A. Examples of Convergence

As mentioned earlier, we use the mean relative absolute error of data point values as our measure of convergence. The convergence is reached when $\frac{1}{n}\sum_{i=1} n\frac{|\phi_i^t - \phi_i^{t-100})}{|\phi_i^t|} < 0.05$. Fig. 6 shows example of data Shapley value convergence for the tasks of cancer prediction discussed in Section 4.1.

B. Robustness to Truncation

For the same data sets in Section 4.2, we do the following experiment: For each iteration of TMC-Shapley, we truncate the computation of marginal contributions at different positions. For instance, a truncation of size 10% means that for a data set of size 1000, in each iteration of Alg. 1, for the sampled permutation of data points we only perform the calculation of marginal contributions for the first 100 elements and approximate the rest with zero. Fig. 7 shows the results. Columns (a), (b), (c), are from the same data sets of the corresponding columns in Fig. 2 and are trained using the same models. For each data set, model, and training set size, we show two plots. The first plot shows how for different levels of truncation, valuation of data points corresponds to the effectiveness of those data points in model's performance. The second plot shows the rank correlation between the valuation of that level of truncation and the valuation without any truncation. As it is seen, all all cases, values derived by truncation level of 25% have rank correlation around 0.8 with that of having no truncation.

C. TMC-Shapley Versus True Shapley

We computed the true Shapley value for synthetic training data with sizes from 4 to 14 data points trained with a logistic regression model. The pearson correlation between the true Shapley value and the approximate TMC-Shapley value is in the range of 98.4% to 99.5%. Fig. 8 depicts examples.

D. G-Shapley Versus TMC-Shapley

One important question is how much the data Shapley values returned by G-Shapley are similar. In this sectio we will report the values for some of the experiments presented in Section 4:

- **Synthetic datasets** For all data sets in section 4.2 and train set size of 100, using logistic regression model, G-Shapley and TMC-Shapley have correlation coefficient between 0.9 and 0.95. Changing the model to neural network reduces correlation coefficient to the range of 0.7 to 0.8. Increasing the train size to 1000, for the logistic regression model, all correlation coefficients are between 0.94 and 0.97 which for the neural network model reduces to the range of 0.8 to 0.88.
- **Disease prediction** For the breast cancer data, G-Shapley and TMC-Shapley have a correlation coefficient of 0.86. For the task of skin cancer prediction, the values of G-Shapley and TMC-Shapley have correlation coefficient of 0.77 and rank correlation of 0.78.
- **Label flip** For the Fashion MNIST data set G-Shapley and TMC-Shapley have correlation coefficient equal 0.62 and for the flower classification data set, the correlation is 0.57.

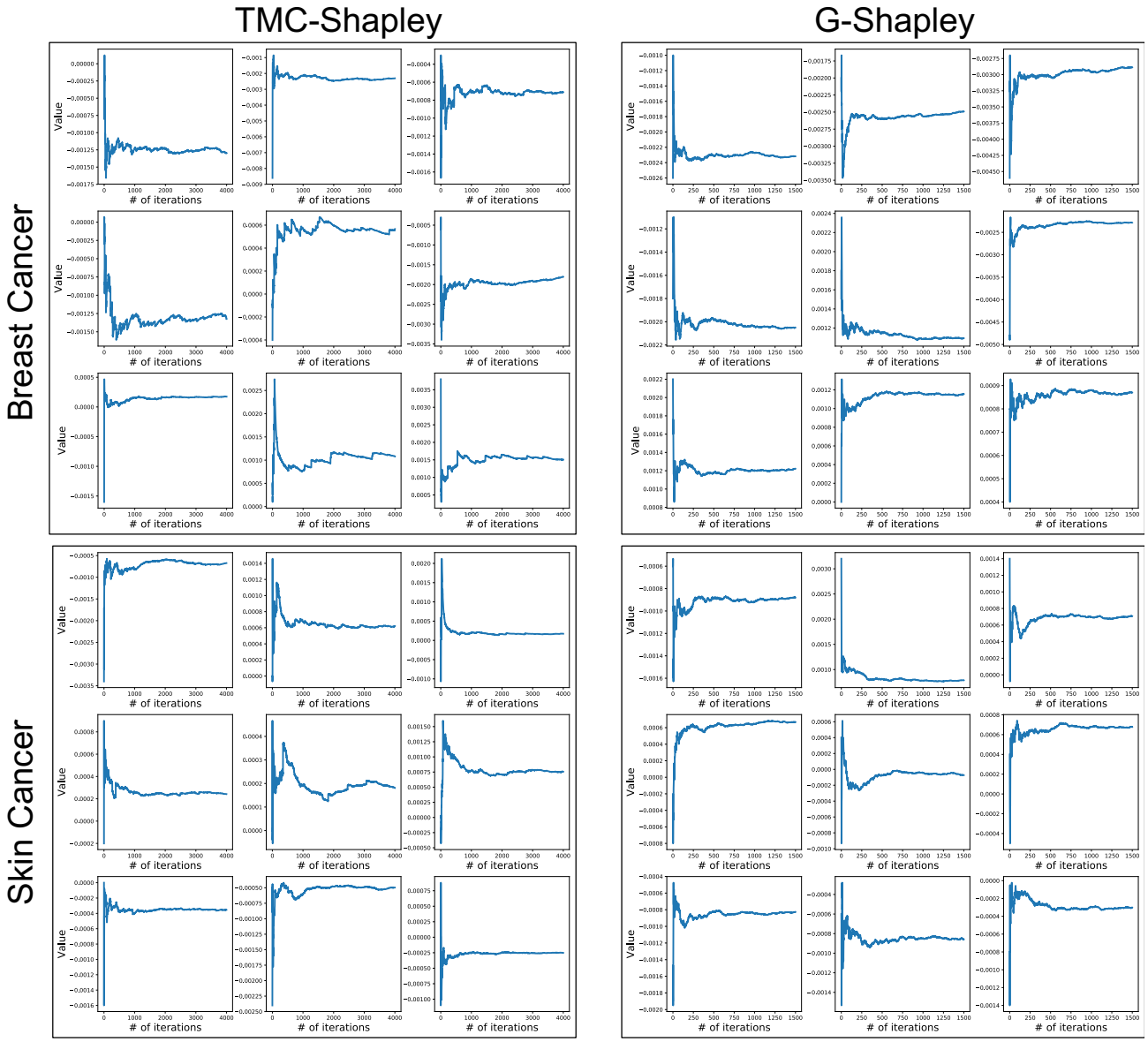

Figure 6. **Convergence** Changes in TMC-Shapley and G-Shapley values of randomly selected training points from the UKB biobank disease prediction data sets discussed in Section 4.2. Nine different random individuals are selected for TMC-Shapley and G-Shapley. The x-axis indicate the number of Monte-Carlo samples and y-axis is the estimate for the data Shapley value.

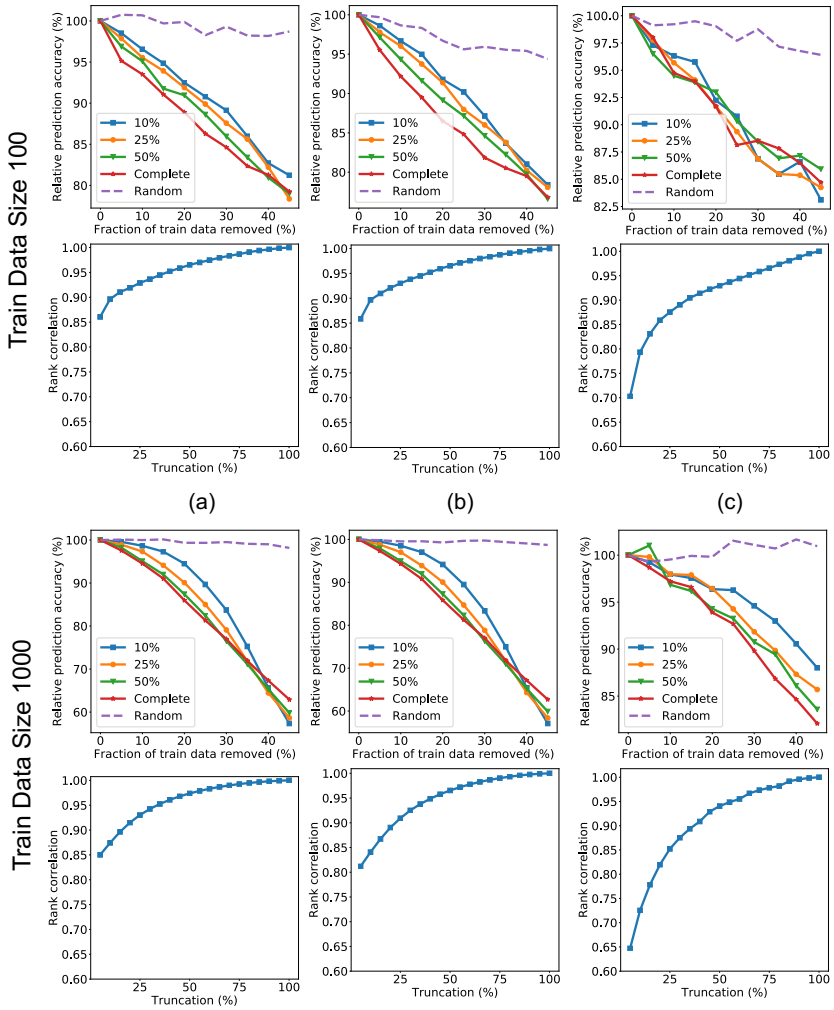

Figure 7. **Truncation approximation** A truncation of 10% means that in Alg. 1, for each sampled permutations, we calculate the marginal contributions of the first 10% elements of that permutation and approximate the remaining by zero marginal contribution. For the datasets generated in Section 4.2. Columns (a), (b), and (c) use the same data sets as their corresponding column in Fig. 2. For each train data size and data set, the upper plot shows the change in model performance as we remove points with high TMC-Shapley values derived by various truncation levels. The bottom plot shows the rank order correlation between TMC-Shapley values derived by a speicif truncation level and the values derived by having no truncation. It can be observed that truncation approximation's bias can be negligible.

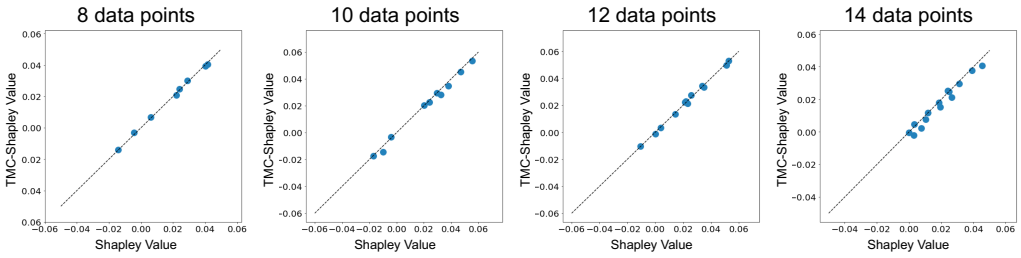

Figure 8. **Approximated vs true value** Four examples of synthetic data sets with their respective true Shapley value and approximated TMC-Shapley value are depicted. As it is shown, the approximation manages to capture the ordering perfectly and also returns similar values.