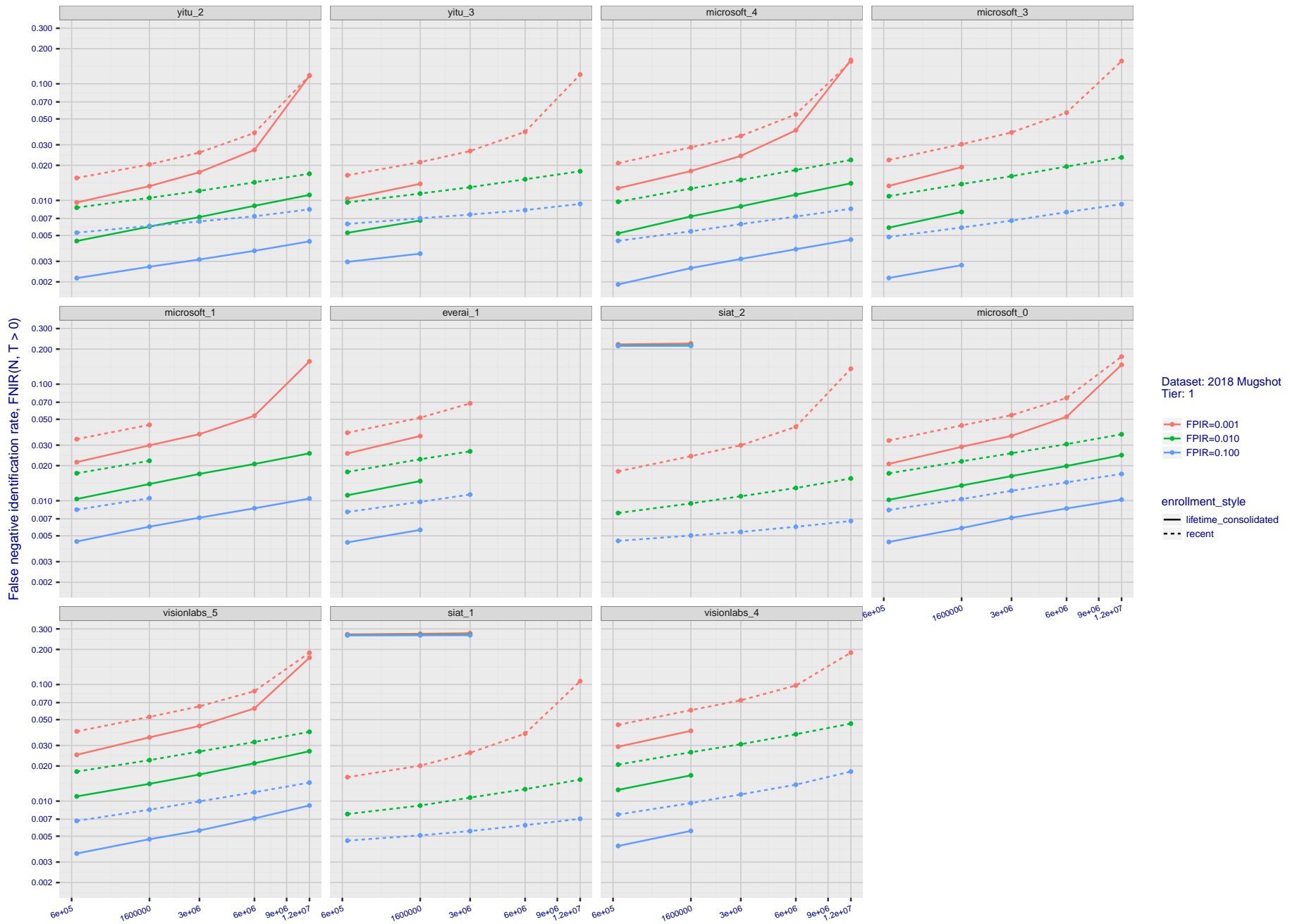
T = 0 → Investigation
T > 0 → Identification

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 $FNIR(N_b, T)$, $N_b = 640\,000$.

2018/11/26
07:24:51FNIR(N, R, T) =
FPIR(N, T) =False neg. identification rate
False pos. identification rateN = Num. enrolled subjects
R = Num. candidates examined

T = Threshold

T = 0 → Investigation
T > 0 → Identification

False negative identification rate, FNIR(N, T > 0)

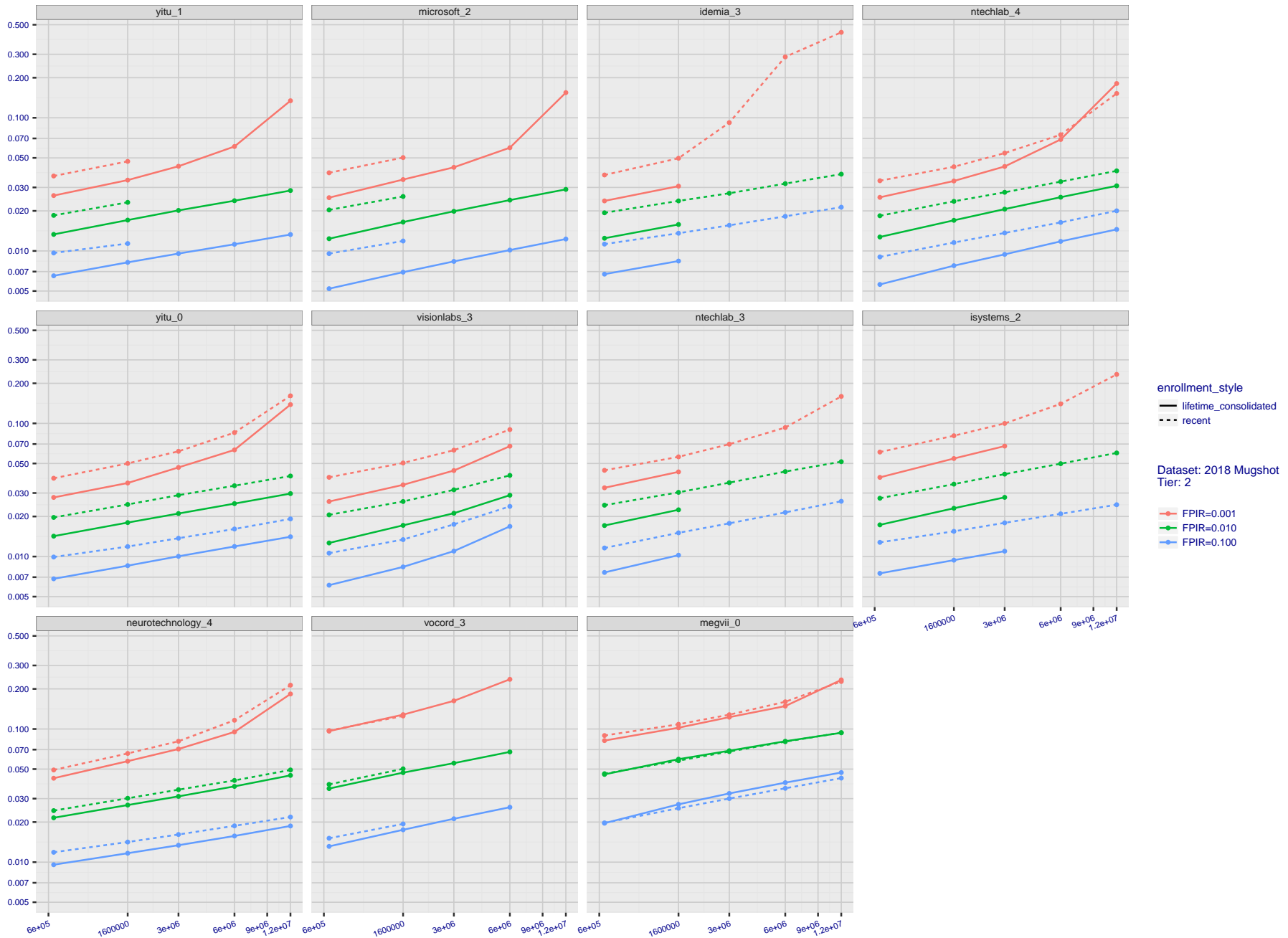
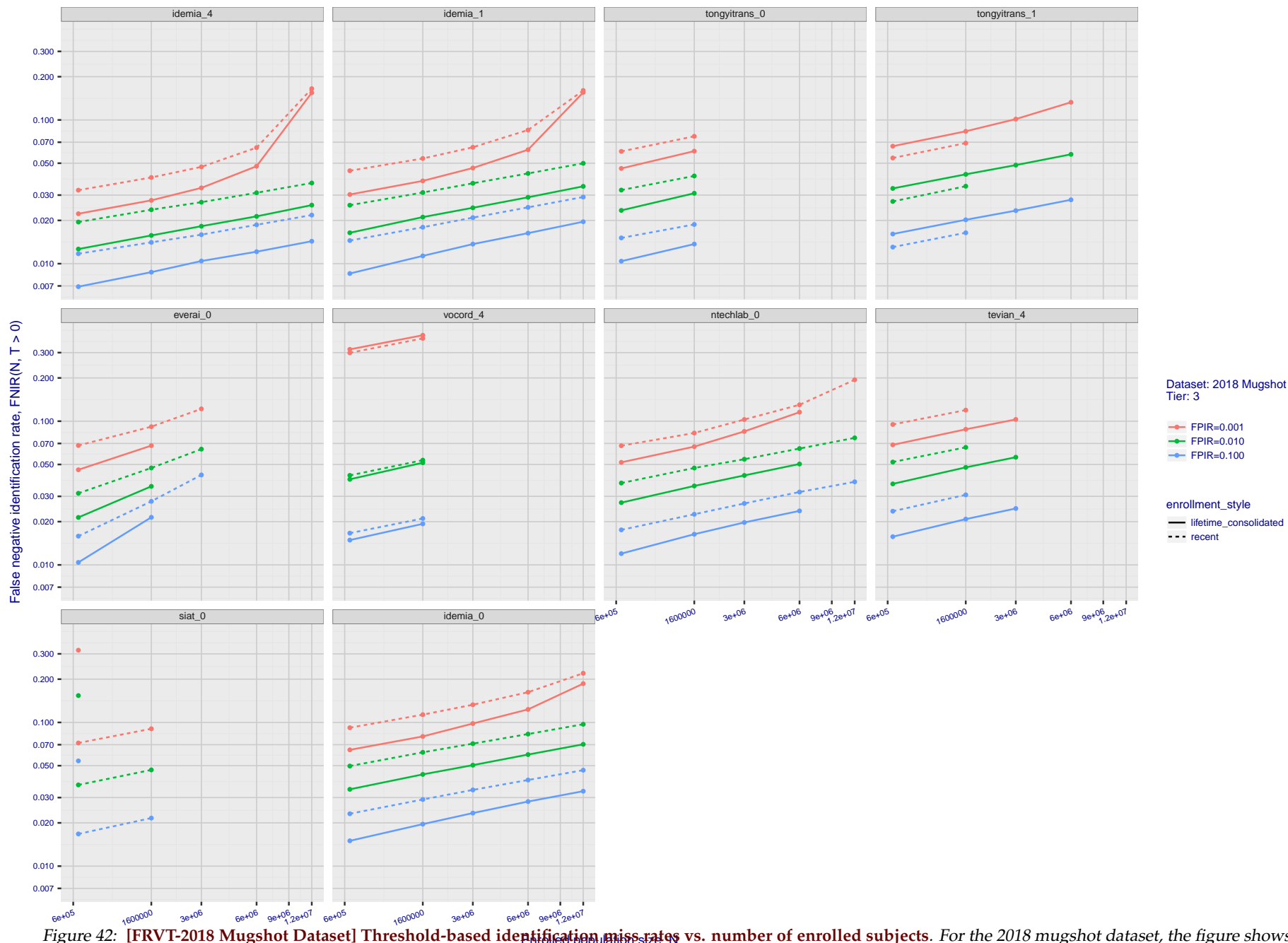


Figure 41: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects.** For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds 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.



2018/11/26
07:24:51FNIR(N, T) =
FPIR(N, T) =False neg. identification rate
False pos. identification rateN = Num. enrolled subjects
R = Num. candidates examined

T = Threshold

T = 0 → Investigation
T > 0 → Identification

False negative identification rate, FNIR(N, T > 0)

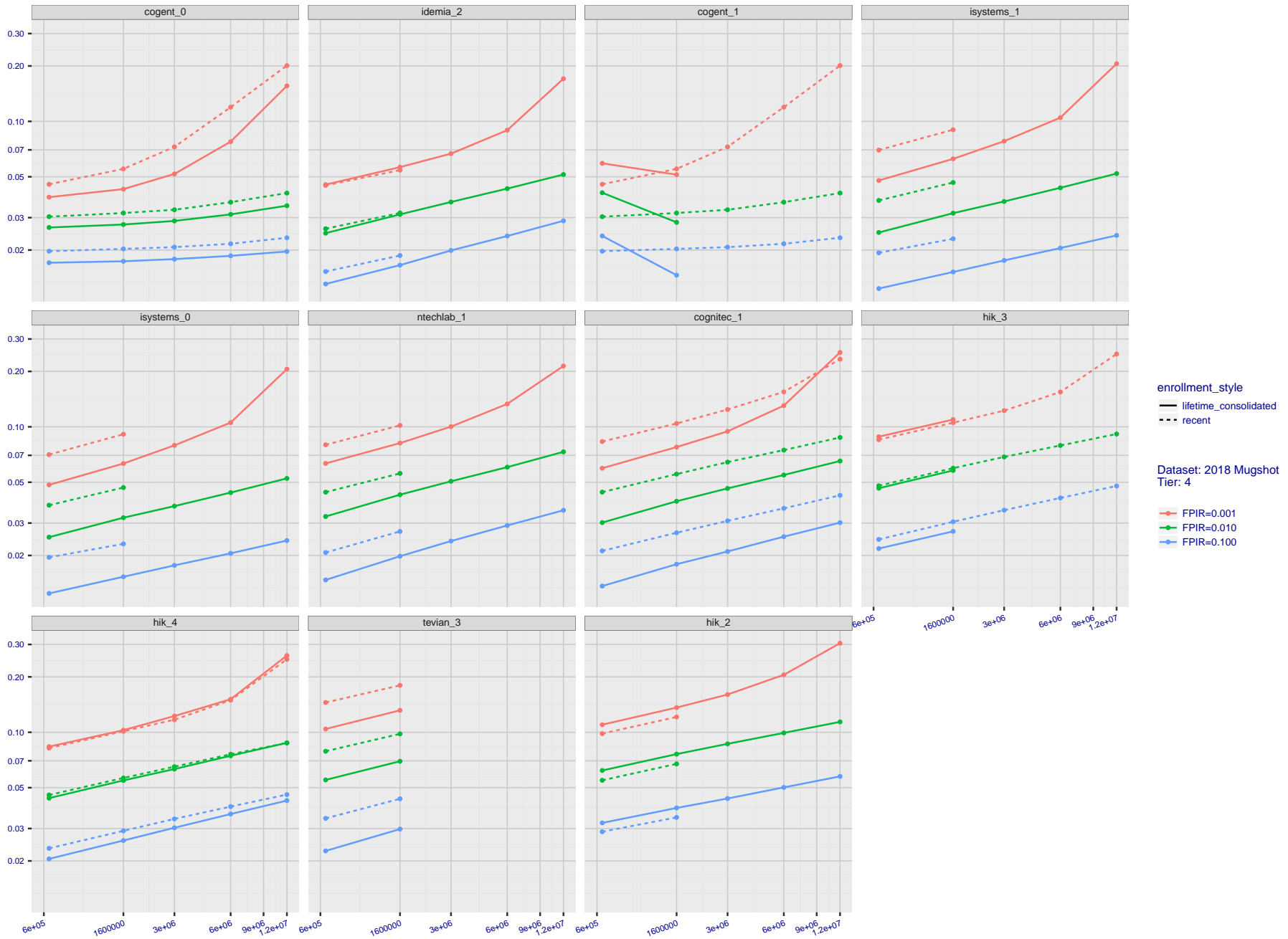


Figure 43: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects.** For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds 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.

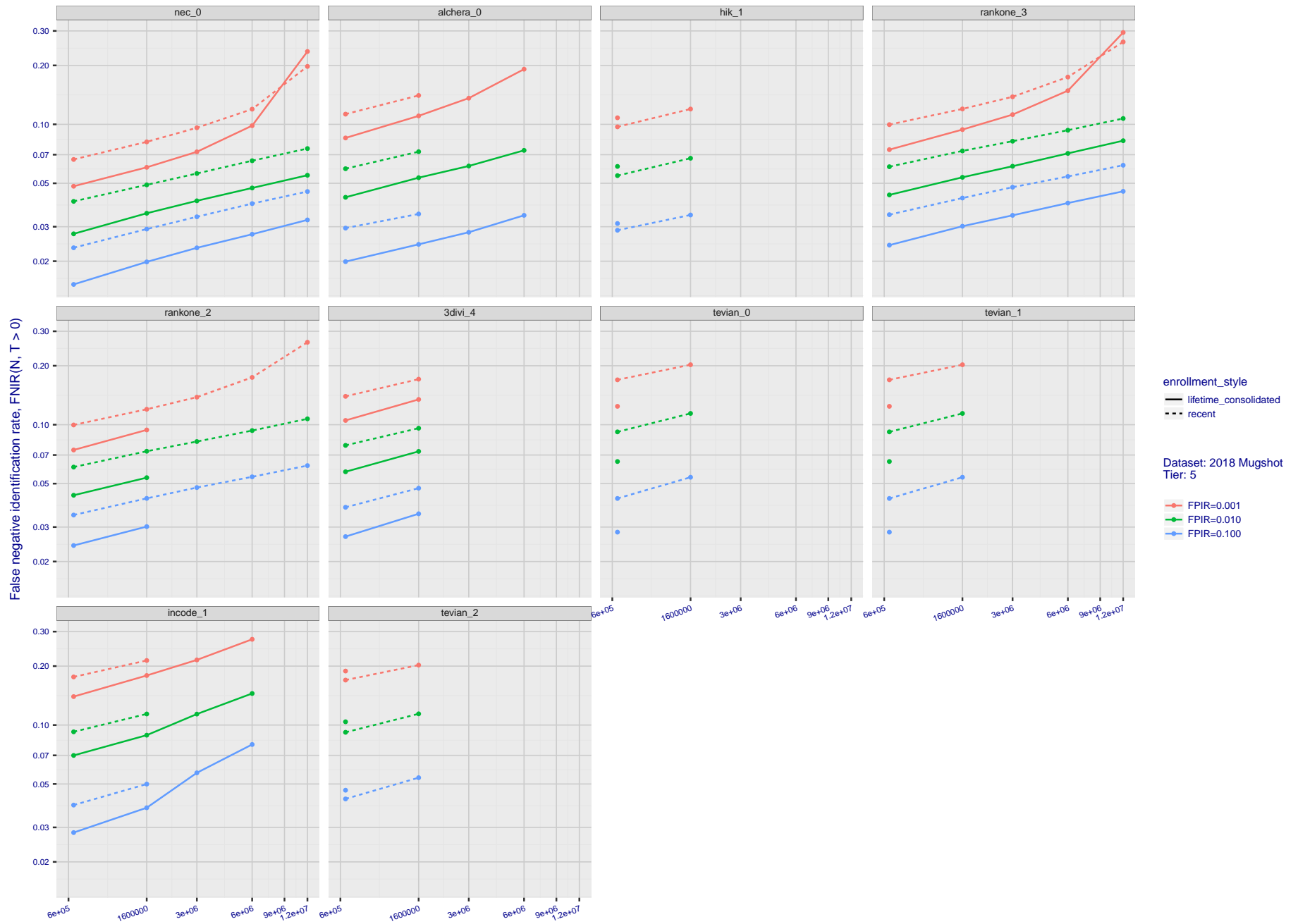


Figure 44: [FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects. For the 2018 mugshot dataset, the figure shows $FNIR(N, T)$ across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds 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 $FNIR(N_b, T)$, $N_b = 640\,000$.

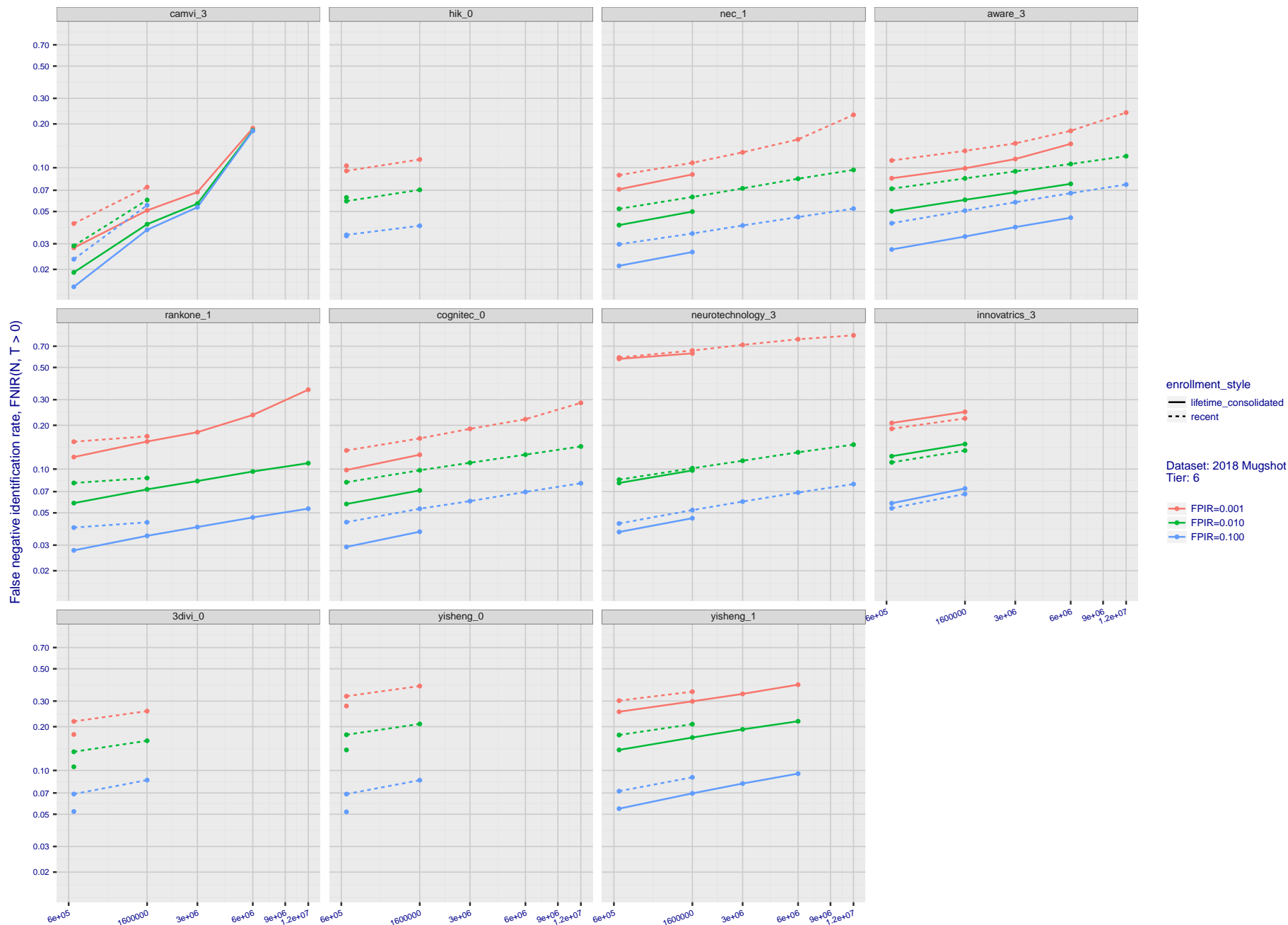


Figure 45: [FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects. For the 2018 mugshot dataset, the figure shows $FNIR(N, T)$ across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds 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 $FNIR(N_b, T)$, $N_b = 640\,000$.

False negative identification rate, FNIR(N, T > 0)

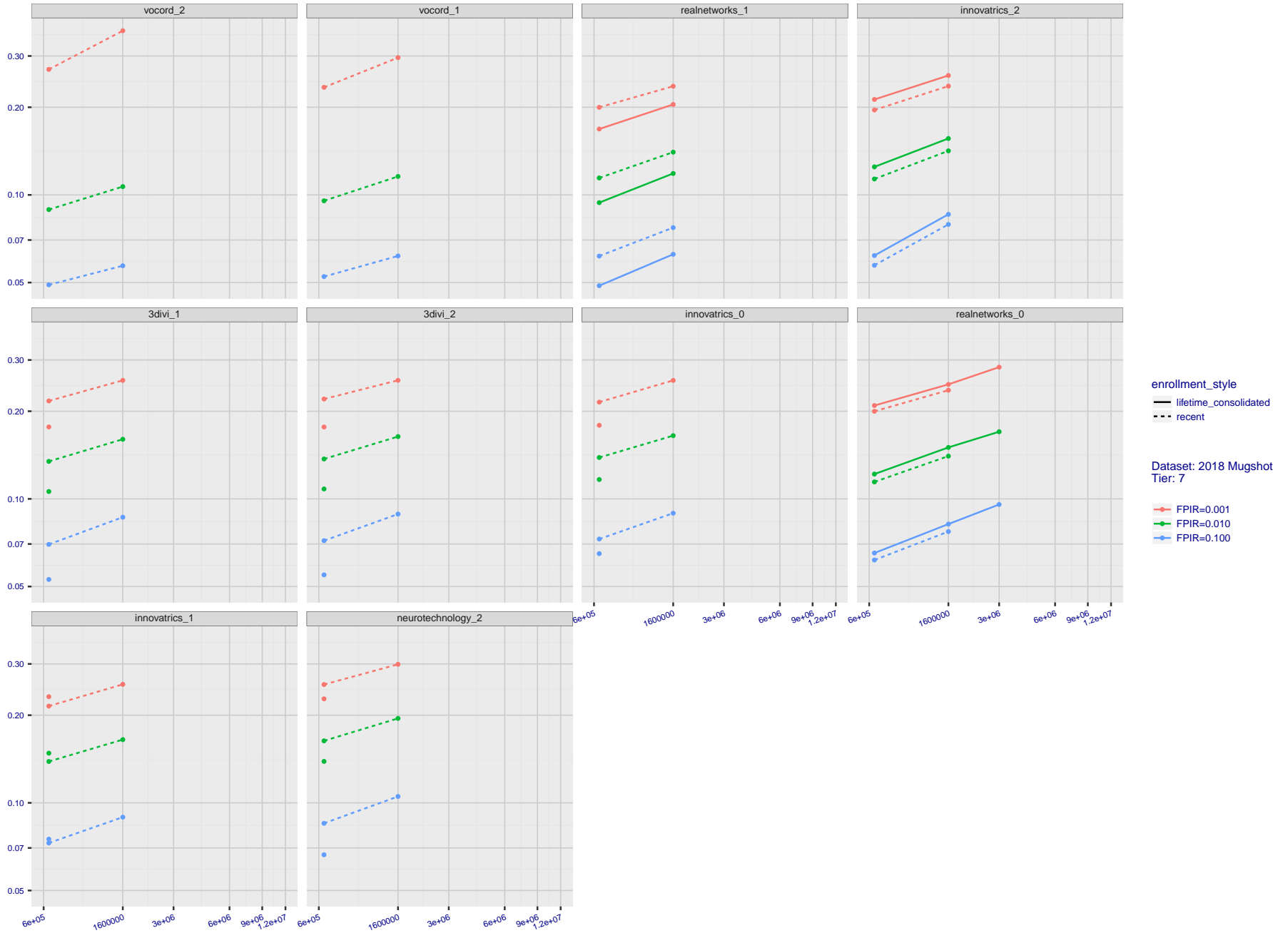


Figure 46: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects.** For the 2018 mugshot dataset, the figure shows FNIR(N, T) across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds 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.

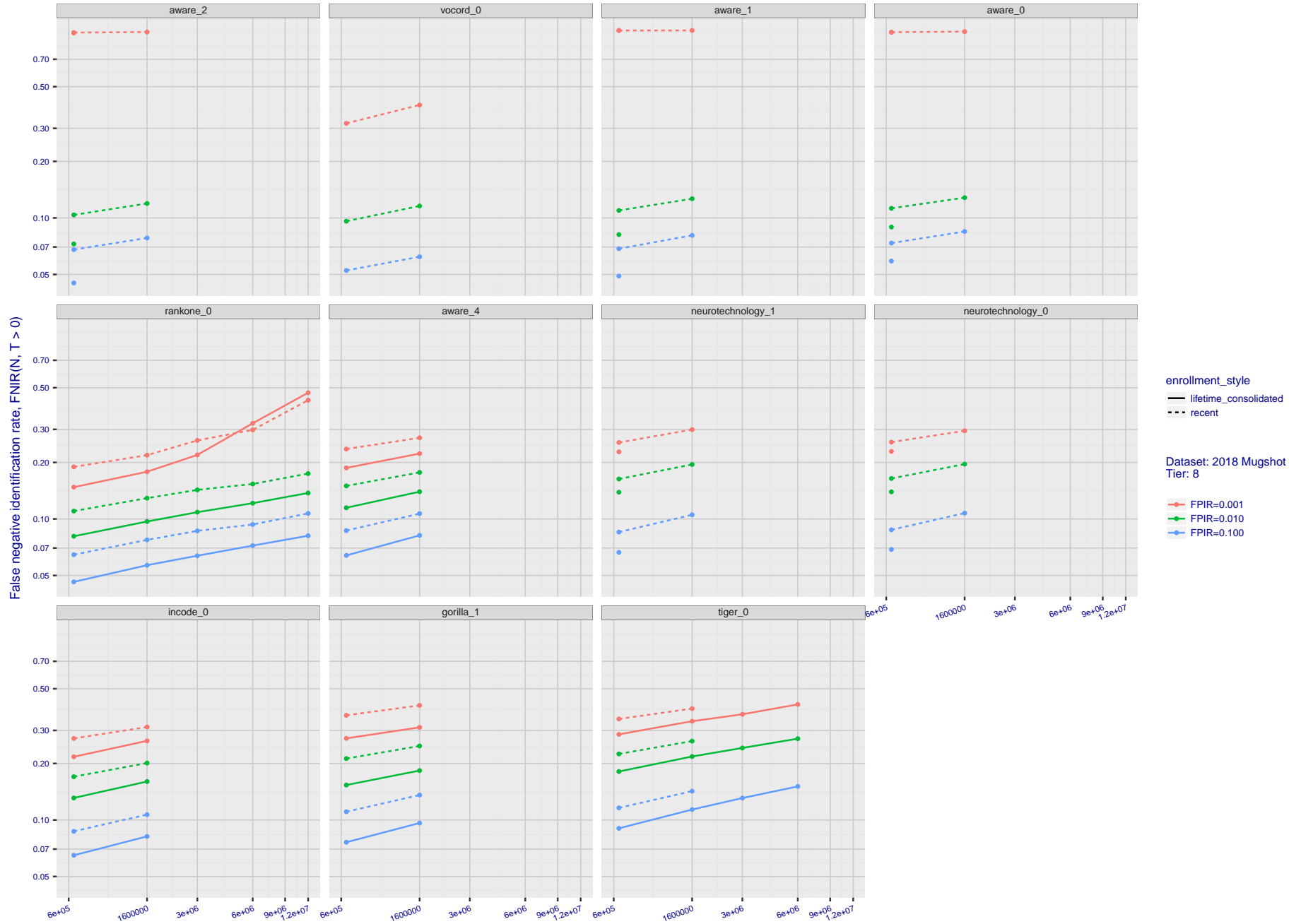


Figure 47: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects.** For the 2018 mugshot dataset, the figure shows $FNIR(N, T)$ across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high 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 $FNIR(N_b, T)$, $N_b = 640\,000$.

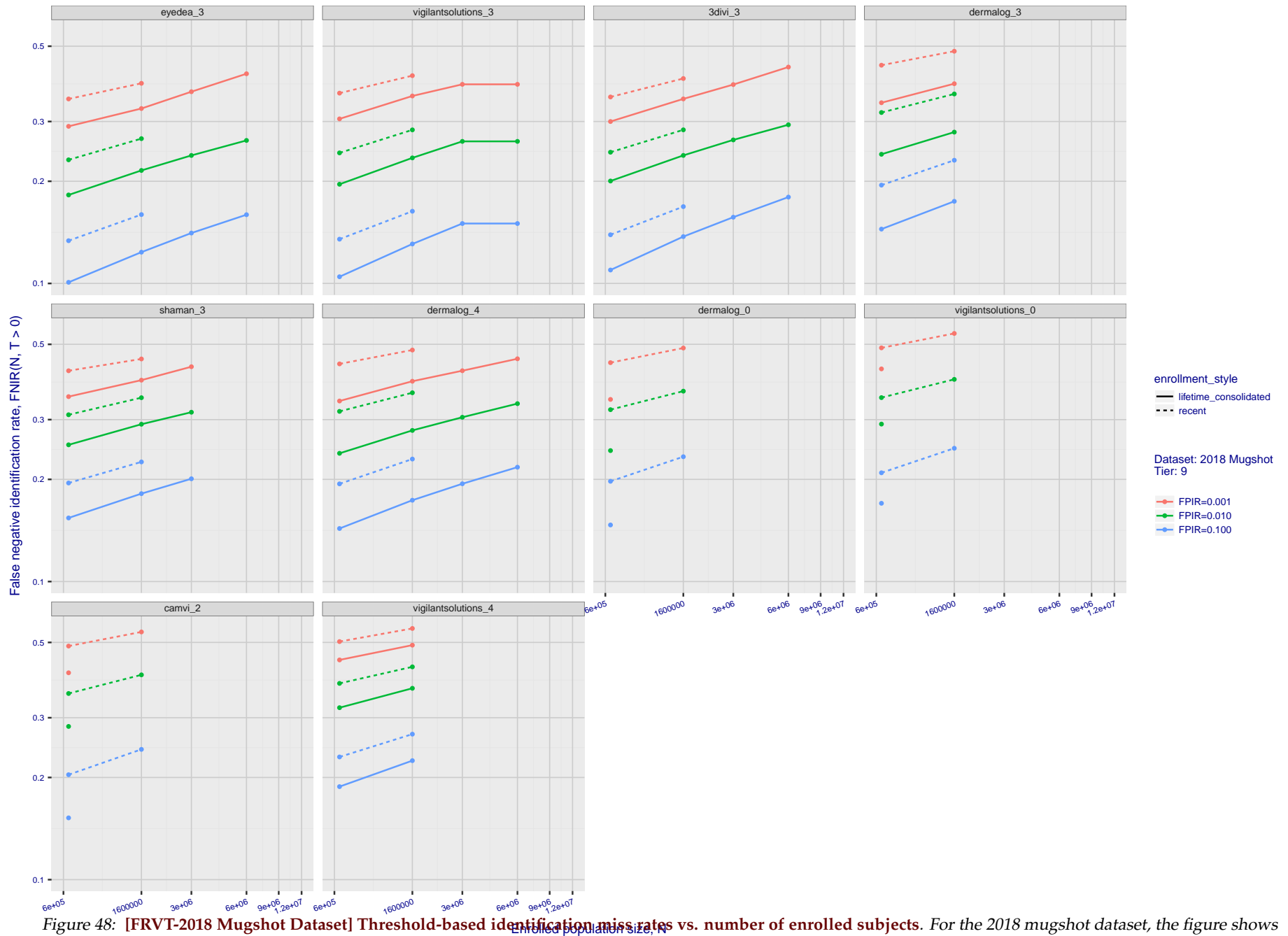
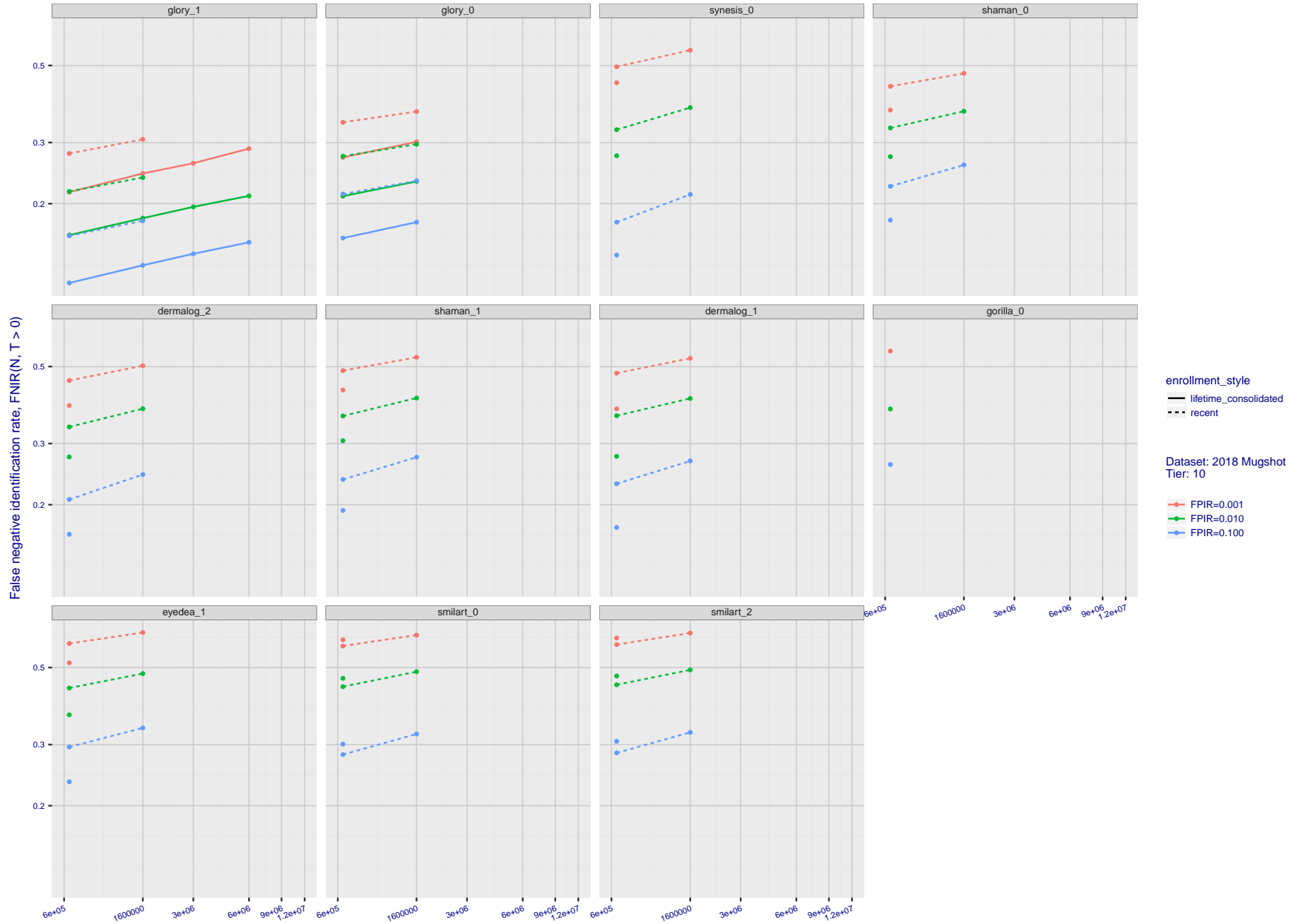
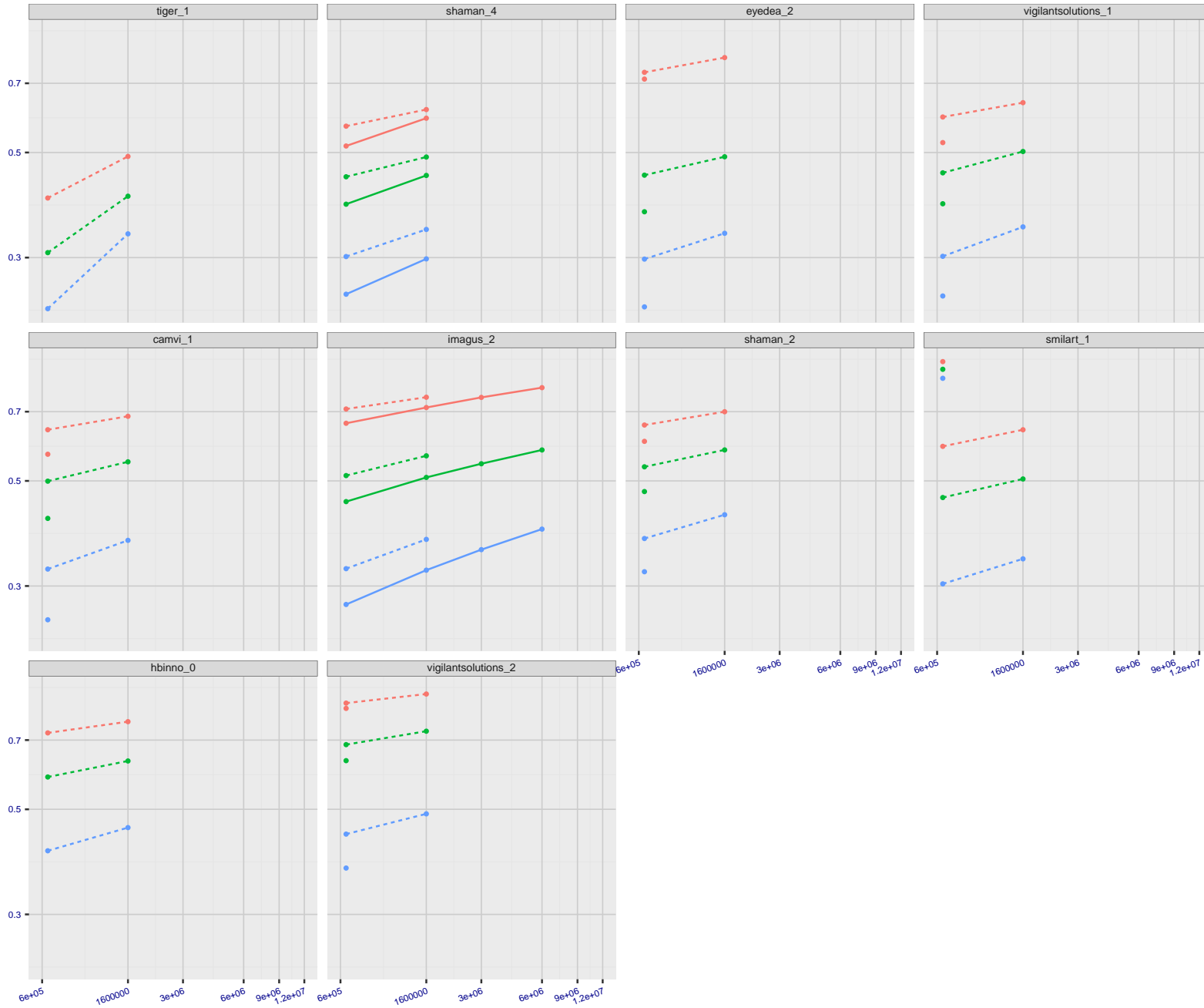


Figure 48: [FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects. For the 2018 mugshot dataset, the figure shows $FNIR(N, T)$ across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high thresholds 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 $FNIR(N_b, T)$, $N_b = 640\,000$.



False negative identification rate, FNIR(N, T > 0)



enrollment_style
— lifetime consolidated
--- recent

Dataset: 2018 Mugshot
Tier: 11

— FPIR=0.001
— FPIR=0.010
— FPIR=0.100

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 $FNIR(N_b, T)$, $N_b = 640\,000$.

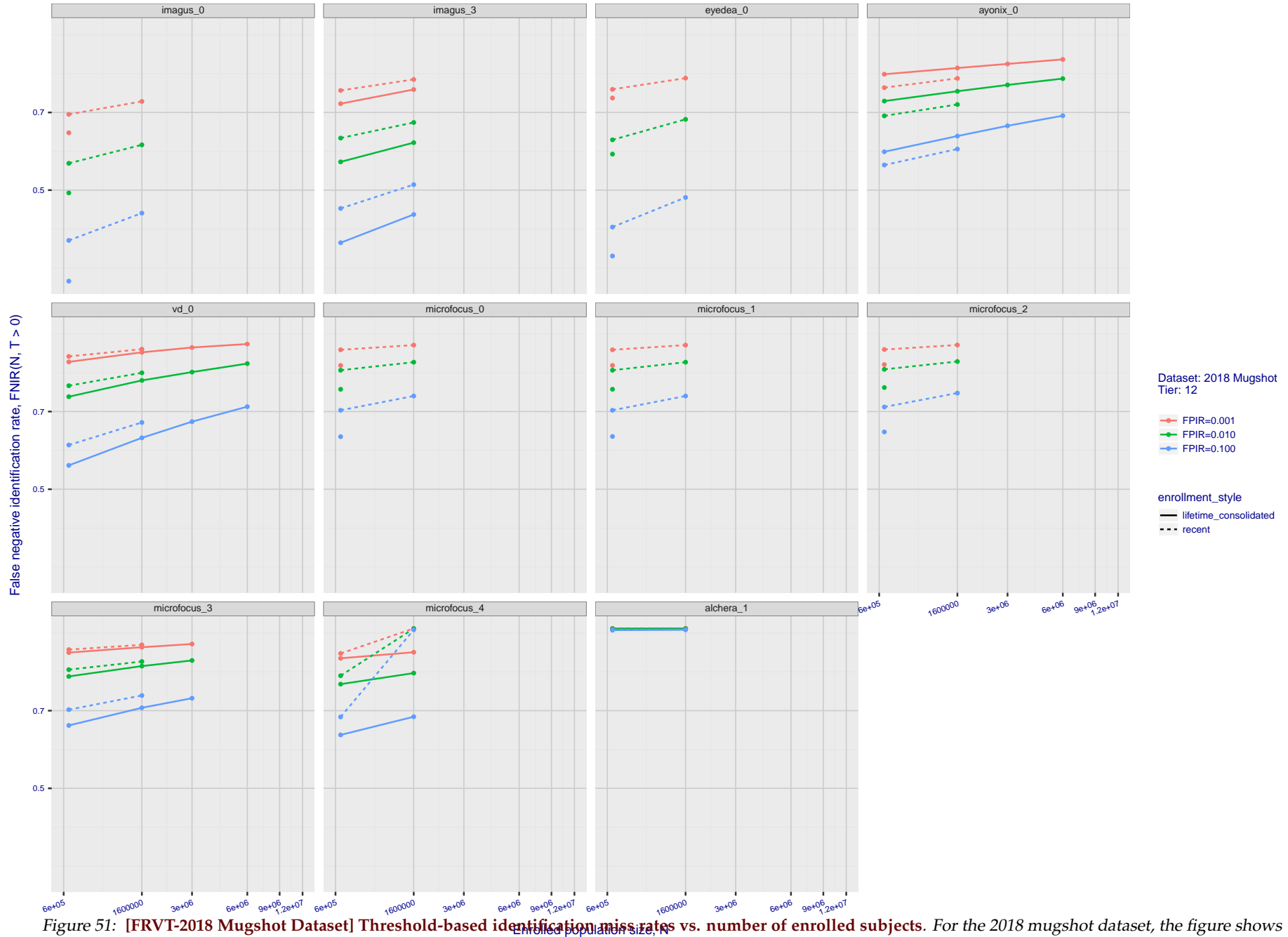


Figure 51: **[FRVT-2018 Mugshot Dataset] Threshold-based identification miss rates vs. number of enrolled subjects.** For the 2018 mugshot dataset, the figure shows $FNIR(N, T)$ across various gallery sizes when the threshold is set to achieve the given FPIRs. The rank criterion is irrelevant at high 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 $FNIR(N_b, T)$, $N_b = 640\,000$.

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

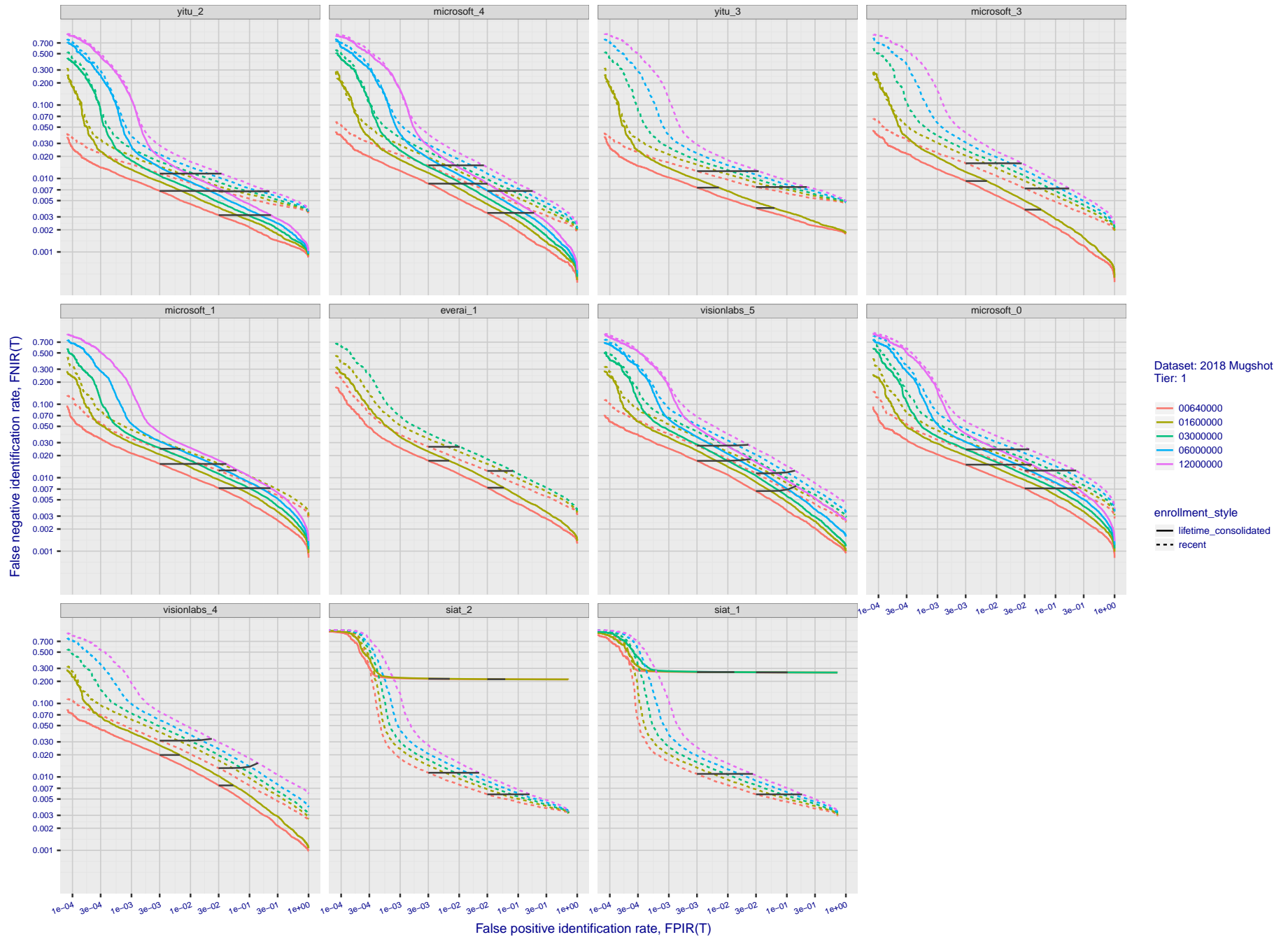


Figure 52: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates.** The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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 .

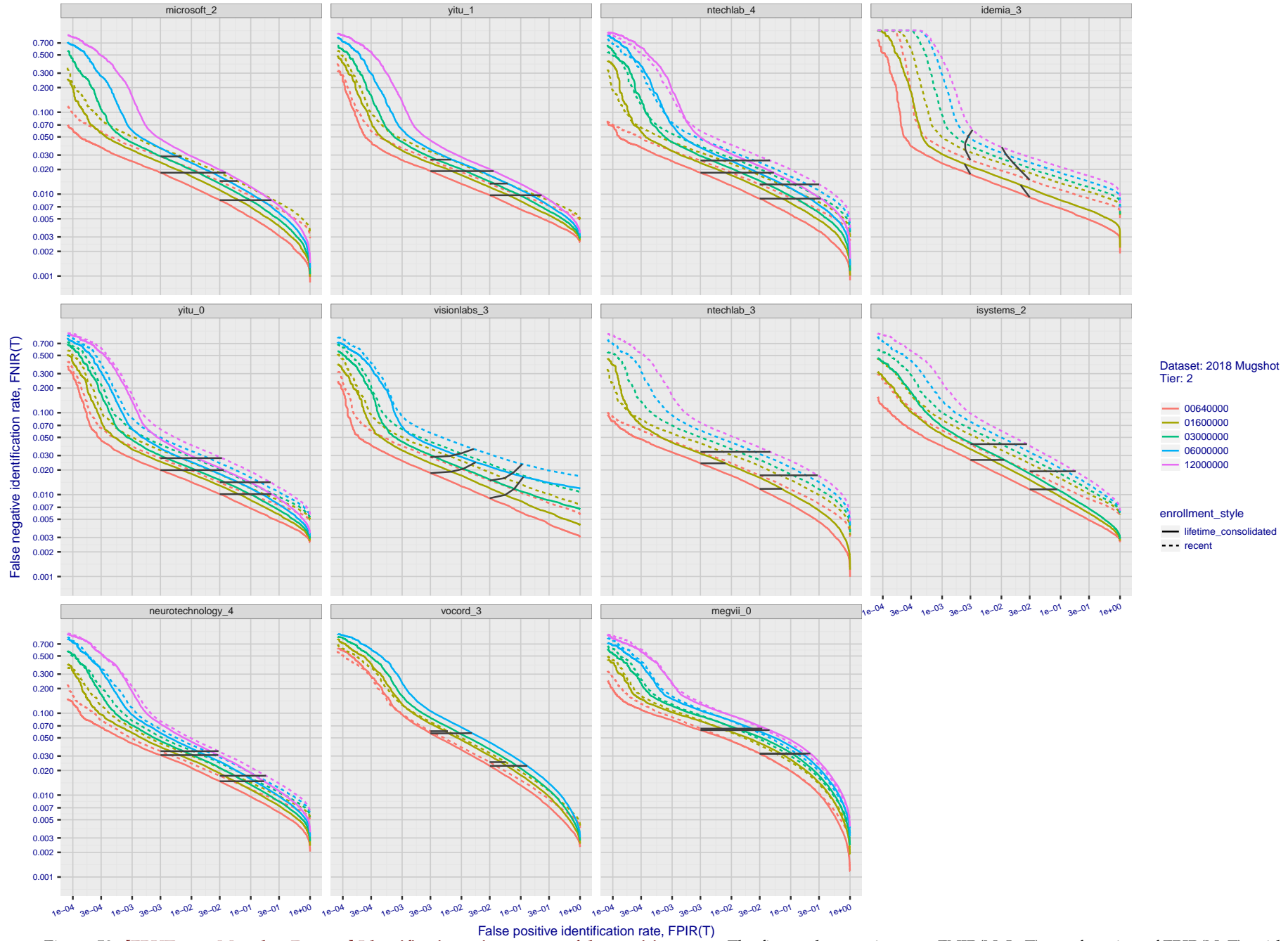
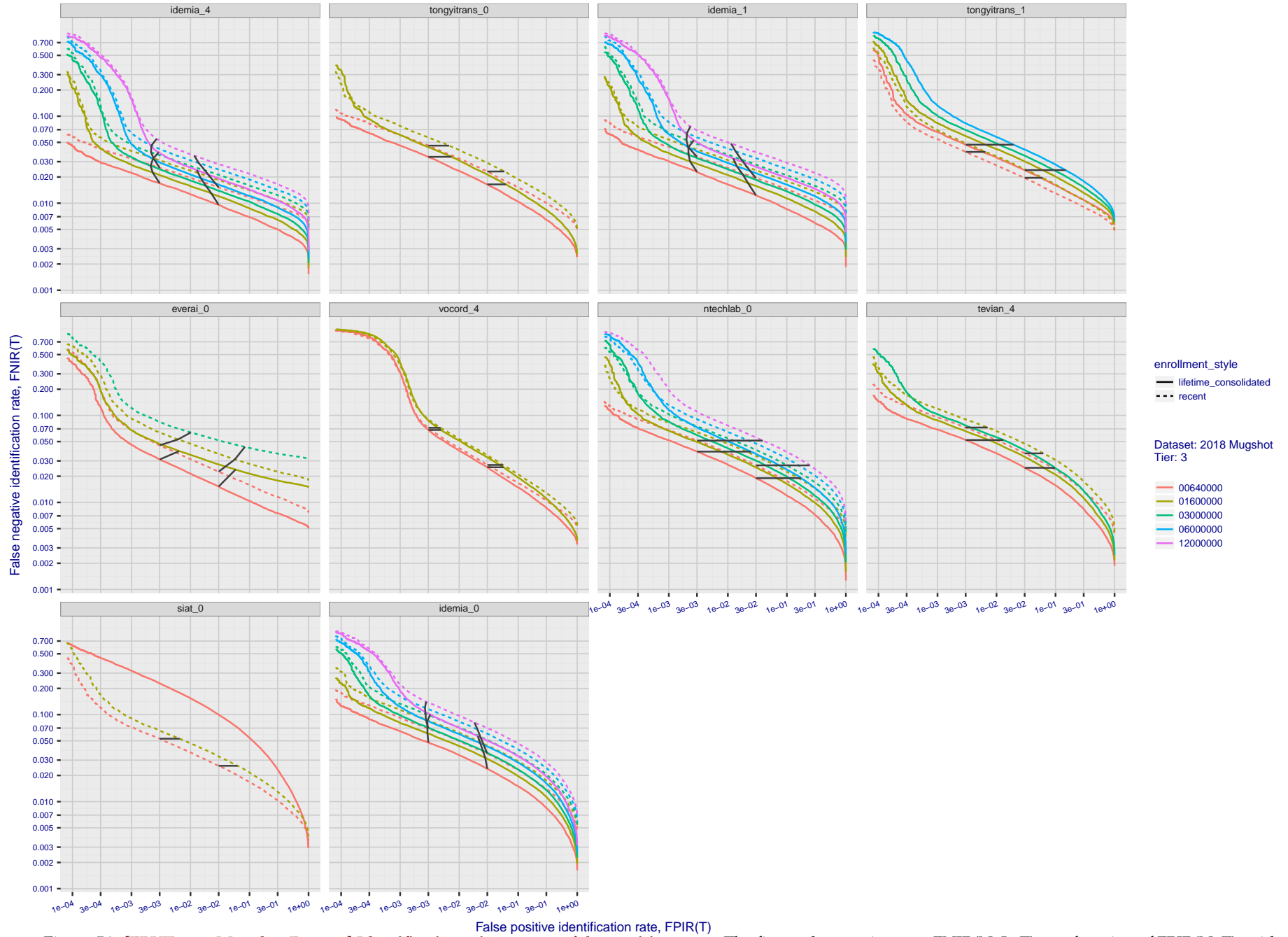


Figure 53: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates.** The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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 .



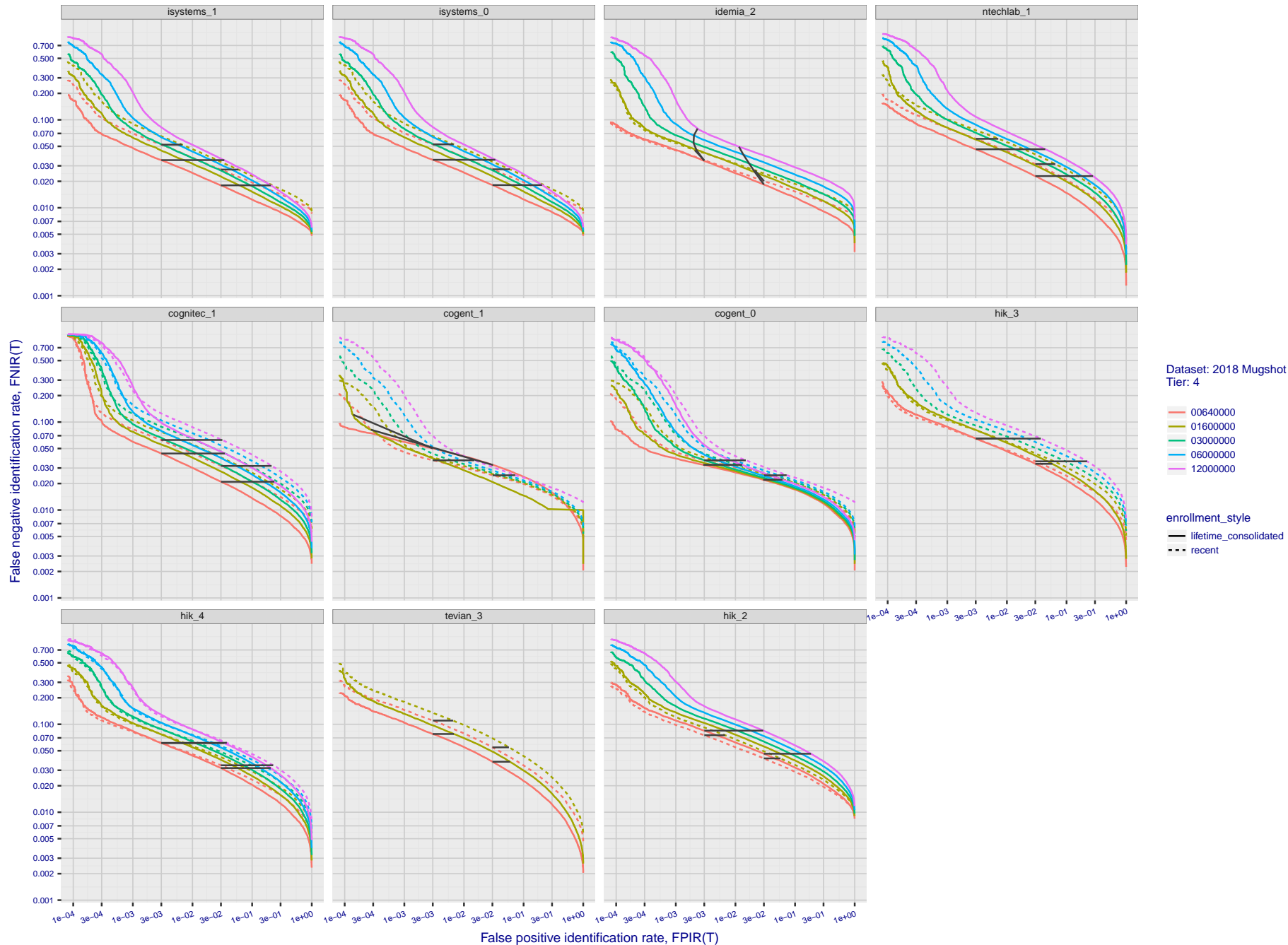


Figure 55: [FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates. The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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 .

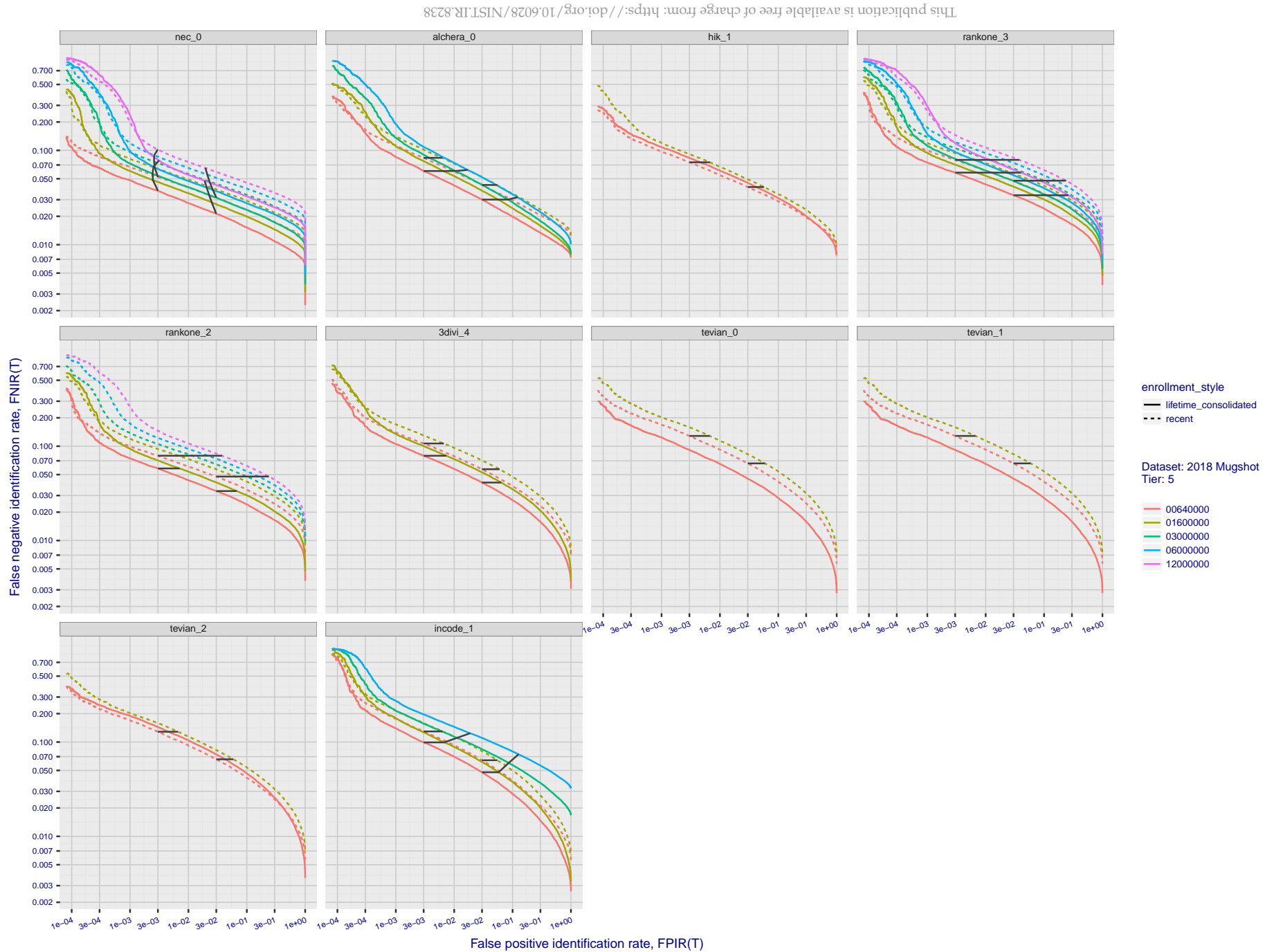


Figure 56: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates.** The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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 .

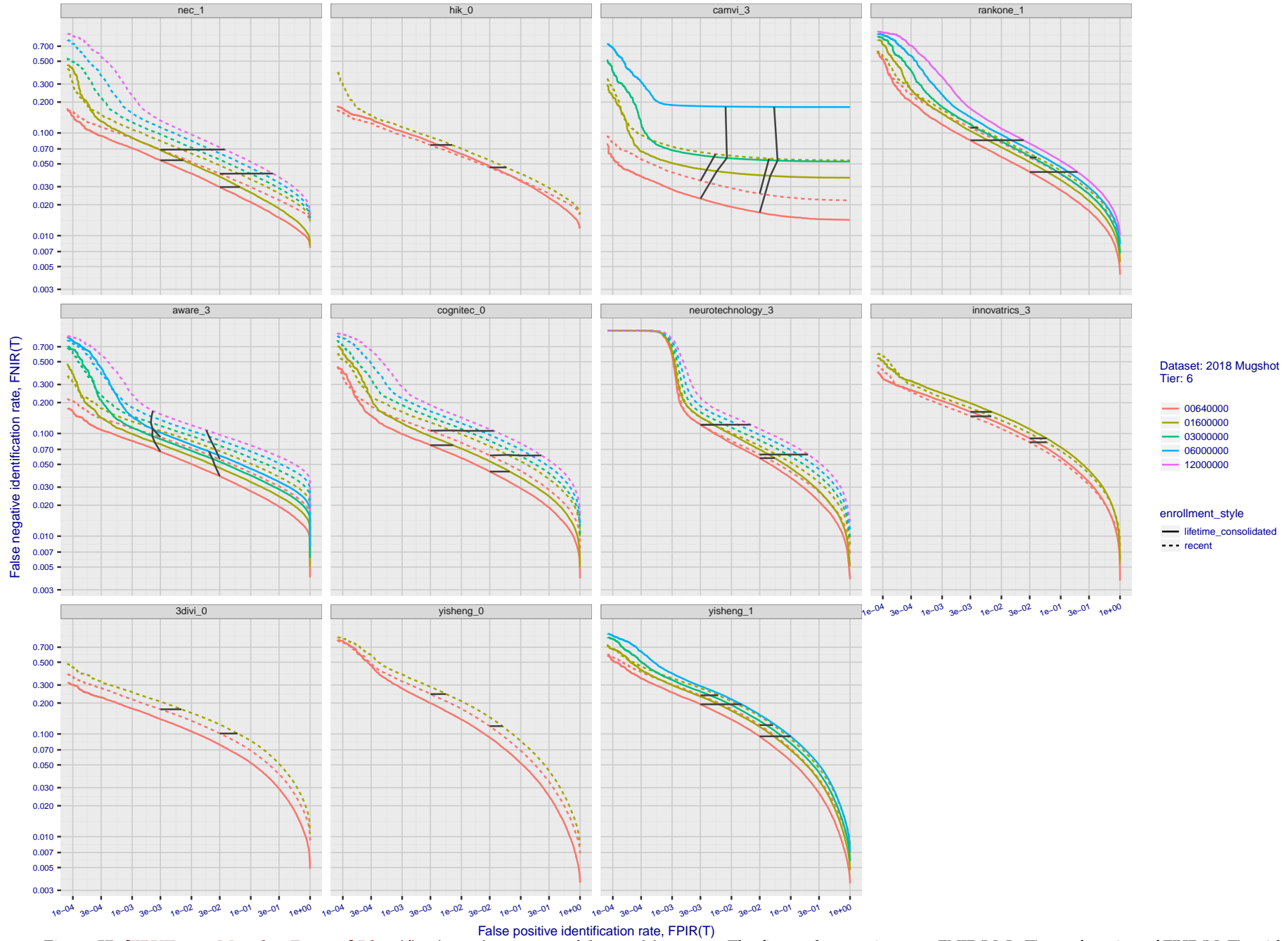


Figure 57: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates.** The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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 .

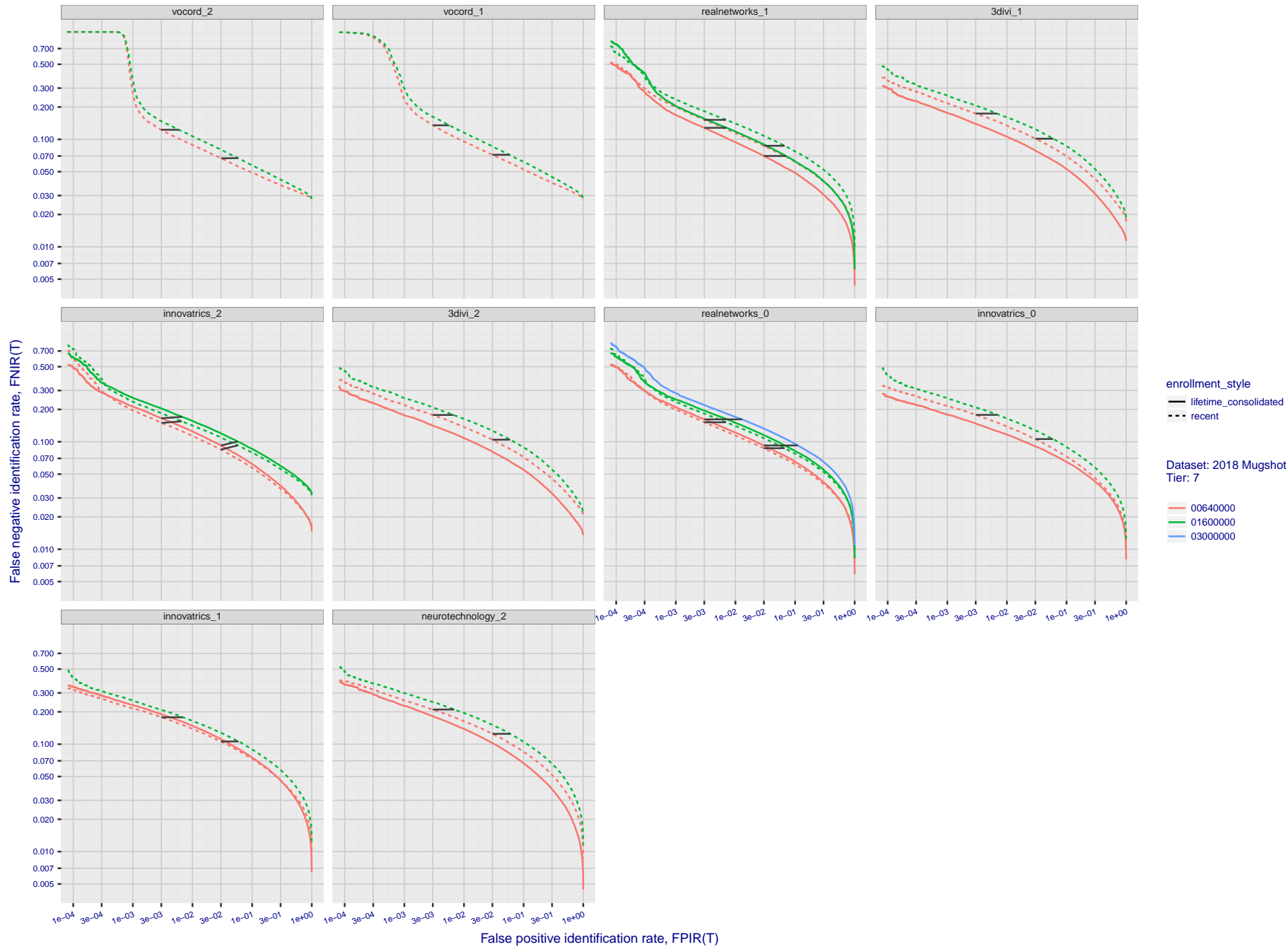


Figure 58: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates.** The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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 .

False negative identification rate, FNIR(T)

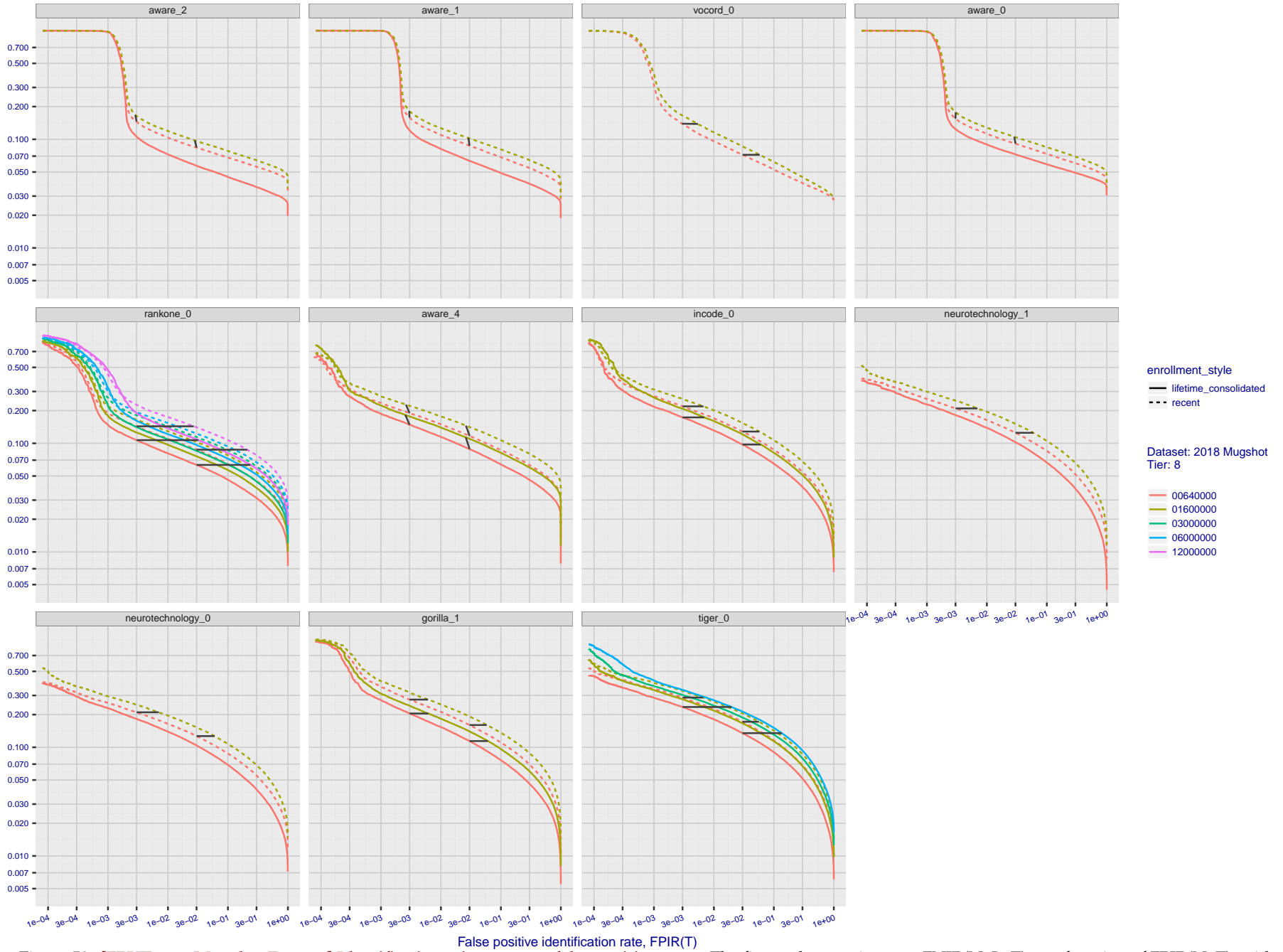


Figure 59: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates.** The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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 .

False negative identification rate, FNIR(T)

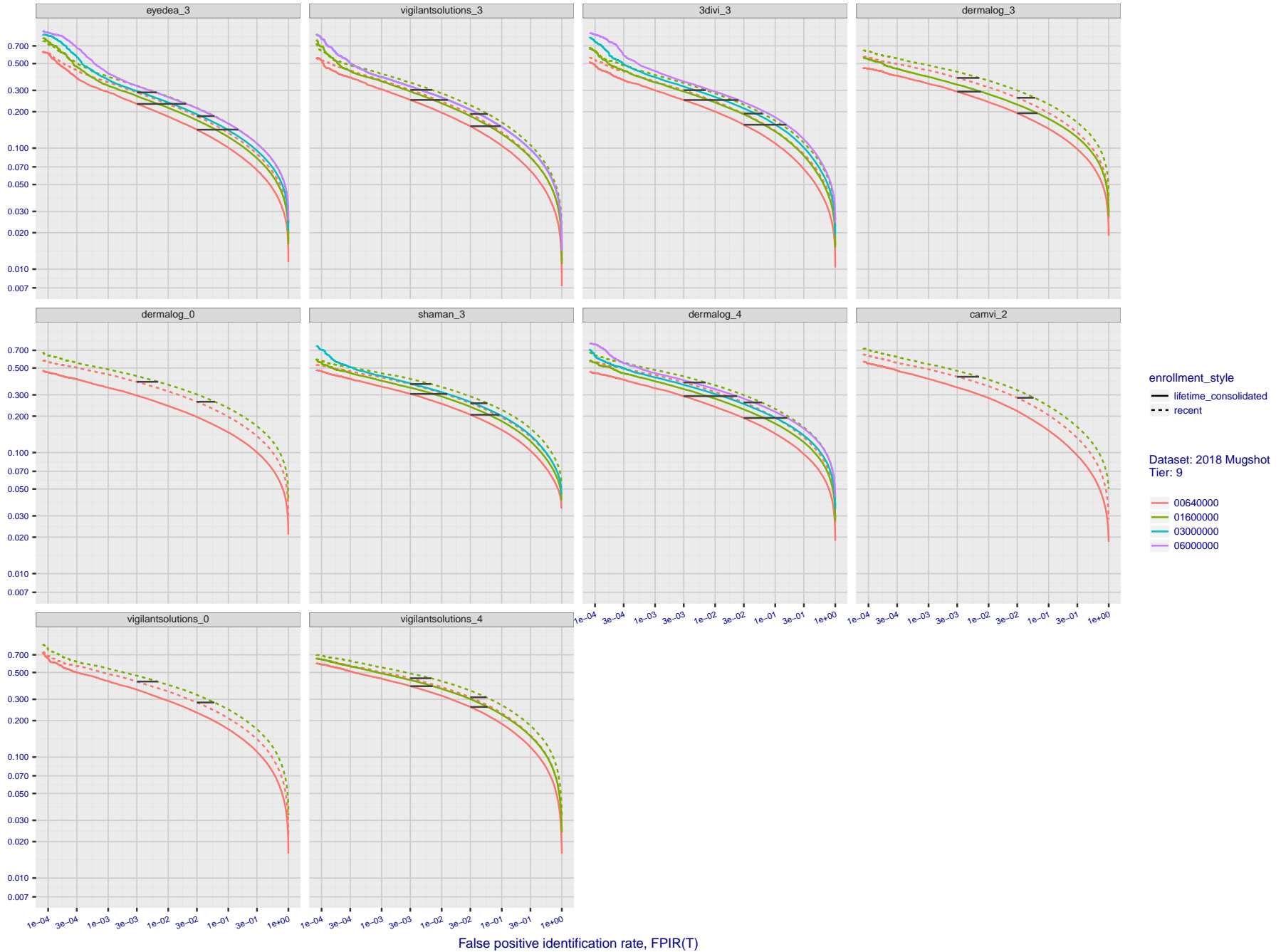


Figure 60: [FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates. The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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 .

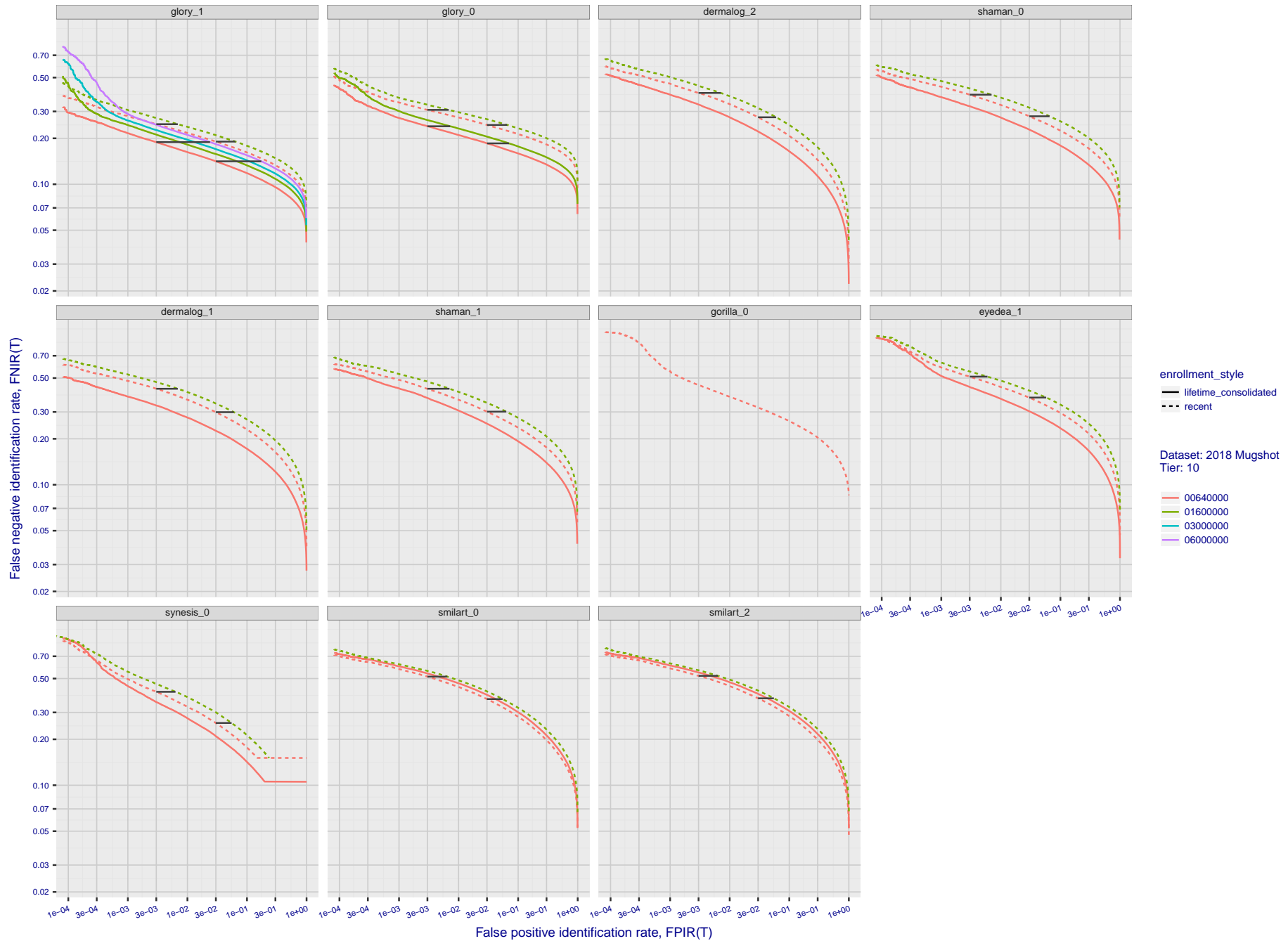


Figure 61: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates.** The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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 .

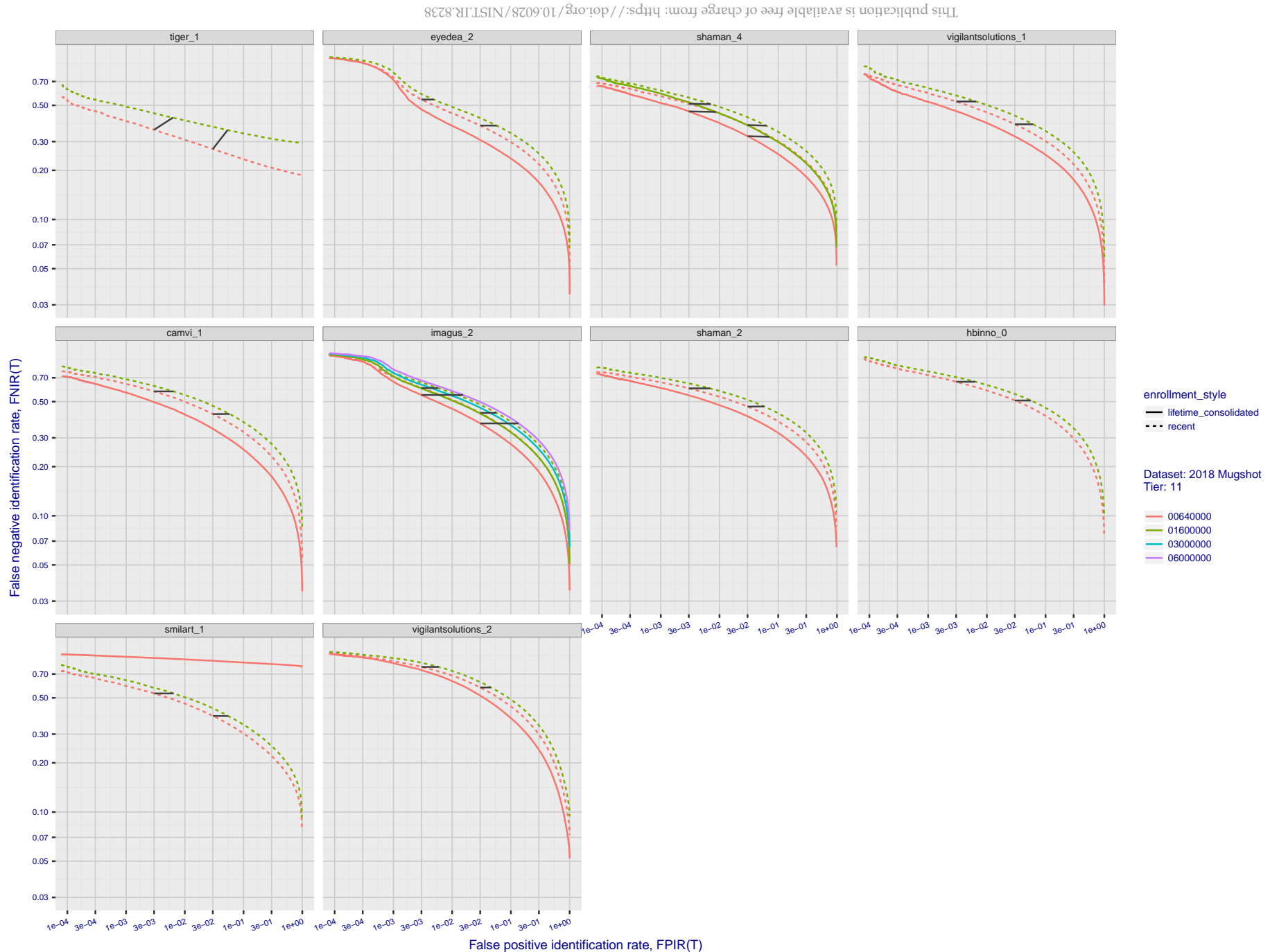


Figure 62: [FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates. The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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 .

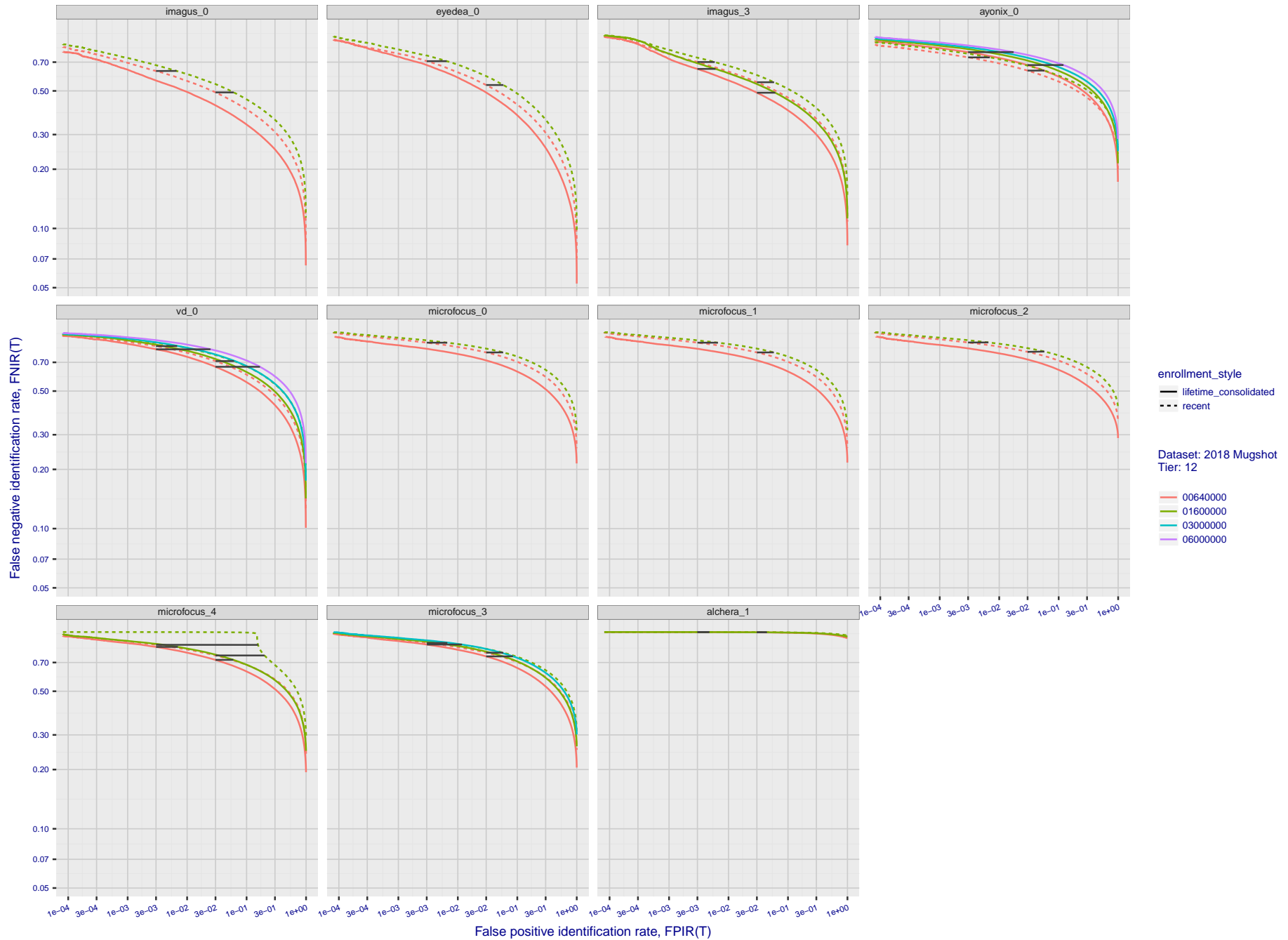


Figure 63: **[FRVT-2018 Mugshot Dataset] Identification miss rates vs. false positive rates.** The figure shows miss rates $FNIR(N, L, T)$ as a function of $FPIR(N, T)$, with N ranging from 640 000 to 12 000 000 as noted in rows 1-10 of Table 6. These error tradeoff characteristics are useful for applications 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

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

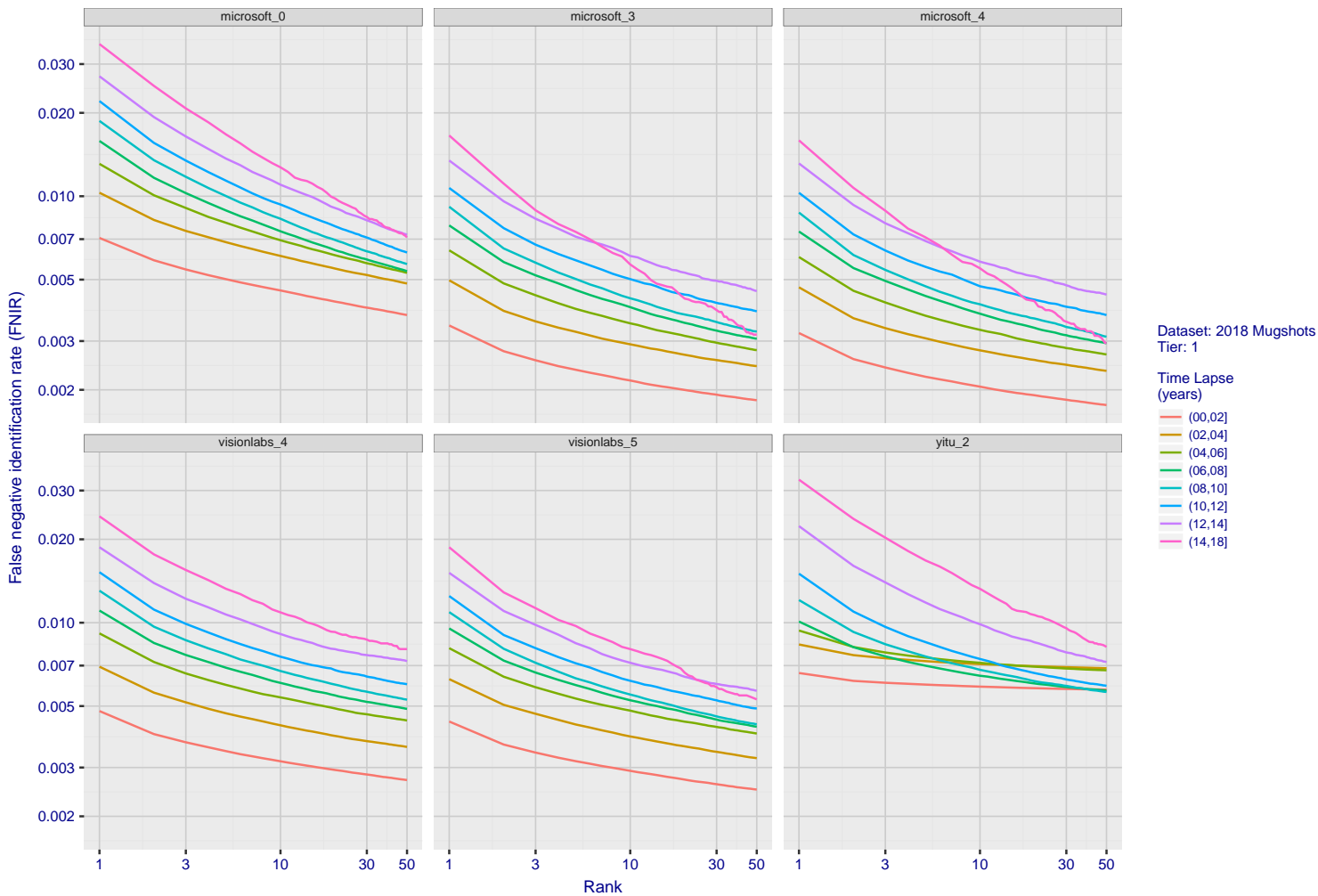


Figure 64: [FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. rank by time-elapsed. The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between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

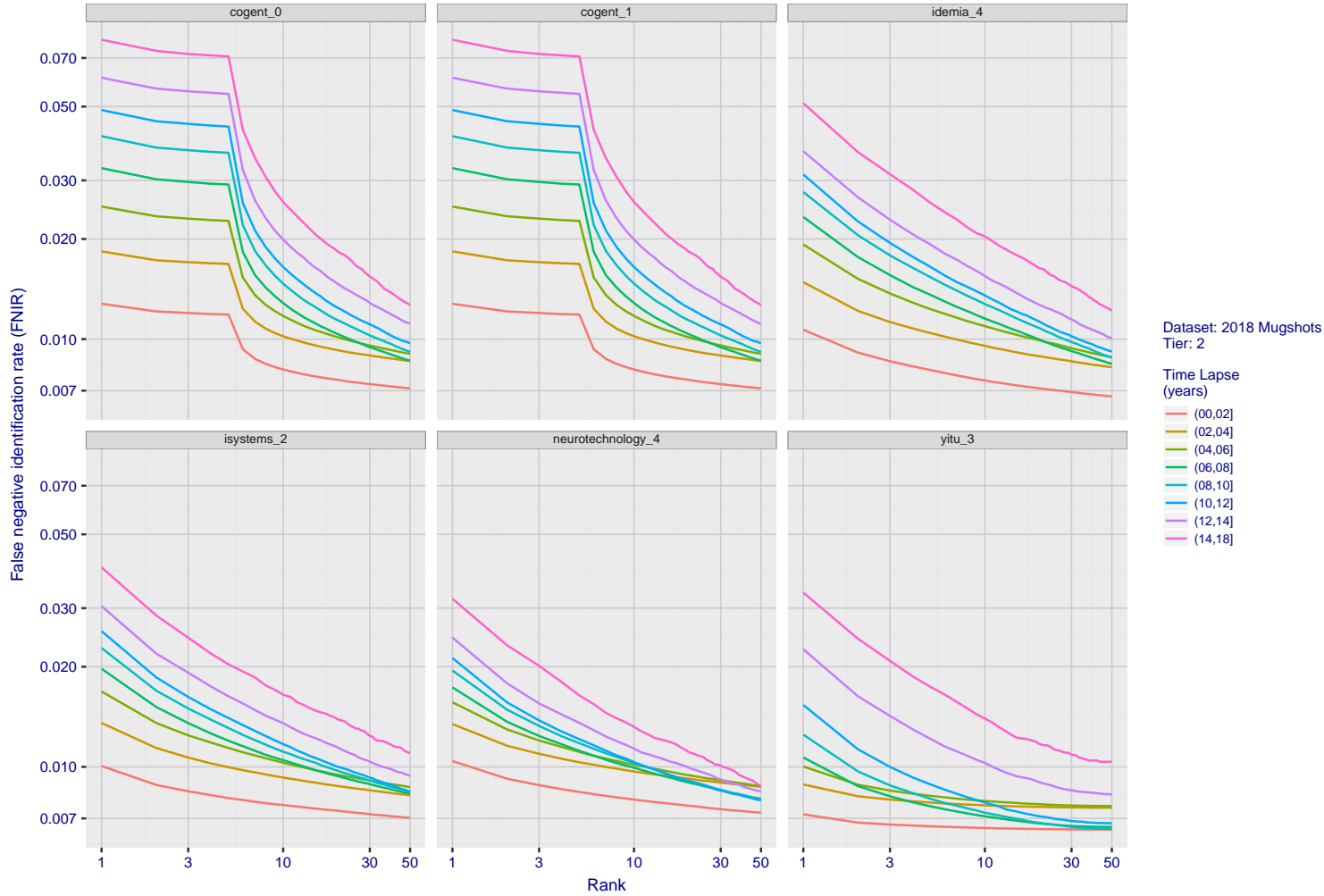


Figure 65: [FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. rank by time-elapsed. The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between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

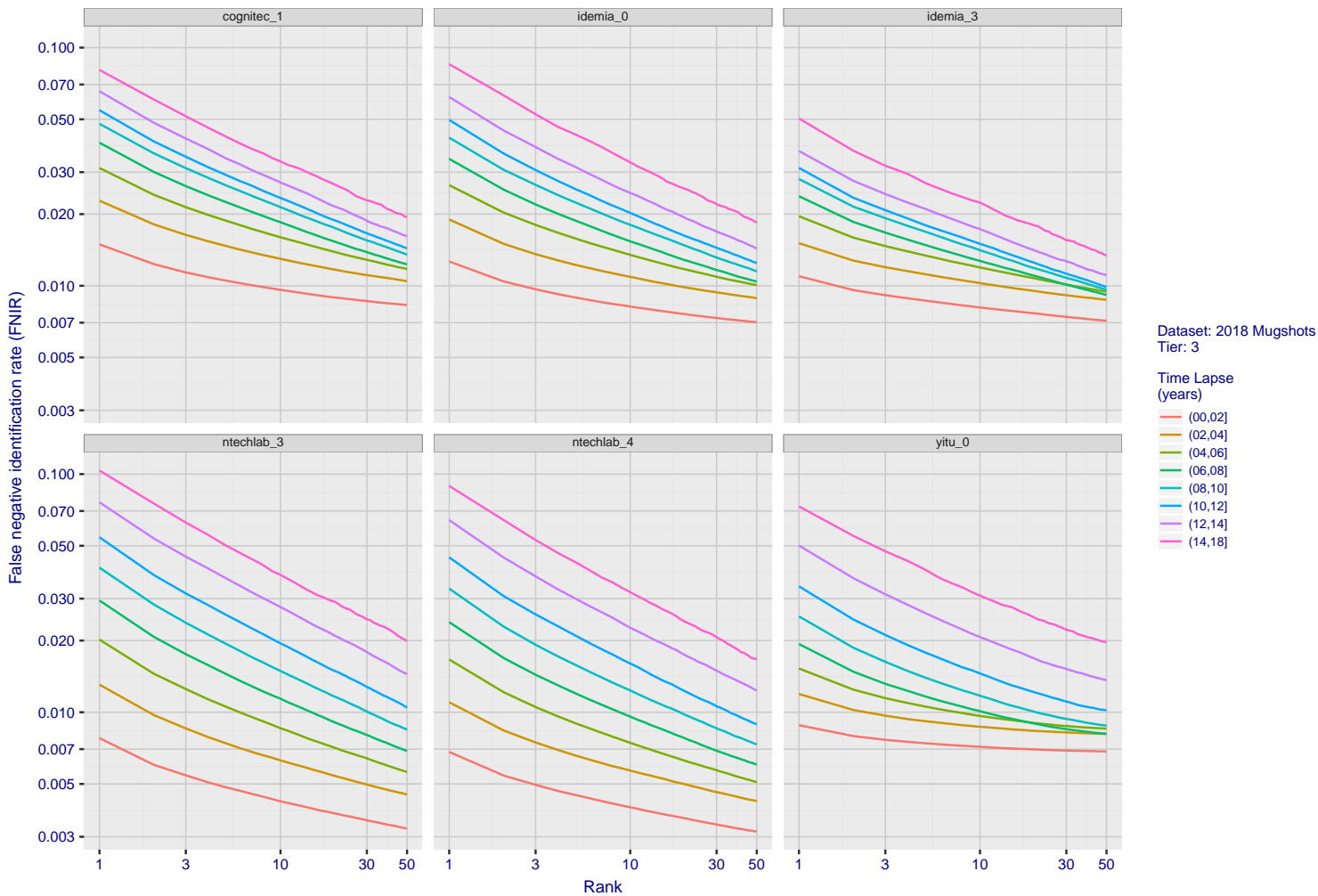


Figure 66: [FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. rank by time-elapsed. The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between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

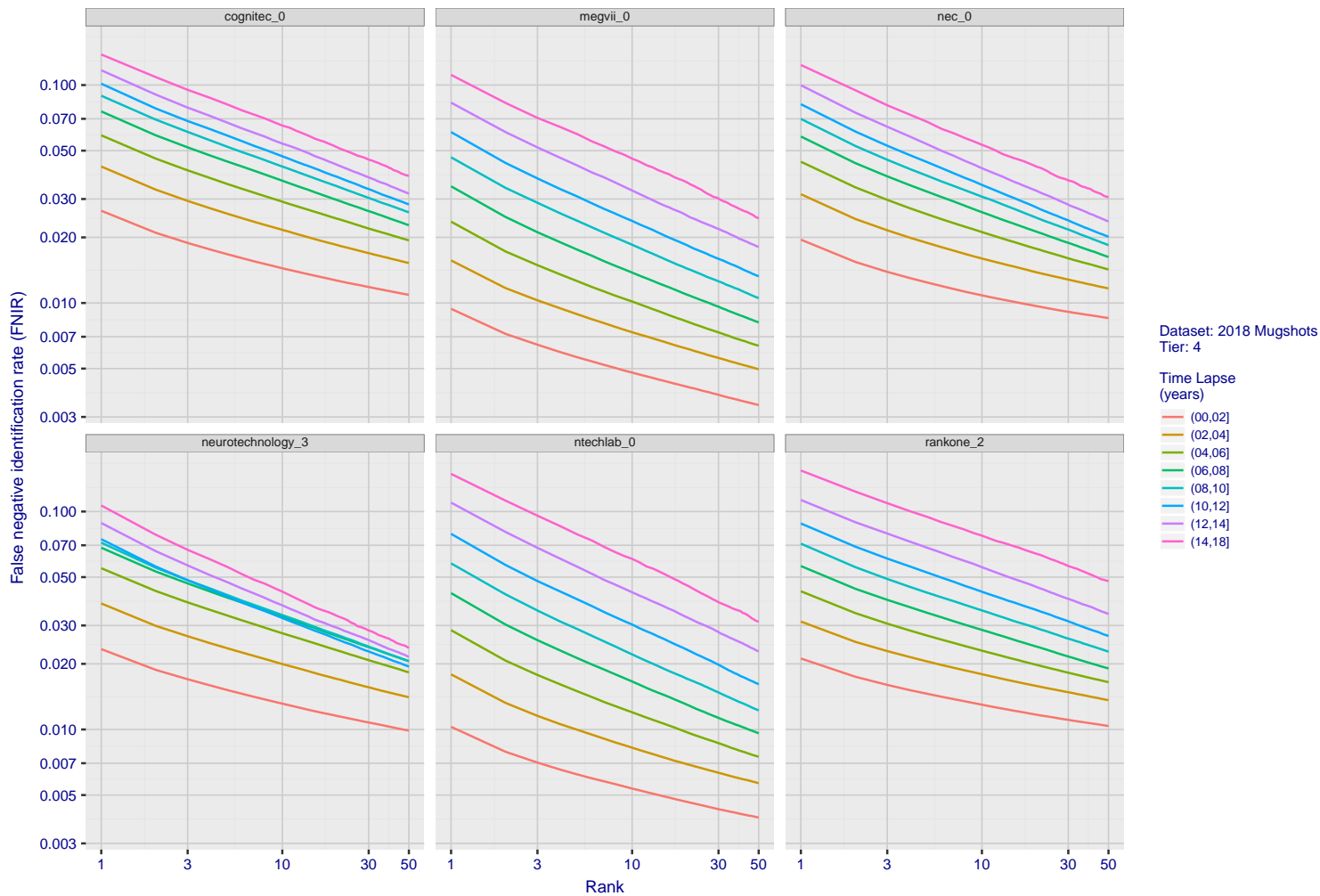


Figure 67: [FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. rank by time-elapsed. The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between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

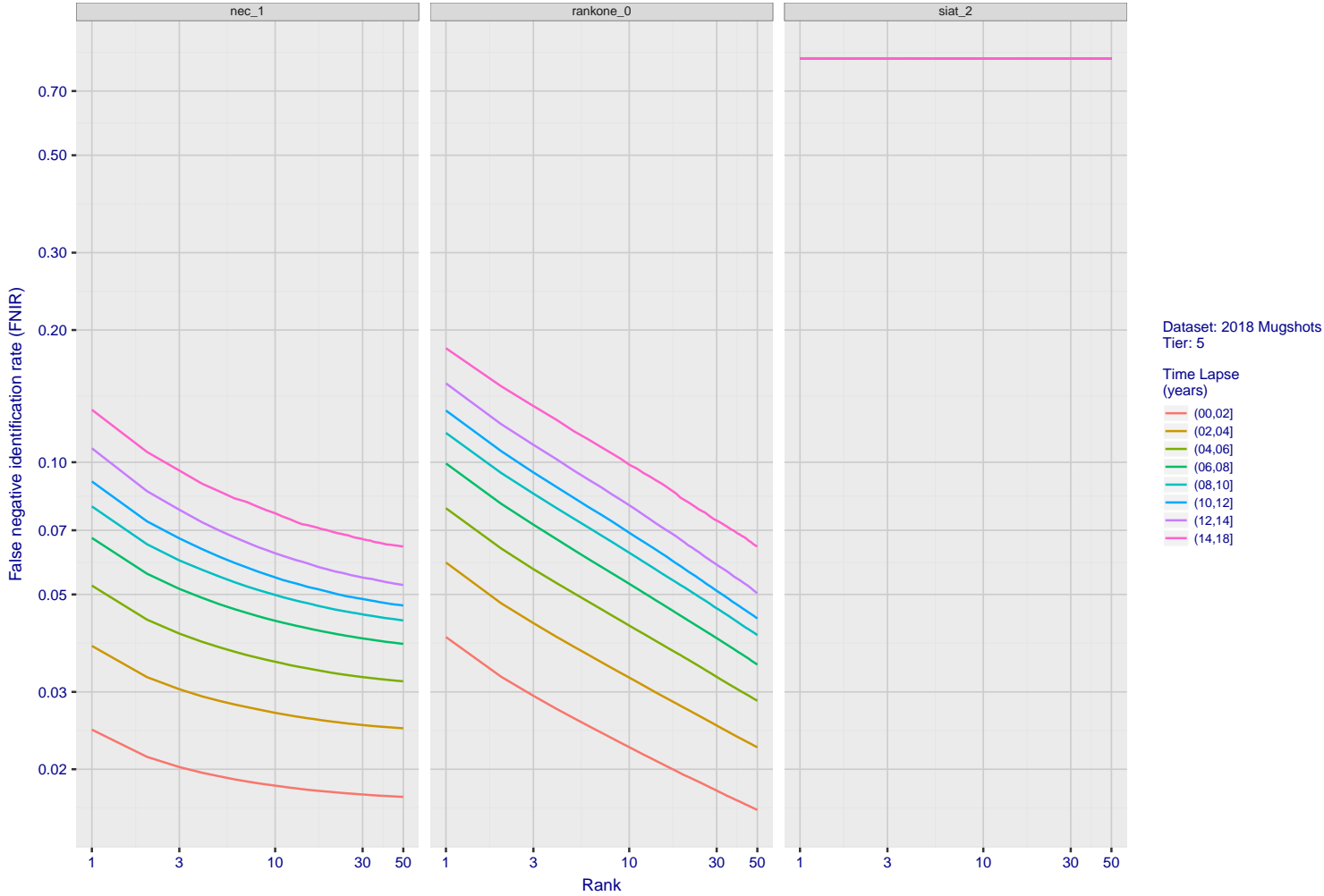


Figure 68: [FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. rank by time-elapsed. The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between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

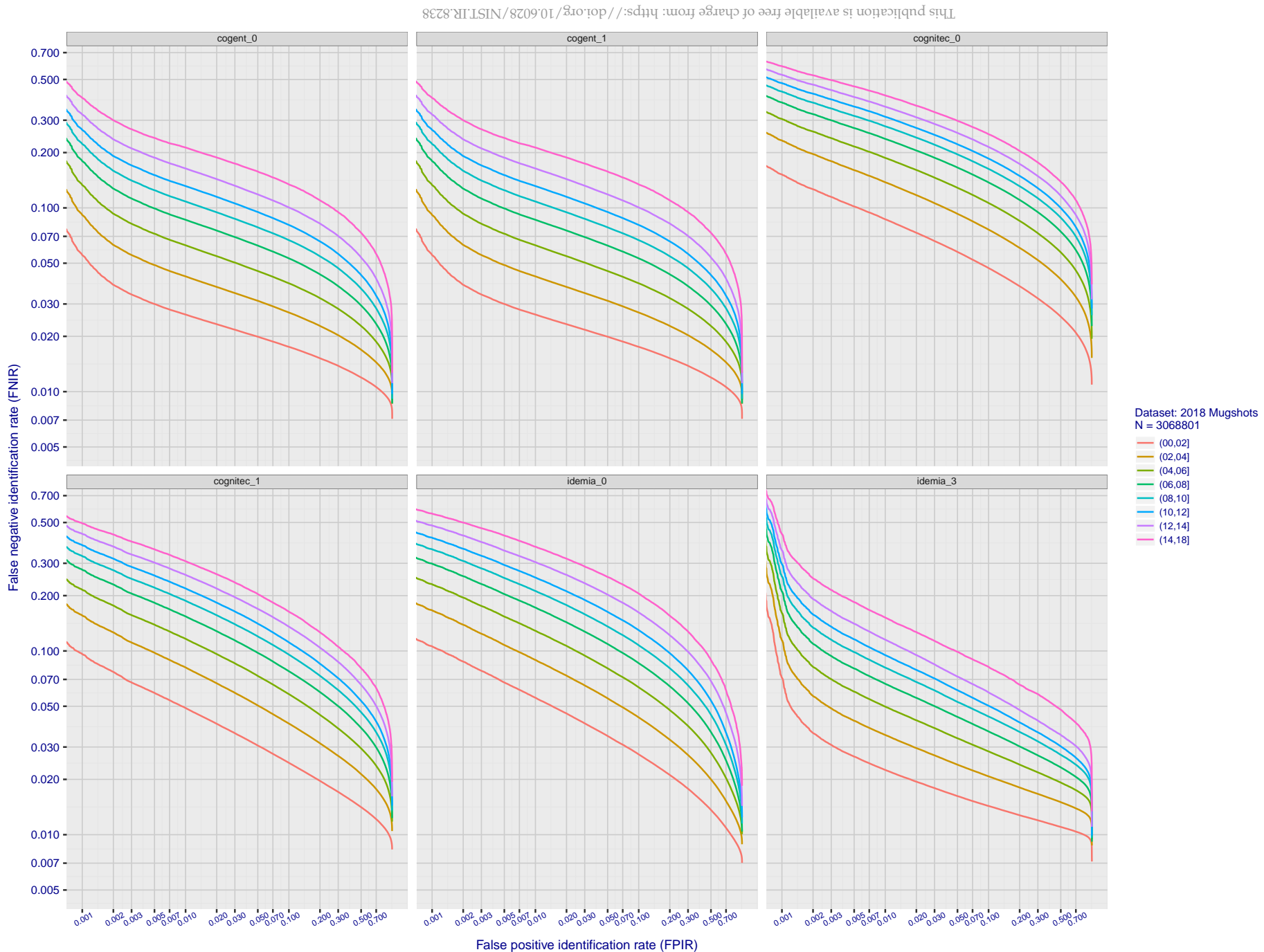


Figure 69: [FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. FPIR by time-elapsed. The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between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

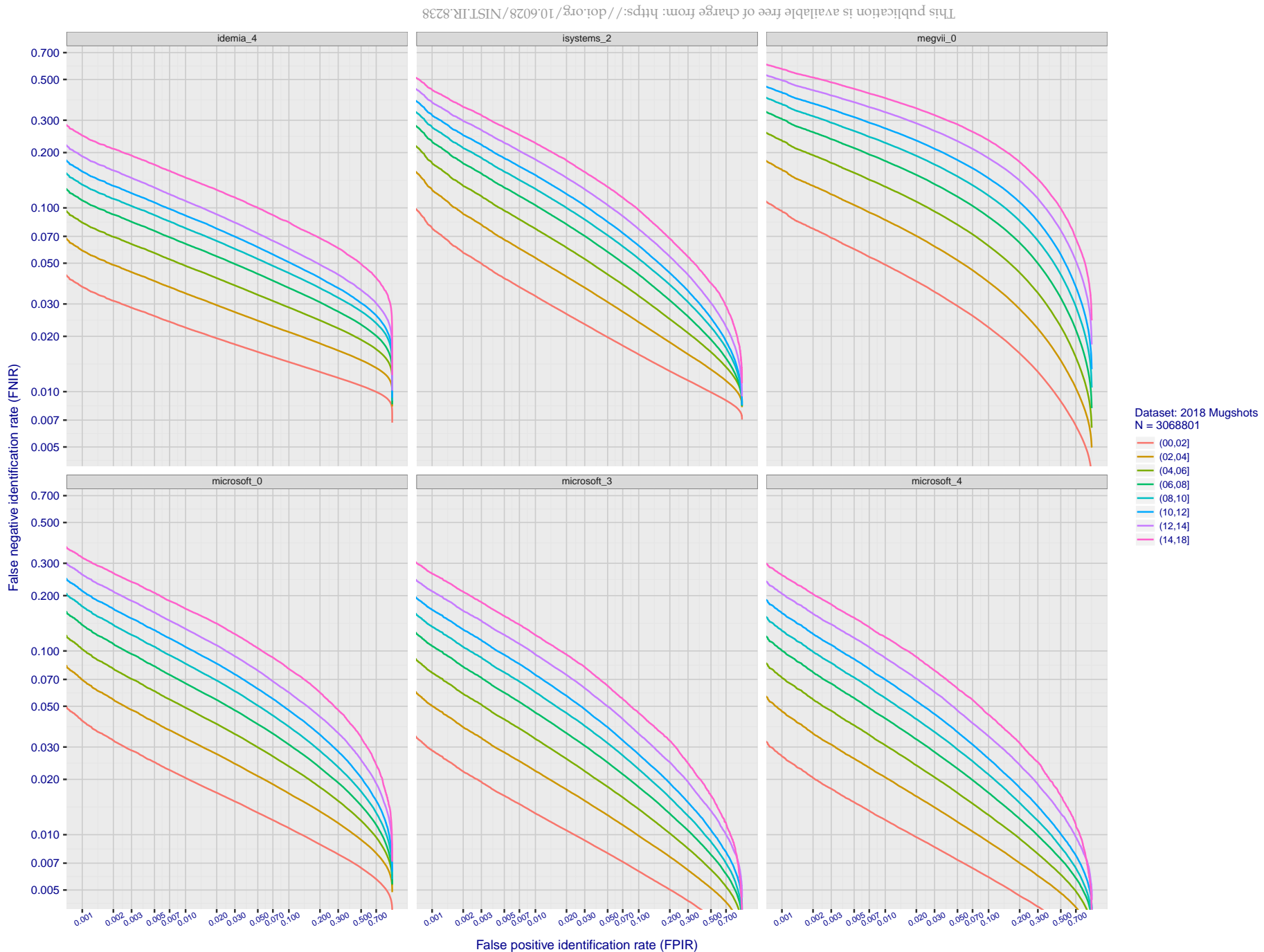


Figure 70: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. FPIR by time-elapsed.** The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between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

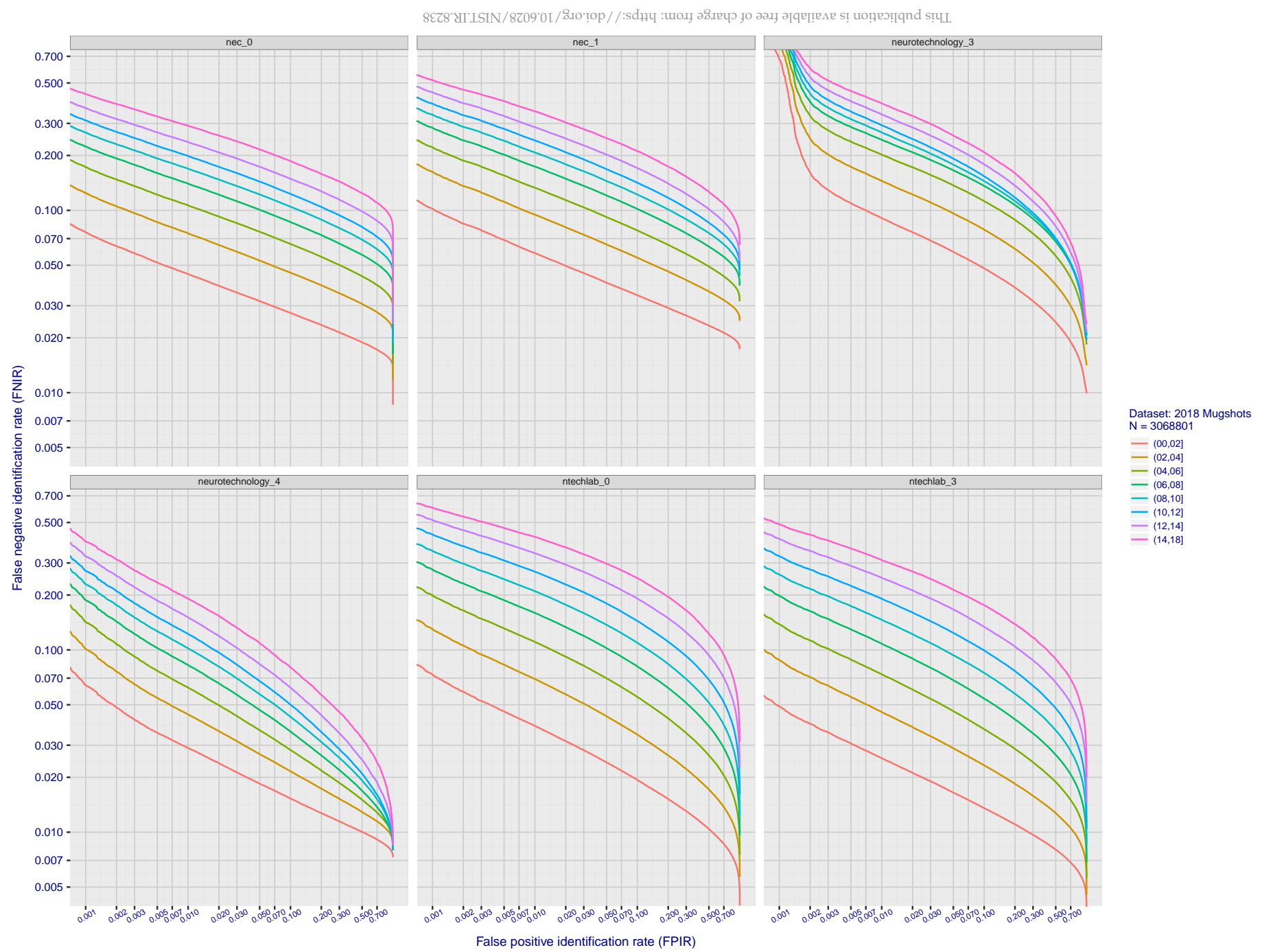


Figure 71: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. FPIR by time-elapsed.** The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between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

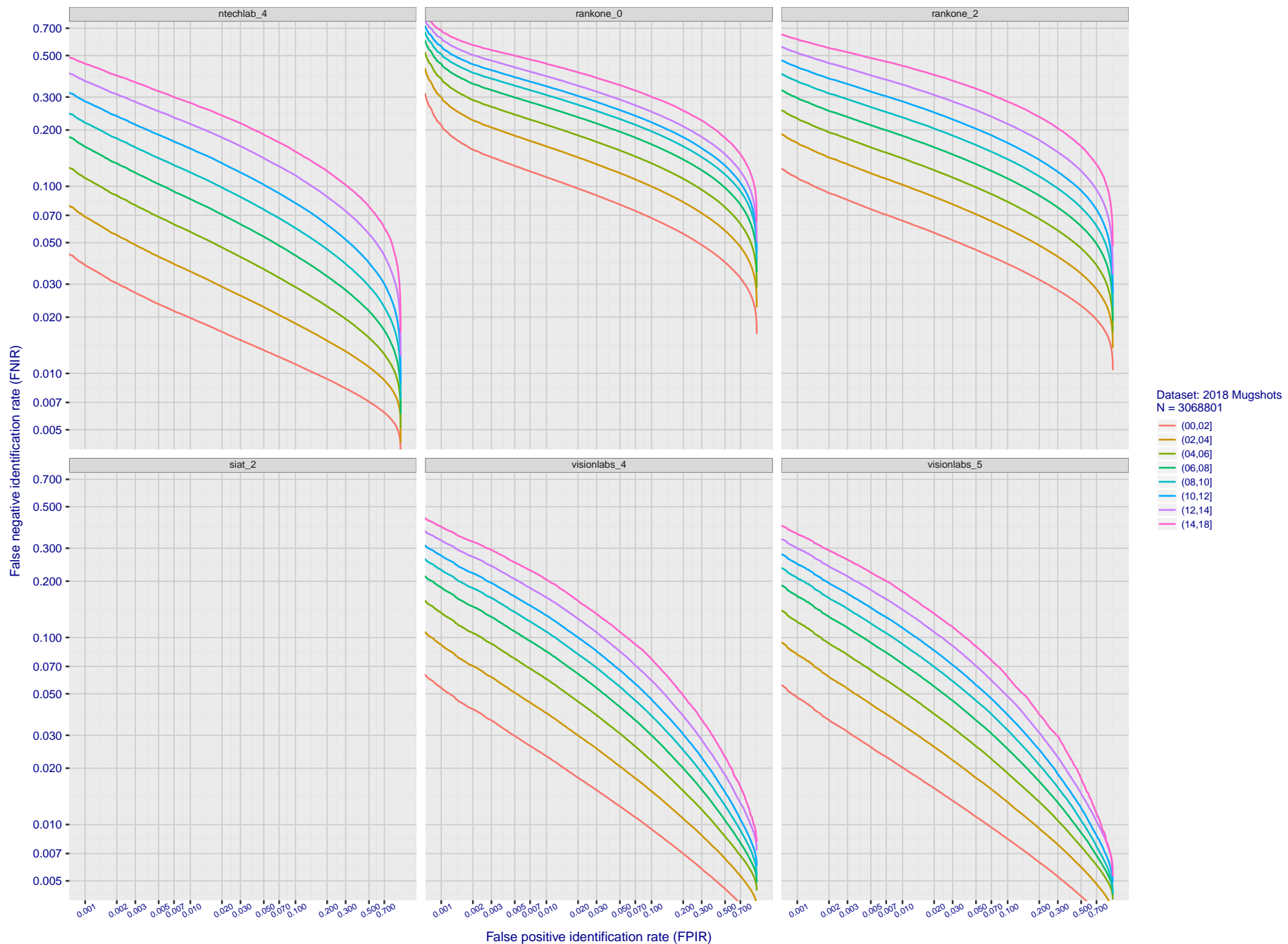


Figure 72: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. FPIR by time-elapsed.** The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between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

2018/11/26
07:24:51FNIR(N, R, T) =
FPIR(N, T) =False neg. identification rate
False pos. identification rateN = Num. enrolled subjects
R = Num. candidates examined

T = Threshold

T = 0 → Investigation
T > 0 → Identification

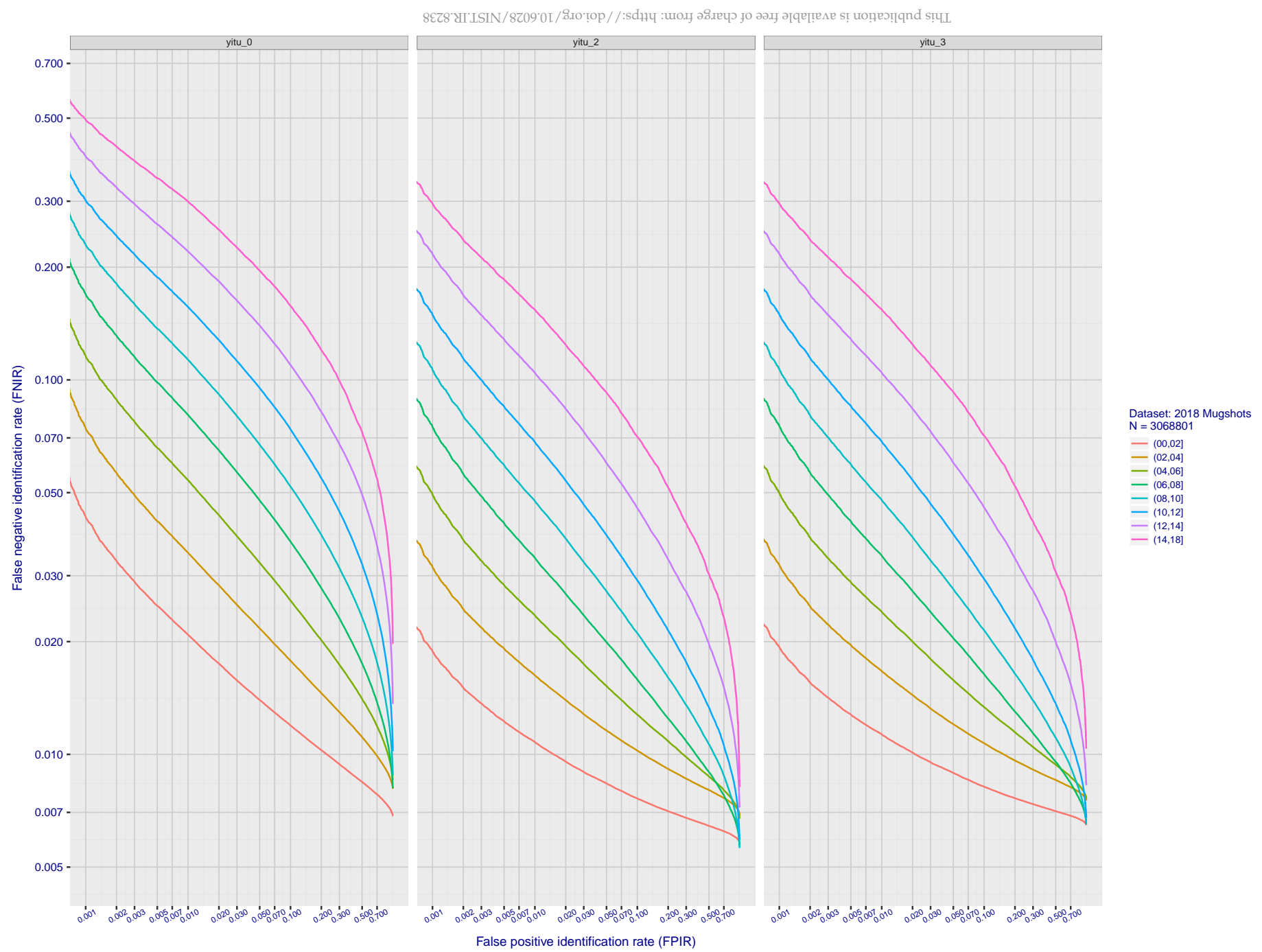


Figure 73: **[FRVT-2018 Mugshot Ageing Dataset] Identification miss rates vs. FPIR by time-elapsd.** The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Miss rates are computed over all searches noted in row 17 of Table 6 and binned by number of years between 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.

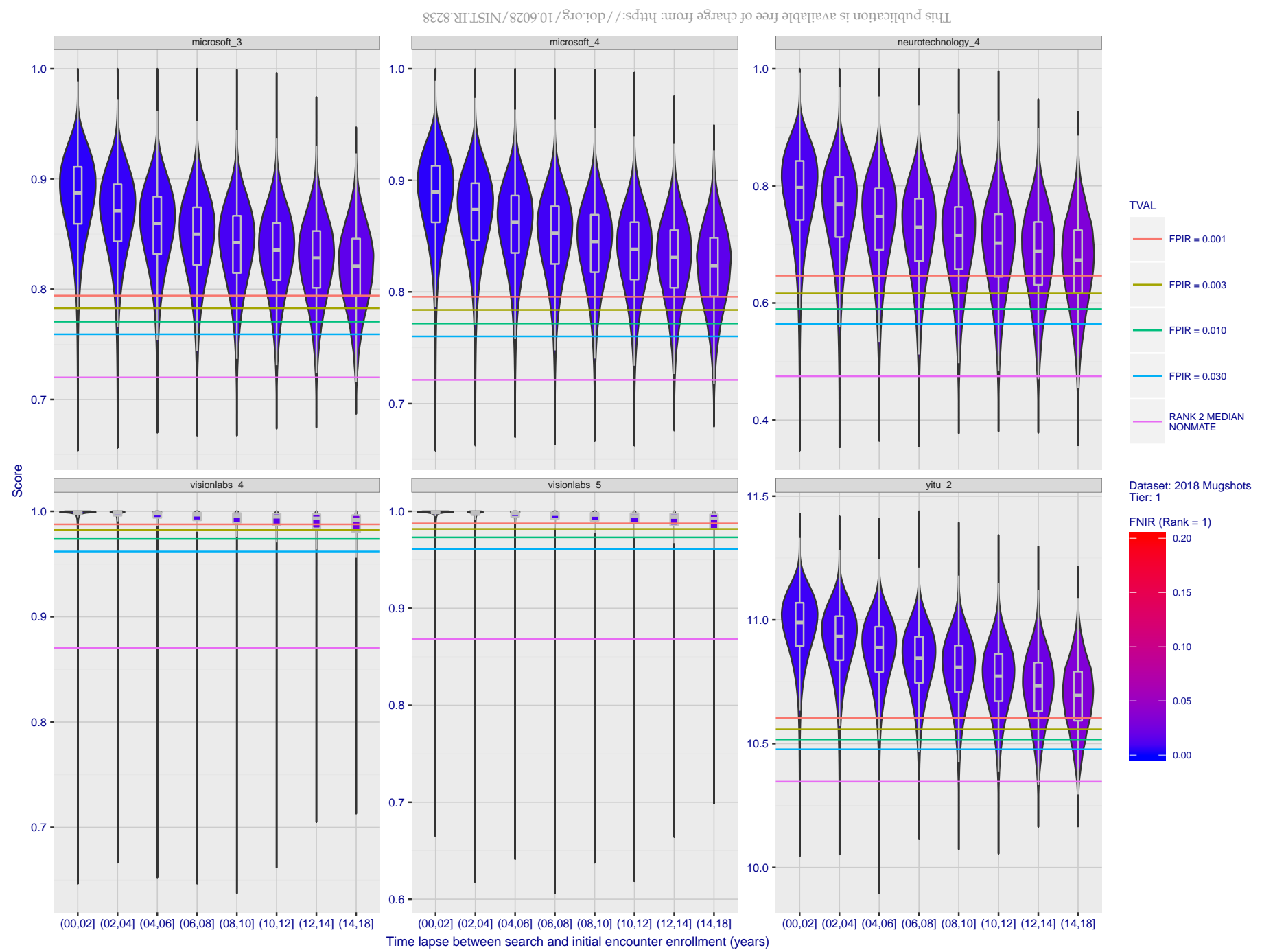


Figure 74: **[FRVT-2018 Mugshot Ageing Dataset] Native mate scores vs. time-elapsed.** The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Mated score distributions are computed over all searches noted in row 17 of Table 6 binned by number of years between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

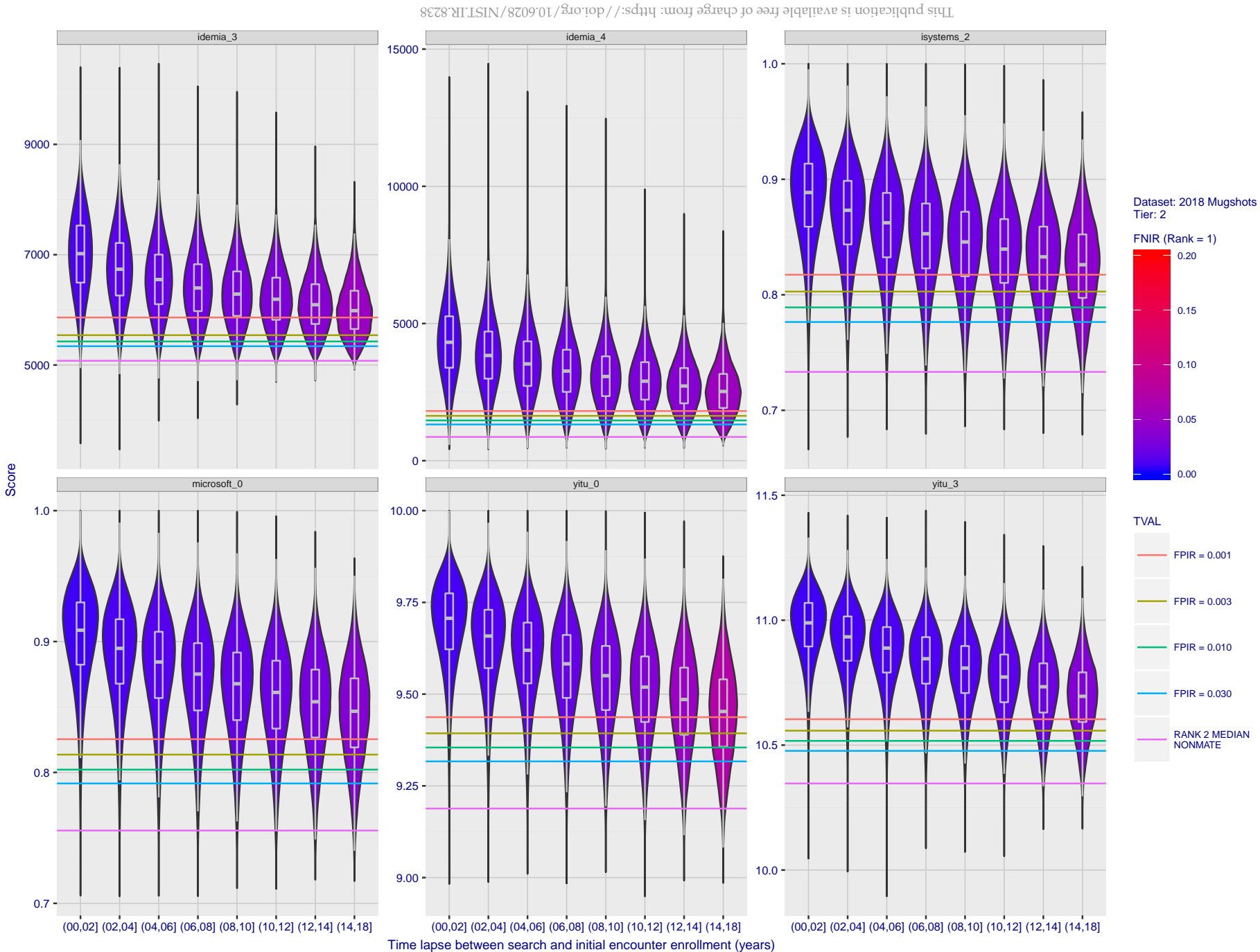


Figure 75: [FRVT-2018 Mugshot Ageing Dataset] Native mate scores vs. time-elapsed. The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Mated score distributions are computed over all searches noted in row 17 of Table 6 binned by number of years between search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

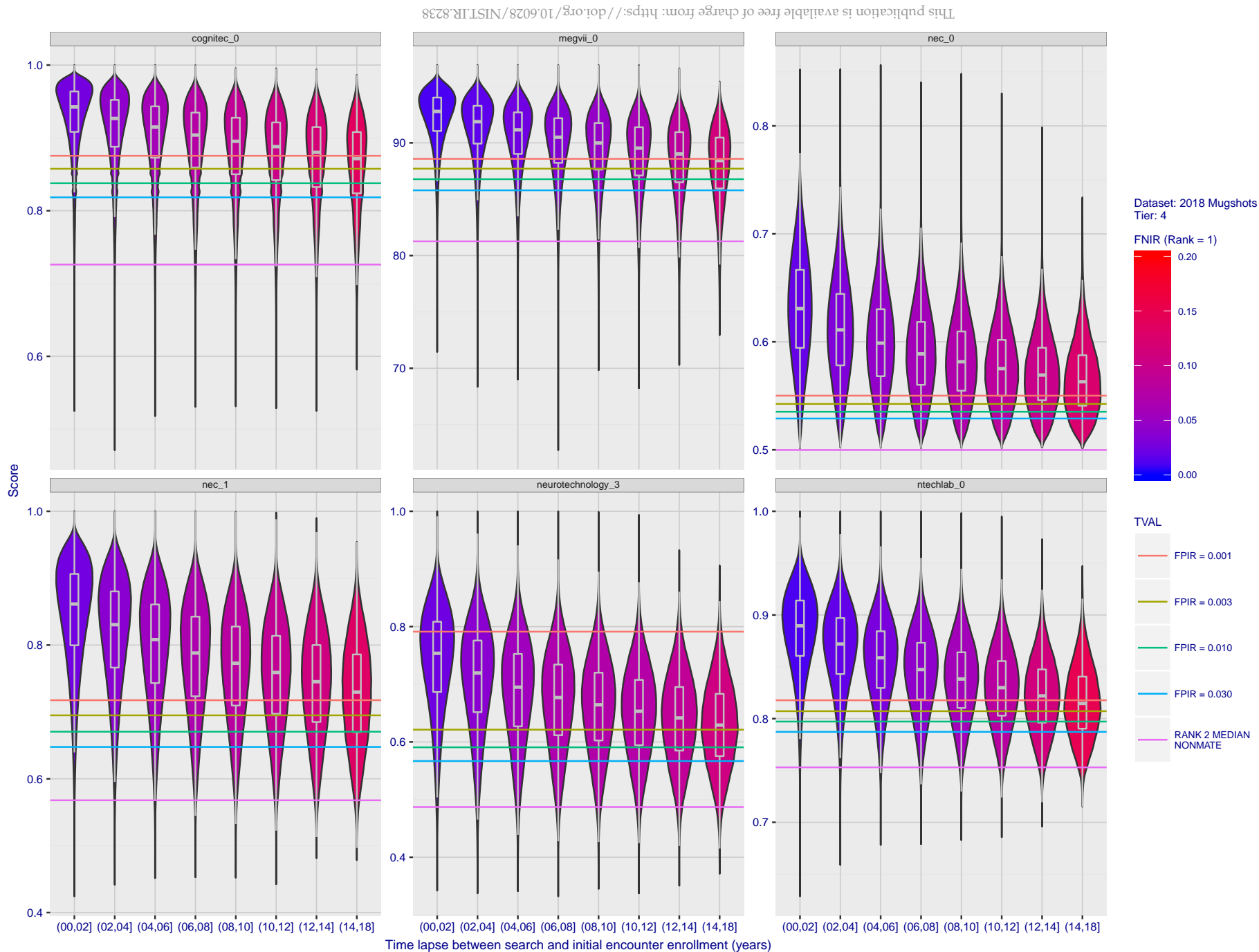


Figure 77: [FRVT-2018 Mugshot Ageing Dataset] Native mate scores vs. time-elapsed. The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Mated score distributions are computed over all searches noted in row 17 of Table 6 binned by number of years between search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

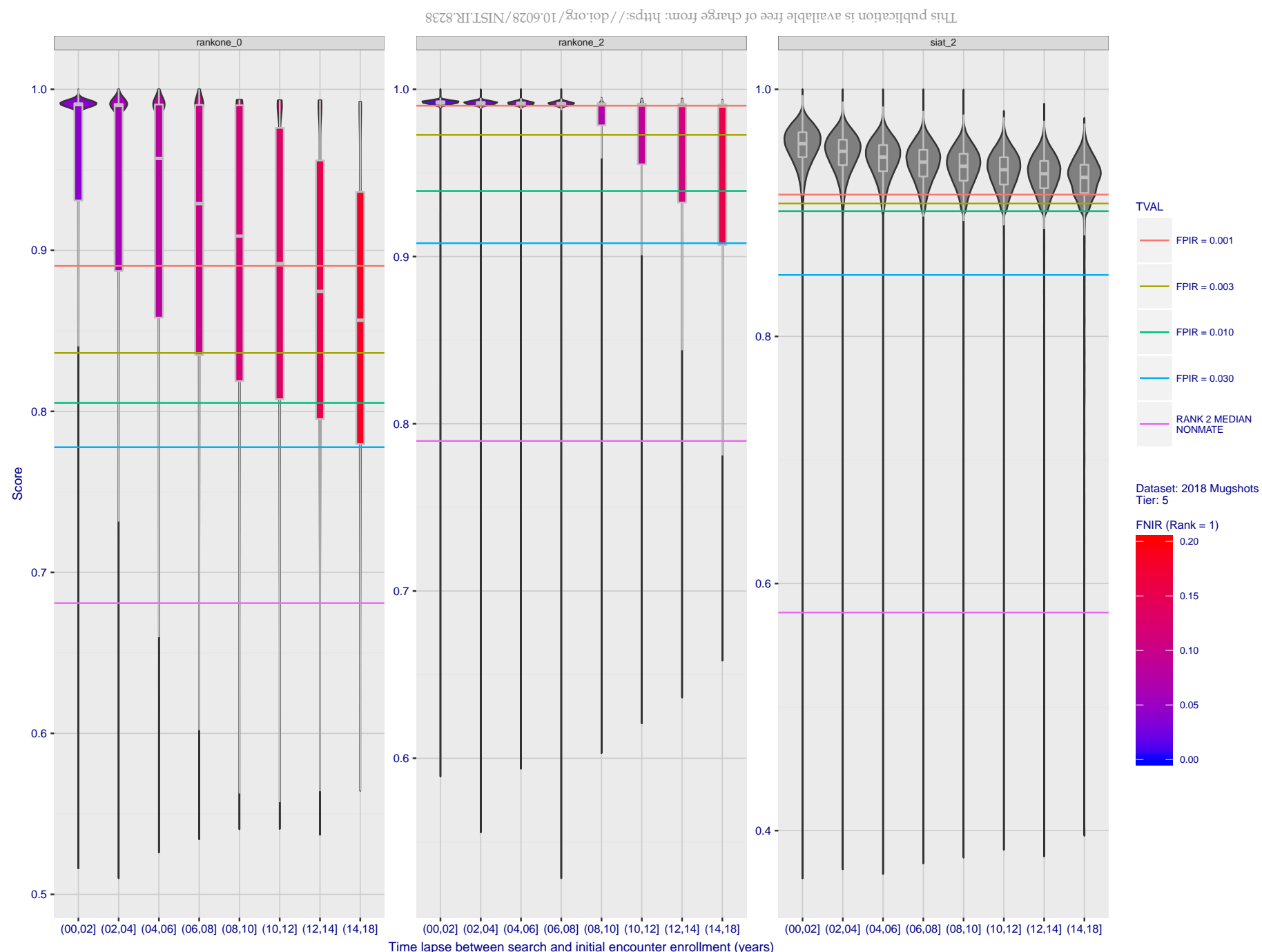


Figure 78: [FRVT-2018 Mugshot Ageing Dataset] Native mate scores vs. time-elapsed. The oldest image of each individual is enrolled. Thereafter, all more recent images are searched. Mated score distributions are computed over all searches noted in row 17 of Table 6 binned by number of years between search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

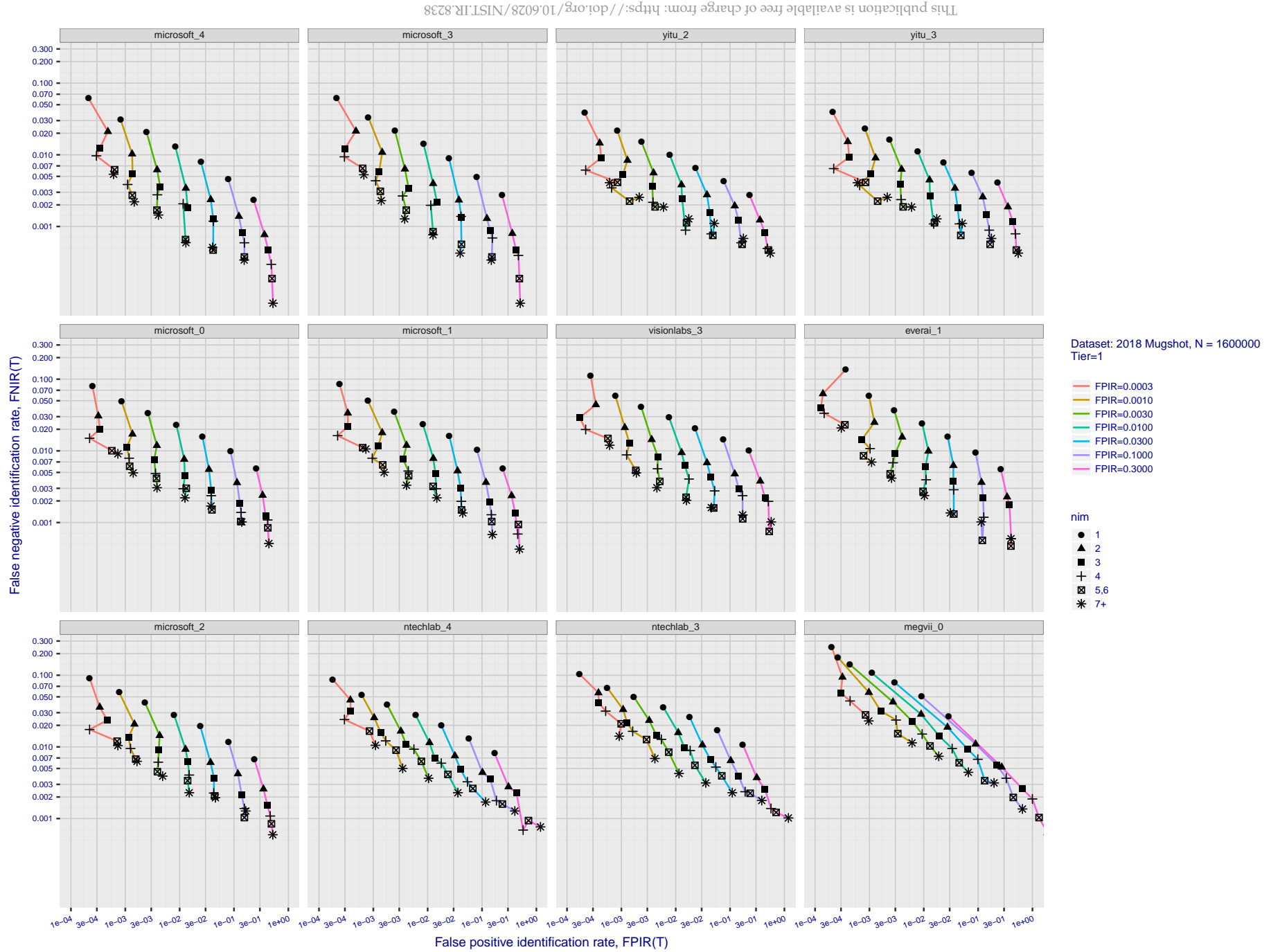


Figure 79: [FRVT-2018 Mugshot Dataset] Effect of enrolling multiple images for each identity. The plot shows an identification miss rates vs. false positive rates, at seven operating thresholds. The enrolled population size is fixed. The images are enrolled with lifetime-consolidation - see section 2.3.

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

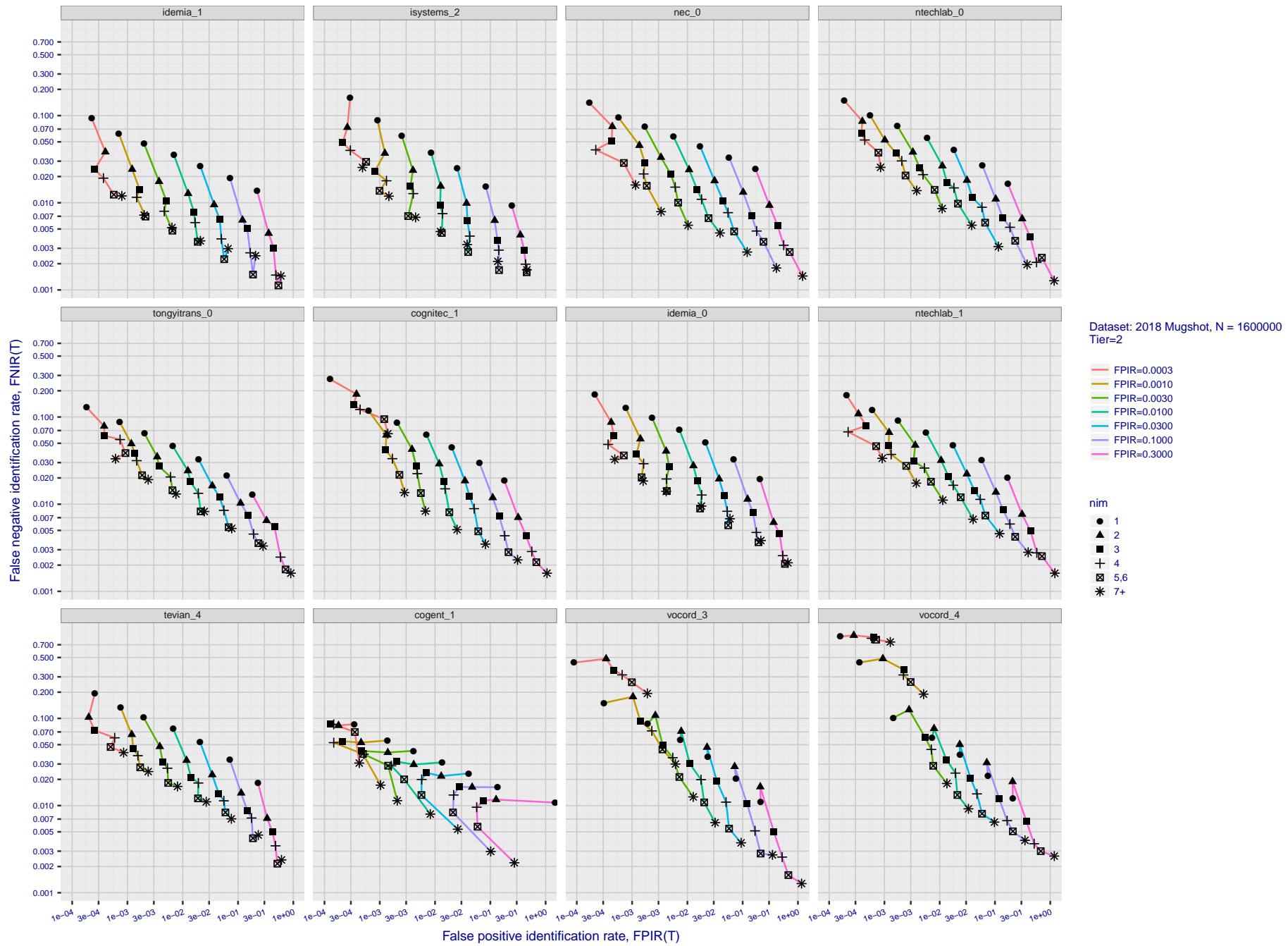
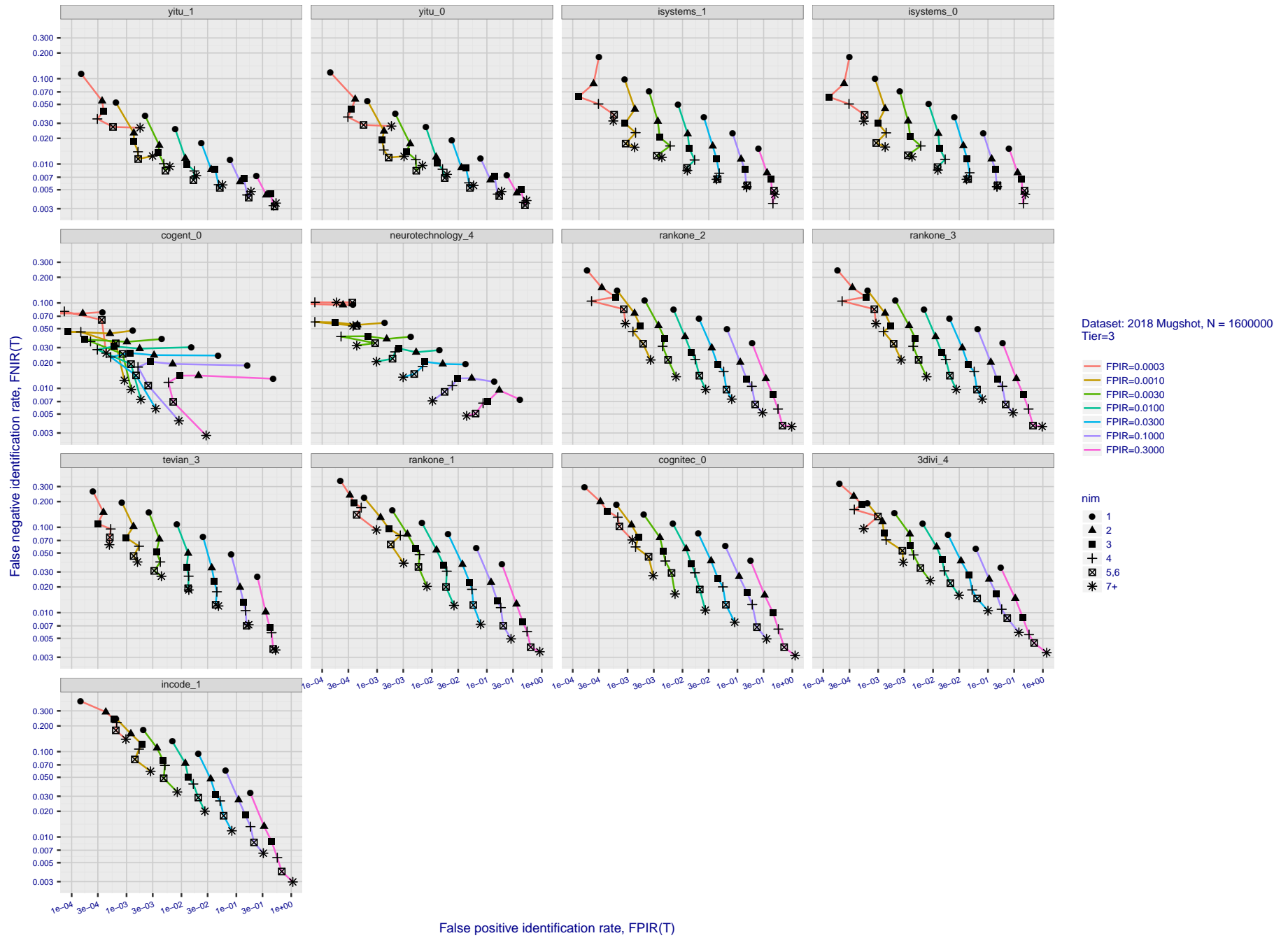


Figure 80: [FRVT-2018 Mugshot Dataset] Effect of enrolling multiple images for each identity. The plot shows an identification miss rates vs. false positive rates, at seven operating thresholds. The enrolled population size is fixed. The images are enrolled with lifetime-consolidation - see section 2.3.



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

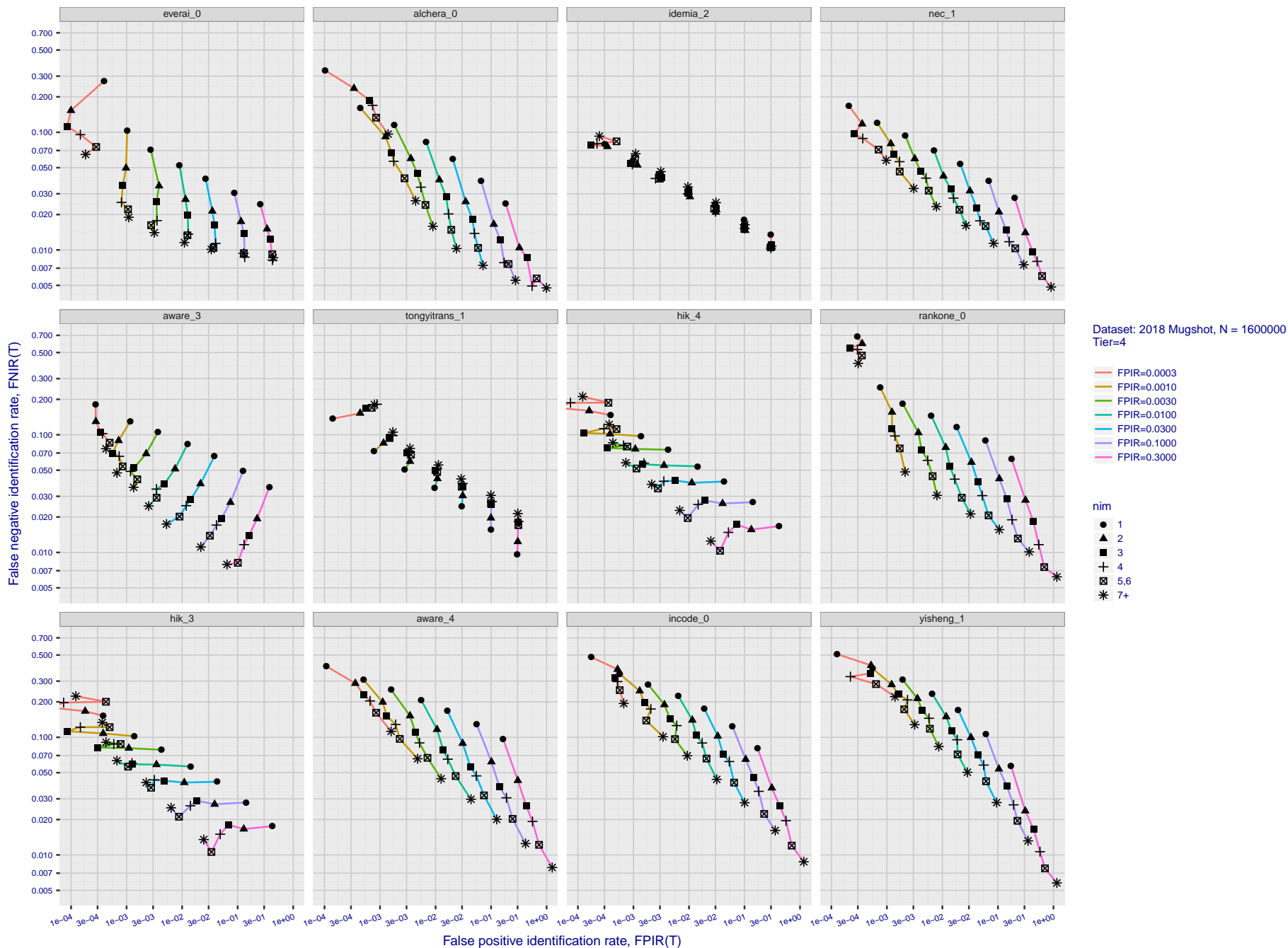


Figure 82: [FRVT-2018 Mugshot Dataset] Effect of enrolling multiple images for each identity. The plot shows an identification miss rates vs. false positive rates, at seven operating thresholds. The enrolled population size is fixed. The images are enrolled with lifetime-consolidation - see section 2.3.

2018/11/26
07:24:51FNIR(N, R, T) =
FPIR(N, T) =False neg. identification rate
False pos. identification rateN = Num. enrolled subjects
R = Num. candidates examined

T = Threshold

T = 0 → Investigation
T > 0 → Identification

False negative identification rate, FNIR(T)

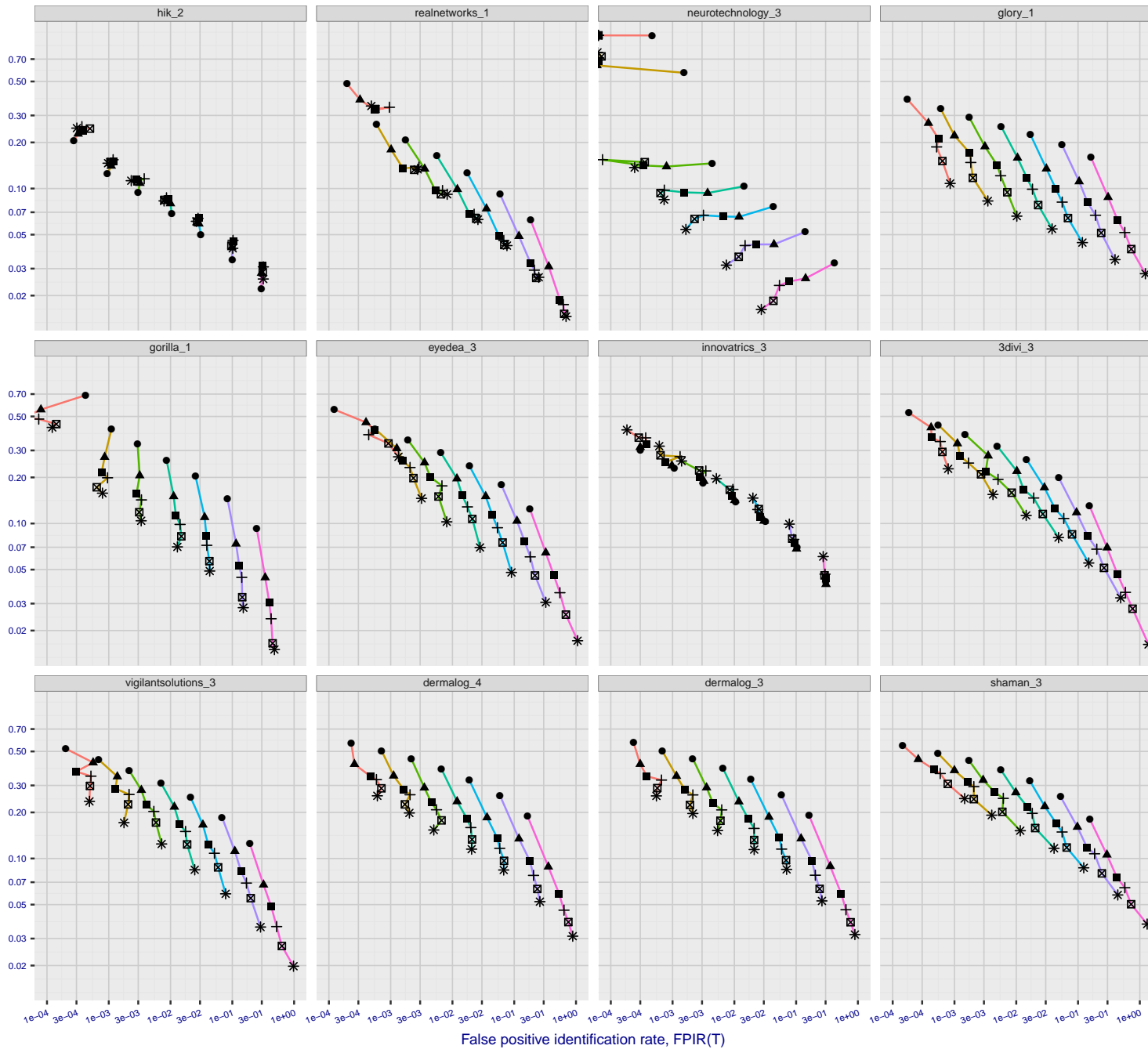
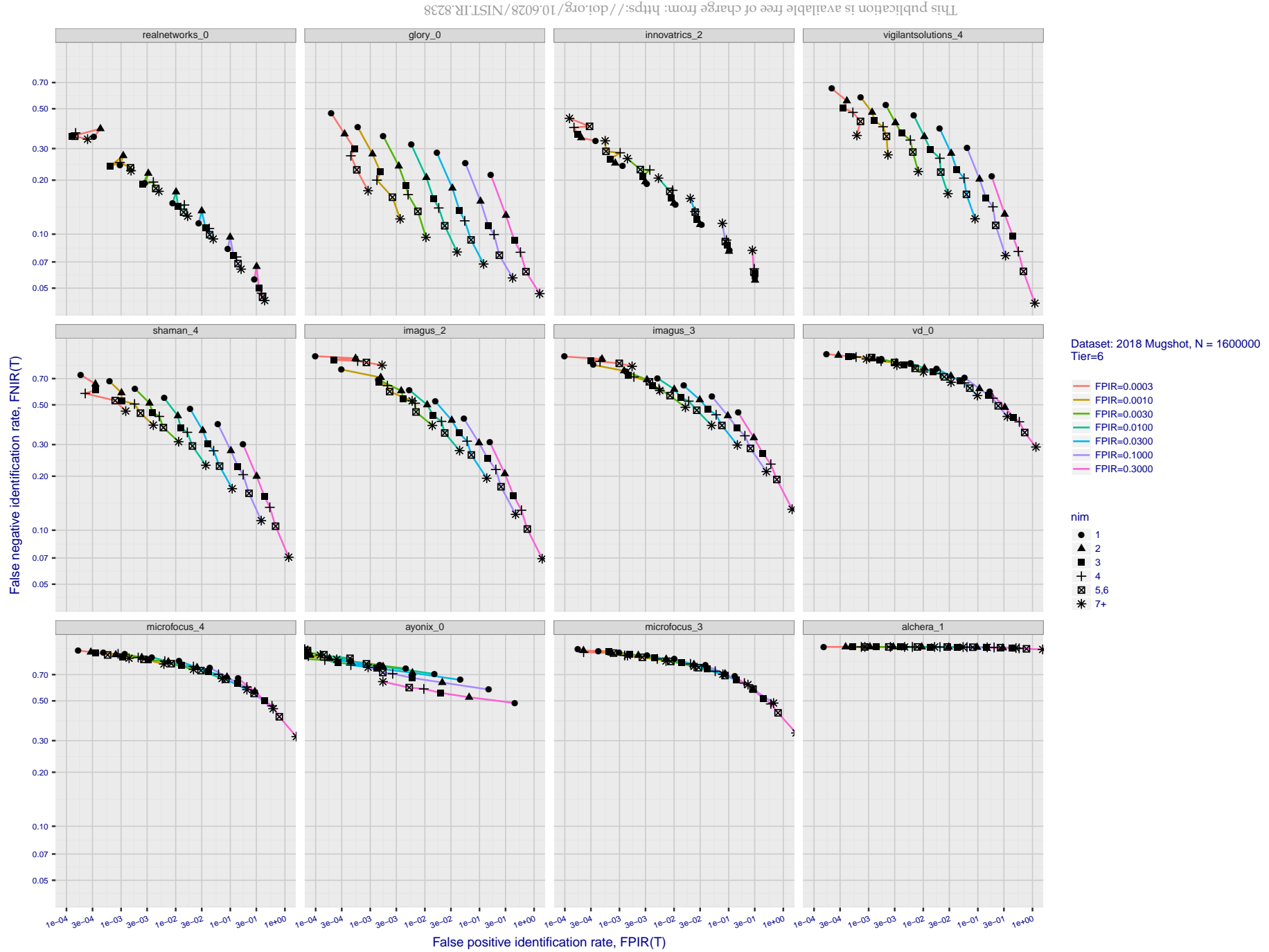


Figure 83: [FRVT-2018 Mugshot Dataset] Effect of enrolling multiple images for each identity. The plot shows an identification miss rates vs. false positive rates, at seven operating thresholds. The enrolled population size is fixed. The images are enrolled with lifetime-consolidation - see section 2.3.



Appendix D Accuracy with poor quality webcam images

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) =
FPIR(N, T) =

False neg. identification rate
False pos. identification rate

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

T = Threshold

T = 0 → Investigation
T > 0 → Identification

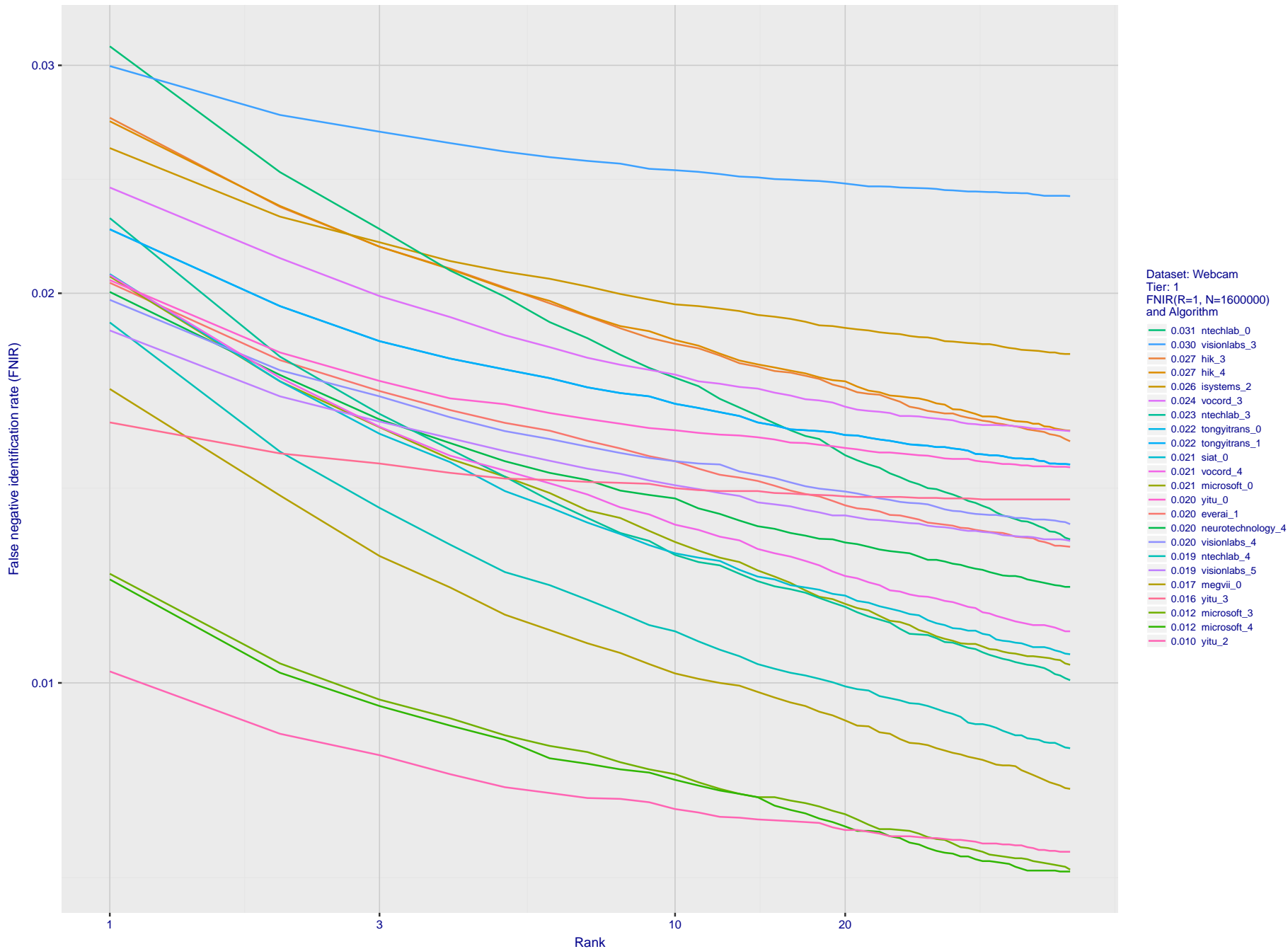
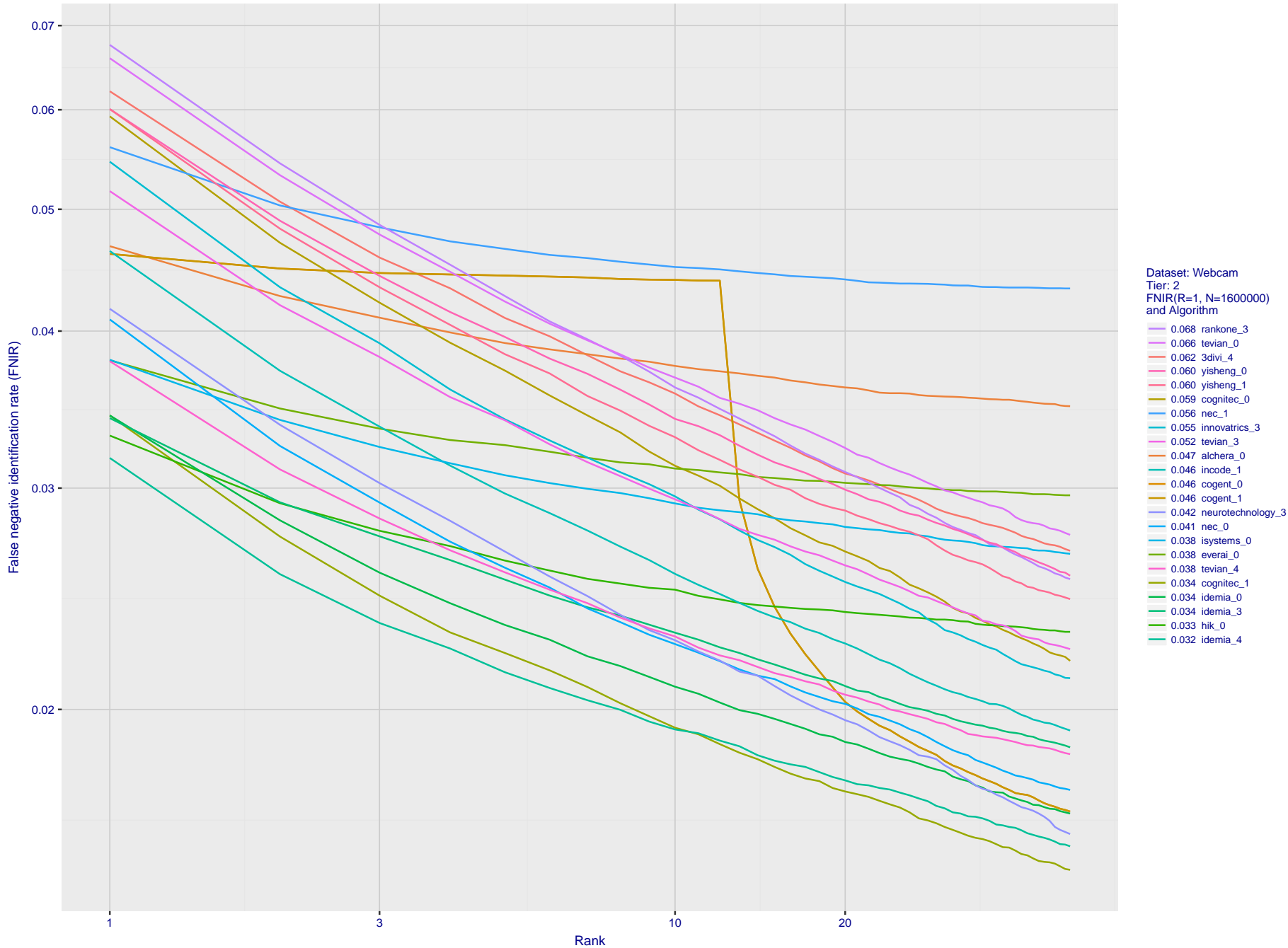


Figure 85: **[Webcam Dataset] Identification miss rates vs. rank.** The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and less constrained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

2018/11/26
07:24:51
FNIR(N, R, T) =
FPIR(N, T) =
False neg. identification rate
False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification



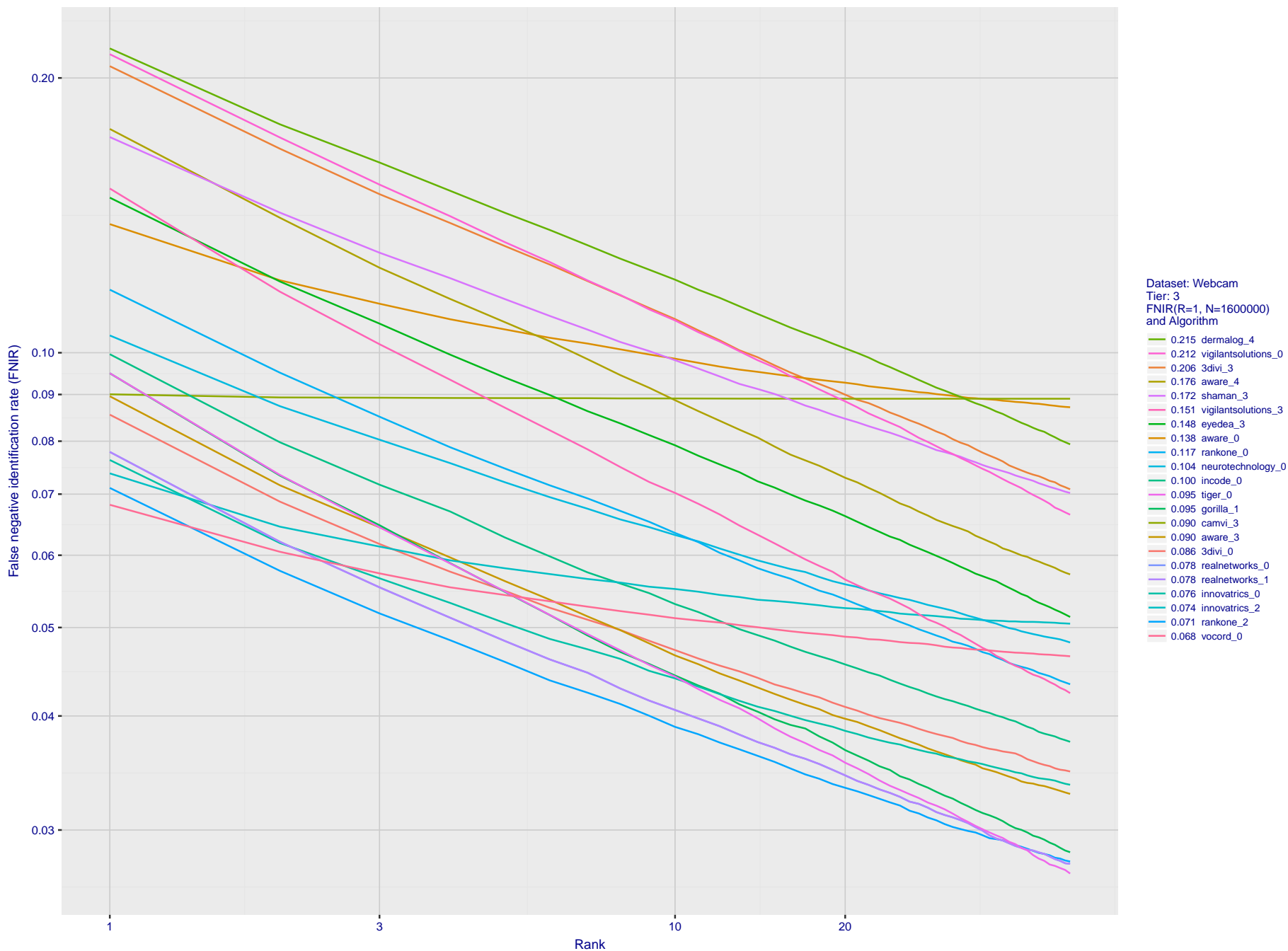


Figure 87: **[Webcam Dataset] Identification miss rates vs. rank.** The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and less constrained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

2018/11/26
07:24:51

FNIR(N, R, T) =
FPIR(N, T) =

False neg. identification rate
False pos. identification rate

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

T = Threshold

T = 0 → Investigation
T > 0 → Identification

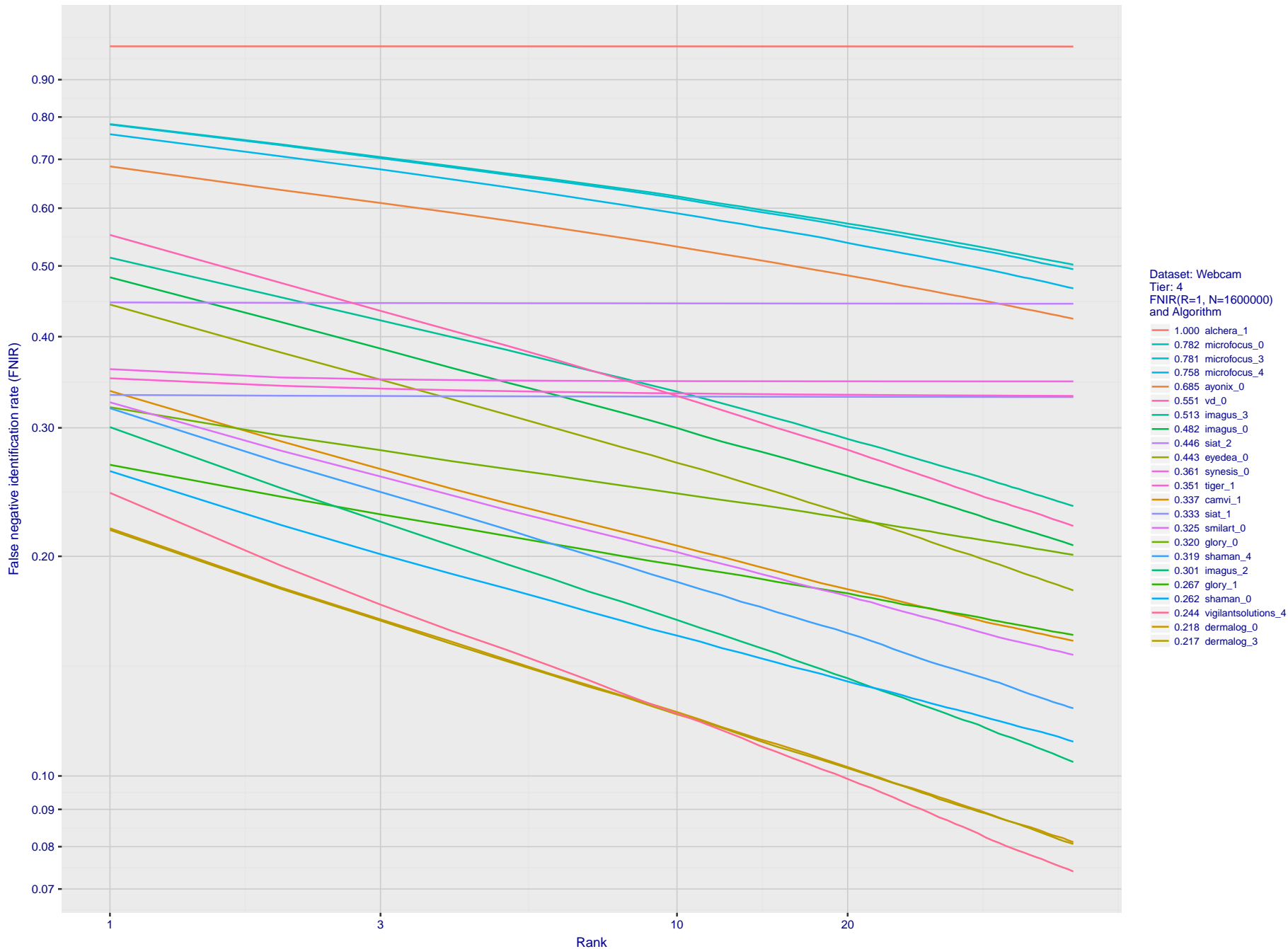


Figure 88: [Webcam Dataset] Identification miss rates vs. rank. The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and less constrained subject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

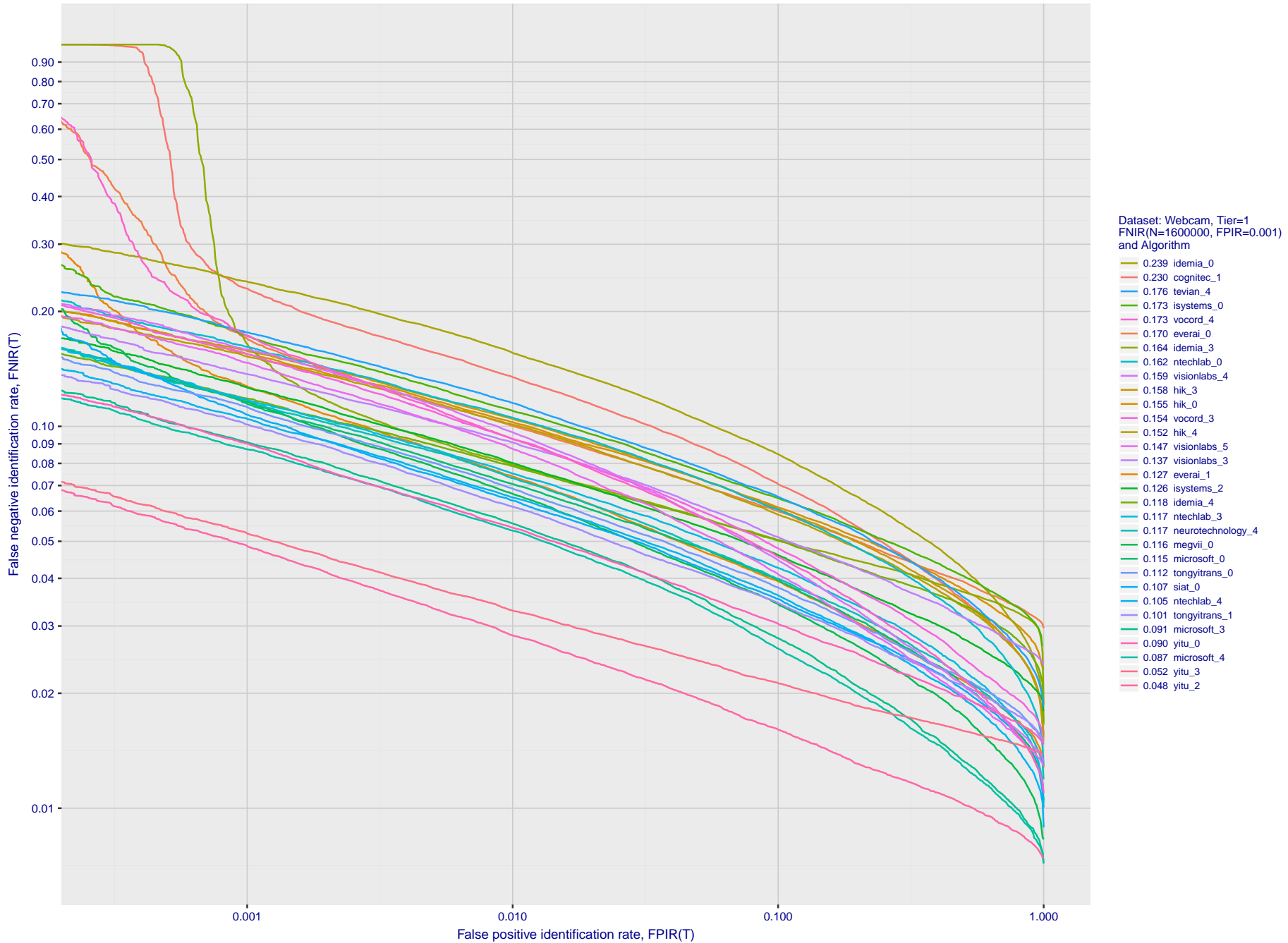


Figure 89: **[Webcam Dataset] Identification miss rates vs. false positive rates.** The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and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

2018/11/26
07:24:51
FNIR(N, R, T) =
FPIR(N, T) =
False neg. identification rate
False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

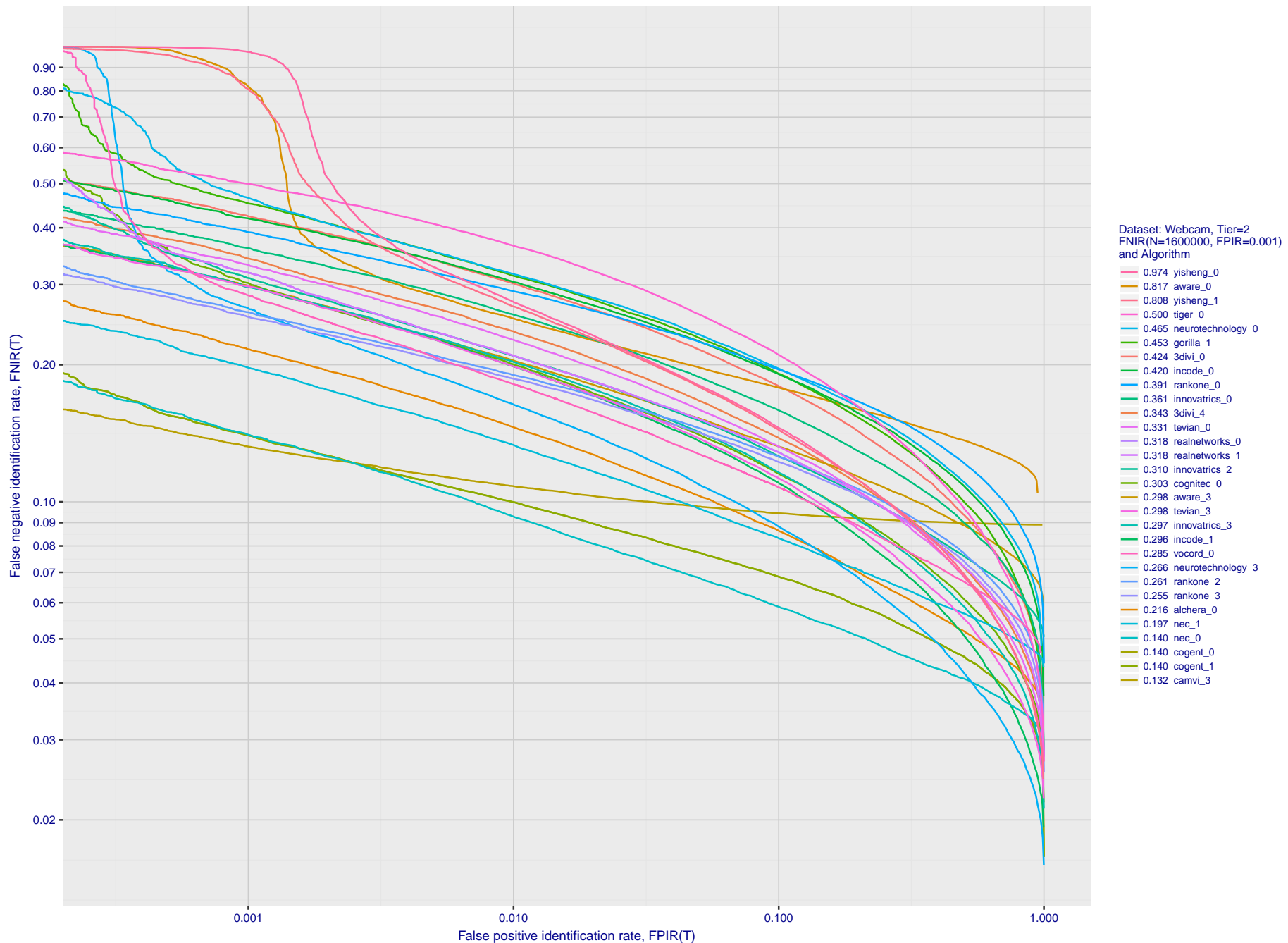


Figure 90: **[Webcam Dataset] Identification miss rates vs. false positive rates.** The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and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

2018/11/26
07:24:51
FNIR(N, R, T) =
FPIR(N, T) =
False neg. identification rate
False pos. identification rate
N = Num. enrolled subjects
R = Num. candidates examined
T = Threshold
T = 0 → Investigation
T > 0 → Identification

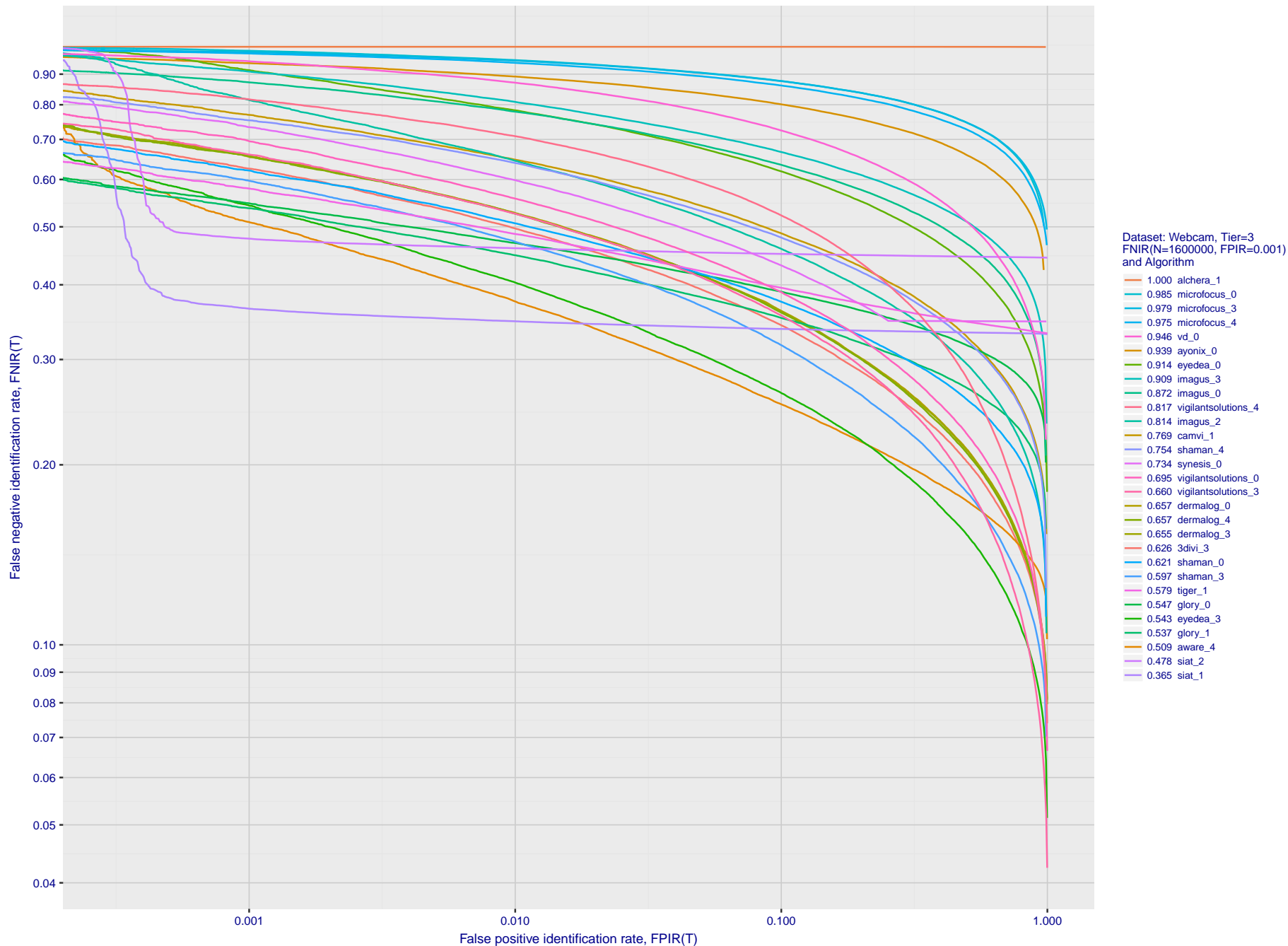


Figure 91: **[Webcam Dataset] Identification miss rates vs. false positive rates.** The results apply to cross-domain recognition in which webcams are searched against enrolled mugshots. The FNIR values are higher than those for mugshot-mugshot identification due to low image resolution, lighting and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

Appendix E Accuracy with non-cooperating subjects

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

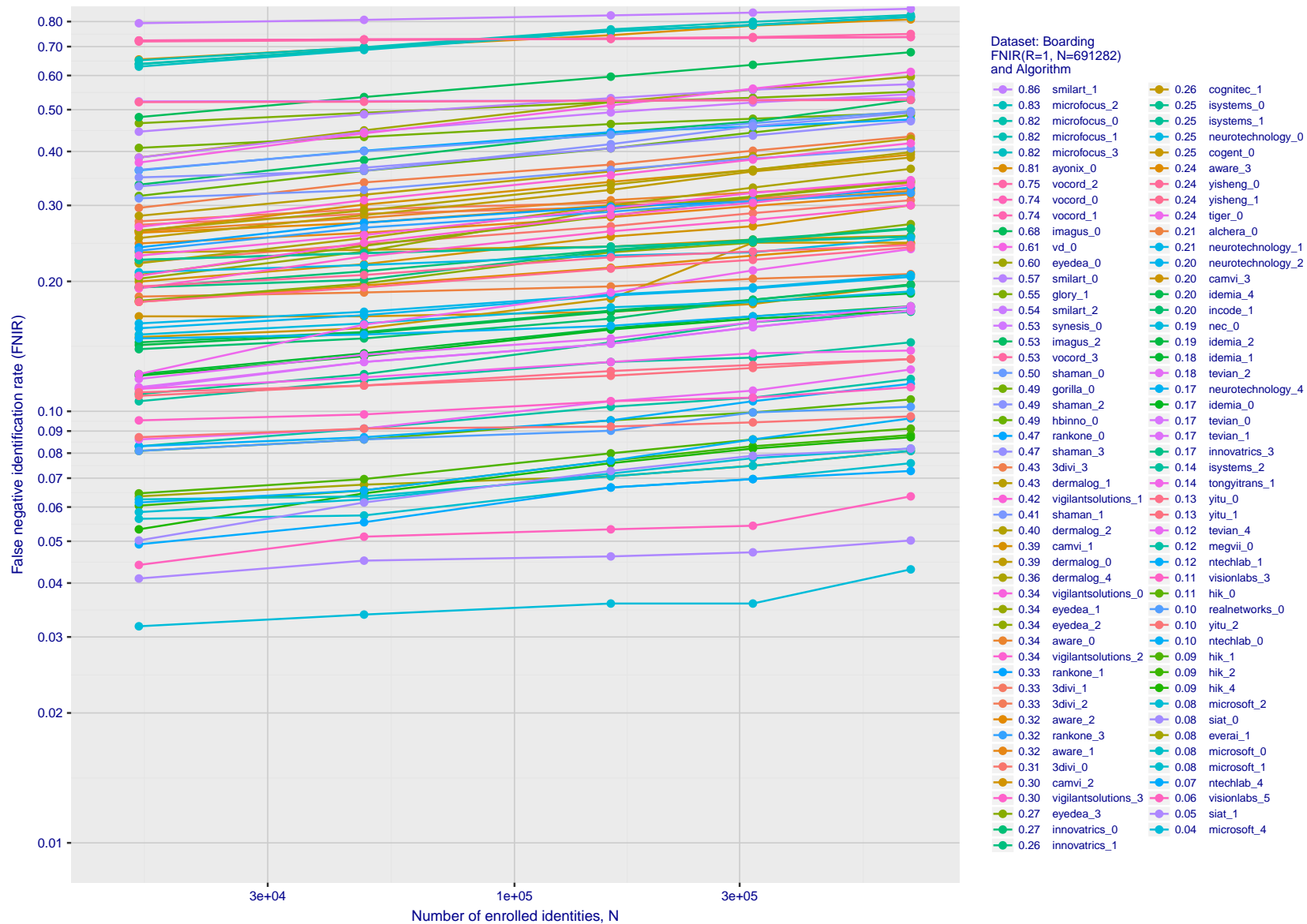


Figure 92: [FRPC Dataset: Boarding] Miss rates vs. number of enrolled identities. The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of people crossing walking toward an aircraft boarding pass reader, using it, then proceeding left across the optical axis 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.

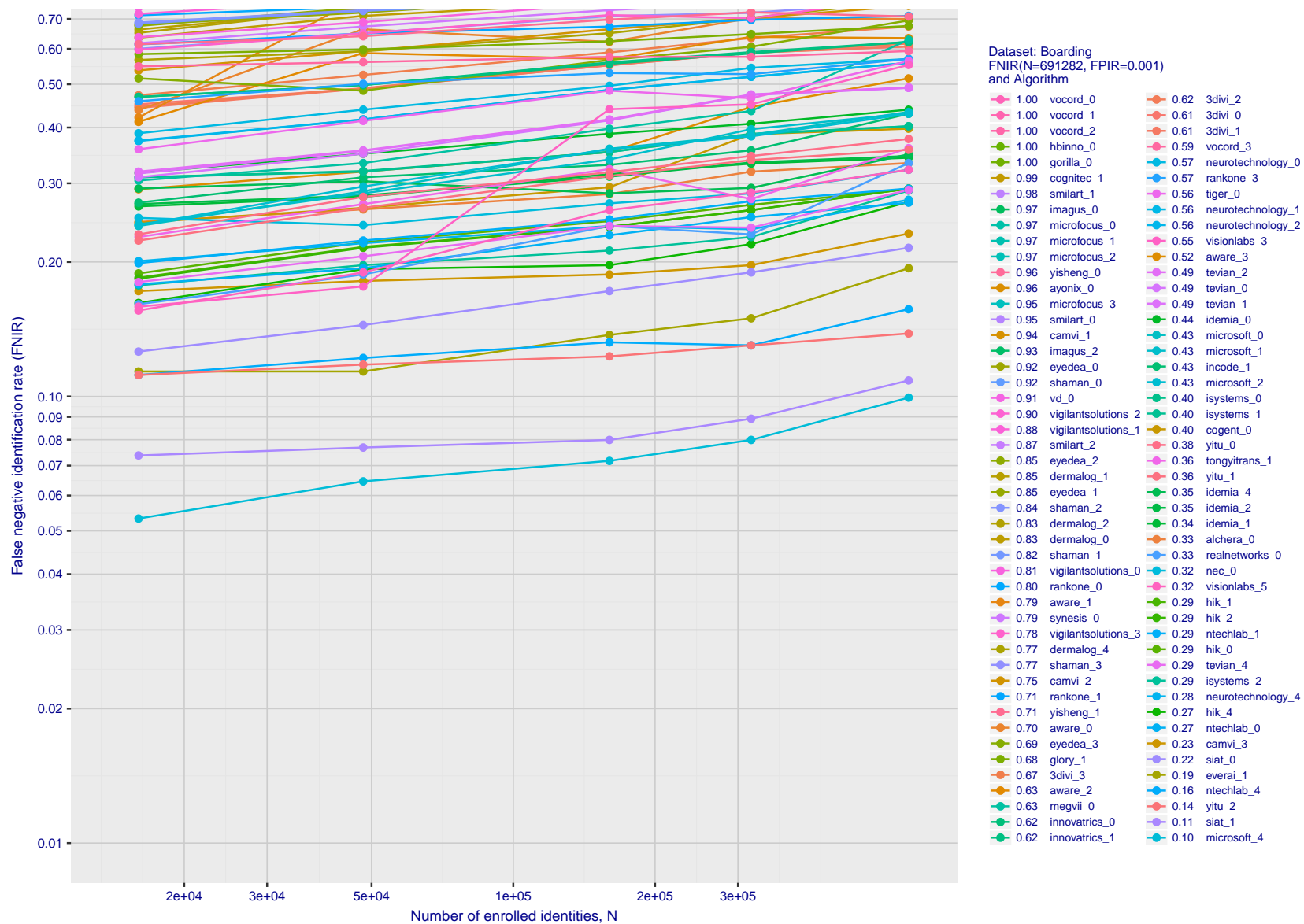


Figure 93: [FRPC Dataset: Boarding] Miss rates vs. number of enrolled identities. The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of people crossing walking toward an aircraft boarding pass reader, using it, then proceeding left across the optical axis 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.

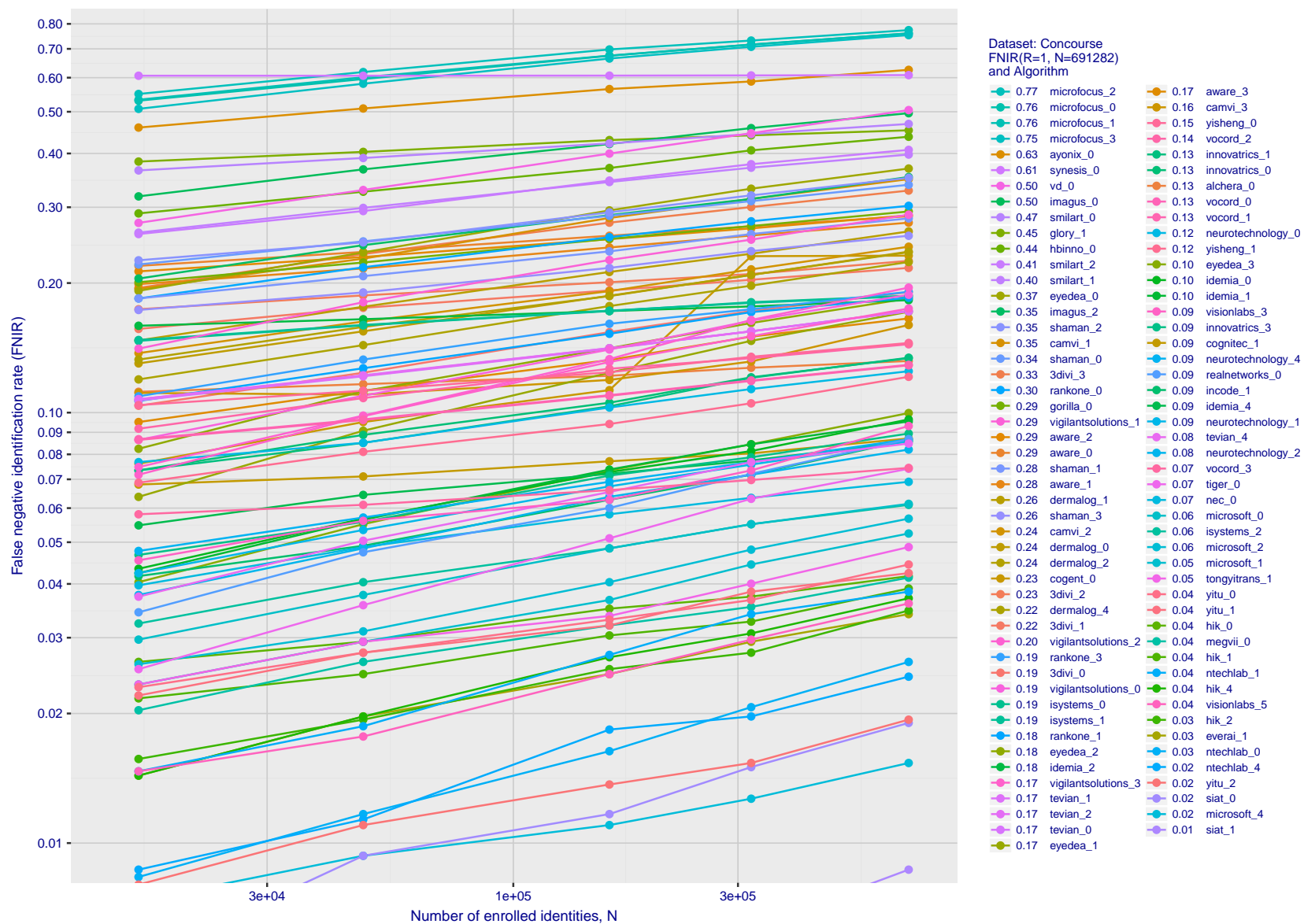


Figure 94: [FRPC Dataset: Concourse] Miss rates vs. number of enrolled identities. The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of people walking down a travel concourse, searched against well-controlled, portrait images of up to 691 282 individuals 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.

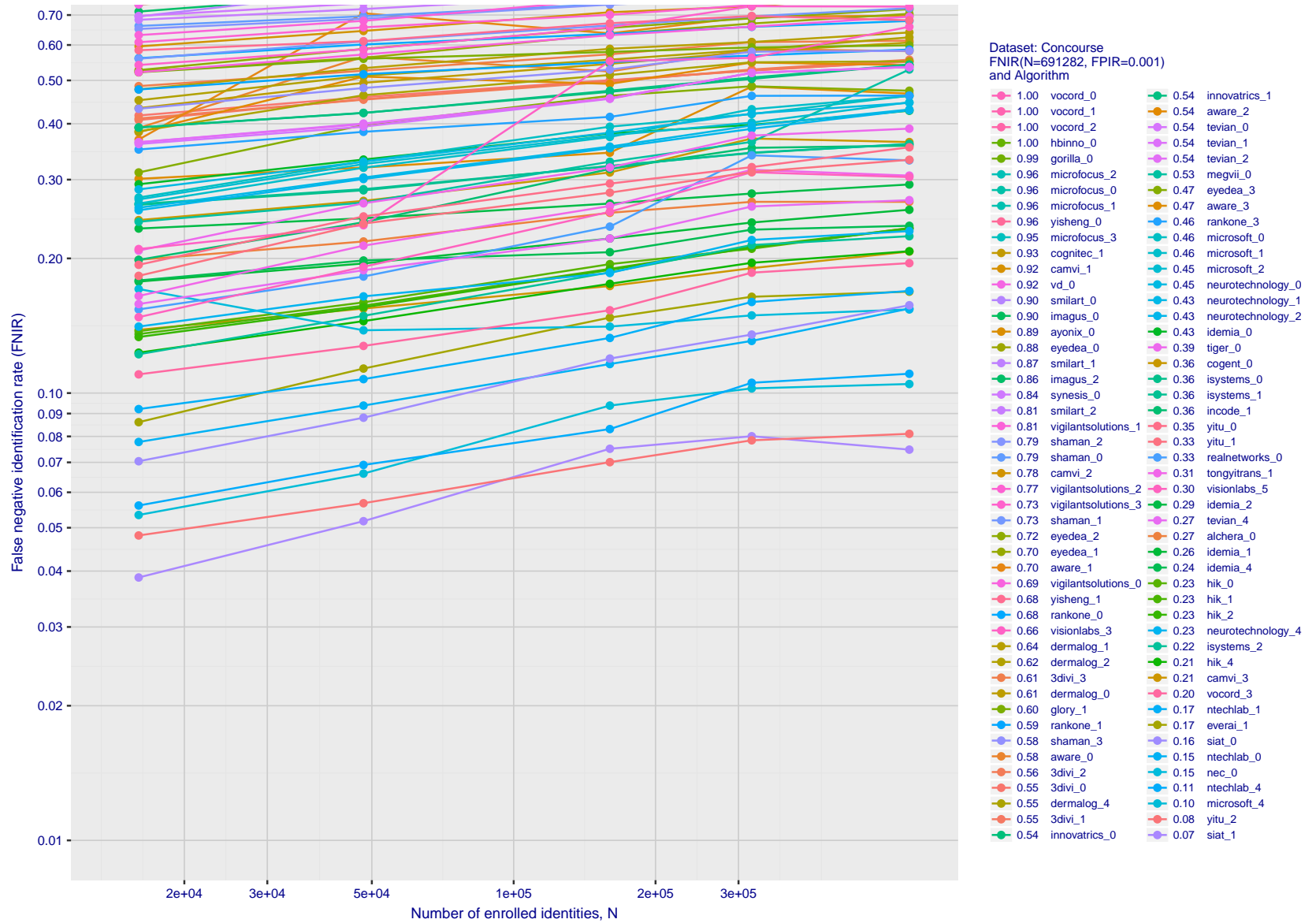


Figure 95: [FRPC Dataset: Concourse] Miss rates vs. number of enrolled identities. The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of people walking down a travel concourse, searched against well-controlled, portrait images of up to 691 282 individuals 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.

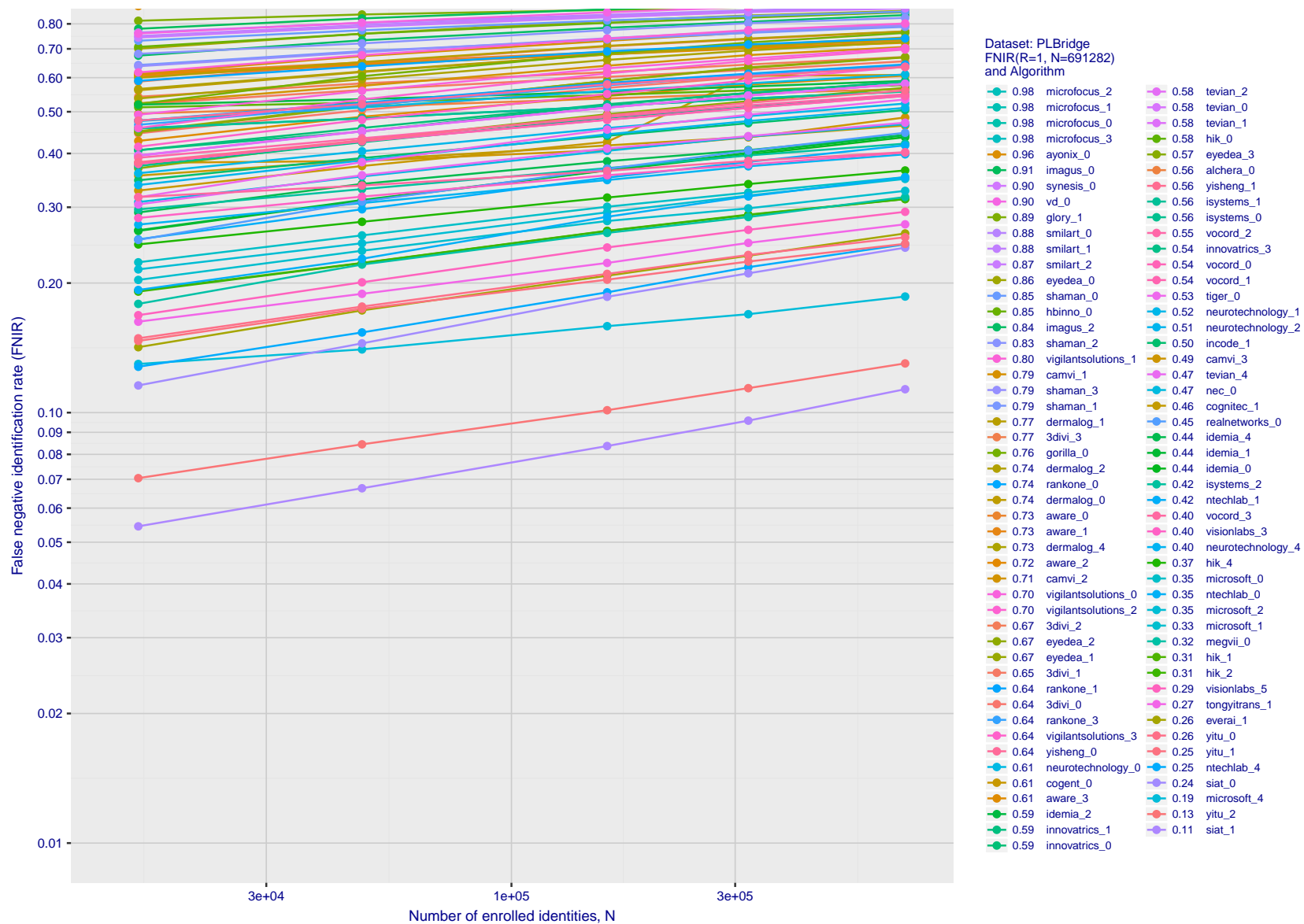


Figure 96: [FRPC Dataset: Passenger Loading Bridge] Miss rates vs. number of enrolled identities. The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of subjects walking along a purpose-built simulated passenger loading bridge, searched against well-controlled, 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.

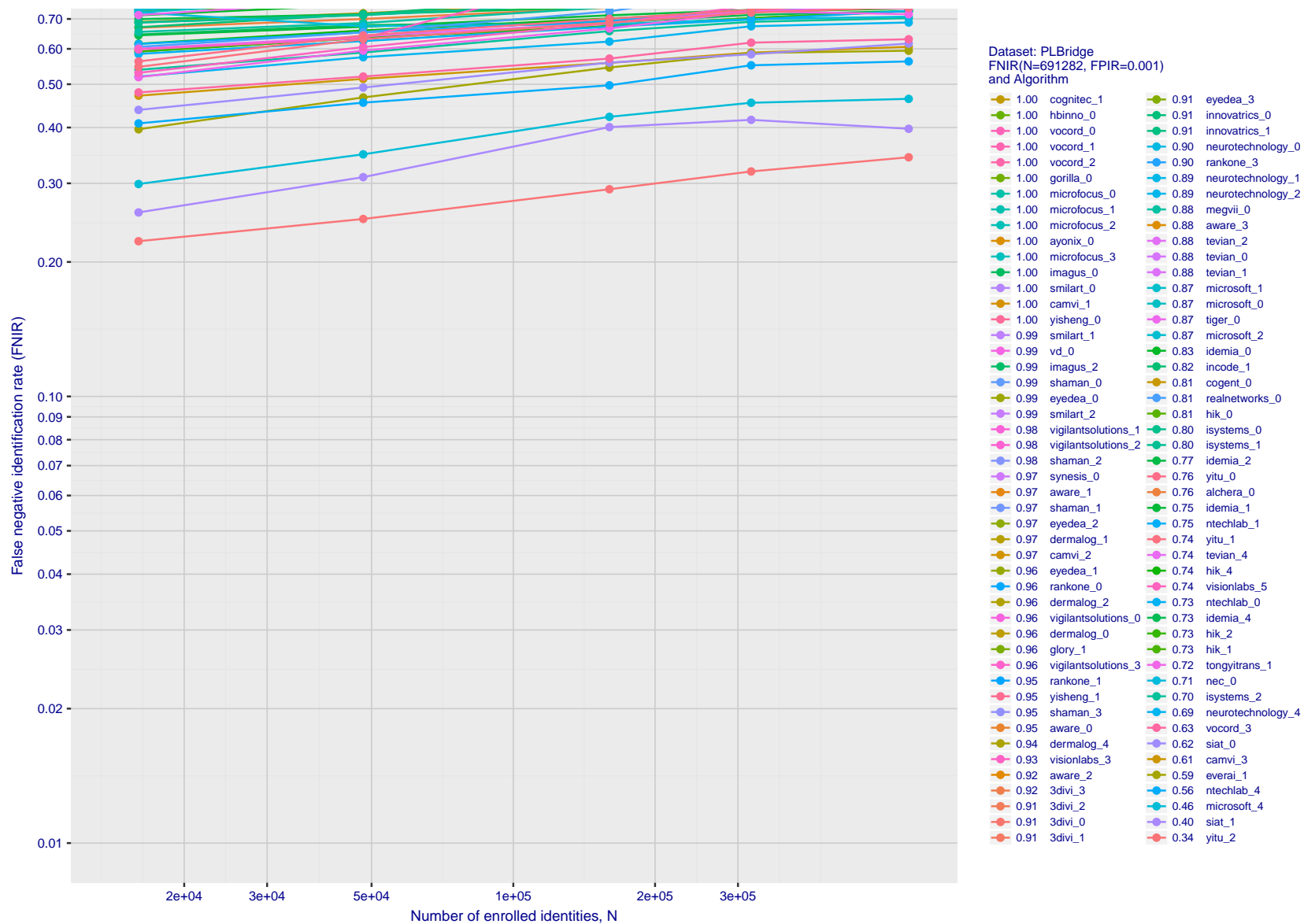


Figure 97: [FRPC Dataset: Passenger Loading Bridge] Miss rates vs. number of enrolled identities. The figure shows accuracy of algorithms on non-cooperative face images cropped from video footage of subjects walking along a purpose-built simulated passenger loading bridge, searched against well-controlled, 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

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

2018/11/26
07:24:51FNIR(N, R, T) =
FPIR(N, T) =False neg. identification rate
False pos. identification rateN = Num. enrolled subjects
R = Num. candidates examined

T = Threshold

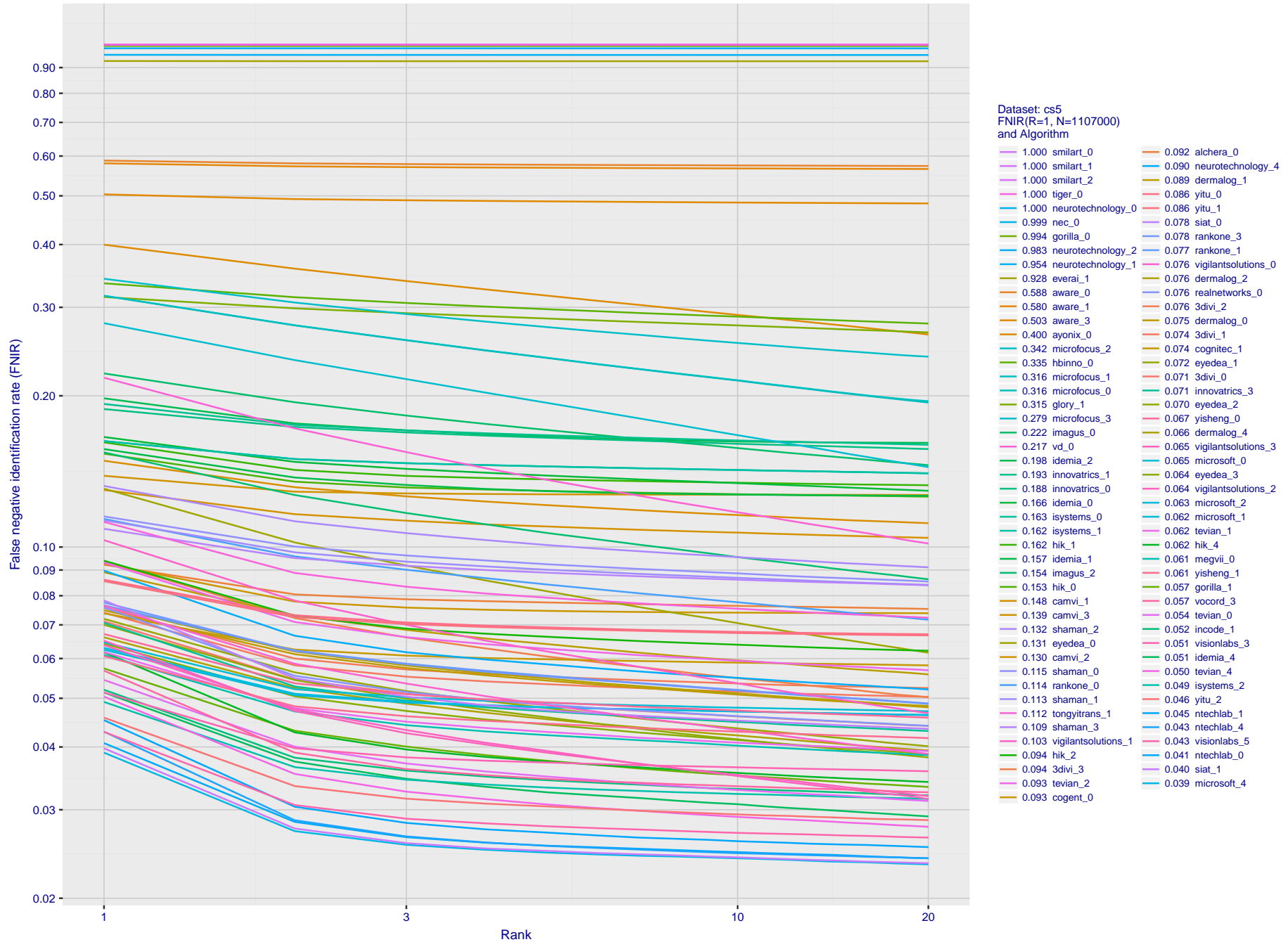
T = 0 → Investigation
T > 0 → Identification

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.

2018/11/26
07:24:51FNIR(N, R, T) =
FPIR(N, T) =False neg. identification rate
False pos. identification rateN = Num. enrolled subjects
R = Num. candidates examined

T = Threshold

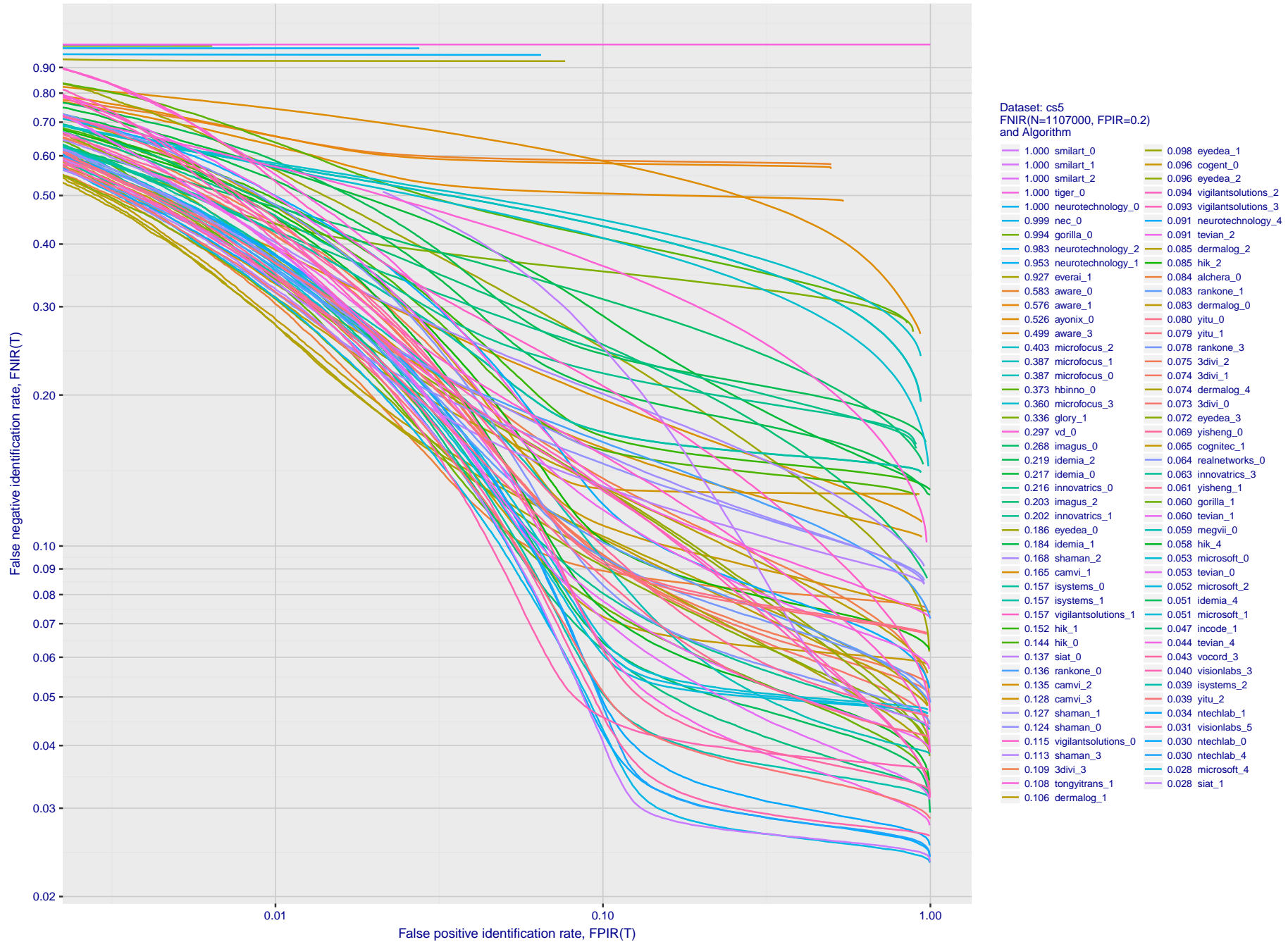
T = 0 → Investigation
T > 0 → Identification

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

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

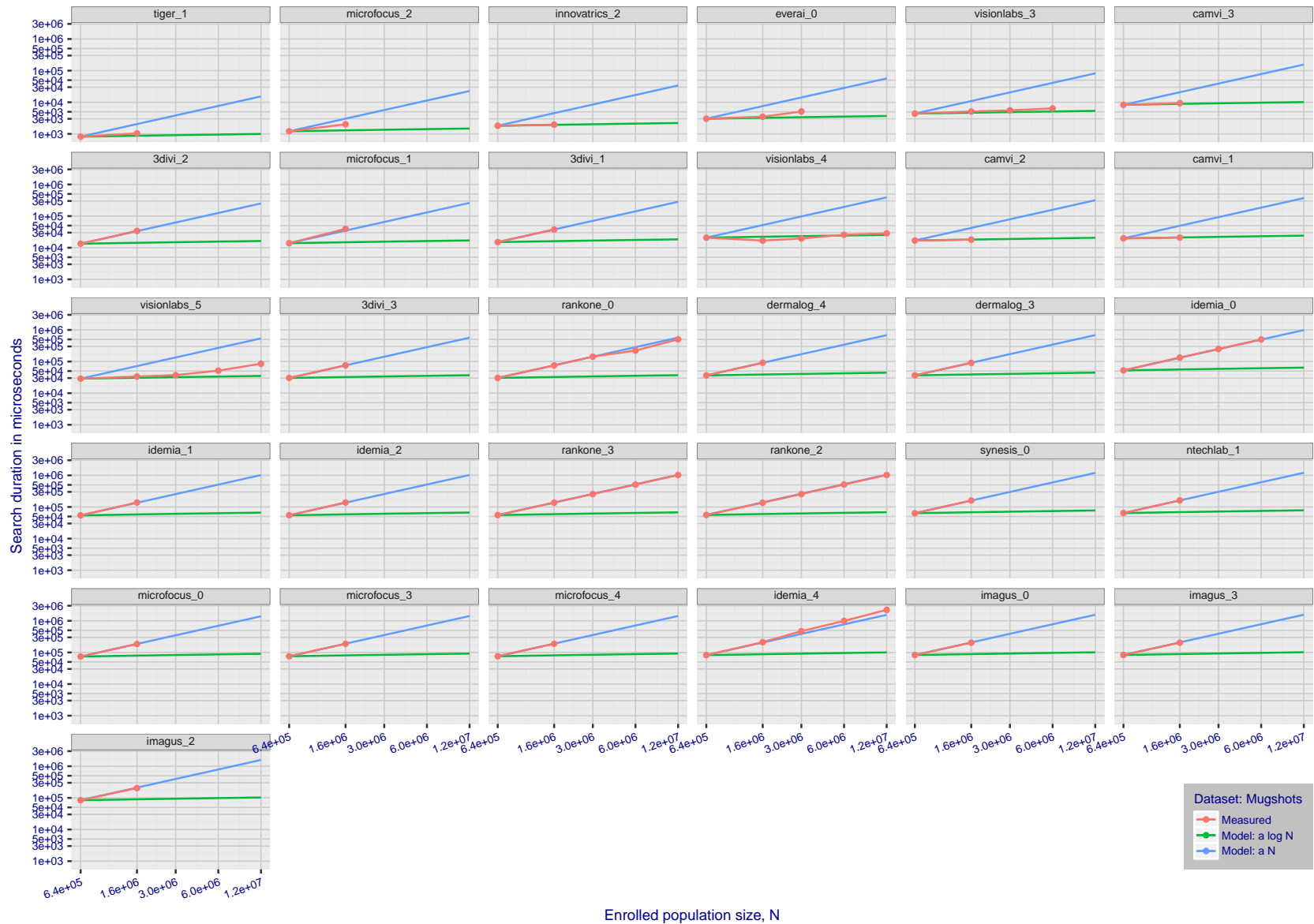


Figure 100: **[Mugshot Dataset] Search duration vs. enrolled population size.** The red line shows actual durations measured on single c. 2016 core. The blue shows linear growth from $N = 640000$. The green line shows logarithmic growth from that point. The red lines often covers blue. Notable sublinear 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.

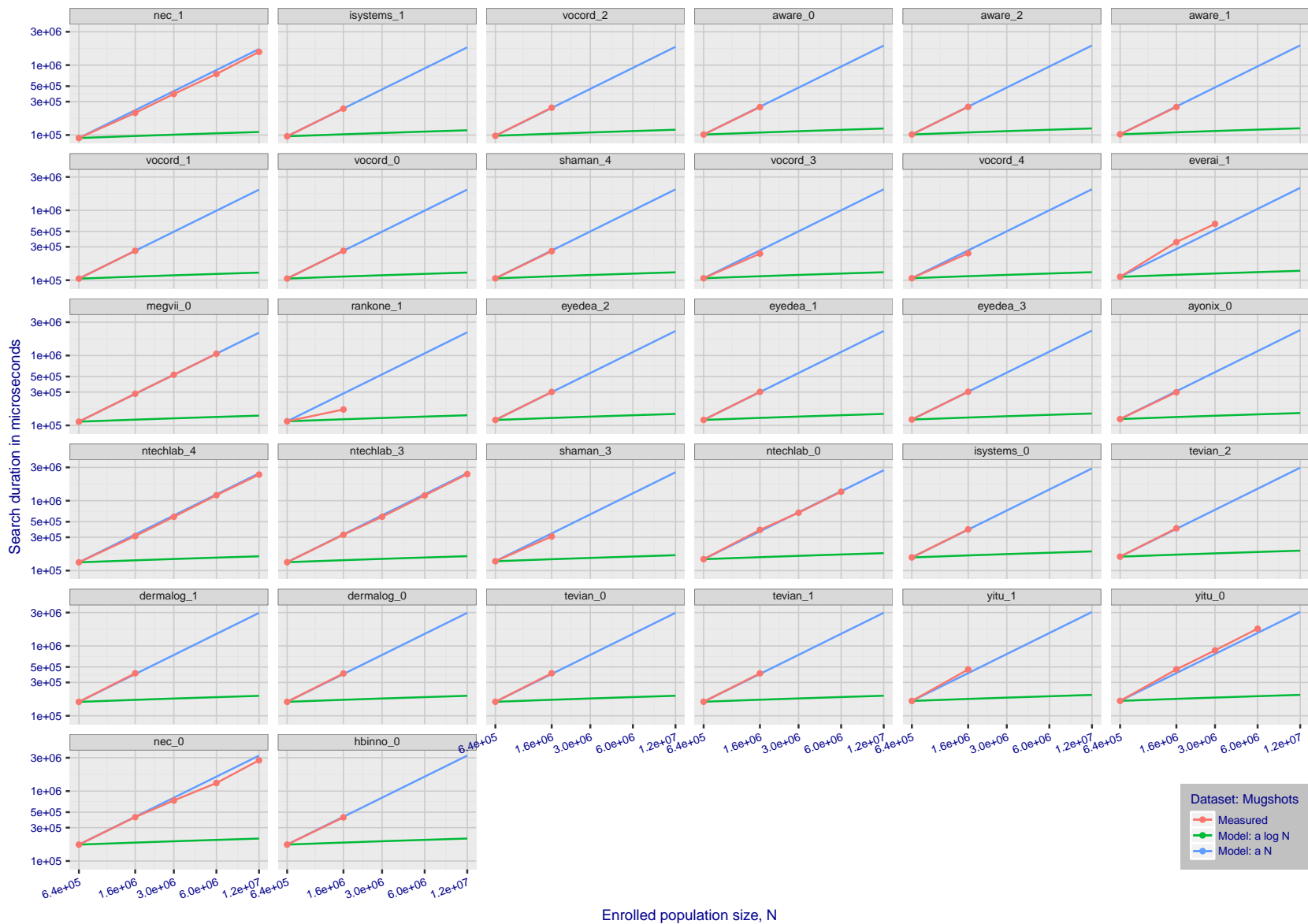


Figure 101: [Mugshot Dataset] Search duration vs. enrolled population size. The red line shows actual durations measured on single c. 2016 core. The blue shows linear growth from $N = 640000$. The green line shows logarithmic growth from that point. The red lines often covers blue. Notable sublinear 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.

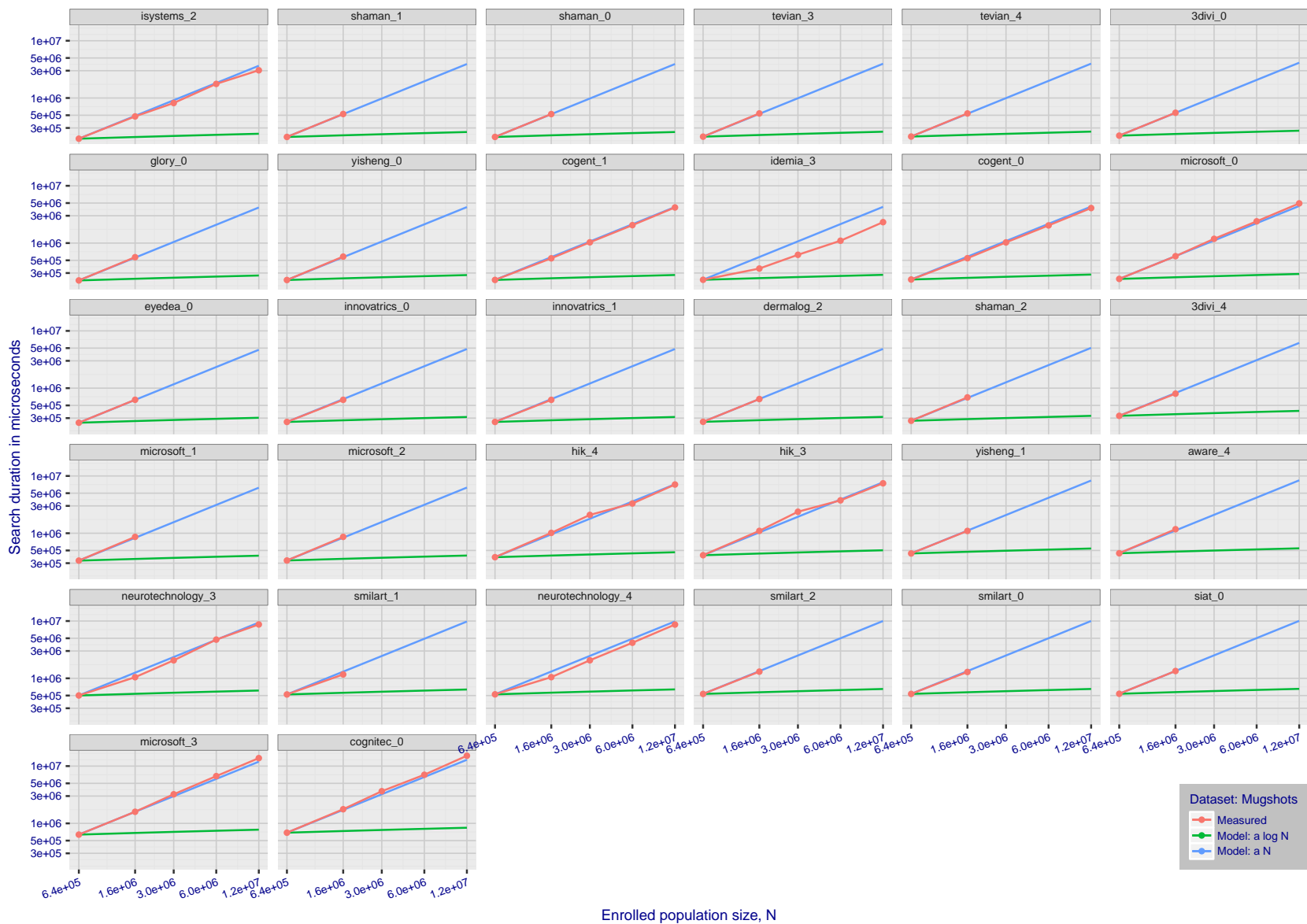


Figure 102: **[Mugshot Dataset] Search duration vs. enrolled population size.** The red line shows actual durations measured on single c. 2016 core. The blue shows linear growth from $N = 640000$. The green line shows logarithmic growth from that point. The red lines often covers blue. Notable sublinear 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.

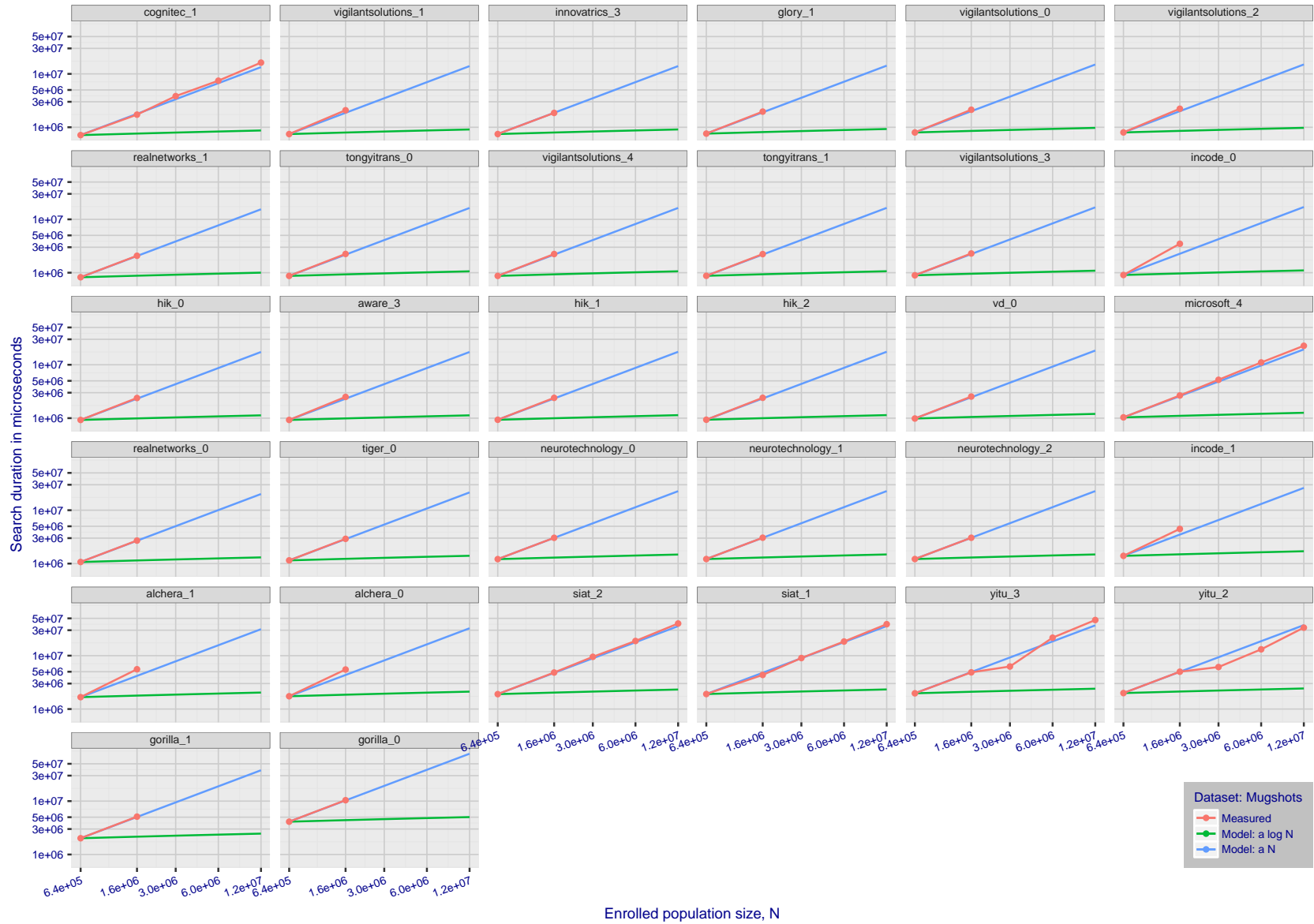


Figure 103: **[Mugshot Dataset] Search duration vs. enrolled population size.** The red line shows actual durations measured on single c. 2016 core. The blue shows linear growth from $N = 640000$. The green line shows logarithmic growth from that point. The red lines often covers blue. Notable sublinear 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.

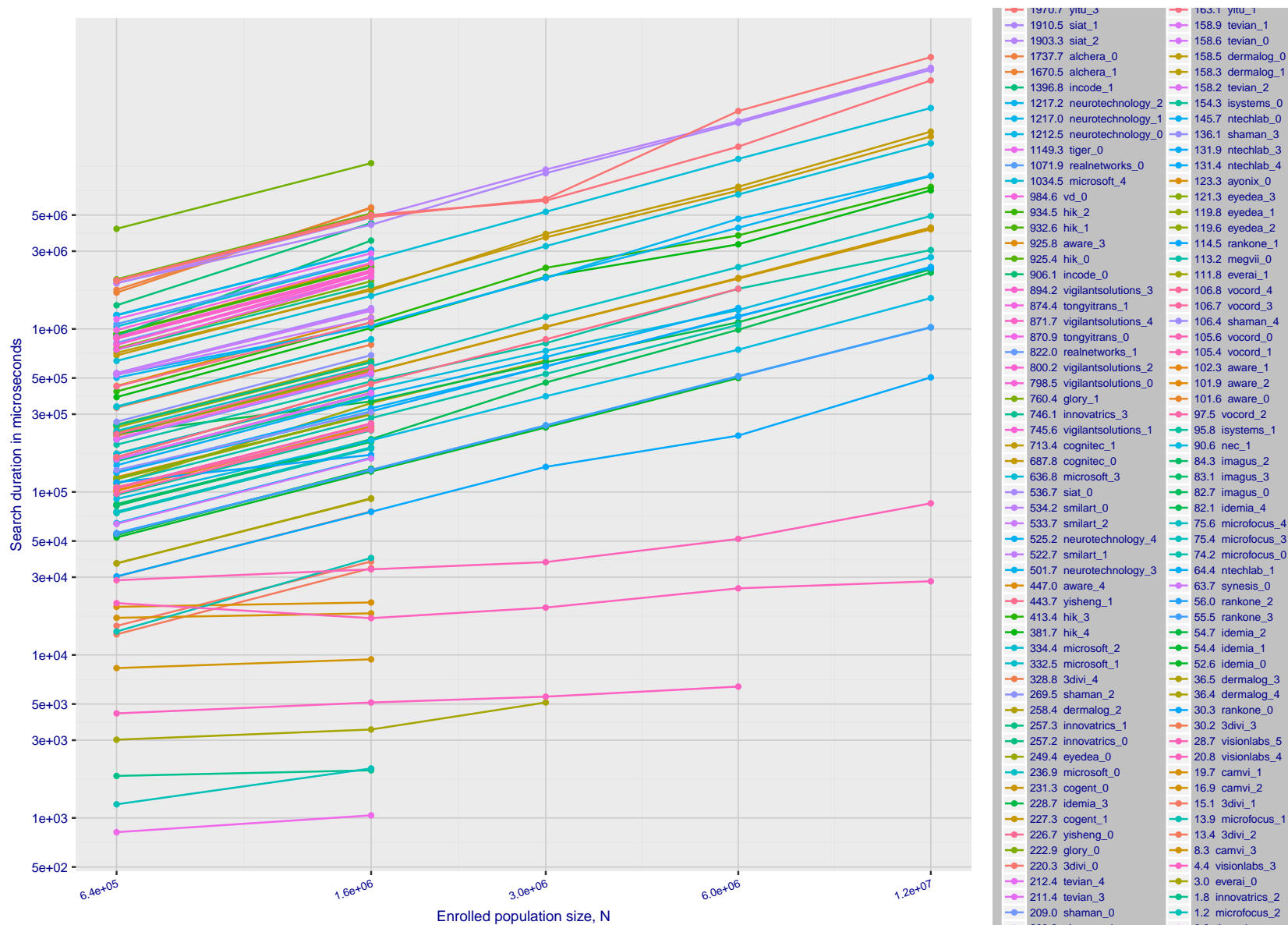


Figure 104: [Mugshot Dataset] Search duration vs. enrolled population size. Alternative visualization of the same data as shown in Figure 103. Generally, only the more accurate algorithms were run on galleries with $N \geq 3\,000\,000$.

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 Visser, 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.