# On Missing Scores in Evolving Multibiometric Systems

Melissa R Dale
Michigan State University
East Lansing, Michigan 48824
Email: dalemeli@msu.edu

Anil Jain
Michigan State University
East Lansing, Michigan 48824
Email: jain@cse.msu.edu

Arun Ross
Michigan State University
East Lansing, Michigan 48824
Email: rossarun@msu.edu

*Abstract*—The use of multiple modalities (e.g., face and fingerprint) or multiple algorithms (e.g., three face comparators) has shown to improve the recognition accuracy of an operational biometric system. Over time a biometric system may evolve to add new modalities, retire old modalities, or be merged with other biometric systems. This can lead to scenarios where there are missing scores corresponding to the input probe set. Previous work on this topic has focused on either the verification or identification tasks, but not both. Further, the proportion of missing data considered has been less than 50%. In this work, we study the impact of missing score data for both the verification and identification tasks. We show that the application of various score imputation methods along with simple sum fusion can improve recognition accuracy, even when the proportion of missing scores increases to 90%. Experiments show that fusion after score imputation outperforms fusion with no imputation. Specifically, iterative imputation with K nearest neighbors consistently surpasses other imputation methods in both the verification and identification tasks, regardless of the amount of scores missing, and provides imputed values that are consistent with the ground truth complete dataset.

## I. INTRODUCTION

Biometrics refers to the task of recognizing individuals based on their unique physical or behavioral traits [1][1]. A typical biometric *identification* system aims to determine a potential set of matching identities, while a *verification* system aims to determine if a probe's claimed identity is valid (i.e., genuine) or not (i.e., imposter). To accomplish either goal, the feature set extracted from the probe is compared against a gallery template to produce a similarity score.

Operational or deployed biometric systems evolve as technology evolves. For example, new more advanced sensors may be added to existing systems or outdated sensors removed. In these updated systems, not all enrolled (gallery) identities will have data pertaining to all the modalities. Additionally, consider a scenario where multiple operational biometric systems are to be merged into a single biometric system. The list of available gallery identities for the different modalities may not be the same. In situations like these, the resulting gallery data is likely to be more sparse than the typical biometric galleries often studied in literature. In this paper we explore

how to handle missing similarity scores of a probe through a combination of imputation [2] and score fusion.

## II. BACKGROUND

### A. Biometric Systems

Biometric systems using a single biometric modality are sometimes referred to as *unimodal* systems. Relying on a single biometric cue can negatively impact biometric system performance when the input data (probe) is of low quality, such as low-resolution faces, highly distorted fingerprints, and occluded irides. Additionally, disabilities or illnesses may prevent some subjects from providing a a biometric cue [3], [4]. The use of multiple biometric cues in a biometric system is often referred to as *multibiometrics* and can help address these issues. Leveraging multibiometric approaches has shown to improve recognition performance, enhance system reliability, combat noisy sensor data, address large variations in user samples, and improve the system's security [5]–[8]. However, there are also potential drawbacks if the multiple biometric sources are not combined carefully, including an increase in recognition error, the need for additional sensors, longer recognition times, and sometimes lower user convenience [9].

Once multiple sources of biometric information are identified for an application, the critical question is how to best leverage the information from these sources. *Information fusion* has been used to improve classification performance in classical machine learning and has been popular in multibiometrics [8], [10], [11]. While fusion can occur at any point in the biometric system, this paper focuses on fusion at the similarity score level. Score fusion with multimodal data has been shown to increase the performance of biometric systems [12]–[14]. There are many existing methods to fuse similarity scores across multiple biometric modalities. One popular fusion technique is the *simple sum rule*, which takes the mean of available scores to produce a fused score [1]. The simple sum fusion approach is a transformation-based approach to fusion, as the scores are first transformed into a common domain prior to fusing (e.g., scores are normalized into the range [0-1]). Simple sum fusion is a popular choice because of its straightforward approach, flexibility in the data

---

[1]The terminology used in this article follows current recommendations for multibiometric fusion defined by International Organization for Standardization [ISO], 2015 [2]

[2]In statistics, imputation refers to the process of replacing missing values with an estimated value

it can be applied to, and it often produces desired results. The work in this paper utilizes the simple sum fusion approach.

When performing multimodal fusion, it should be noted that auxiliary information may also be combined with score data to improve performance. This auxiliary information can include quality measures [15], [16], soft biometrics [17], and demographic information [18], [19]. Incorporating additional information into the fusion can require additional design decisions and computational resources.

The papers discussed above focus on the verification task (one-to-one comparison), which is a binary classification problem (e.g., "accept" or "reject"). Identification tasks (one-to-many comparisons) are often more complex due to the multi-class nature of the problem. Nandakumar et al. successfully extend the likelihood ratio fusion from verification tasks to identification tasks [20]. However, to achieve competitive performances on the identification task, the authors propose a hybrid fusion approach which not only leverages the similarity scores but also incorporates the rank information in the fusion. The authors demonstrate the ability of this approach to handle sparse data by randomly dropping 5%, 10%, and 25% of all partitions in the NIST BSSR1 dataset.

### B. Fusion with Missing Scores

Missing scores can present a challenge when designing a multibiometric system since many fusion methods require score data to be complete. In deciding how to handle missing data, it is important to consider *why* the data is missing. Patterns of missing values are defined by Rubin in [21]:

- **Missing Completely at Random (MCAR):** MCAR describes missing values where the probability that a value is missing is unaffected by other data, whether observed or unobserved. For example, a patient's missing weight value cannot be explained by observed data such as their sex or age, or unobserved data such as a scale's battery malfunctioning. This missing value is MCAR.
- **Missing at Random (MAR):** Values that are MAR are influenced by observed data. Bhaskaran and Smeeth highlight MAR by providing the example of blood pressure records [22]. Records for older people are more likely to be documented because it is more often a regimen of their care. While blood pressure records for younger people may be more sparse compared to their elderly counterparts, this difference can be explained by the observable data of age.
- **Missing Not at Random (MNAR):** If values are neither MCAR nor MAR, missing values are MNAR. An example of a value MNAR is if a patient's drug test is missing *because* they intentionally skipped in order to prevent a positive test value.

Many methods of addressing missing data require the data to be either MCAR or MAR because MNAR has the potential to skew the data. If data can be assumed to be either MCAR or MAR, the following approaches may be applied.

One option for missing data is to simply ignore probes with incomplete scores, i.e., probes which do not have scores

TABLE I: A simple example of a score dataset with missing values, denoted as ?.

| Subject | Face | Fingerprint | Iris |
|---------|------|-------------|------|
| Subject 1 | ? | 0.74 | 1.00 |
| Subject 2 | 0.41 | 0.89 | 0.47 |
| Subject 3 | 0.27 | ? | 0.03 |
| Subject 4 | 0.85 | 0.00 | 0.31 |

corresponding to *all* the modalities. This approach works if there are only a small proportion of missing data and that missing data is truly MCAR. Additionally, this approach may not be an effective use of available data if otherwise usable scores are dropped. For example in the multibiometrics context, consider a probe that obtains a similarity score for the face and iris modalities, but is missing a fingerprint similarity score. When the entire probe is removed from consideration, a valid face similarity score and a valid iris similarity score are dropped from the dataset. Ignoring an entire probe from analysis is referred to as *Listwise Deletion* [23]. When deleting data, it is also important to consider if deleting said data would result in bias. For example, if a sensor in a biometric system is more likely to fail than other sensors in the system, missing scores are not missing at random, but rather likely missing due to the sensor used to collect the data.

Another option to handling missing data is to simply fill missing scores with the modality's mean or median score value. This univariate approach only requires information about the missing modality's scores and is unaffected by other modalities. Again consider a situation like that described above and presented in Table I. Using mean substitution, for example, would replace the face modality's missing score with $0.51$ (the mean of the given score: $0.51 = \frac{0.41 + 0.27 + 0.85}{3}$) and the missing fingerprint score would be replaced with $0.54$. Likewise, the missing scores may be imputed with the median. For this example the face modality's missing score is replaced with the median of the face scores $0.25$ and the missing fingerprint score would be replaced with $0.74$.

Multivariate imputation schemes attempt to map relationships between the modalities. *Multiple Imputations by Chained Equations* (MICE) is a popular approach to multiple imputations, where missing values are temporarily filled with a placeholder value and then iteratively updated using a trained machine learning model [24]. In the given example, shown in Table I, both the face and iris missing values are initially filled with each modality's mean or median. The scores of individual modalities are sequentially and iteratively updated with a specified machine learning classifier. Once the classifier has been trained, the missing values are updated from the initial placeholder value to the value predicted by the trained classifier, and then the next modality's scores are fixed and the classifier is trained again to update the placeholder values. This process is repeated for a specified number of iterations, or until the imputed values stop changing between iterations.

Missing data within individual biometrics can occur in many different ways, as described in the introduction. In order to better understand the role of missing scores in multibiometric

**983**

TABLE II: Summary of the scores present in set1 of the NIST BSSR1 multimodal dataset.

| | Per Modality | Total |
|---|---|---|
| **Total Number of Scores** | 267,289 | 1,069,156 |
| **Genuine Scores** | 517 | 2,068 |
| **Imposter Scores** | 266,772 | 1,067,088 |

settings, researchers often are required to simulate datasets with missing score values [20], [25]–[29]. In each of these prior studies, missing scores are simulated by randomly dropping scores from complete datasets. The percentage of dropped scores range from 5% to 50%. Simulating missing scores allows researchers to know that missing values truly are MAR, and allows for a fine degree of control over the amount of missing data. However, in real world situations, such as the example of incorporating a new modality to an existing biometric system, the proportion of missing data is likely to exceed 50%. In addition to the likelihood estimation approach proposed by [12], additional proposed multimodal imputation methods include classifier based methods such as SVM [30], or Naïve Bayes [31], and deep learning approaches [32].

In this work, we focus on Listwise Deletion, Mean (and Median) Substitution, and MICE to understand how performances may vary between simulated missing score data and real world datasets that naturally contain missing score values. We explore how these methods perform as the amount of missing score data increases in multiple biometric system evolution scenarios.

## III. Experiments

We present 3 sets of experiments modelling 3 real-world biometric scenarios: adding a new modality to an existing multimodal biometric system, merging separate biometric systems into one, and retiring a modality from an existing multi-modal system. To facilitate these experiments, we leverage the NIST BSSR1 dataset [33].[3] This publicly available multimodal dataset is comprised of similarity scores for 4 modalities: one score for the comparison of the user's right index finger to each gallery identity's right index finger, one score for comparison of the user's left index finger to each gallery identity's left index finger, and two scores reported from two facial comparators (referred to as Face Algorithm C and Face Algorithm G).

The NIST BSSR1 Set 1 dataset contains complete score vectors for each of the 517 user comparisons to the 517 templates in the gallery. Table II provides a summary of the scores in this dataset. Figure 3 shows the receiver operating characteristic (ROC) curves for the complete NIST BSSR1 dataset.

In addition to the NIST BSSR1 dataset, we generate from the BIOCOP2008 Ocular dataset. Similarity scores were obtained for both the ocular image and iris image using a trained

Multi-Channel CNN.[4] Figure 1 summarizes this breakdown of the provided images.
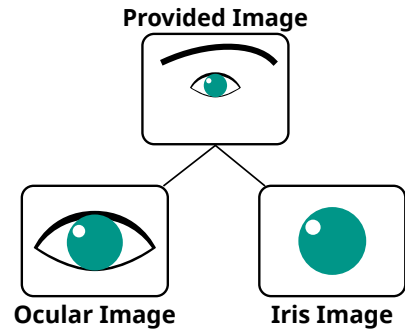


Fig. 1: Breakdown of provided Ocular image in the BIO-COP2008 dataset

If we consider "Ocular" and "Iris" as the two modalities in this dataset, the resulting score dataset is complete. That is, for every iris image, there is a corresponding ocular image and vice versa. This is logical, as each iris image is a subset of a corresponding ocular image. This 2-modality formatting is described in Table III. Figure 4 shows the ROC curves for the complete BIOCOP2008 dataset.

TABLE III: Summary of the scores present in the BIO-COP2008 ocular dataset.

| | Per Modality | Total |
|---|---|---|
| **Total Number of Scores** | 312,606 | 625,212 |
| **Genuine Scores** | 998 | 1,996 |
| **Imposter Scores** | 311,608 | 623,216 |

Each experiment is described in detail below. The experimental design for all experiments is summarized in Table IV and a diagram of the experiments is presented in Figure 2. To perform score fusion, we opt to apply the simple sum fusion as described in Section II. This fusion technique allows us to compare the performance of fusion with imputed scores to the performance of incomplete score vectors. Many other fusion methods (e.g., score fusion with SVM [30], or Naïve Bayes [31]) a complete score vector is required and incomplete score vectors would have to be ignored (listwise deletion). We divide the data into train and test sets, and randomly drop the specified proportion of scores 5 times to obtain mean and standard deviation (s.d.) values.

### A. Experiment 1: Adding a New Modality to an Existing Biometric System

To simulate adding a new modality to a biometric system, we set all but one modality with complete score data, and incrementally introduce scores from the remaining modality. We consider the scenario where we introduce the best performing modality, or the worst performing modality. For the NIST BSSR1, the best modality for both the identification and

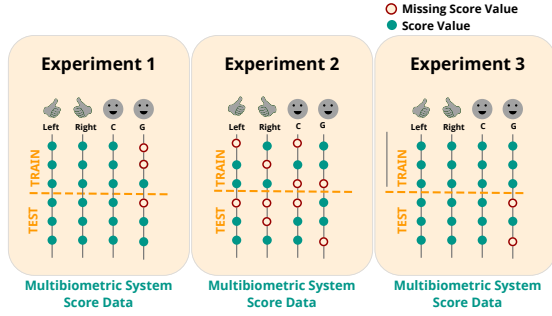Fig. 2: Experiment contexts for NIST BSSR1 dataset.

TABLE IV: Summary of settings used in the experiments.

| Experimental Parameter | Settings |
|---|---|
| Training, Testing Split | 80%, 20% |
| # of Missing Score Simulations | 5 |
| % Missing | [0, 10, 20, 30, 40, 50, 60, 70, 80, 90] |
| Univariate Imputations | Mean Median |
| Multivariate Imputations | Bayesian Regression [34] Decision Tree [35] K Nearest Neighbors [36] |
| Fusion Applied | Simple Sum Fusion |

verification tasks is the right fingerprint. While Face Algorithm C is the poorest performing modality on the verification task, it is the second best on the identification task. Face Algorithm G performs relatively poorly on both verification and identification tasks. So we select Face Algorithm G as the worst modality. For the BIOCOP2008 dataset, performance between the 2 modalities are comparable; we consider the Ocular modality the best and the Iris modality the worst performing.

### B. Experiment 2: Merging Biometric Systems

When merging separate biometric systems, it is very unlikely that the systems share the same gallery of subjects. For this reason, the merged score data can be very sparse and scores can be missing randomly across all modalities. For this
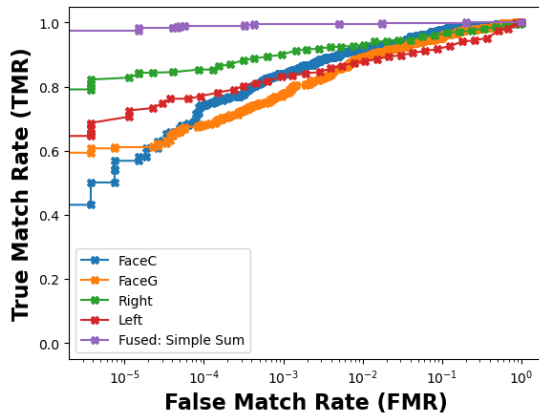


Fig. 3: ROC Curves for the complete NIST BSSR1 dataset
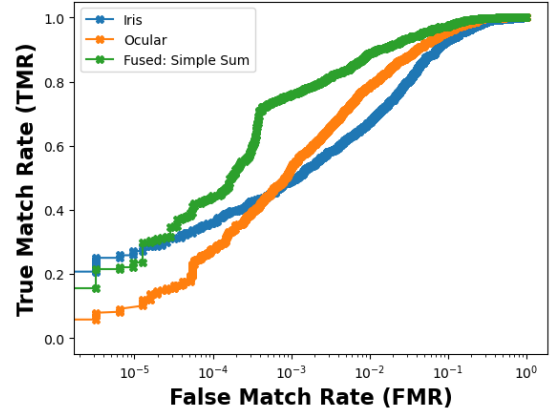


Fig. 4: ROC Curves for the complete BIOCOP2008 dataset

experiment, there is no constraint on the number of missing modalities other than to ensure that each probe has at least one score. For context, in the NIST BSSR1 with 90% missing scores, out of the 1,069,156 probes only 145,383 ($\pm$) 257 probes contain scores for all 4 modalities, and up to 296 probes contain only 1 score.

### C. Experiment 3: Retiring a Modality

The first 2 scenarios describe situations where biometric systems grow larger either through the addition of a new modality or by combining separate biometric systems into one. Conversely, experiment 3 models a scenario where a biometric system loses a modality. An old sensor might begin malfunctioning or the modality data itself might be outdated. To simulate this situation, we train models on a full training score dataset and use these models to impute missing scores for the missing modalities in the testing set.

The following describes the questions addressed with this experiment:

- If a modality is retired, is it better to continue using imputed scores or retrain the biometric system without the retired modality?
- When a modality is starting to malfunction, is there a limit on the number of times imputation can be applied before it is no longer helpful?

## IV. RESULTS

In our experiments, we find that imputation and fusion drastically boosts biometric performance for the NIST BSSR1 dataset, both for the verification and identification tasks. Even when 90% of probes are incomplete, NIST BSSR1 recognition performance is drastically improved in all three scenarios. The BIOCOP2008 dataset did not show such strong recognition performance gains, however imputation does appear to bridge the gap between a complete score dataset and analysis performed on an incomplete score dataset. We highlight a few of the results below [5].

[5]Complete results can be viewed https://melissadale.github.io/ICPR2022/
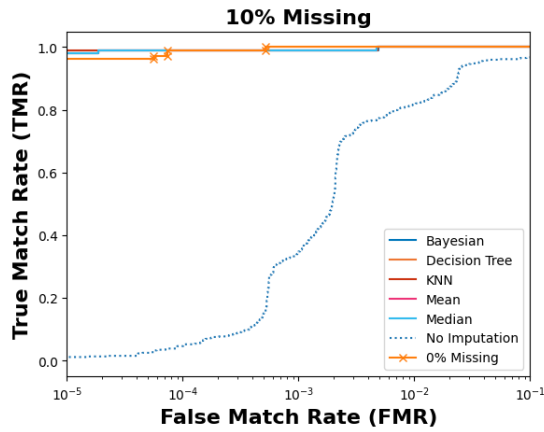
Fig. 5: ROC curves for imputation methods on the 10% missing in the NIST BSSR1 dataset scenario in Experiment 1. The curves corresponding to fusion with imputed values overlap near the top with the 0% missing baseline curve, which is shown in pink.

## A. Experiment 1: Adding a New Modality to an Existing Biometric System

**Verification Tasks** For the NIST BSSR1 dataset, all imputation methods provide recognition performance competitive to the complete dataset, even when 90% of scores are missing from either the best performing modality (Right Index Fingerprint) or the worst performing modality (Face Comparator G), while the fusion performance on the incomplete dataset is very poor across all amounts of missing data, as can be seen in Figure 5, where only 10% of the Right Fingerprint modality's scores are missing.

Figure 6 provides the ROC plot for the BIOCOP2008 dataset analysis for experiment 1 when 40% of scores are missing from the ocular modality. We see that no one approach is significantly better than the other, however we see that generally the performance of imputed score values lies between the complete score dataset and the incomplete, non-imputed score dataset.

These findings suggest that verification performance can be improved when adding a new modality to an existing biometric system, regardless of the individual performance of the modality, by imputing the modality's missing scores before performing fusion.

**Identification Tasks** Identification tasks in experiment 1 in particular appear to benefit from analysis with imputed values for both the NIST BSSR1 dataset and the BIOCOP2008 dataset. In fact, the Rank 1 identification accuracy for the BIOCOP2008 dataset is improved over the performance of the complete dataset, as shown in Figure 7.

While performance is not maintained across increasing proportions of missing score data, fusion with imputed values consistently outperforms fusion with incomplete score data.
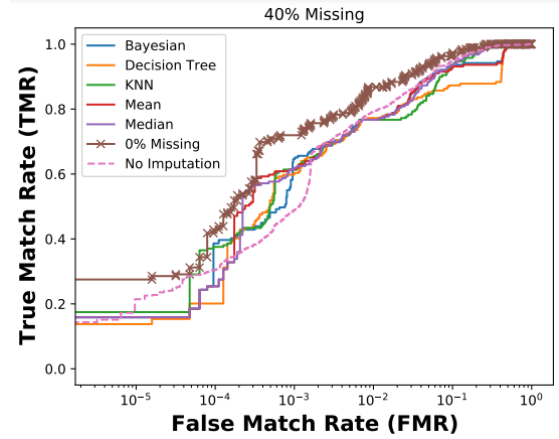


Fig. 6: ROC curves for imputation methods on the 40% missing scenario on the BIOCOP2008 dataset in Experiment 1.
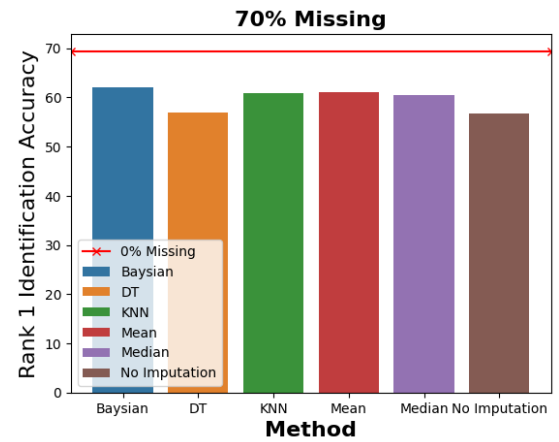


Fig. 7: Rank 1 accuracies for imputation methods on the 10% missing scenario for the BIOCOP2008 dataset in Experiment 1.

## B. Experiment 2: Merging Biometric Systems

The score data simulated in Experiment 2 turns out to be more complicated than the score data in Experiment 1, because each probe may have up to $N - 1$ scores missing, where $N$ is the number of modalities in the biometric system. While the recognition performance is not as strong as the performance in Experiment 1, we can see that both verification and identification tasks benefit from applying imputation methods, even when 90% of the probe data is incomplete.

**Verification Tasks** The verification performance for experiment 2 closely mirrors the verification performance seen in experiment 1 for both NIST BSSR1 dataset and BIOCOP2008 dataset. For example, Figure 8 shows the ROC plot for the BIOCOP2008 dataset when 30% of the probes are incomplete.

**Identification Tasks** Figure 9 provides the identification accuracies across the first 10 ranks in the NIST BSSR1 dataset. We can see that when 90% of probes are incomplete, with
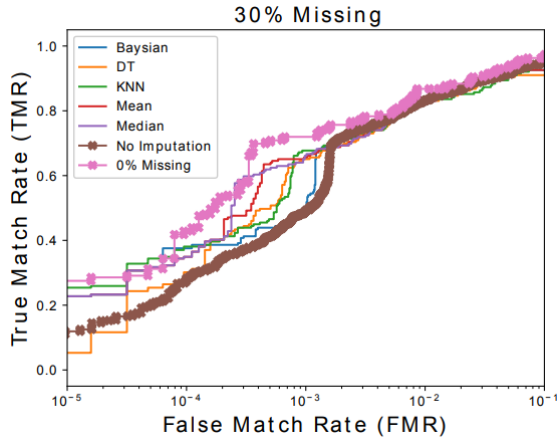
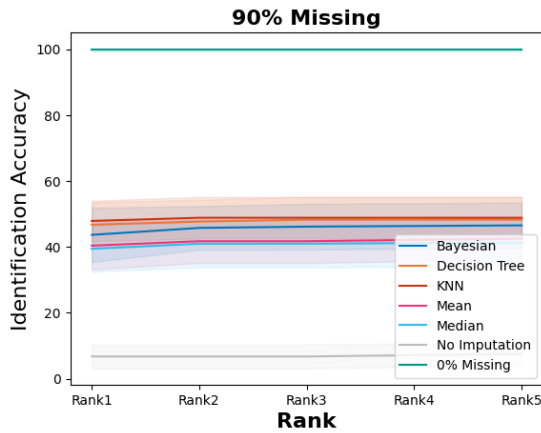Fig. 8: ROC curves for imputation methods on the 30% missing scenario on the BIOCOP2008 in Experiment 2.



Fig. 9: Identification accuracies over the first 10 Ranks for the 90% missing scenario on NIST BSSR1 in Experiment2.

some incomplete probes containing only 1 similarity score out of a possible 4 modalities, performance is improved with imputation, especially for the KNN and Decision Tree iterative imputation techniques.

### C. Experiment 3: Retiring a Modality

When removing a modality from the biometric system, we again see that using imputed scores perform better than applying simple sum fusion without imputation. However, we also see that retraining without the retired modality often yields performances similar to the complete dataset. , as seen in Figure 10 for verification tasks and Table V for identification tasks. This suggest that it may be prudent to retrain a biometric system when retiring a modality.

## V. CONCLUSIONS

In this study, we show that applying imputation to incomplete score data in a multimodal biometric system can improve the performance of both verification and identification tasks. The small, well-groomed multimodal NIST BSSR1
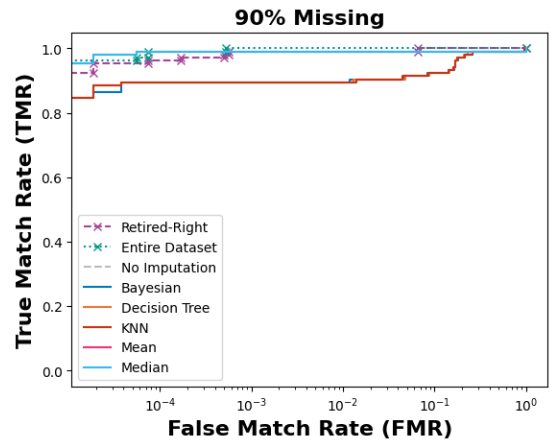


Fig. 10: ROC plot for NIST BSSR1 dataset for experiment 3 when 90% of the probes are incomplete.

TABLE V: The Rank 1, Rank 2, and Rank 3 identification accuracies for the imputation methods on 90% missing data in Experiment 3 on the NIST BSSR1 dataset.

| Method | Rank1 | Rank2 | Rank3 |
|---|---|---|---|
| Bayesian | 97.12% ± 1.52 | 98.85% ± 1.05 | 98.85% ± 1.05 |
| Decision Tree | 95.38 % ± 2.49 | 98.85% ± 1.05 | 98.85 % ± 1.05 |
| KNN | 97.12 % ± 1.52 | 98.85 % ± 1.05 | 98.85% ± 1.05 |
| Mean | 98.27% ± 1.43 | 98.85% ± 1.05 | 98.85 % ± 1.05 |
| Median | 98.27 % ± 1.43 | 98.85 % ± 1.05 | 98.85% ± 1.05 |
| No Imputation | 11.35% ± 3.15 | 12.12% ± 3.09 | 12.31% ± 3.22 |
| Retraining | 99.81 | 99.81 | 99.81 |
| 0% Missing | 100 | 100 | 100 |

dataset benefits greatly from simple sum fusion, with nearly 100% identification Rank 1 accuracy on the complete dataset. While all imputation methods produced strong recognition performances, iterative imputation with KNN provided the best results in both verification and identification tasks. For the BIOCOP2008 dataset, recognition performance gains over the complete data are not as pronounced, however the recognition performance does not decrease due to the imputed values. Furthermore, in a few situations, fusion with imputed similarity score values outperforms fusion with incomplete score vectors. To better understand how biometric systems are impacted by sparse data, larger multimodal datasets are required. Additionally, multimodal datasets that naturally contain missing score data can better illuminate how these methods perform in situations where the data may not be missing at random.

## VI. FUTURE WORK

This study was applied to missing scores, however as mentioned in Section II, it is possible scores are considered missing due poor quality data. Further research is needed to understand the impact of score quality when invoking imputation in multibiometric systems. Additionally, this research has focused on score-level fusion. Further research should be conducted on other levels of the biometric system.

## REFERENCES

[1] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*. Springer Science & Business Media, 2006.

[2] International Organization for Standardization. (2015), *Information technology — Biometrics — Multimodal and other multibiometric fusion*, (ISO/IEC Standard No. 24722) Retrieved from https://www.iso.org/standard/64061.html.

[3] N. V. Boulgouris, K. N. Plataniotis, and E. Micheli-Tzanakou, *Biometrics: Theory, Methods, and Applications*. John Wiley & Sons, 2009.

[4] A. Mishra, "Multimodal biometrics it is: Need for future systems," *International Journal of Computer Applications*, vol. 3, no. 4, pp. 28–33, 2010.

[5] R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 10, pp. 955–966, 1995.

[6] S. Prabhakar and A. K. Jain, "Decision-level fusion in fingerprint verification," *Pattern Recognition*, vol. 35, no. 4, pp. 861–874, 2002.

[7] K.-A. Toh, X. Jiang, and W.-Y. Yau, "Exploiting global and local decisions for multimodal biometrics verification," *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 3059–3072, 2004.

[8] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern recognition letters*, vol. 24, no. 13, pp. 2115–2125, 2003.

[9] L. Allano, B. Dorizzi, and S. Garcia-Salicetti, "Tuning cost and performance in multi-biometric systems: A novel and consistent view of fusion strategies based on the sequential probability ratio test (sprt)," *Pattern Recognition Letters*, vol. 31, no. 9, pp. 884–890, 2010.

[10] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[11] M. Singh, R. Singh, and A. Ross, "A comprehensive overview of biometric fusion," *Information Fusion*, vol. 52, pp. 187–205, 2019.

[12] K. Nandakumar, Y. Chen, S. C. Dass, and A. Jain, "Likelihood ratio-based biometric score fusion," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 342–347, 2007.

[13] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain, "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 450–455, 2005.

[14] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[15] Y. Ding, A. Rattani, and A. Ross, "Bayesian belief models for integrating match scores with liveness and quality measures in a fingerprint verification system," in *International Conference on Biometrics*. IEEE, 2016, pp. 1–8.

[16] Y. Liu, J. Yan, and W. Ouyang, "Quality aware network for set to set recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5790–5799.

[17] H. Kazemi, S. Soleymani, A. Dabouei, M. Iranmanesh, and N. M. Nasrabadi, "Attribute-centered loss for soft-biometrics guided face sketch-photo recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 499–507.

[18] M. Sultana, P. P. Paul, and M. L. Gavrilova, "Social behavioral information fusion in multimodal biometrics," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2176–2187, 2017.

[19] P. P. Paul, M. L. Gavrilova, and R. Alhajj, "Decision fusion for multimodal biometrics using social network analysis," *IEEE Transactions on Systems, Man, and Cybernetics: systems*, vol. 44, no. 11, pp. 1522–1533, 2014.

[20] K. Nandakumar, A. K. Jain, and A. Ross, "Fusion in multibiometric identification systems: What about the missing data?" in *International Conference on Biometrics*. Springer, 2009, pp. 743–752.

[21] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.

[22] K. Bhaskaran and L. Smeeth, "What is the difference between missing completely at random and missing at random?" *International Journal of Epidemiology*, vol. 43, no. 4, pp. 1336–1339, 2014.

[23] H. Kang, "The prevention and handling of the missing data," *Korean Journal of Anesthesiology*, vol. 64, no. 5, p. 402, 2013.

[24] S. Van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in r," *Journal of Statistical Software*, vol. 45, no. 1, pp. 1–67, 2011.

[25] Q. D. Tran, P. Liatsis, B. Zhu, and C. He, "An approach for multimodal biometric fusion under the missing data scenario," in *International Conference on Uncertainty Reasoning and Knowledge Engineering*, vol. 1. IEEE, 2011, pp. 185–188.

[26] Y. Ding and A. Ross, "A comparison of imputation methods for handling missing scores in biometric fusion," *Pattern Recognition*, vol. 45, no. 3, pp. 919–933, 2012.

[27] N. Poh, T. Bourlai, J. Kittler, A. Salah, O. A. John, B. H. Ganster, L. Allano, O. Fatukasi, and O. Ambekar, "A biosecure (ds2) report on the technological evaluation of score-level quality-dependent and cost-sensitive multimodal biometric performance," 2007, citeseer.

[28] O. Fatukasi, J. Kittler, and N. Poh, "Estimation of missing values in multimodal biometric fusion," in *IEEE Second International Conference on Biometrics: Theory, Applications and Systems*, 2008.

[29] N. Damer, B. Führer, and A. Kuijper, "Missing data estimation in multi-biometric identification and verification," in *2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*. IEEE, 2013, pp. 41–45.

[30] H.-A. Park and K. R. Park, "Iris recognition based on score level fusion by using svm," *Pattern Recognition Letters*, vol. 28, no. 15, pp. 2019–2028, 2007.

[31] J. Aravinth and S. Valarmathy, "Multi classifier-based score level fusion of multi-modal biometric recognition and its application to remote biometrics authentication," *The Imaging Science Journal*, vol. 64, no. 1, pp. 1–14, 2016.

[32] S. Maity, M. Abdel-Mottaleb, and S. S. Asfour, "Multimodal biometrics recognition from facial video with missing modalities using deep learning," *Journal of Information Processing Systems*, vol. 16, no. 1, pp. 6–29, 2020.

[33] "NIST biometric scores set," https://www.nist.gov/itl/iad/ig/biometricscores, 2004.

[34] D. J. MacKay, "Bayesian interpolation," *Neural computation*, vol. 4, no. 3, pp. 415–447, 1992.

[35] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Routledge, 2017.

[36] O. Kramer, "K-nearest neighbors," in *Dimensionality reduction with unsupervised nearest neighbors*. Springer, 2013, pp. 13–23.