# IR and Raman spectra of Ukrainian beers: machine learning exploratory study

**K. S. Yablochkova [1], D.Daniliuk [1], A.I. Yezhola[1]**

*[1]Taras Shevchenko National University of Kyiv , Kyiv, Ukraine, yablochkova@knu.ua*

## Introduction

Beer is one of the world's favorite drinks. Both producers and distributors of beer, require cost-effective and reliable methods for its control and authentication. Traditional methods to assess specific properties of beer are time-consuming and costly [1]. However, IR and Raman spectra of beer are likely sensitive to the differences in the fermentation, production, and storage of this beverage. Recently the robust machine learning classification and clustering methods drew attention of the industry as rapid, non-destructive, and automated beer quality assessment tools [2].

## Method and Dataset

The FTIR and Raman (excitation wavelength 532 nm) spectra for the total of 30 samples of Ukrainian beer were collected in the range from 500 to 3000 $cm^{-1}$. The dataset for Raman spectrum contained the weighted intensities of the spectrum sampled at 0.5 $cm^{-1}$ in order to harvest additional data due to the luminescence of the samples. The dataset for FTIR spectrum was preprocessed to remove the parts of the spectrum where no absorption bands are present.

The dataset contained beers that are either top or bottom-fermented, bottled and canned, and those from large industrial and micro- breweries.

ML classification models (KNN, Decision Tree, Random Forest) were applied to classify the samples according to such properties as the scale of the production (industrial vs. craft), storage condition (can vs. bottle), and fermentation type (top vs. bottom). The performance of models was assessed via cross-validation using well-known metrics, such as accuracy and sensitivity.

## Results

- Not all models demonstrate decent accuracy, although the performance of the decision tree algorithm in discrimination between the mass-produced and microbrewery-produced beers based on FTIR spectra is quite high (accuracy 0.887, sensitivity 0.932). Similarly, the KNN algorithm discriminates well between high- and low- type fermentation beers based on their Raman spectra (accuracy 0.85).
- Based on the ratio of the test and train accuracy metrics the performance of the models is restricted by a small number of samples tested.

## Conclusions

- ML methods can be useful for classifying fermented beverages, according to some of its properties.
- Further measurements with a better choice of hyperparameters is required.

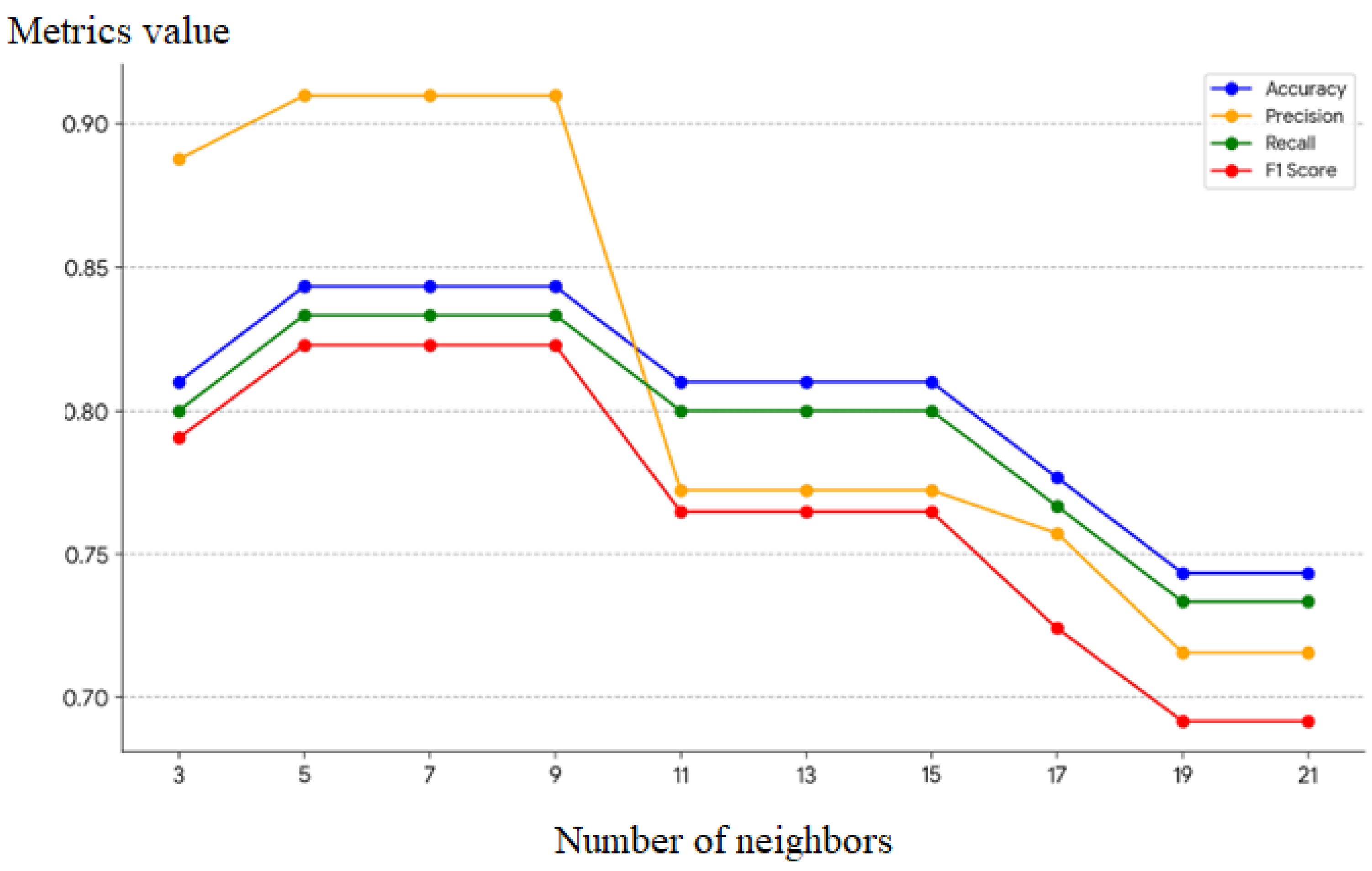


**Figure 1** The choice of the number of neighbors for the Raman spectra- based KNN model and the resulting values of metrics

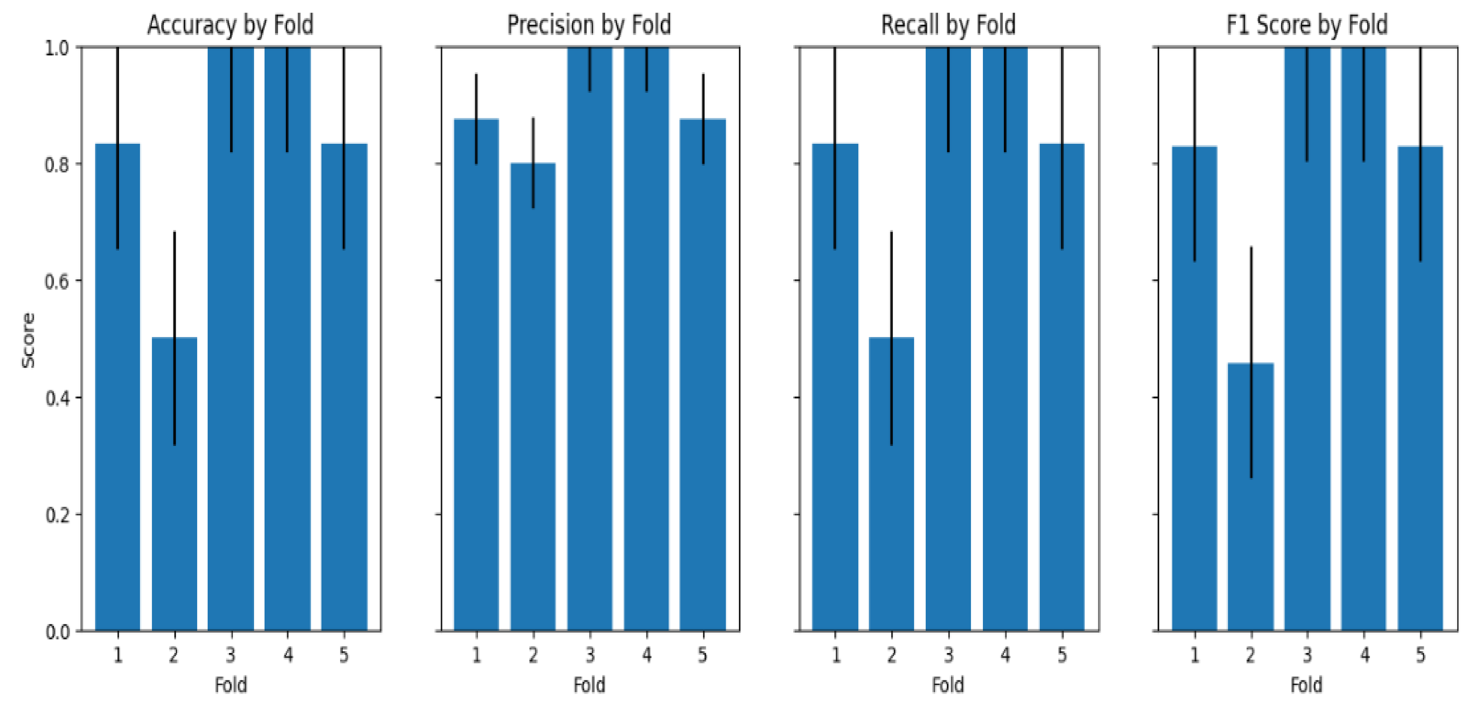


**Figure 2** Cross-validation values of metrics for the Raman spectra- based KNN model

| | | Type of beer (High vs. Low fermentation) | Scale of production (Craft vs. Mass Produced) | The packaging (Glass vs. Al can) |
|---|---|---|---|---|
| **KNN** | Accuracy (Train) | 0.802 | 0.895 | 0.874 |
| | Accuracy (Test) | 0.541 | 0.708 | 0.745 |
| | Sensitivity | 0.546 | 0.708 | 0.765 |
| **Decision Tree** | Accuracy (Train) | 1.0 | 1.0 | 1.0 |
| | Accuracy (Test) | 0.721 | **0.887** | 0.666 |
| | Sensitivity | 0.998 | **0.932** | 0.833 |
| **Random forest** | Accuracy (Train) | 1.0 | 1.0 | 1.0 |
| | Accuracy (Test) | 0.500 | 0.690 | 0.642 |
| | Sensitivity | 0.873 | 0.854 | 0.799 |

**Table 1** Accuracy and sensitivity metrics for the FTIR spectra-based KNN, Decision Tree, Random Forest models

**References:**

[1] Gonzaler V.C et al, Journal of the Science of Food and Agriculture , 2018, 98.2, pp. 618-627.

[2] Gao, Yi-Fang, et al. Food Chemistry, 2024, X 22, pp. 101300.