To compromise for the loss of information due to truncating all residual values larger than T, for each remaining type, they considered several submodels with different amounts of q. It allowed their model to "See" dependencies among residual samples whose values lie beyond the threshold.

Based on sample experiments with different algorithms and submodels, they determined that the best performance of each submodel is always achieved when q [c, 2c], where c is the residual order.

The union of all submodels, including their differently quantized versions, has a total dimension of 2 3 ×= 34, 671.5.They applied their feature selection strategies to the entire submodels based on the estimate of their performance, which would not be possible if they were selecting personal features.

Since the dimensionality of submodels that originated from the 'spam' type residual is 169, to make the ranking by OOB fair, they merged the vertical and horizontal 'spam' submodels into a single 2\*169 = 338-dimensional submodel.

They merged the two spam11 submodels from class 'SQUARE' into one 338dimensional submodel.

Formally, they formed the model by merging M submodels with the M smallest OOB error estimate Ei out of all 106 submodels.

BEST-q: Since two differently quantized versions of one submodel provide less diversity than two submodels built from different residuals, in this strategy, they forced difference among the selected submodels while hoping to improve the performance-to-dimensionality ratio.

As a result, they obtained 39 submodels, Mi=i, with the corresponding OOB error estimates Ei=i. Now, they applied the forward feature selection to these 39 submodels as in ALL. In particular, they merge M submodels Mi=I with M lowest OOB errors Ei i. BEST-q-CLASS. In this strategy, they forced diversity even more than in BEST-q by first selecting one submodel from each class before choosing a different submodel from the same category again.

They started with 39 submodels, just like in strategies BEST-q and BEST-q-CLASS. The first submodel selected is the one with the lowest OOB error.

The CLASS-q strategy corresponds to merging all submodels with the best q from one chosen class, while the Q1 approach corresponds to joining all 39 submodels with a fixed quantization q = 1c. The purpose of these two simple heuristic merging strategies is somewhat investigative. All experiments in this paper are carried out on three steganographic algorithms with different embedding mechanisms: 1) Non-adaptive ±1 embedding implemented with ternary matrix embedding that is optimally coded to minimize the number of embedding modifications.

2) HUGO, designed to minimize embedding distortion in a high-dimensional space function, calculated from the differences of four adjacent pixels. The best individual submodels for the larger payload and ±1 embedding, HUGO, 6. Experiment 1 was carried out on the training set for one fixed split into 8,074 training and 1000 testing images.

3) Edge-Adaptive (EA) algorithm restricts the embedding of changes to pixel pairs whose absolute value difference is as significant as possible (e.g., around edges).

Both HUGO and EA place the embedding changes on those parts of the picture that are hard to model and are now therefore supposed to be more stable than the non-adaptive ±1 embedding.

Three of the images showed the average OOB error estimates Ei for all i = 1, ... , 39 and for all values of q. The dashed lines separate the 'spam' submodels from submodels of type 'min-max'.

While there exist apparent differences among the performance of each submodel across algorithms, it is worth noting that certain submodels rank the same w.r.t. Each other for all three algorithms, both payloads, and all quantization steps.

The gain between using M = 39 submodels of BESTq-CLASS and all 106 quantized submodels is, however, rather negligible, indicating saturation of performance.

The residuals shown have been selected based on the rule of simplicity and are by no means intended as the ultimate result since there are undoubtedly many other possibilities. They consider the model building as an open-ended method because there are likely to be other predictors that will further boost detection after applying them to the proposed model. They observed a "saturation" of output in the sense that further enrichment of the model with other types of predictors leads to an insignificant increase in detection accuracy for all the algorithms tested.

From the machine learning point of view, the first three strategies, ALL, BEST-q, and BEST-q-CLASS, could be classified as filters. They are based solely on the initial OOB ranking of each submodel and therefore ignore the dependence of the submodels. They vary primarily in the way that they promote diversity. The ITERATIVE-BEST-q technique, on the other hand, continually utilizes the classification input of the ensemble as it greedily minimizes the error of the OOB in each iteration, taking into account the mutual dependency of the individual submodels.

The proposed detectors provide a significant improvement in detection accuracy over the 1234-dimensional CDF array with the Gaussian SVM even when the smallest version (TOP1 with a dimensionality slightly above 300) is used. This increase is much higher again for the two adaptive stego algorithms.

The steganalysis and models proposed in this paper consist of several procedures and modules whose development is undoubtedly worthy of further study and optimization, which could lead to further improvement in performance.

It involves, for instance, the quantifier of multidimensional co-occurrences. It is certainly possible to use the speed and convenience of the system to construct the template and then use it to create a classifier using a specific machine learning method. Better differentiation between groups may be made where there is a strongly non-linear boundary that may not be well captured by the set equipment.