The Incremental Multiresolution Matrix Factorization Algorithm Supplementary Material

Vamsi K. Ithapu[†], Risi Kondor[§], Sterling C. Johnson[†], Vikas Singh[†]
[†]University of Wisconsin-Madison, [§]University of Chicago
http://pages.cs.wisc.edu/~vamsi/projects/incmmf.html

2. Multiresolution Matrix Factorization

Explanation for Equation 3: This directly follws from Proposition 1 and 2 in [4].

3. Incremental MMF

```
Algorithm 1  InsertRow($(\mathbf{C}, \mathbf{w}, \{t^\ell, s^\ell\}_{\ell=1}^{L})$)
Output: $\{\tilde{t}^\ell, \tilde{s}^\ell\}_{\ell=1}^{L+1}$
  $\tilde{\mathbf{C}}^0 \leftarrow \tilde{\mathbf{C}}$
  $z^1 \leftarrow m+1$
  for $\ell = 1$ to $L-1$ do
    $\{\tilde{t}^\ell, \tilde{s}^\ell, z^{\ell+1}, \mathbf{Q}^\ell\} \leftarrow$ CheckInsert($(\tilde{\mathbf{C}}^{\ell-1}; t^\ell, s^\ell, z^\ell)$)
    $\tilde{\mathbf{C}}^\ell = \mathbf{Q}^\ell \tilde{\mathbf{C}}^{\ell-1} (\mathbf{Q}^\ell)^T$
  end for
  $\mathcal{T} \leftarrow$ GenerateTuples($([m+1] \setminus \cup_{\ell=1}^{L-1} \tilde{s}^\ell(\tilde{\mathbf{C}}))$)
  $\{\tilde{\mathbf{O}}, \tilde{t}^L, \tilde{s}^L\} \leftarrow \text{argmin}_{\mathbf{O}, t \in \mathcal{T}, s \in t}\, \mathcal{E}(\tilde{\mathbf{C}}^{L-1}; \mathbf{O}; t, s)$
  $\mathbf{Q}^L = \mathbf{I}_{m+1}, \quad \mathbf{Q}^L_{\tilde{t}^L, \tilde{t}^L} = \tilde{\mathbf{O}}$
  $\tilde{\mathbf{C}}^L = \mathbf{Q}^L \tilde{\mathbf{C}}^{L-1} (\mathbf{Q}^L)^T$
```

```
Algorithm 2  CheckInsert($(\mathbf{A}, \hat{t}, \hat{s}, z)$)
Output: $\tilde{t}, \tilde{s}, z, \mathbf{Q}$
  $\mathcal{T} \leftarrow$ GenerateTuples($(\hat{t}, z)$)
  $\{\tilde{\mathbf{O}}, \tilde{t}, \tilde{s}\} \leftarrow \text{argmin}_{\mathbf{O}, t \in \mathcal{T}, s \in t}\, \mathcal{E}(\mathbf{A}; \mathbf{O}; t, s)$
  if $\tilde{s} \in z$ then
    $z \leftarrow (z \cup \hat{s}) \setminus \tilde{s}$
  end if
  $\mathbf{Q} = \mathbf{I}_{m+1}, \ \mathbf{Q}_{\tilde{t}, \tilde{t}} = \tilde{\mathbf{O}}$
```

Algorithm 1 adds one extra row/column is to a given MMF, and clearly, the incremental procedure can be repeated as more and more rows/columns are added. Algorithm 3 summarizes this incremental factorization for arbitrarily large and dense matrices. It has two components: an initialization on some randomly chosen small block (of size $\tilde{m} \times \tilde{m}$) of the entire matrix $\mathbf{C}$; followed by insertion of the remaining $m - \tilde{m}$ rows/columns using Algorithm 1 in a streaming fashion The initialization entails computing a batch-wise MMF on this small block ($\tilde{m} \geq k$).

BatchMMF – Exhaustive: Note that at each level ℓ, the error criterion can be explicitly minimized via an exhaustive search over all possible k-tuples from $\mathcal{S}_{\ell-1}$ (the active set) and a randomly chosen (using properties of QR decomposition

Algorithm 3 INCREMENTAL MMF($\mathbf{C}$)

```
Output: $\mathcal{M}(\mathbf{C})$
  $\bar{\mathbf{C}} = \mathbf{C}_{[\tilde{m}],[\tilde{m}]}$, $L = m - k + 1$
  $\{t^\ell, s^\ell\}_1^{\tilde{m}-k+1} \leftarrow$ BATCHMMF($\bar{\mathbf{C}}$)
  for $j \in \{\tilde{m}+1, \ldots, m\}$ do
    $\{t^\ell, s^\ell\}_1^{j-k+1} \leftarrow$ INSERTROW($\bar{\mathbf{C}}, \mathbf{C}_{j,:}, \{t^\ell, s^\ell\}_1^{j-k}$)
    $\bar{\mathbf{C}} = \mathbf{C}_{[j],[j]}$
  end for
  $\mathcal{M}(\mathbf{C}) := \{t^\ell, s^\ell\}_1^L$
```

[7]) dictionary of k^{th} order rotations. If the dictionary is large enough, the exhaustive procedure would lead to the smallest possible decomposition error. However, it is easy to see that this is combinatorially large, with an overall complexity of $\mathcal{O}(n^k)$ [5] and will not scale well beyond $k = 4$ or so. Note from Algorithm 1 that the error criterion $\mathcal{E}(\cdot)$ in this second stage which inserts the rest of the $m - \tilde{m}$ rows is performing an exhaustive search as well.

Other Variants: The are two alternatives that avoid this exhaustive search.

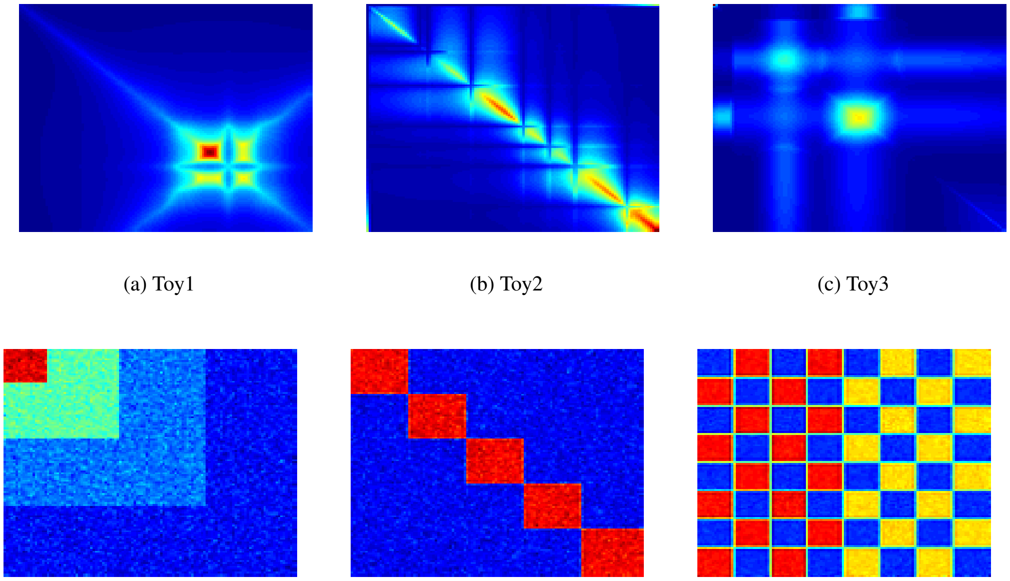
- **BATCHMMF – EIGEN:** Since $\mathbf{Q}^\ell$'s job is to diagonalize some k rows/columns.one can simply pick the relevant $k \times k$ block of $\mathbf{C}^\ell$ and compute the best $\mathbf{O}$ (for a given t^ℓ). Hence the first alternative is to bypass the search over $\mathcal{O}$, and simply use the eigen-vectors of $\mathbf{C}^\ell_{t^\ell, t^\ell}$ for some tuple t^ℓ. Nevertheless, the search over $\mathcal{S}_{\ell-1}$ for t^ℓ still makes this approximation reasonably costly.
- **BATCHMMF – RANDOM:** Instead, the k-tuple selection may be approximated while keeping the exhaustive search over $\mathcal{O}$ intact [5]. Since diagonalization effectively nullifies correlated dimensions, the best k-tuple can be the k rows/columns that are maximally correlated. This is done by choosing some $s_1 \sim \mathcal{S}_{\ell-1}$ (from the current active set), and picking the rest by

$$s_2, \ldots, s_k \leftarrow \operatorname*{argmin}_{s_i \sim \mathcal{S}_{\ell-1} \setminus s_1} \sum_{i=2}^{k} \frac{(\mathbf{C}^{\ell-1}_{:,s_1})^T \mathbf{C}^{\ell-1}_{:,s_i}}{\|\mathbf{C}^{\ell-1}_{:,s_1}\| \|\mathbf{C}^{\ell-1}_{:,s_i}\|} \tag{1}$$

  This second heuristic has been shown to be robust [5], however, for large k it might miss some k-tuples that are vital to the quality of the factorization. Depending on $\tilde{m}$, and the available computational resources at hand, these alternatives can be used instead of the earlier proposed exhaustive procedure for the initialization.

4. Experiments

4.0. Data



(a) Toy1

(b) Toy2

(c) Toy3

(d) Toy4

(e) Toy5

(f) Toy6

Figure 1. Symmetric matrices from six different toy datasets.

We used multiple different toy and real data sets for our experiments. The evaluations in Section 4.1 used 6 different toy/example datasets as shown in Figure 1. These cover a wide-variety of structures in typical symmetric matrices.

The evaluations in Section 4.2 uses two different datasets as described below.

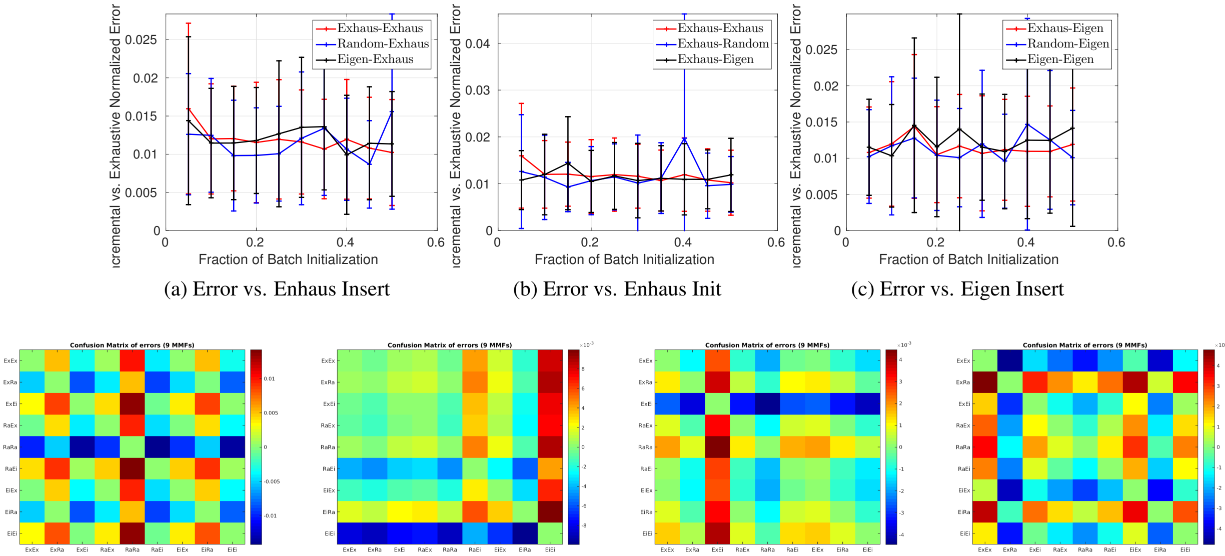
- **PET-ADNI** Positron Emission Tomographic brain images from approximately 1300 subjects are used to extract summaries corresponding to 80 region-of-interests (ROI). This data is publicly available from Alzheimer's Disease Neoruimaging Initiatice [1]. These 80 ROIs are used as predictors (or independent) variables in the evaluations in Section 4.2. The subject age, the corresponding disease status (a discrete ordinal variable) and the genotype for the disease characteristic (binary) are used as the responses (or dependent) variables.
- **Early-Alzheimers** This data comes from healthy middle-aged adults (collected as a part of a local Alzheimer's disease research center). The features include a total of 78 different brain ROI summaries, age, cognitive scores, vascular summaires (like body mass index, cholesterol etc.), demographic information, family history for Alzheimer's disease and genetic status for Apolipoprotein 4E. Here, unlike the **PET-ADNI** case, from all the available covariates (a total of 140), the subject age, the family history and genotype (both are binary), the disease summary (discrete ordinal) and cumulative cognitive summary (discrete ordinal) are used as responses, while the remaining are the predictors.

The four models Mdl1-Mdl4 that were used in Figure 3 from Section 4.2 of the main paper correspond to **PET-ADNI** and **Early-Alzheimers** with disease-status and age as predictors respectively.

The experiments in Section 4.3 uses the *learned* deep networks – AlexNet [6] and VGG-S [3]. The networks were learned on ImageNet data [8], and we will be using the corresponding hidden representations from an overall 55 different classes – coming from the popular synsets listed on the ImageNet repository [2]. These 55 different categories as as follows; and they will also be listed out appropriately via the visualizations (both in this document and the parent webpage).

55 *classes:* bag, ball, basket, basketball, bathtub, bench, blackboard, bottle, box, building, chalkboard, edifice, saute, margarine, bannock, chapati, pita, limpa, shortcake, strawberry, salad, ketchup, chutney, salsa, puppy, green lizard, garden spider, ptarmigan, kangaroo, possum, cow, insectivore, killer whale, hound, male horse, warhorse, pony, mule, zebra, bison, sheep, goat, blackbuck, deer, elk, pot, rack, roller coaster, rule, sail, sheet, ski, couch, racket, stick, table, toilet seat.

4.1. Incremental vs. Batch MMF



(a) Error vs. Enhaus Insert

(b) Error vs. Enhaus Init

(c) Error vs. Eigen Insert

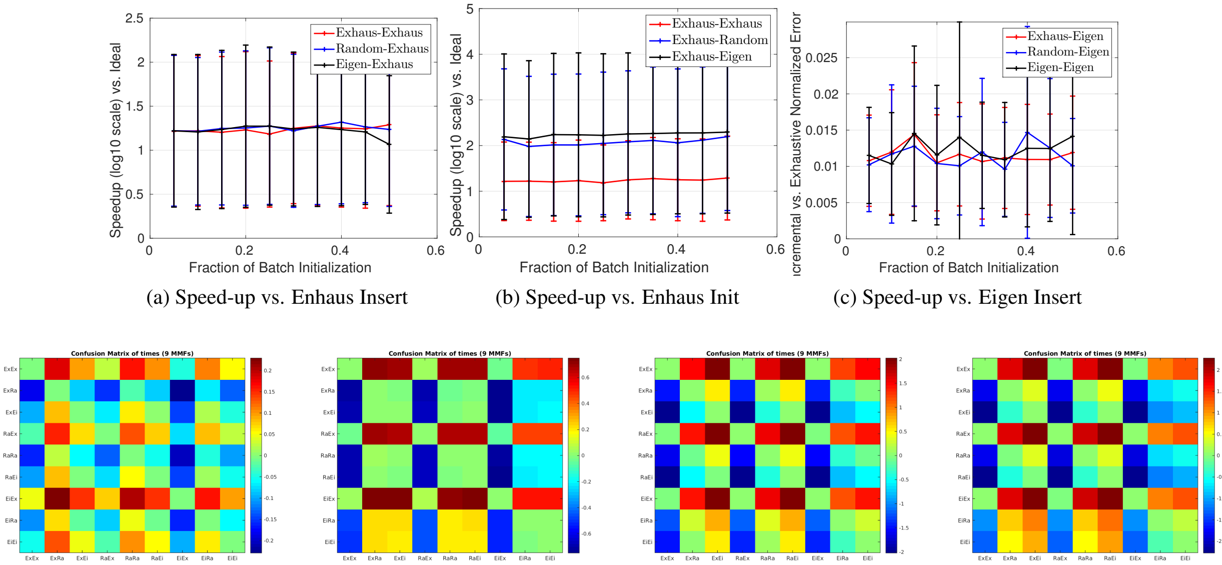
(d) Error Conf. Mat.

(e) Error Conf. Mat.

(f) Error Conf. Mat.

(g) Error Conf. Mat.

Figure 2. **Incremental versus Batch MMFs on Toy1 dataset [ERROR]:** Comparing factorization errors between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization errors between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Speed-up vs. Enhaus Insert

(b) Speed-up vs. Enhaus Init

(c) Speed-up vs. Eigen Insert

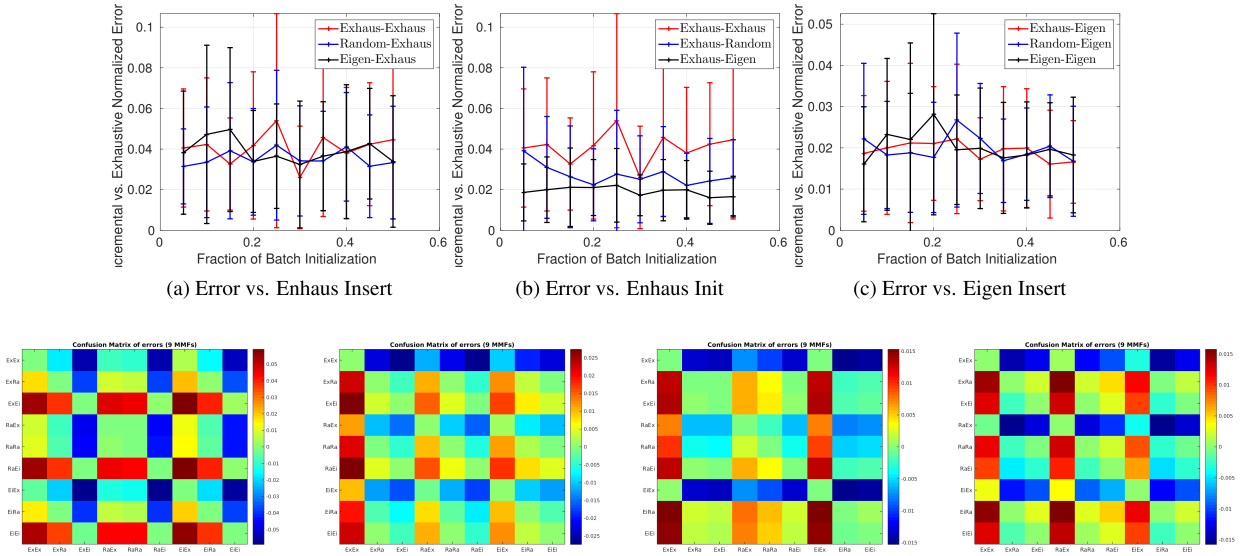
(d) Time Conf. Mat.

(e) Time Conf. Mat.

(f) Time Conf. Mat.

(g) Time Conf. Mat.

Figure 3. **Incremental versus Batch MMFs on Toy1 dataset [SPEED-UP]:** Comparing factorization speed-ups between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization times between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Error vs. Enhaus Insert

(b) Error vs. Enhaus Init

(c) Error vs. Eigen Insert

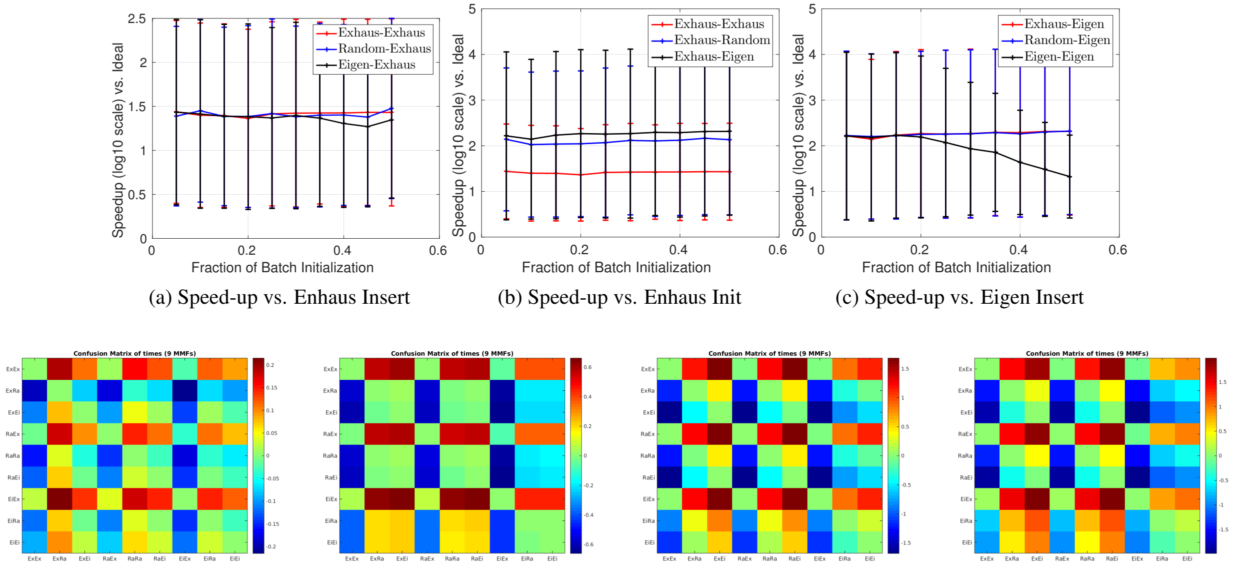
(d) Error Conf. Mat.

(e) Error Conf. Mat.

(f) Error Conf. Mat.

(g) Error Conf. Mat.

Figure 4. **Incremental versus Batch MMFs on Toy2 dataset [ERROR]:** Comparing factorization errors between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization errors between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Speed-up vs. Enhaus Insert

(b) Speed-up vs. Enhaus Init

(c) Speed-up vs. Eigen Insert

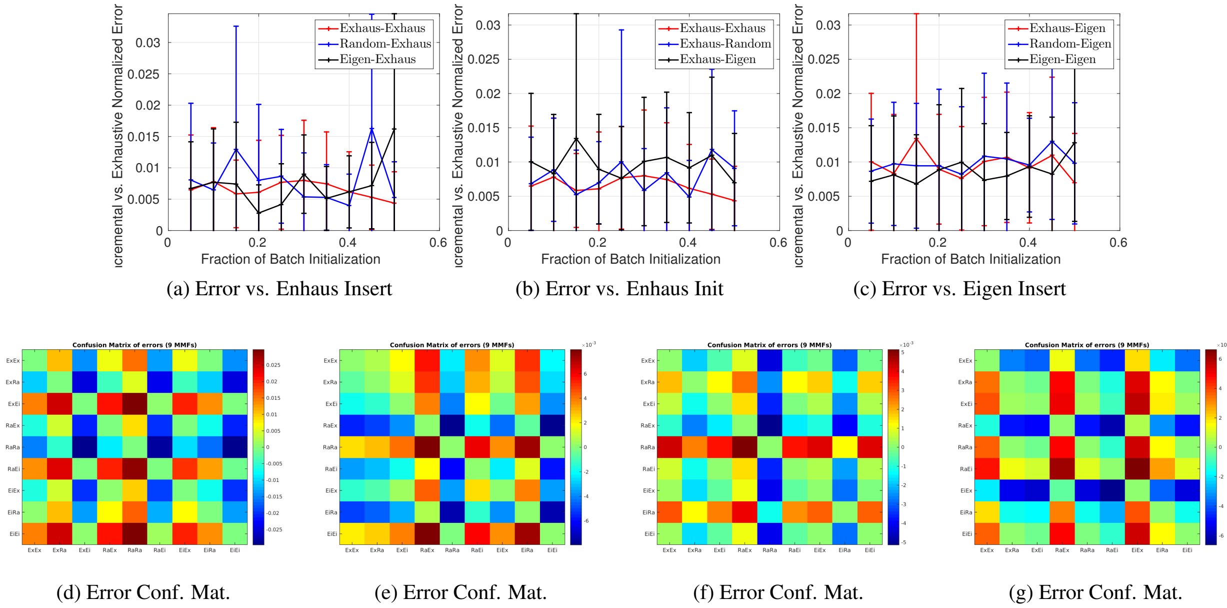
(d) Time Conf. Mat.

(e) Time Conf. Mat.

(f) Time Conf. Mat.

(g) Time Conf. Mat.

Figure 5. **Incremental versus Batch MMFs on Toy2 dataset [SPEED-UP]:** Comparing factorization speed-ups between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization times between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Error vs. Enhaus Insert

(b) Error vs. Enhaus Init

(c) Error vs. Eigen Insert

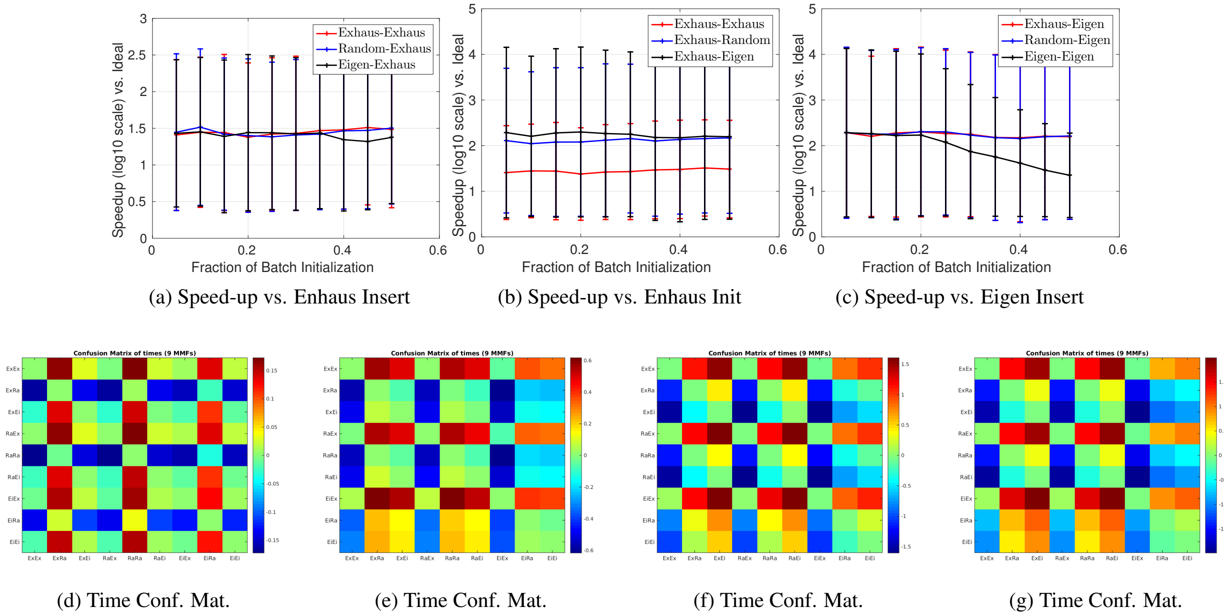
(d) Error Conf. Mat.

(e) Error Conf. Mat.

(f) Error Conf. Mat.

(g) Error Conf. Mat.

Figure 6. **Incremental versus Batch MMFs on Toy3 dataset [ERROR]:** Comparing factorization errors between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization errors between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Speed-up vs. Enhaus Insert

(b) Speed-up vs. Enhaus Init

(c) Speed-up vs. Eigen Insert

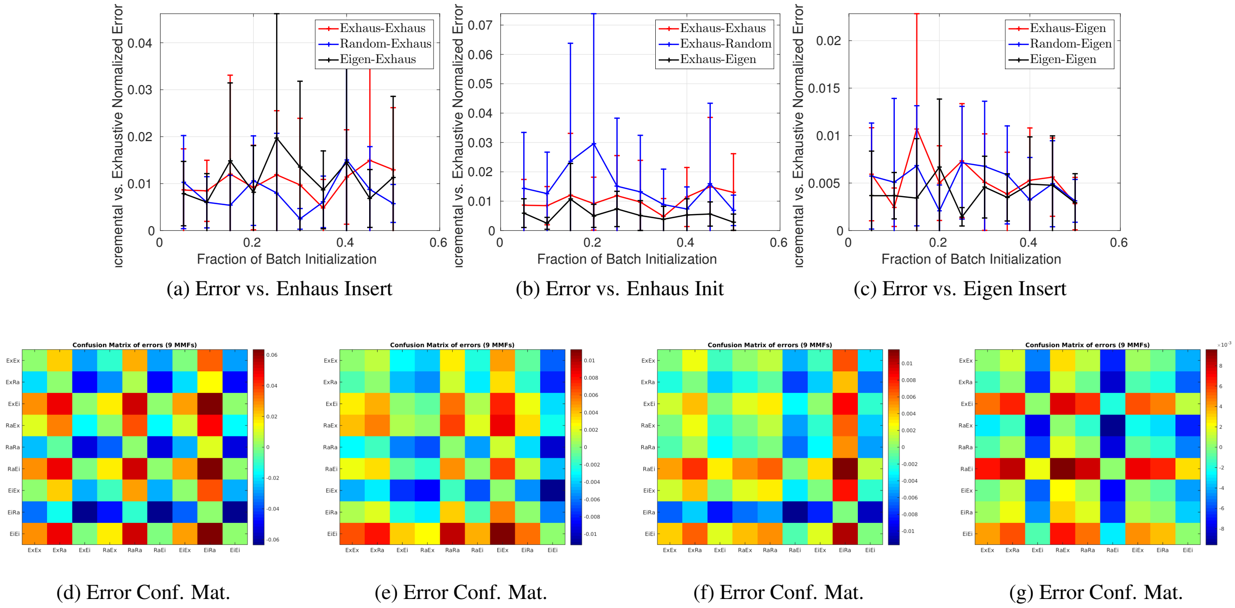
(d) Time Conf. Mat.

(e) Time Conf. Mat.

(f) Time Conf. Mat.

(g) Time Conf. Mat.

Figure 7. **Incremental versus Batch MMFs on Toy3 dataset [SPEED-UP]:** Comparing factorization speed-ups between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization times between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Error vs. Enhaus Insert

(b) Error vs. Enhaus Init

(c) Error vs. Eigen Insert

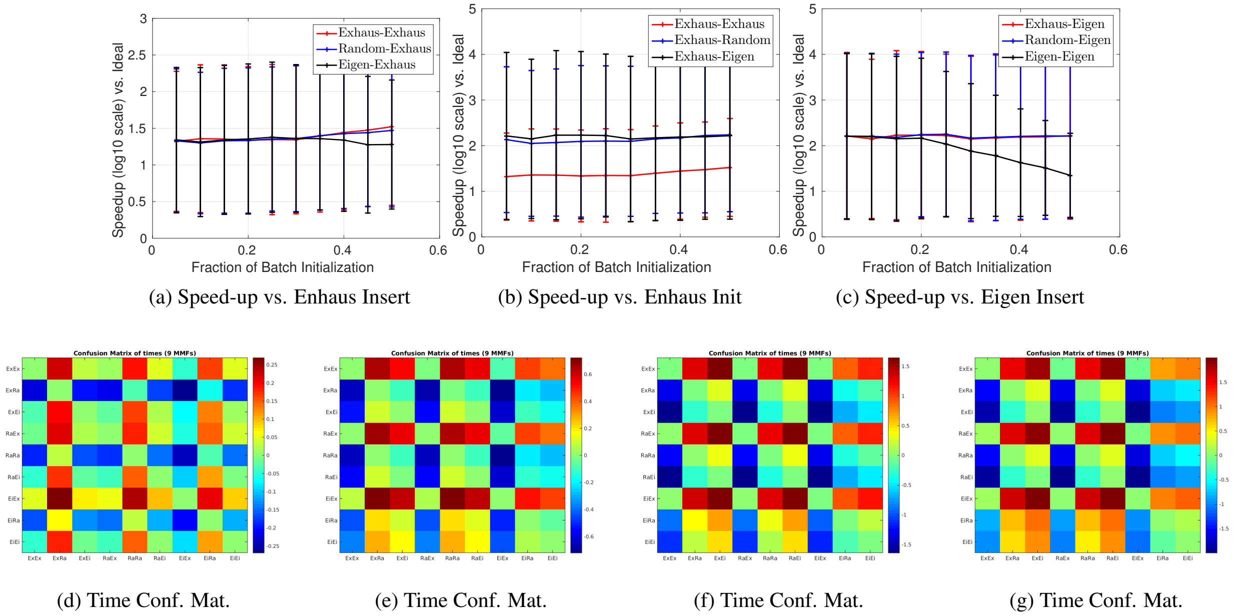
(d) Error Conf. Mat.

(e) Error Conf. Mat.

(f) Error Conf. Mat.

(g) Error Conf. Mat.

Figure 8. **Incremental versus Batch MMFs on Toy4 dataset [ERROR]:** Comparing factorization errors between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization errors between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Speed-up vs. Enhaus Insert

(b) Speed-up vs. Enhaus Init

(c) Speed-up vs. Eigen Insert

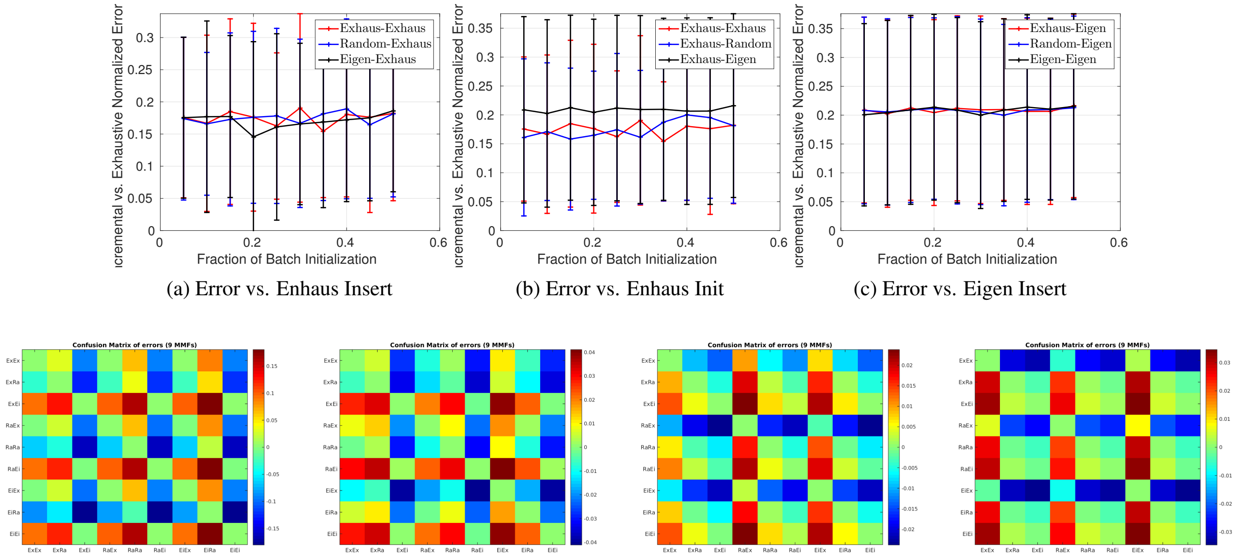
(d) Time Conf. Mat.

(e) Time Conf. Mat.

(f) Time Conf. Mat.

(g) Time Conf. Mat.

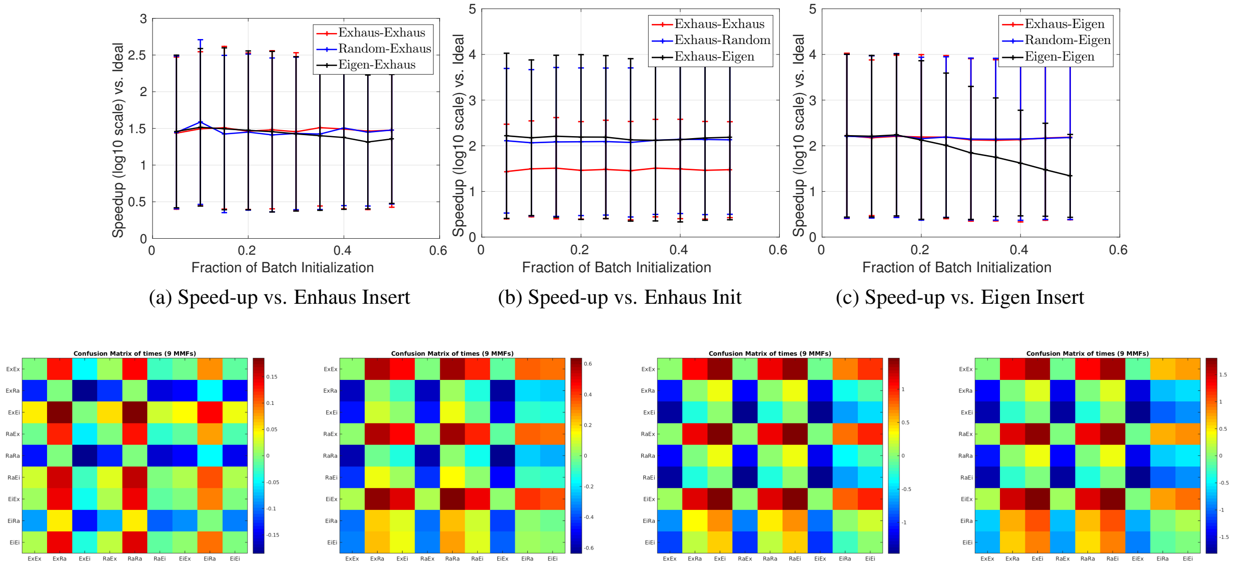
Figure 9. **Incremental versus Batch MMFs on Toy4 dataset [SPEED-UP]:** Comparing factorization speed-ups between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization times between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Error vs. Enhaus Insert (b) Error vs. Enhaus Init (c) Error vs. Eigen Insert

(d) Error Conf. Mat. (e) Error Conf. Mat. (f) Error Conf. Mat. (g) Error Conf. Mat.

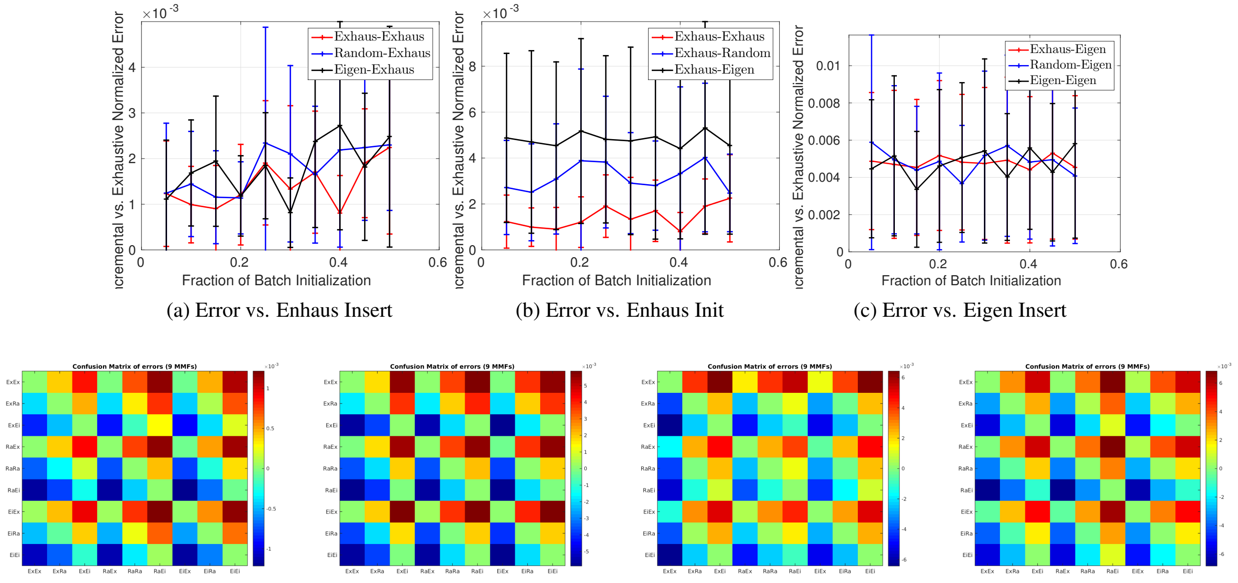
Figure 10. **Incremental versus Batch MMFs on Toy5 dataset [ERROR]:** Comparing factorization errors between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization errors between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Speed-up vs. Enhaus Insert (b) Speed-up vs. Enhaus Init (c) Speed-up vs. Eigen Insert

(d) Time Conf. Mat. (e) Time Conf. Mat. (f) Time Conf. Mat. (g) Time Conf. Mat.

Figure 11. **Incremental versus Batch MMFs on Toy5 dataset [SPEED-UP]:** Comparing factorization speed-ups between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization times between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Error vs. Enhaus Insert

(b) Error vs. Enhaus Init

(c) Error vs. Eigen Insert

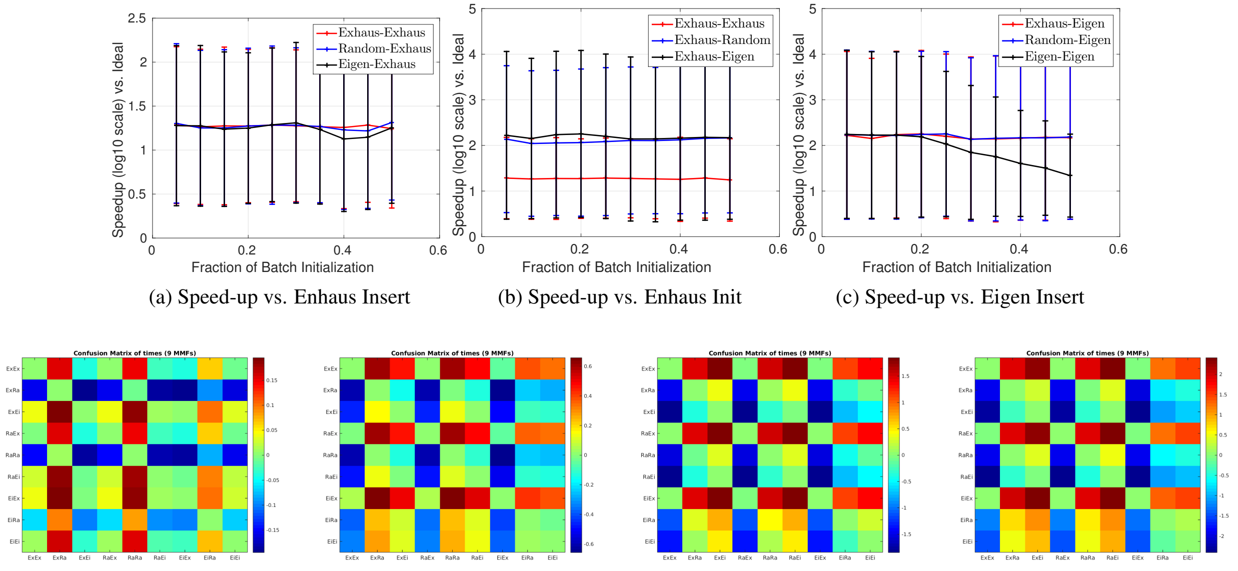
(d) Error Conf. Mat.

(e) Error Conf. Mat.

(f) Error Conf. Mat.

(g) Error Conf. Mat.

Figure 12. **Incremental versus Batch MMFs on Toy6 dataset [ERROR]:** Comparing factorization errors between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization errors between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).



(a) Speed-up vs. Enhaus Insert

(b) Speed-up vs. Enhaus Init

(c) Speed-up vs. Eigen Insert

(d) Time Conf. Mat.

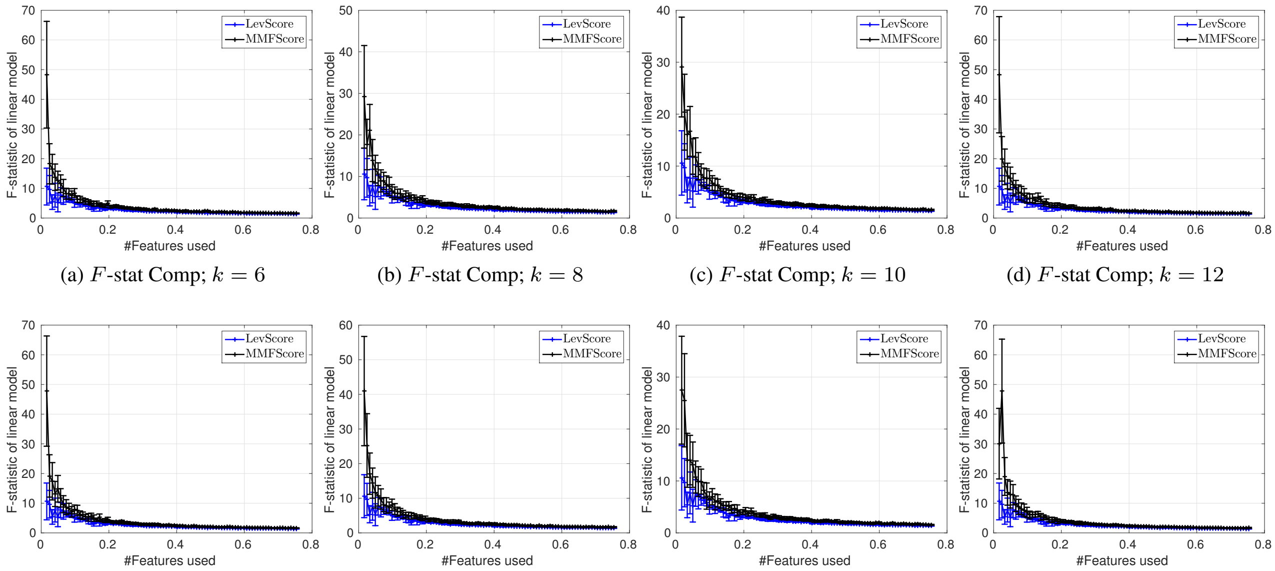
(e) Time Conf. Mat.

(f) Time Conf. Mat.

(g) Time Conf. Mat.

Figure 13. **Incremental versus Batch MMFs on Toy6 dataset [SPEED-UP]:** Comparing factorization speed-ups between the exhaustive batch version and different versions of incremental versions (a–c), including incremental MMFs with exhaustive insertion step (a), exhaustive initialization step (b) and eigen insertion step (c). Confusion matrix of factorization times between the 9 versions of incremental MMFs – three initilizations and three insertions (exhaustive, random and eigen).

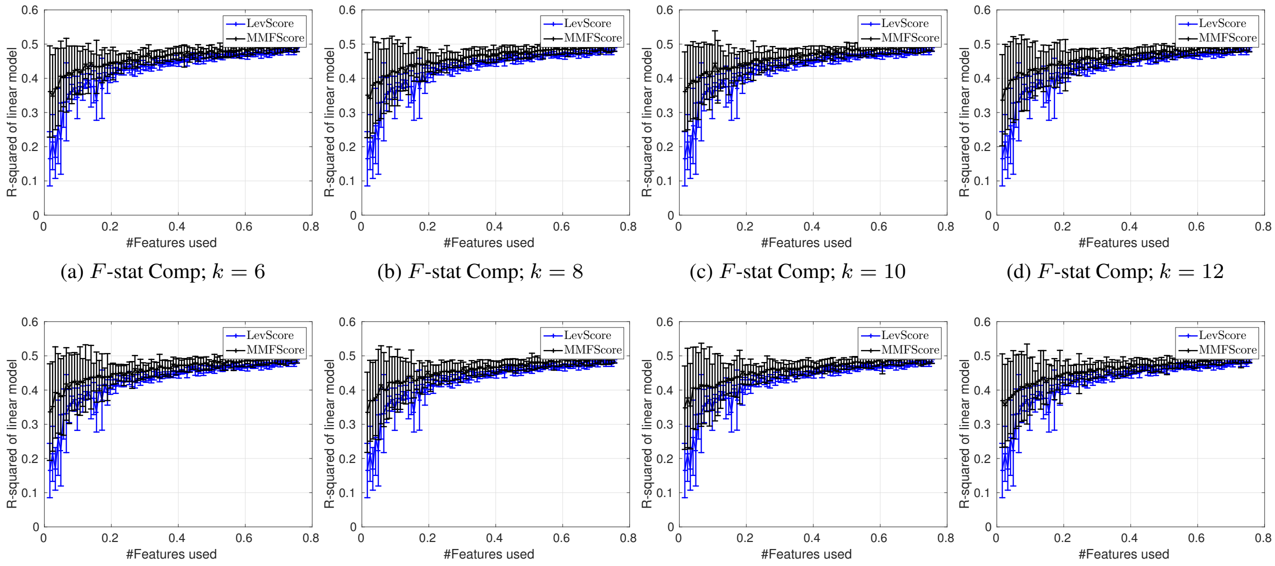
4.2. MMF Scores



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

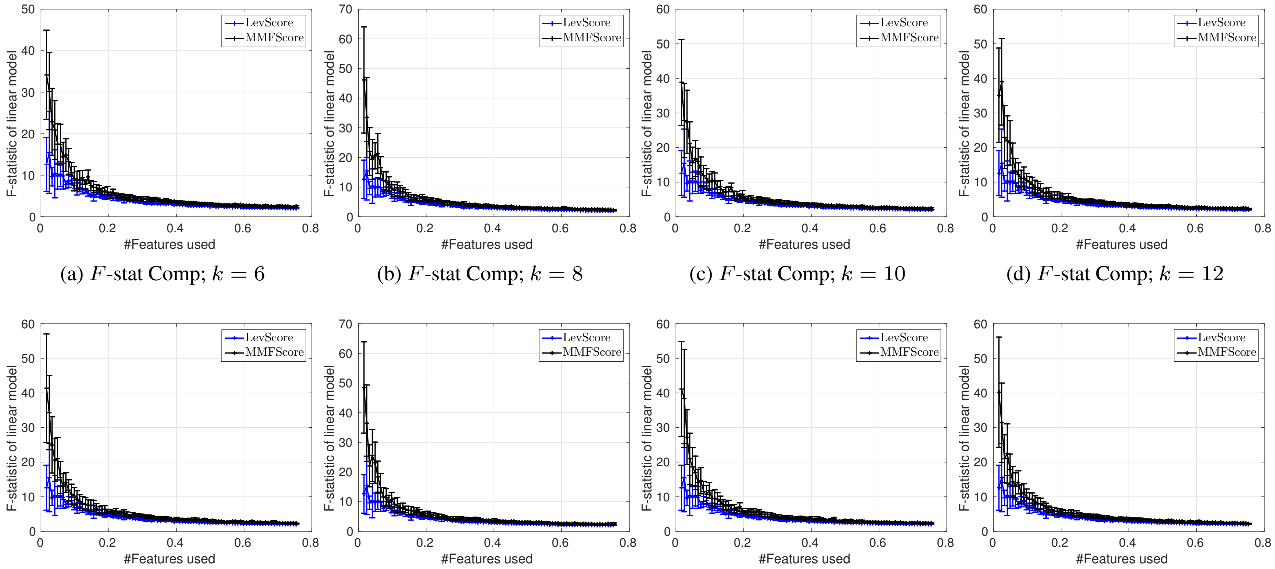
Figure 14. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – F-statistic:** The F-statistic of the linear model 'fitted' using the features (ROIs) from **PET-ADNI** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *disease status*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

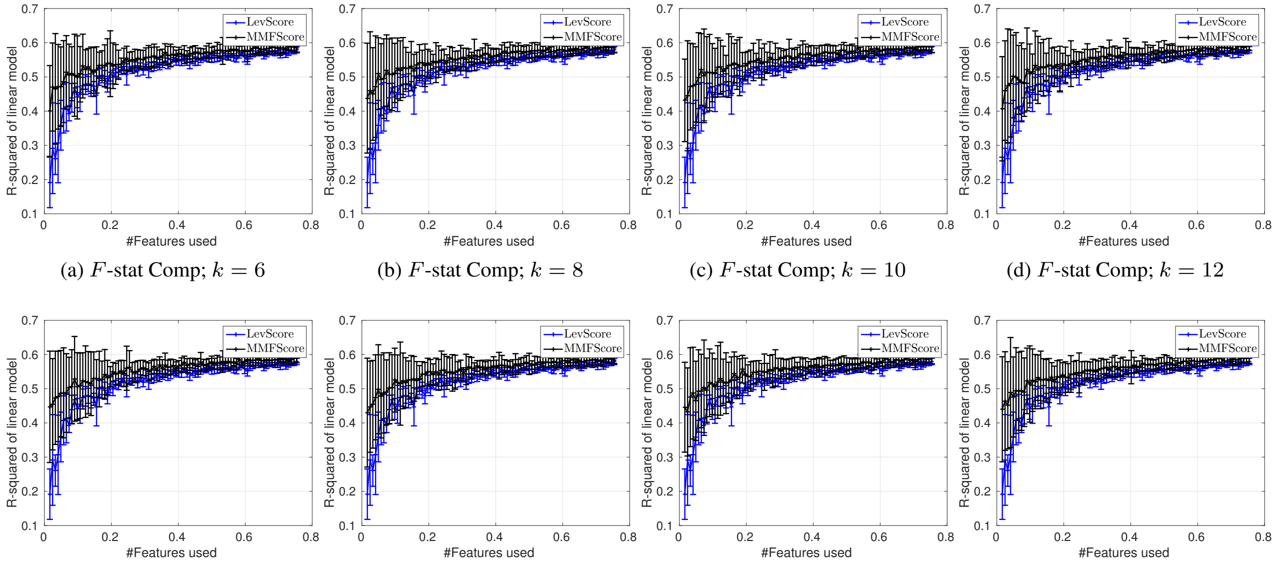
Figure 15. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – R^2:** The R^2 of the linear model 'fitted' using the features (ROIs) from **PET-ADNI** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *disease status*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

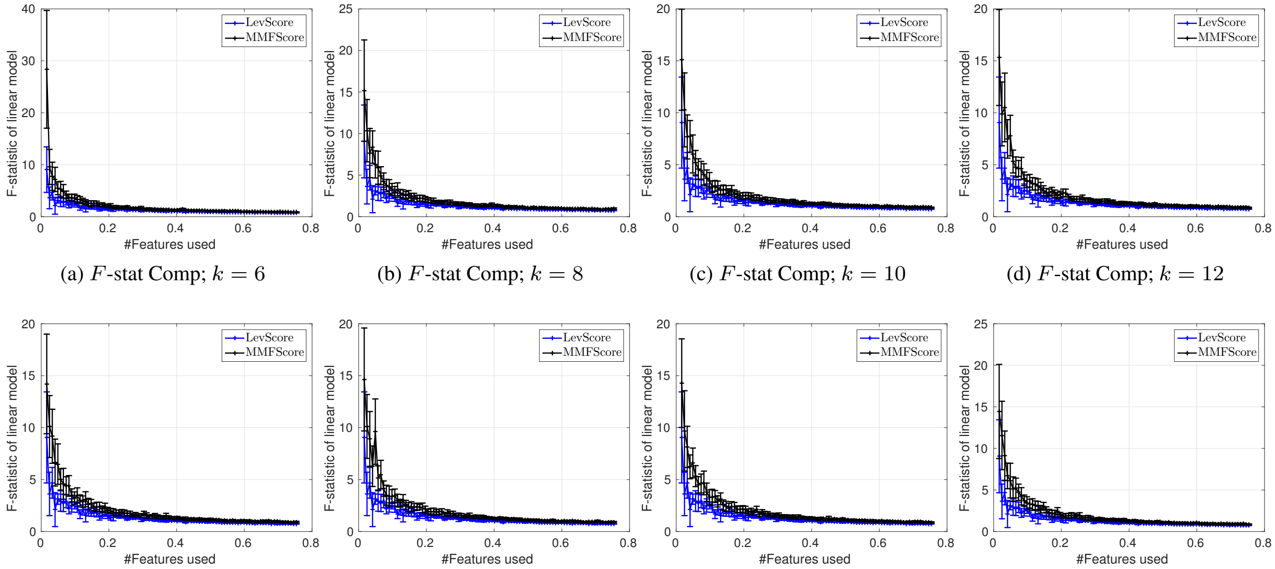
Figure 16. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – F-statistic:** The F-statistic of the linear model 'fitted' using the features (ROIs) from **PET-ADNI** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *age*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

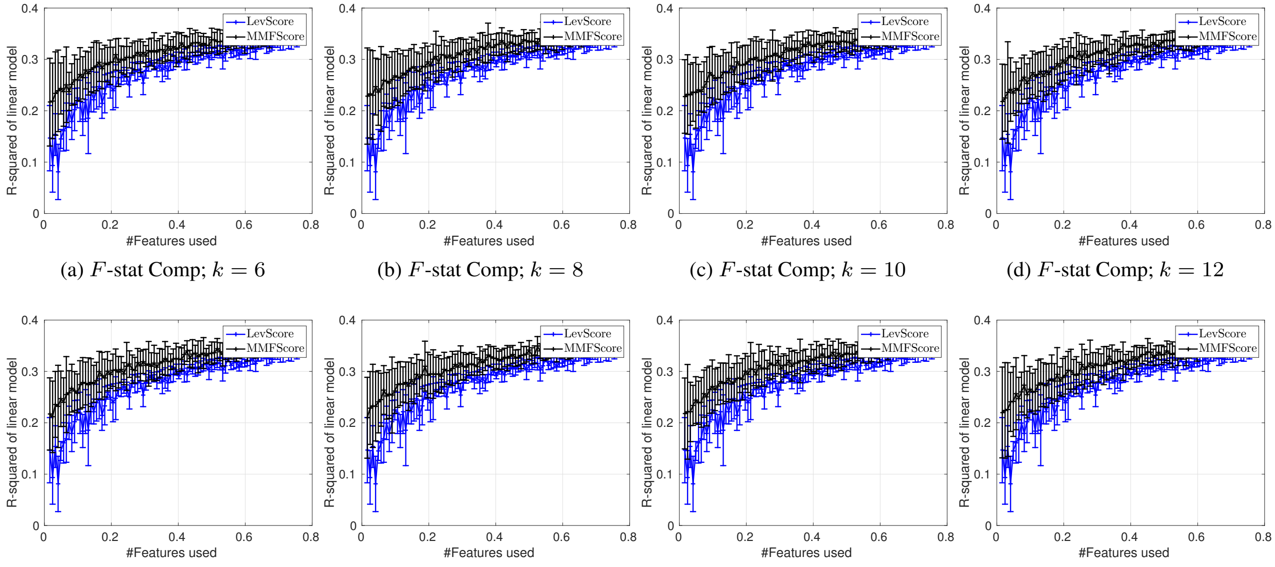
Figure 17. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – R^2:** The R^2 of the linear model 'fitted' using the features (ROIs) from **PET-ADNI** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *age*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

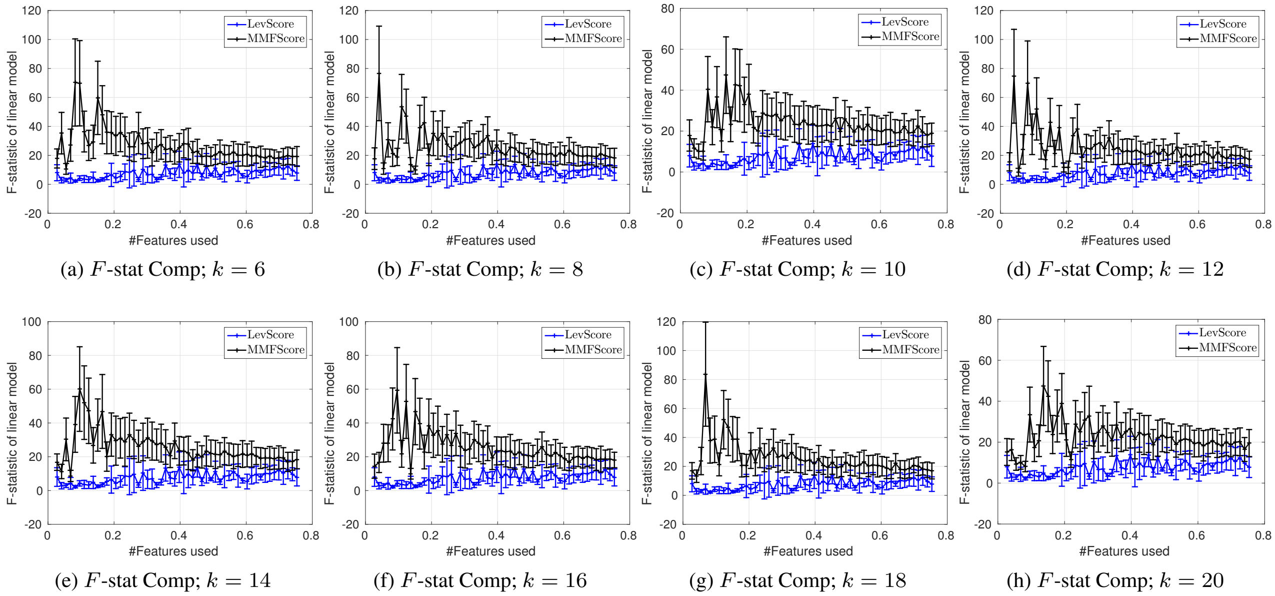
Figure 18. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – F-statistic:** The F-statistic of the linear model 'fitted' using the features (ROIs) from **PET-ADNI** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *genotype*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

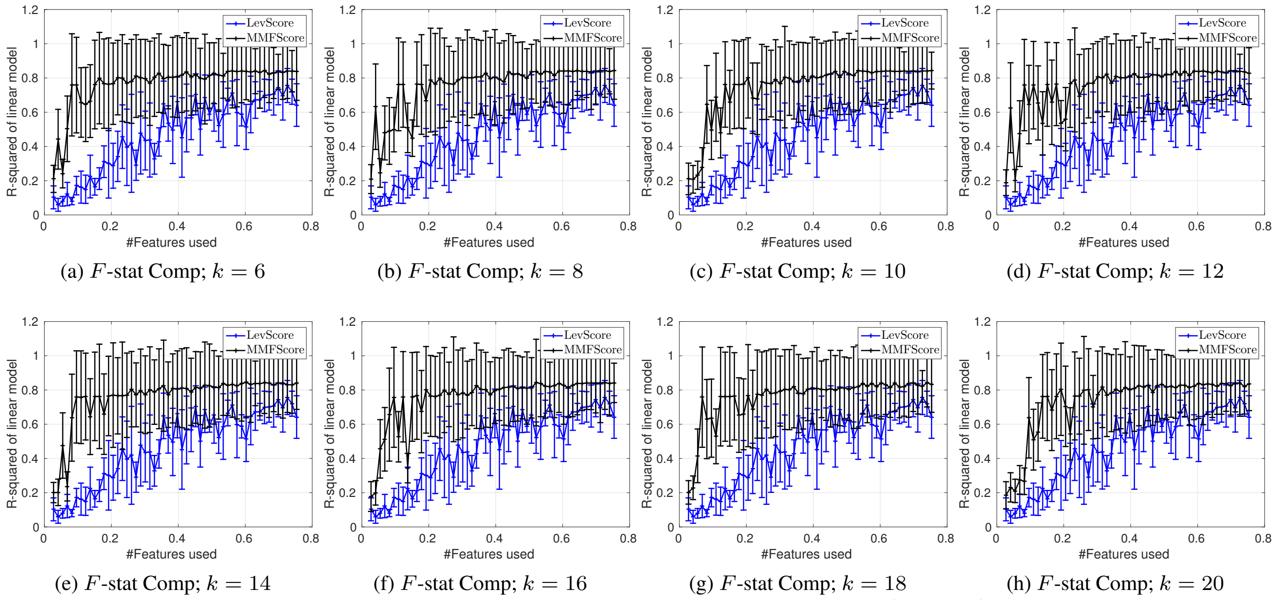
Figure 19. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – R^2:** The R^2 of the linear model 'fitted' using the features (ROIs) from **PET-ADNI** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *genotype*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

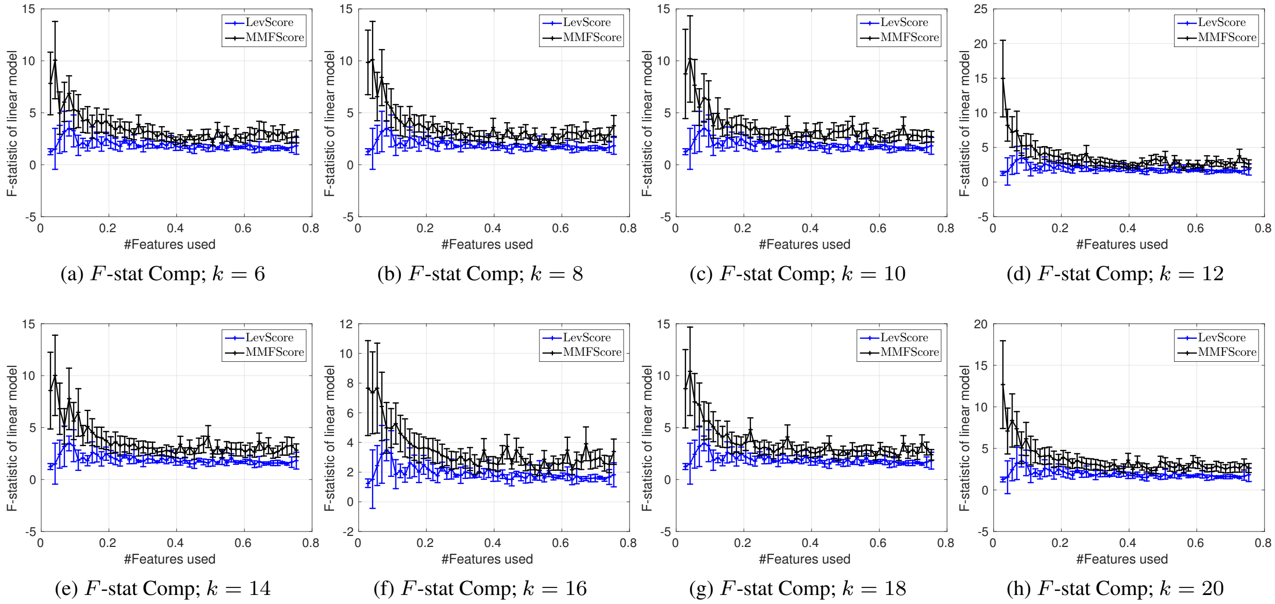
Figure 20. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – F-statistic:** The F-statistic of the linear model 'fitted' using the features from **Early-Alzheimers** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *disease status*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

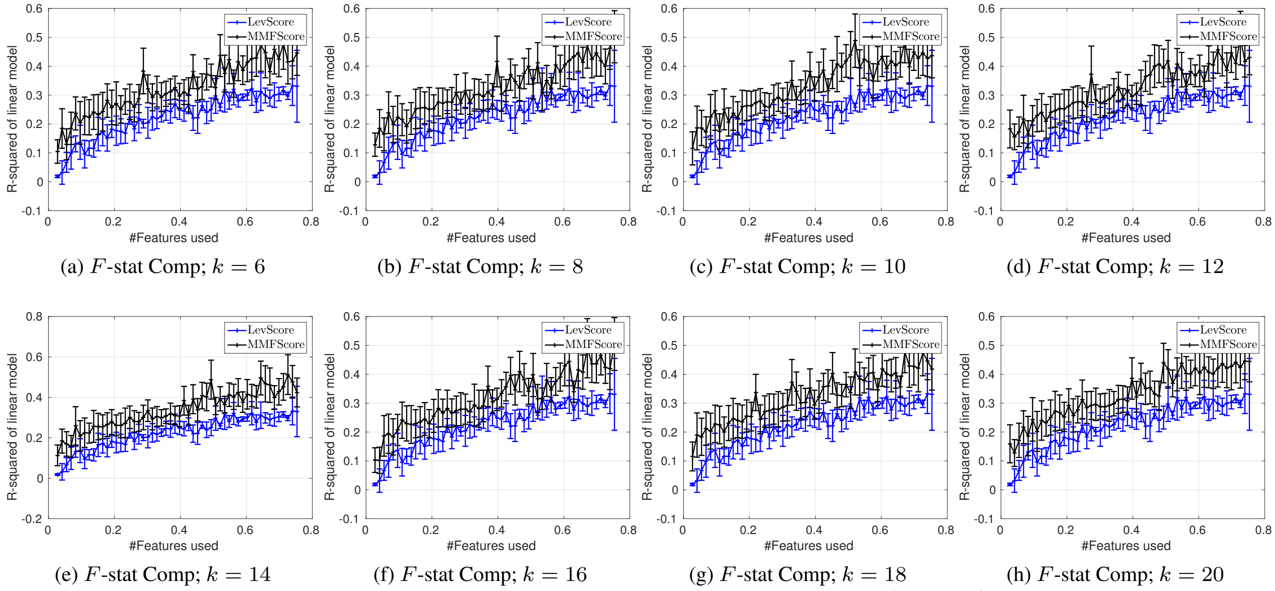
Figure 21. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – R^2:** The R^2 of the linear model 'fitted' using the features from **Early-Alzheimers** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *disease status*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

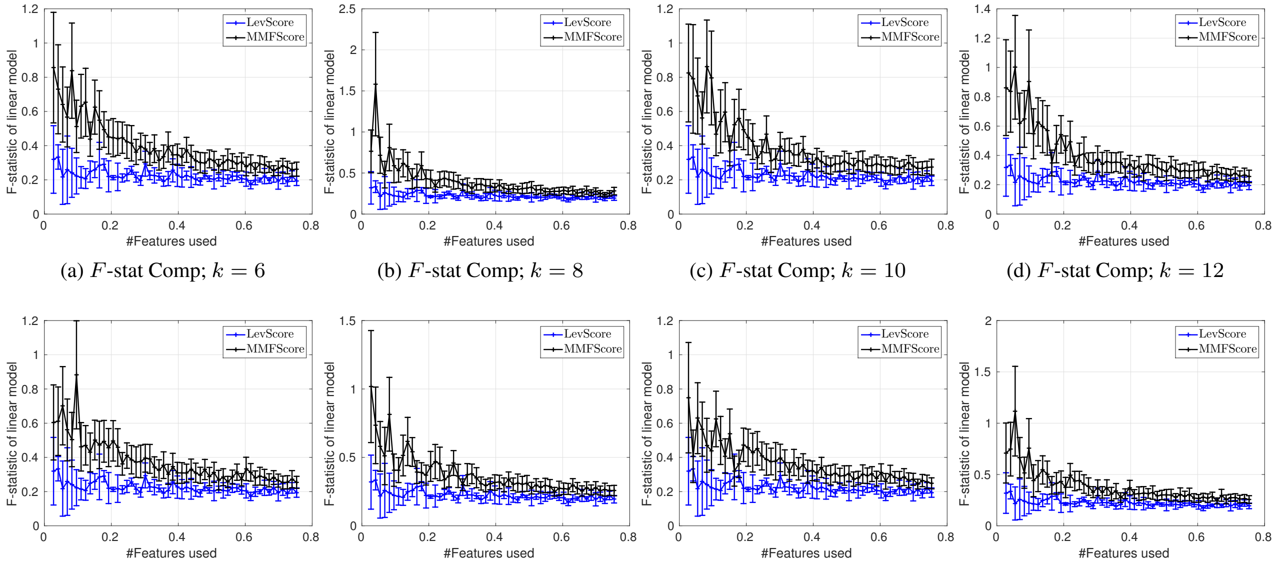
Figure 22. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – F-statistic:** The F-statistic of the linear model 'fitted' using the features from **Early-Alzheimers** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *age*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

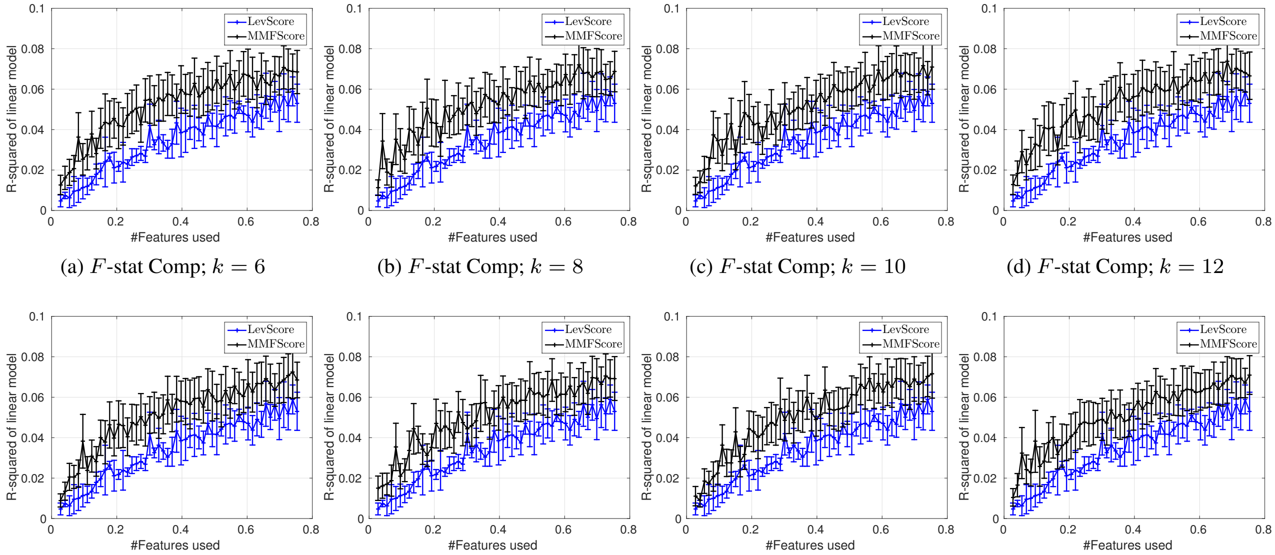
Figure 23. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – R^2:** The R^2 of the linear model 'fitted' using the features from **Early-Alzheimers** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *age*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

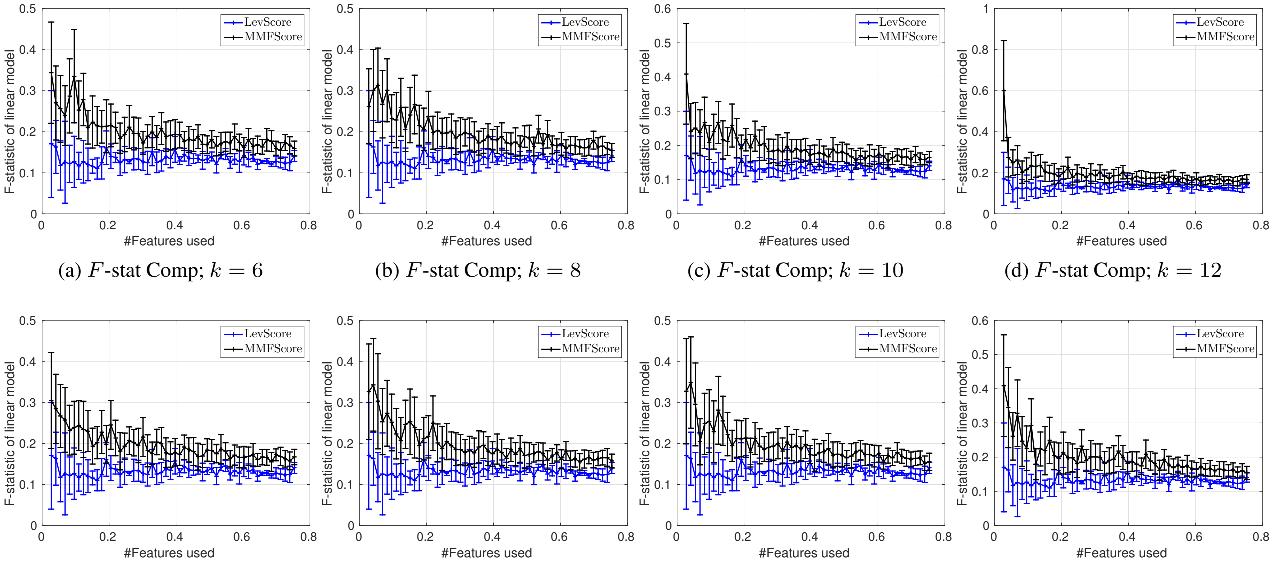
Figure 24. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – F-statistic:** The F-statistic of the linear model 'fitted' using the features from **Early-Alzheimers** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *genotype*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

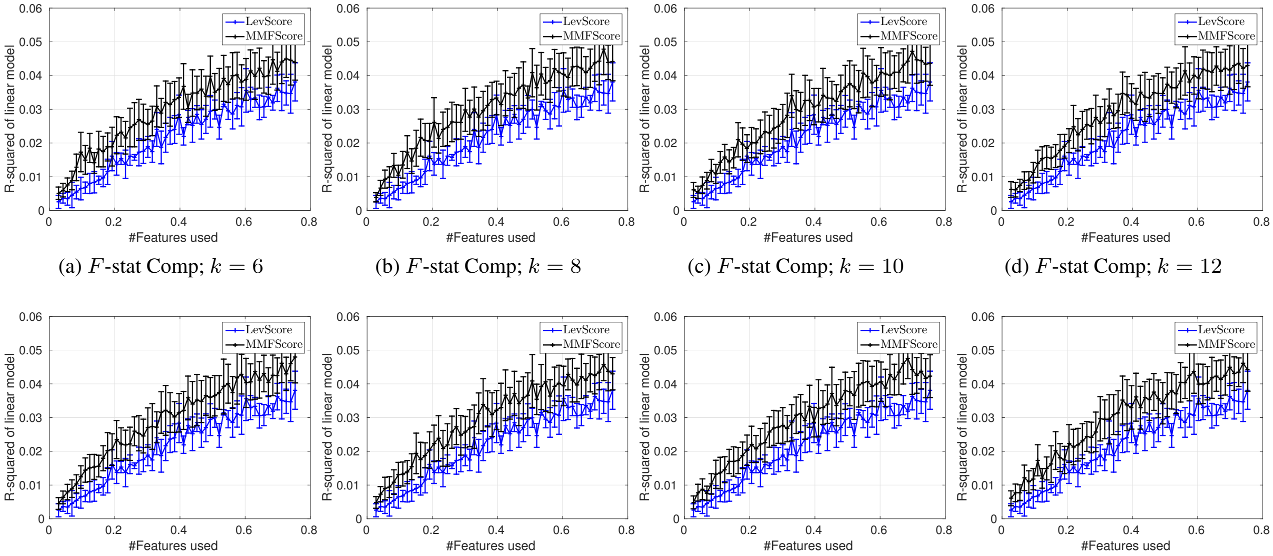
Figure 25. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – R^2:** The R^2 of the linear model 'fitted' using the features from **Early-Alzheimers** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *genotype*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

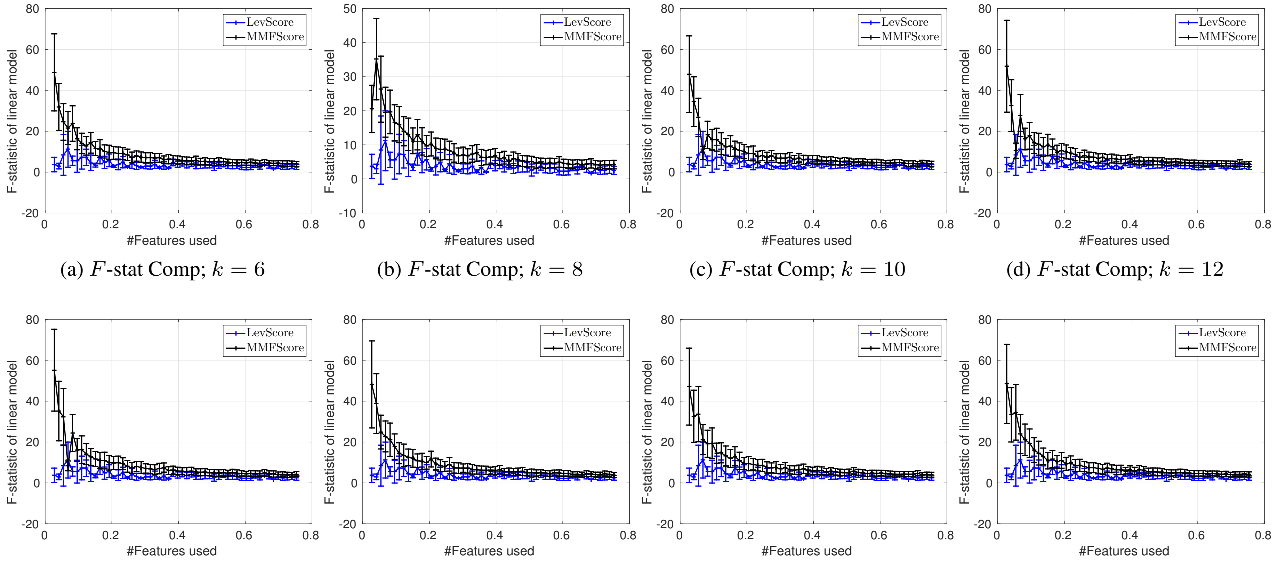
Figure 26. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – F-statistic:** The F-statistic of the linear model 'fitted' using the features from **Early-Alzheimers** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *cumulative cognition*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

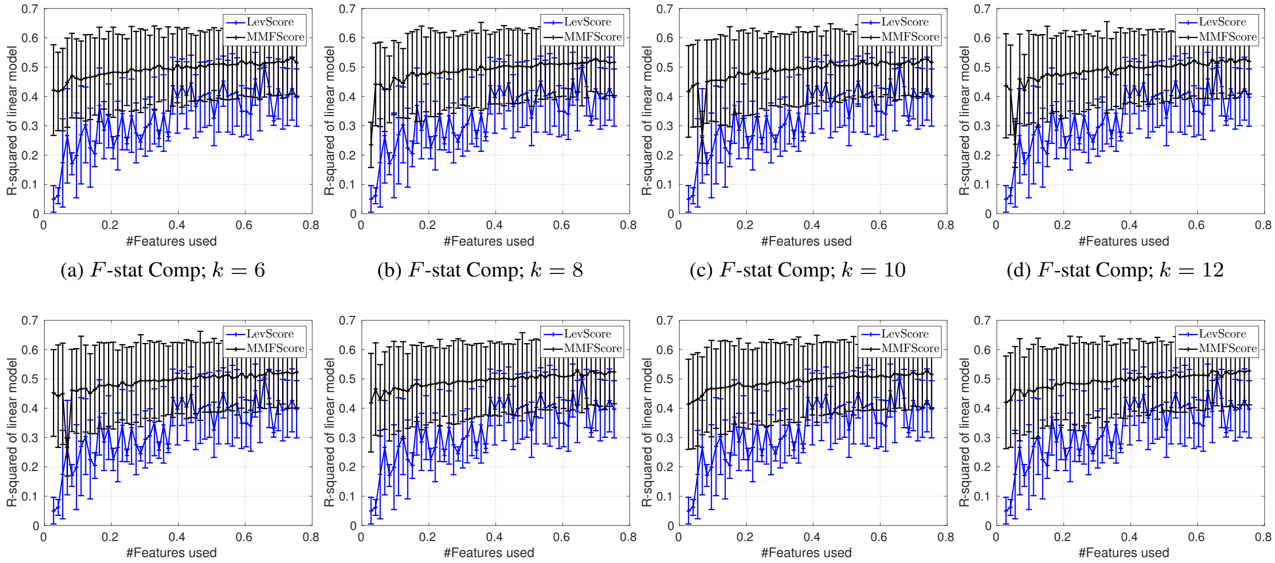
Figure 27. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – R^2:** The R^2 of the linear model 'fitted' using the features from **Early-Alzheimers** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *cumulative cognition*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).



(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

Figure 28. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – F-statistic:** The F-statistic of the linear model 'fitted' using the features from **Early-Alzheimers** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *family history*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).

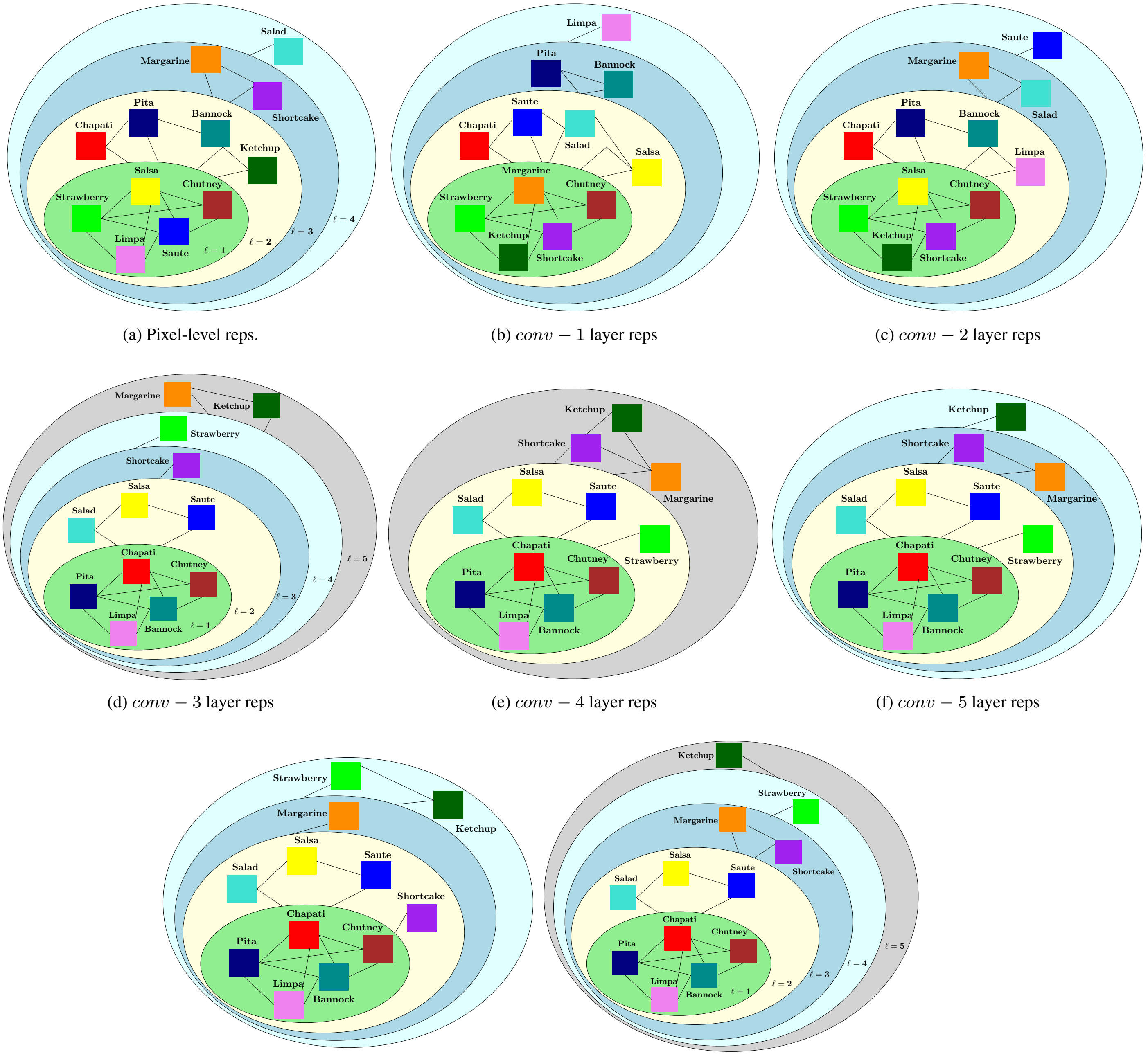


(a) F-stat Comp; $k = 6$ (b) F-stat Comp; $k = 8$ (c) F-stat Comp; $k = 10$ (d) F-stat Comp; $k = 12$

(e) F-stat Comp; $k = 14$ (f) F-stat Comp; $k = 16$ (g) F-stat Comp; $k = 18$ (h) F-stat Comp; $k = 20$

Figure 29. **Importance Sampling from MMF Scores (black) vs. Leverage Scores (blue) – R^2:** The R^2 of the linear model 'fitted' using the features from **Early-Alzheimers** data, selected according to the MMF Score (black) or Leverage Score (blue) importance samplers. The response is the *family history*. The x-axis denotes the fraction of such features selected. The errorbars on the plots correspond to 10 repetitions of the linear model. Each plot corresponds to different order of the MMFs ($k = 6$ to 20 at steps of 2).

4.3. MMF Graphs



(a) Pixel-level reps.

(b) $conv - 1$ layer reps

(c) $conv - 2$ layer reps

(d) $conv - 3$ layer reps

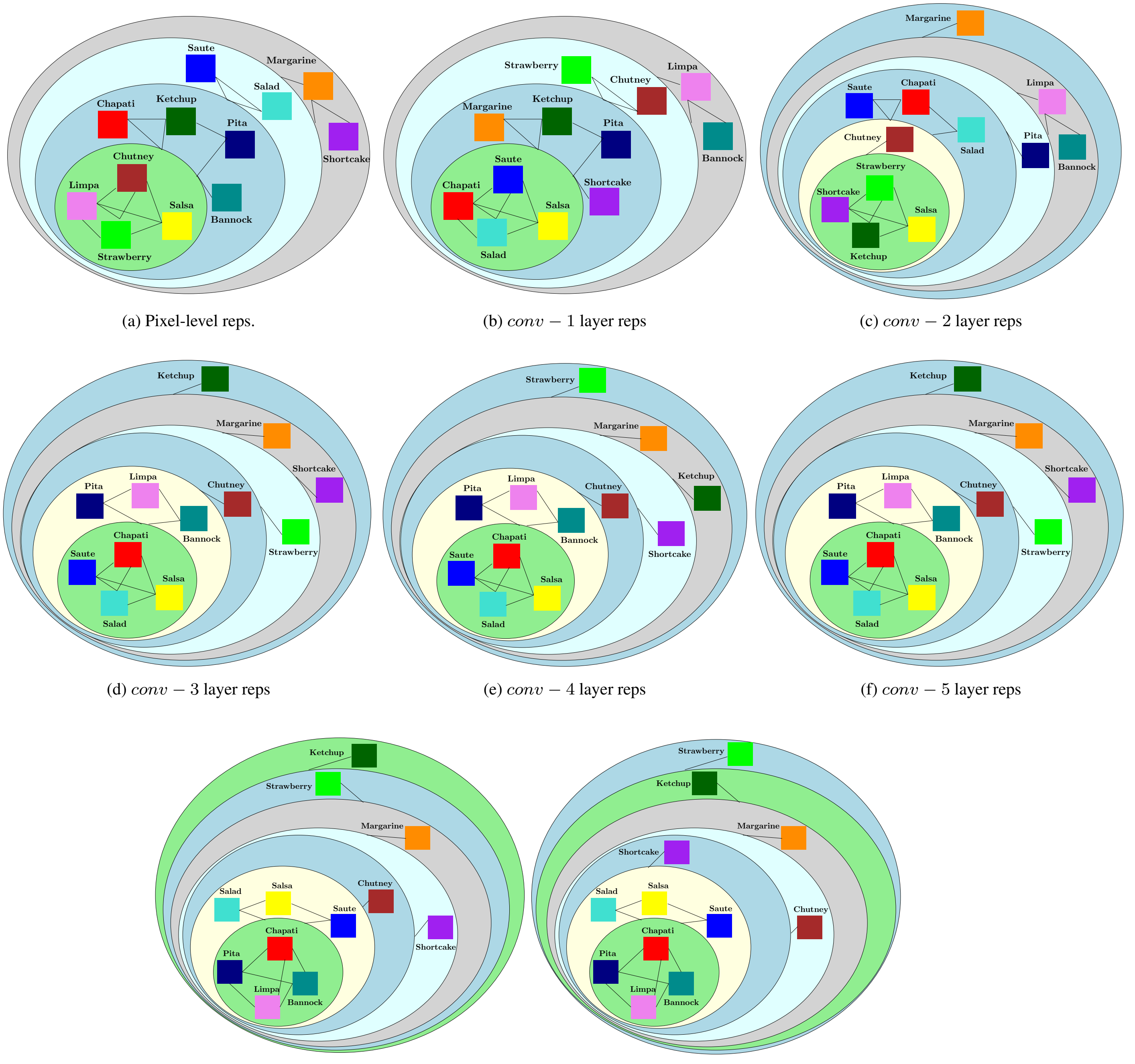
(e) $conv - 4$ layer reps

(f) $conv - 5$ layer reps

(g) $FC6$ layer reps

(h) $FC7$ layer reps

Figure 30. **Hierarchical Compositions from 5^{th}-order MMF constructed on VGG-S Network representations:** Each plot corresponds to the representations coming from different layers of the VGG-S network ($conv$ denotes convolutional, and FC denotes fully-connected). The compositions are for the 12 classes considered in Figure 4 of the main paper.



(a) Pixel-level reps.

(b) $conv - 1$ layer reps

(c) $conv - 2$ layer reps

(d) $conv - 3$ layer reps

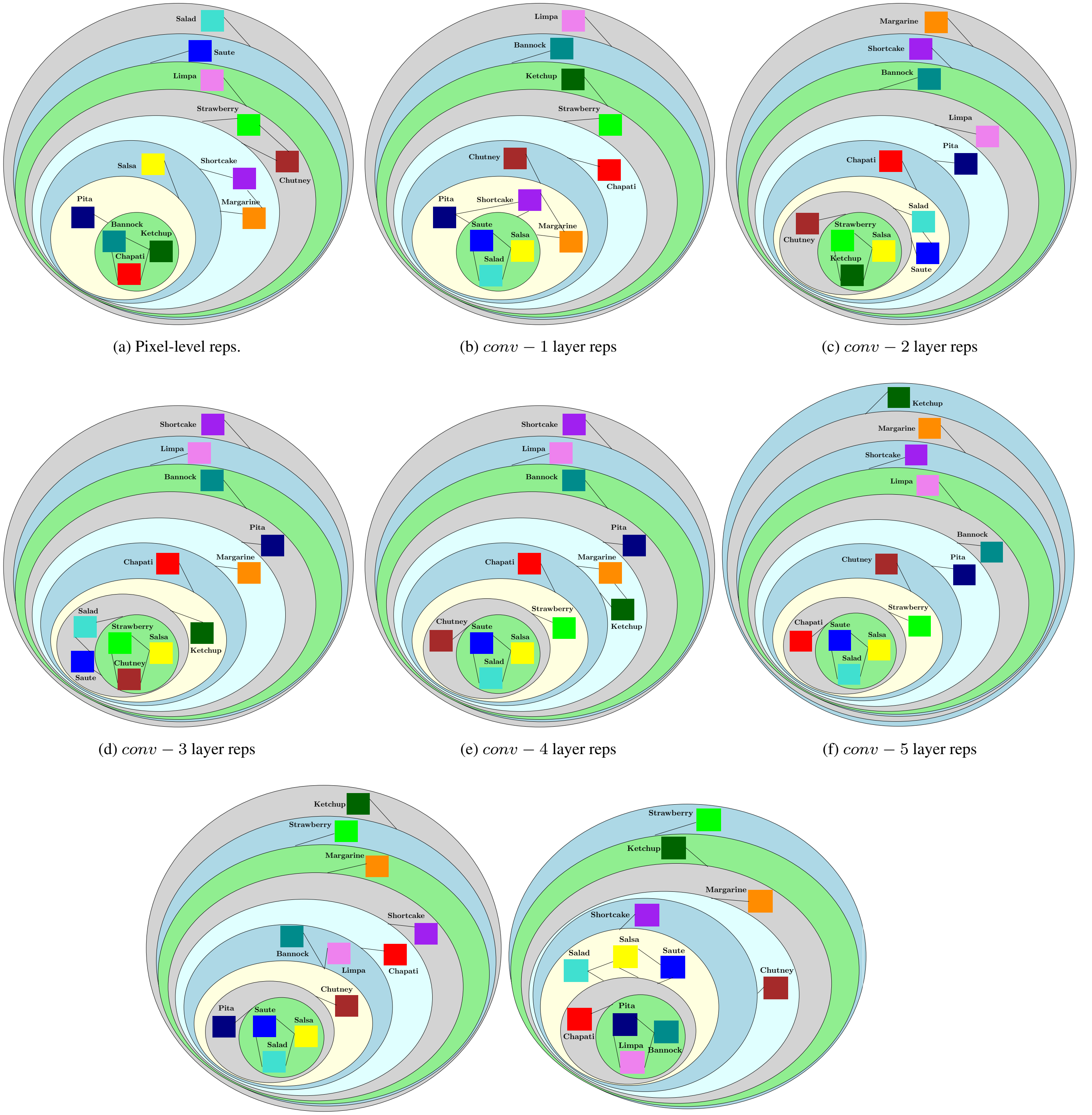
(e) $conv - 4$ layer reps

(f) $conv - 5$ layer reps

(g) $FC6$ layer reps

(h) $FC7$ layer reps

Figure 31. **Hierarchical Compositions from $4th$-order MMF constructed on VGG-S Network representations:** Each plot corresponds to the representations coming from different layers of the VGG-S network ($conv$ denotes convolutional, and FC denotes fully-connected). The compositions are for the 12 classes considered in Figure 4 of the main paper.



(a) Pixel-level reps.

(b) $conv - 1$ layer reps

(c) $conv - 2$ layer reps

(d) $conv - 3$ layer reps

(e) $conv - 4$ layer reps

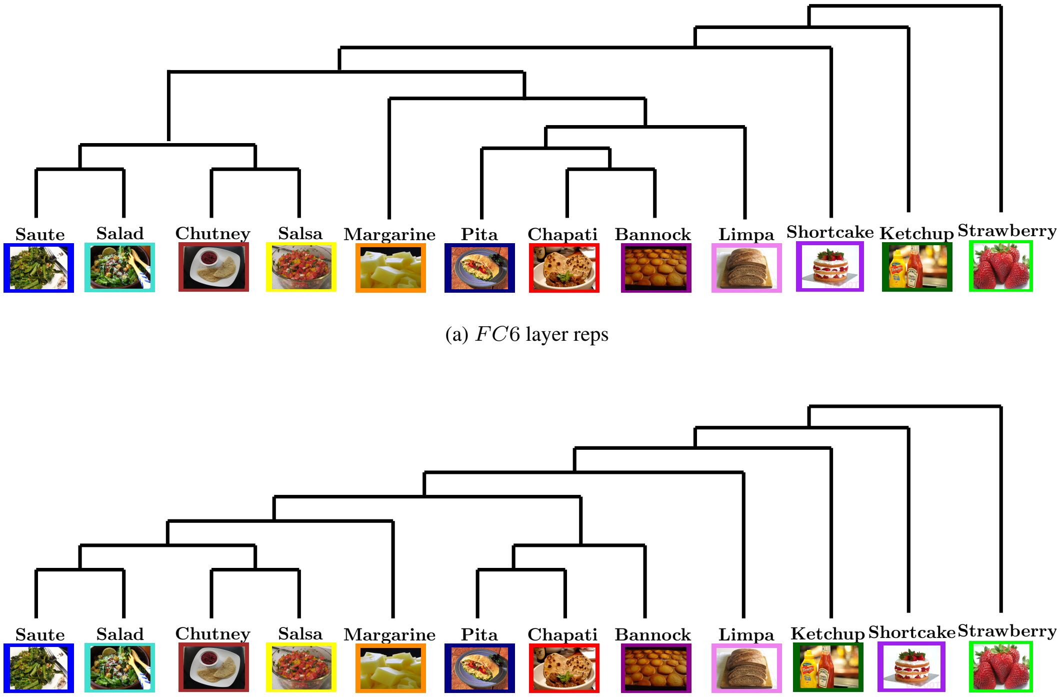
(f) $conv - 5$ layer reps

(g) $FC6$ layer reps

(h) $FC7$ layer reps

Figure 32. **Hierarchical Compositions from 3^{rd}-order MMF constructed on VGG-S Network representations:** Each plot corresponds to the representations coming from different layers of the VGG-S network ($conv$ denotes convolutional, and FC denotes fully-connected). The compositions are for the 12 classes considered in Figure 4 of the main paper.

Hierarchical Agglomerative Clustering: The following figure shows the dendrogram corresponding to the agglomerative clustering (based on mean distance of the clusters) on the $FC6$ and $FC7$ representations of the 12 classes from Figures 30–32. Observe that the categories like *bannok* and *chapati*, or *saute* and *salad* are closer to each other here in hierarchical clustering. However, the compositional relationships inferred from the MMF graphs, see Figures 30–32, especially in the $FC7$ and $FC6$ layers, are not apparent in these dendrogram outputs (also see Section 4.3.2 in the main paper). Similar such cluster trees can be produced using other layers' representations, where in the compositions from MMF based graphs are much more informatice and contextual.



(a) $FC6$ layer reps

(b) $FC7$ layer reps

Figure 33. **Hierarchical clustering (Dendrograms) constructed on VGG-S Network representations:** Plots correspond to the representations coming from $FC6$ and $FC7$ layers of the VGG-S network (FC denotes fully-connected). The clustering is for the 12 classes considered in Figure 4 of the main paper (and Figures 30–32 above).

References

[1] http://adni.loni.usc.edu. 3

[2] http://image-net.org/explore_popular.php. 3

[3] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014. 3

[4] R. Kondor, N. Teneva, and V. Garg. Multiresolution matrix factorization. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1620–1628, 2014. 1

[5] R. Kondor, N. Teneva, and P. K. Mudrakarta. Parallel mmf: a multiresolution approach to matrix computation. *arXiv preprint arXiv:1507.04396*, 2015. 2

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 3

[7] F. Mezzadri. How to generate random matrices from the classical compact groups. *arXiv preprint math-ph/0609050*, 2006. 2

[8] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. 3