The authors regret to inform that there was an unintentional train-test leakage in our machine learning experiment. The reason for such leakage was not taking into account highly correlated LC-MS repetions of individual physical samples. In order to correct train-test leakage, splitting into train and test was done such that either all repetitions from one physical sample were put in train set or one (random) of them was put in test set while the rest repetitions were ignored. Overall, accuracy and F1-score of our models was reduced by around 2% to 7%. All the results that were affected by this leakage were recalculated the following corrections are suggested (in addition to tables and figures) We used blue font color for our corrections to the relevant parts of the original text.

Corrigenda

1. (abstract) For most of used machine learning algorithms, classification accuracy of 95% higher were obtained on cross-validation dataset.
   For most of used machine learning algorithms, classification accuracy of 87% or higher were obtained on cross-validation dataset.
2. (conclusion) Logistic regression, SVM, and Random Forest were implemented to classify 36 species of medicinal plants with sufficient (>95%) accuracy.
   Logistic regression, SVM, and Random Forest were implemented to classify 36 species of medicinal plants with sufficient (up to 95%) accuracy.
3. (results and discussion) There are two values in cells of the table, mean percentage for training and validation set consequently. All approaches showed $> 95\%$ accuracy on validation set; however, eliminated behaviour of learning curves indicates high variance trend and the problem of generalization appears.
   Values in each sell were calculated for validation set. All approaches showed 87% or higher accuracy on validation set; however, eliminated behaviour of learning curves indicates high variance trend and the problem of generalization appears.
4. (experiment description) In the second part the same randomized approach (100 starts with random permutations inside each class) have been used to plot learning curve dependent on the training and validation set size. Permuted input had been separated into equal sized sets due to the lack of experimental data. Training and validation sets have been incremented to show how it affects the loss functional.
   In the second part the same randomized approach (100 starts with random permutations inside each class) have been used to plot learning curve dependent on the training set size. Training set have been incremented to show how it affects the loss functional.
5. In order to this requirement, depth of each decision tree varies from 15 to 25 and number of elementary decision trees should be greater than 20.
   In order to meet this requirement, depth of each decision tree and number of elementary decision trees should be greater than 20.
6. Table 1 , Figure 2, Figure 3, Figure 4, Figure 5,
   Were remade according to the recalculation results and are included in this corrigendum.

We look forward to hearing from you and will try to respond to any further questions and comments you may have. Many thanks for your consideration,

Respectfully,

Pavel Kharyuk
E-mail: kharyuk.pavel@gmail.com
Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow 143026, Russia; Institute of Numerical Mathematics of Russian Academy of Sciences

Dmitry Nazarenko
E-mail: dmitro.nazarenko@gmail.com
Lomonosov Moscow State University, Faculty of Chemistry, Moscow, 119991, Russia

Ivan Oseledets
E-mail: i.oseledets@skoltech.ru
Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow 143026, Russia. Institute of Numerical Mathematics, Russian Academy of Sciences. Gubkina St. 8, Moscow 119333, Russia.

Igor Rodin
E-mail: rodin@analyt.chem.msu.ru
Lomonosov Moscow State University, Faculty of Chemistry, Moscow, 119991, Russia

Oleg Shpigun
E-mail: shpigun@analyt.chem.msu.ru
Lomonosov Moscow State University, Faculty of Chemistry, Moscow, 119991, Russia