Nat. Volatiles & Essent. Oils, 2021; 8(4): 3607-3617

NVEO
Natural Volatiles &
Essential Oils

Classification Of Chemicals Present In Essential Oils Using Deep Learning Algorithm

Bindu Krishnan

Department of Data Science, Jain University, Kerala, India 682 042. bindusruthy@gmail.com

Abstract

Classification of chemical compounds present in the essential oils is considered vital to check the genuineness of the chemical oils, its smell and odour. Hence, it is essential in utilising gas chromatography to find the retention indices using the structure of the chemical compounds present in the chemical oil in gaseous state. The utilisation of deep learning model cloud help in determining the retention index of the gases compounds to test the genuineness. In this paper, a model is developed using convolutional neural network (CNN) to detect the presence of molecules while processing the essential oils. In Gas chromatography, the CNN involves in predicting the retention index of the GC on polar and mid-polar phases. The mean square error is measured over the stationary phases on the test datasets to validate the prediction accuracy of the model. The comparison with experimental observation shows that the proposed state-of-art model achieves near optimal training and testing accuracy than other methods.

Keywords: Classification, Prediction, Chemicals, Essential Oil, Deep Learning

1. Introduction

Essential oils are defined as the mostly volatile and odorous fraction of vegetable materials that has been isolated by some physical method. In the alchemical lexicon, the term 'essential oil' refers to the essence or whole flavour and aroma of a plant species, as coined by Paracelsus and other alchemists. Essential oils in plants have a distinct and sometimes diagnostic odour that helps identify the plant. Each essential oil is made up of organic chemicals whose type and relative amounts depend on a variety of agricultural circumstances, including the surrounding climate and soil conditions, as well as when the plant was harvested and how it was handled afterward [1].

The number of essential oils discovered and identified from various plant species has risen to over 3000, with hundreds of them being commercially manufactured. The price of any commercial essential oil is determined by the percentage of oil yield from the plant species, the rate of production, and, most importantly, the intended use of the oil in the product [2] [3].

Many plant parts, such as fruit, leaves, roots, bark, and heartwood, are used to make essential oils. Other plant parts that are used to make essential oils are balsam and gum [4] [5]. Essential oils are made by processing plant materials to remove glycerides, cellulose, sugars, starches, salts, tannins, and other minerals, resulting in oils that are mainly free of these substances. From 0.05–18%, the production of essential oils from plants varies greatly [6].

The essential oil is found in the oil sacs of various plant parts and is extracted using a combination of comminution, heat, water, and solvents to create the final product. The three fundamental procedures used for essential oil isolation are distillation, selective solvent extraction, and mechanical expression, with enhancements or alterations introduced for each when accessible. Resinoids, concretes, absolutes, distillates, and other derivatives are all considered essential oils [7].

The gas chromatographic retention index (RI) measures how well a given stationary phase (SP) retains a specific molecule without being heavily dependent on the chromatographic conditions in use. RI can be used in a variety of gas chromatographic settings and instrument configurations [8].

A crucial part of SP forecasting is the ability to accurately anticipate RI for the poles and mid-poles. For polar SP, reference experimental RI is only accessible for a small subset of the total number of substances. Large adequate RI databases are not available for the mid-polar SP region. For volatile GC-MS analysis, RI prediction for mid-polar SP can be used, as these SP are commonly used in analytical practise and have a wide range of applications [9].

It is possible to pick SP for given analytes and plan an experiment based on RI predictions for a variety of SP. Deep learning approaches can be employed for both polar SP and non-polar SP. Such approaches can attain a decent level of accuracy with only a few thousand compounds. Small data sets hinder prediction for mid-polar SP, and new techniques are likely required.

Human-engineered features are generally referred to as deep learning when used in machine learning models that use deep neural networks with complicated architectures to derive information directly from raw features. Among all scientific disciplines, deep learning is one of the most quickly evolving approaches [10]. Deep learning algorithms consistently outperform more traditional approaches in all of these scenarios.

To detect the presence of compounds in essential oils during processing, a model is created utilising a convolutional neural network (CNN). By estimating the retention index of GC on polar and mid-polar phases, the CNN aids in Gas Chromatography (GC). To verify the model accuracy in making predictions, the mean square error is calculated over the dataset’s stationary phases.

2. Background

The RI prediction, also known as retention relationships of molecular structure, has been the subject of numerous earlier studies. The majority of these studies take into account only a few dozen chemically homogeneous chemicals from limited data sets. For relatively tiny test sets, good accuracy is often obtained in these situations [11]-[14].

Such models, on the other hand, are not very adaptable, and their applicability boundaries are often a mystery. References [14] - [16] evaluate earlier QSRR efforts that focused on small data sets. Developing flexible RI prediction models that may be used on virtually any structure is a critical goal to be completed soon. Multiple publications [17–19] on RI prediction for various compounds use data sets ranging from hundreds to tens of thousands of chemicals.All but a few of these pieces have already been thoroughly examined in our prior work [17].

When large enough training sets are available, deep learning models generally outperform models based on hard-coded characteristics. At least four studies [17] through [19] - [21] have used huge data sets and deep learning algorithms for RI prediction in recent years. The prediction of GC retention time by various steroids is done using deep learning in another study [22].

3. Proposed Method

With the same training data as the authors of the original article, we trained our second-level model to predict RI (in ms). Using a large enough data set, it was possible to create models for second-dimension retention times and indices.

DB-1701 was studied with a set of 36 chemicals, while DB-210 was studied with 130 compounds. There are multiple series of homologues in the later data set, so it is not small either. There are compounds in it from several classes, but it is not extremely diversified. These sets of data can be utilised to train the model, but the model applicability and generalizability cannot be determined based on them. When dealing with tiny data sets, it is impossible to know whether a complicated model would not "overfit" for specific classes of molecules in the data set or whether it will have adequate accuracy in general.

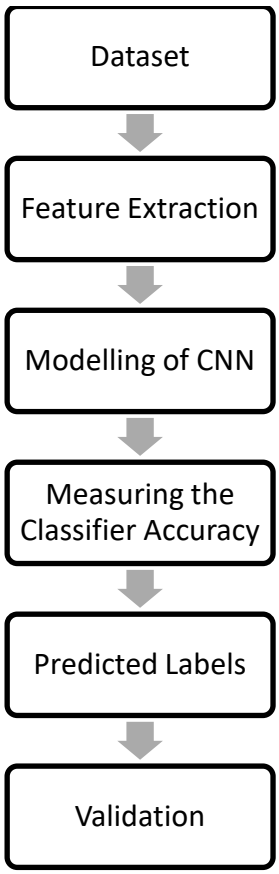

Figure 1: Proposed Methodology

3.1. CNN Prediction

A CNN is a feedforward neural network of this type. Pattern recognition and image processing both benefit from the usage of CNN, a fast and accurate recognition technique. It has a lot of advantages, such as a straightforward structure, fewer training requirements, and adaptability. As a result, voice analysis and image recognition have both seen a spike in interest in the technology. Biological neural networks are more like it because of its weight-shared network structure. It reduces the number of weights and the complexity of the network model.

CNN has two levels of structure. A feature extraction layer connects each neuron input to the previous layer local receptive fields and extracts the local feature from those fields. Once the local features have been extracted, the relationship between their positions in reference to other characteristics can be calculated. Feature map layers are another network component; each computing layer is made up of a number of feature maps. Feature maps are made up of planes, with each plane neurons having equal weight. It is because of this that the feature map has shift invariance because its structure uses the sigmoid function as the activation function for the convolution network. Furthermore, the number of network parameters is minimised because neurons in the same mapping plane share weight. There are two distinct extraction structures in the

convolution neural network: one that calculates the local average and a second that extracts additional features. This additional layer reduces the resolution by one for each convolution layer in the neural network.

For the majority of its applications, CNN is used to detect distortion in two-dimensional visuals. If we utilise CNN, we can avoid explicitly extracting features because the feature detection layer of CNN learns from training data instead of doing so explicitly. In addition, the weights of the neurons in the same feature map plane are the same, allowing the network to conduct many studies at once. As compared to a neural network, the convolution network has a significant advantage in this area. Because of the unique structure of the CNN local shared weights, voice recognition and image processing benefit greatly from CNN technology. Its design is more in line with the structure of the human brain genuine neural network. By using shared weights, the network becomes simpler. Feature extraction and classification are simplified since multi-dimensional input vector images can enter the network immediately, avoiding the need for data reconstruction.

Convolution Process: The deconvolution of the input picture is performed using a trainable filter Fx, followed by the addition of a bias bx, to produce a convolution layer Cx.

Sampling Process: In this process, n pixels from each neighbourhood become a pixel, which is then scalar-weighted with Wx + 1 weighted, bias-added with bx + 1, and activated to produce a narrow Sx + 1 times feature map of n times pixels in each neighbourhood.

Local receptivity, weight sharing, and subsampling by time or space are all significant features of CNN technology. These techniques help to extract features while reducing the number of training parameters required. To eliminate explicit feature extraction and implicitly learn from the training data, CNN uses the same neuron weights on the feature mapping surface, which allows it to learn in parallel and so lower the network overall complexity. It is possible to acquire some robustness, scale, and deformation displacement by using a sub sampling structure based on time or space. With the right input information and network design, speech recognition and image processing can benefit from distinct benefits

4. Results and Discussions

The study used 50% trifluoropropylmethyl 50% dimethyl polysiloxane (DB-210) and 14% cyanopropylphenyl 86% dimethyl polysiloxane (DB-1701) as its chemical compounds to be detected in large enough data sets.

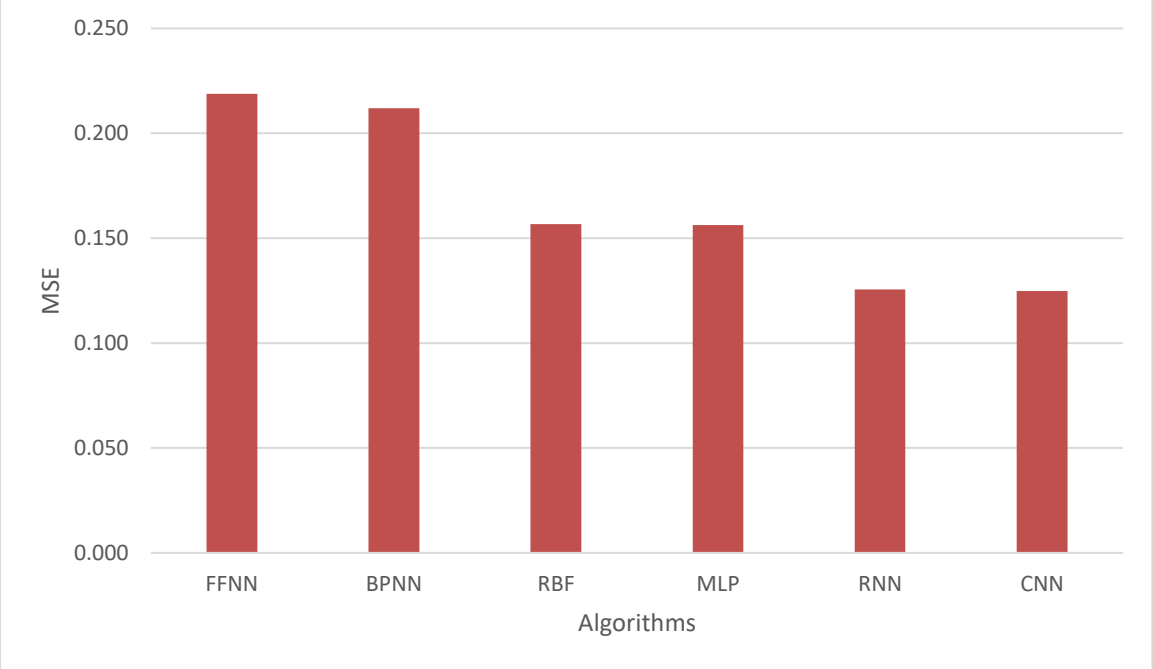

Figure 2: Training Loss

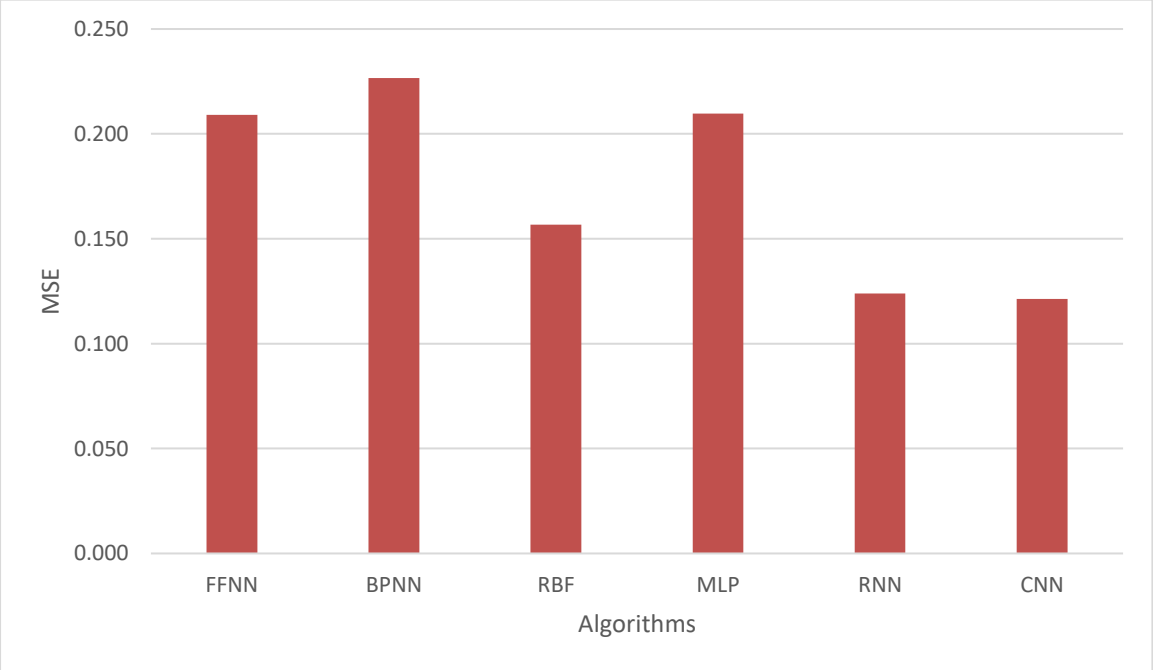

Figure 3: Testing Loss

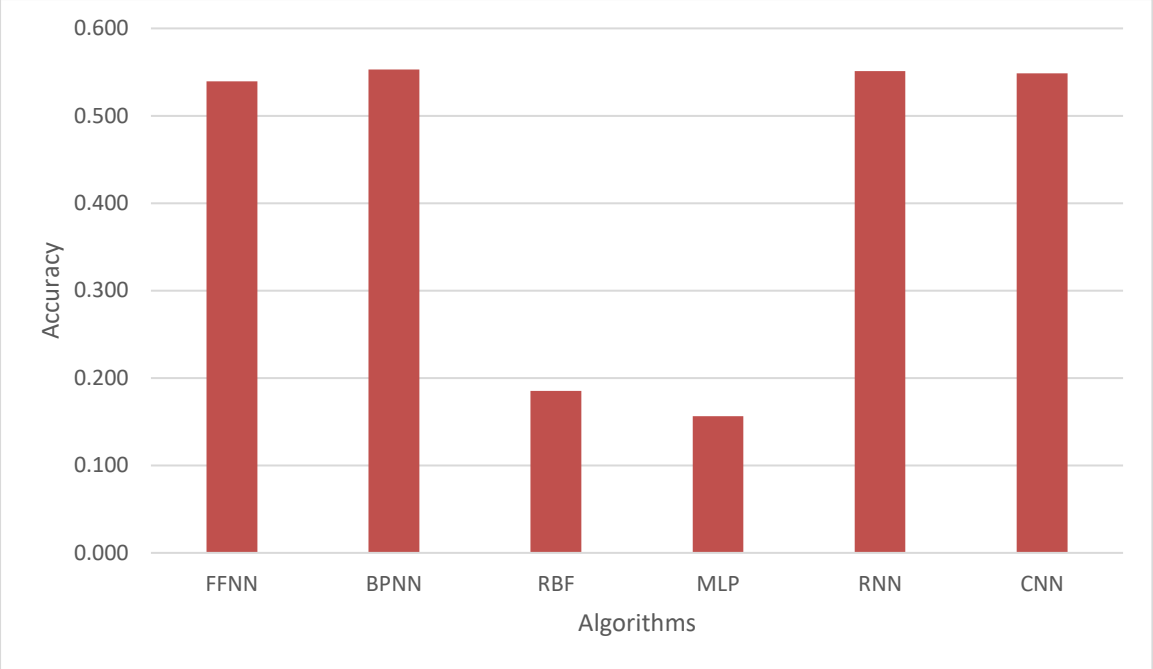

Figure 4: Train Accuracy

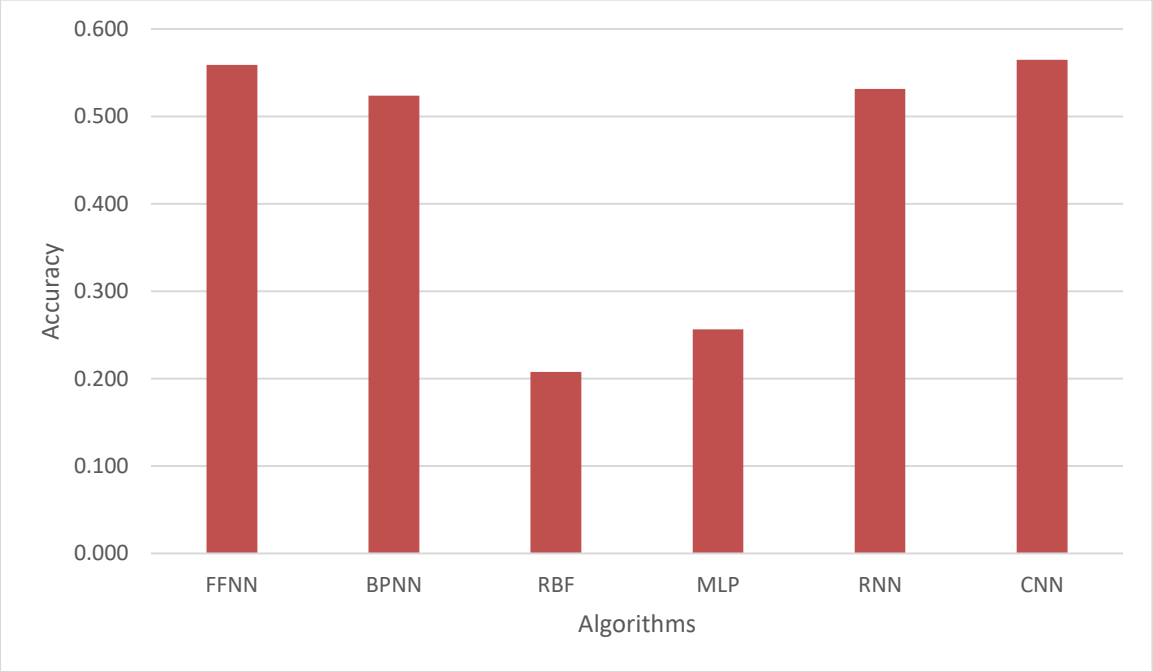

Figure 5: Testing Accuracy

Figure 2-4 shows a schematic representation of the machine learning and deep learning models that were used in this study. Data from non-polar standard/semi-standard SP is used to train CNN. There are a lot of similarities between the structure and hyperparameters of these neural networks and those employed in this research. The fundamental concept is to employ non-polar and polar SP RI values as input features for mid-polar SP RI prediction.

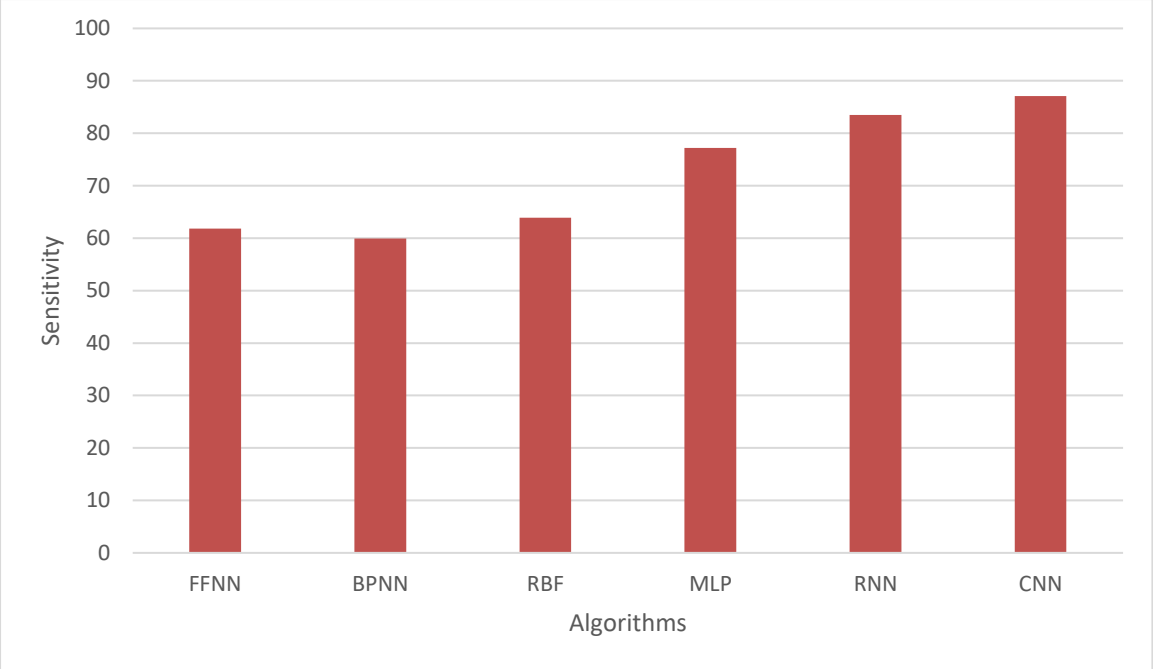

Figure 5: Sensitivity

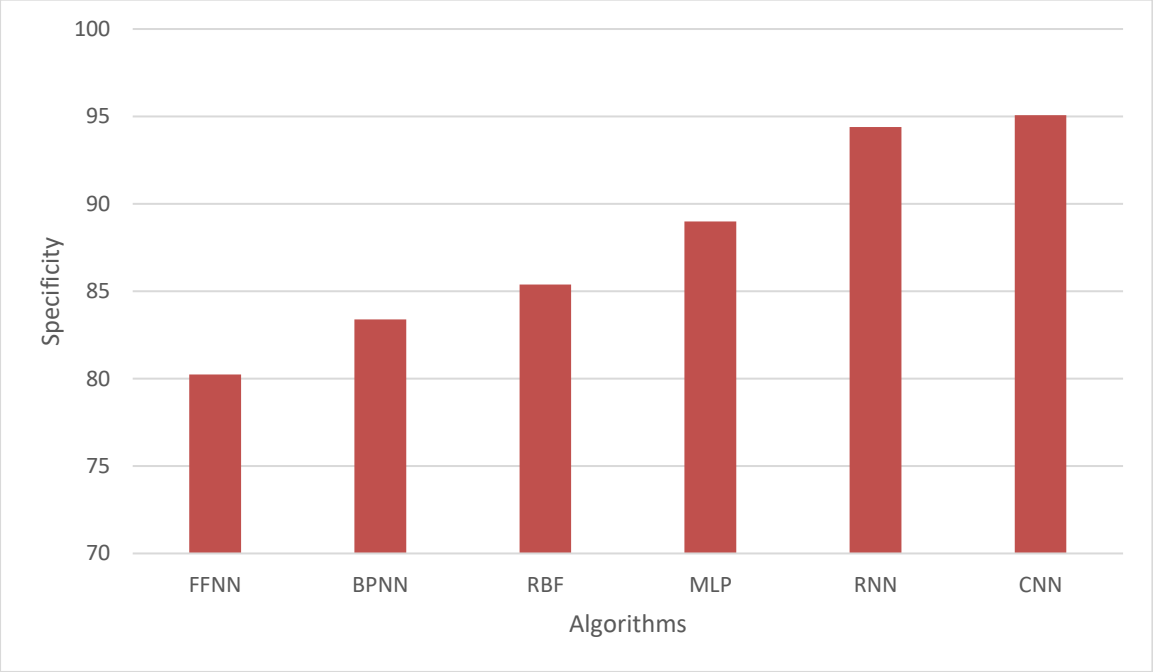

Figure 6: Specificity

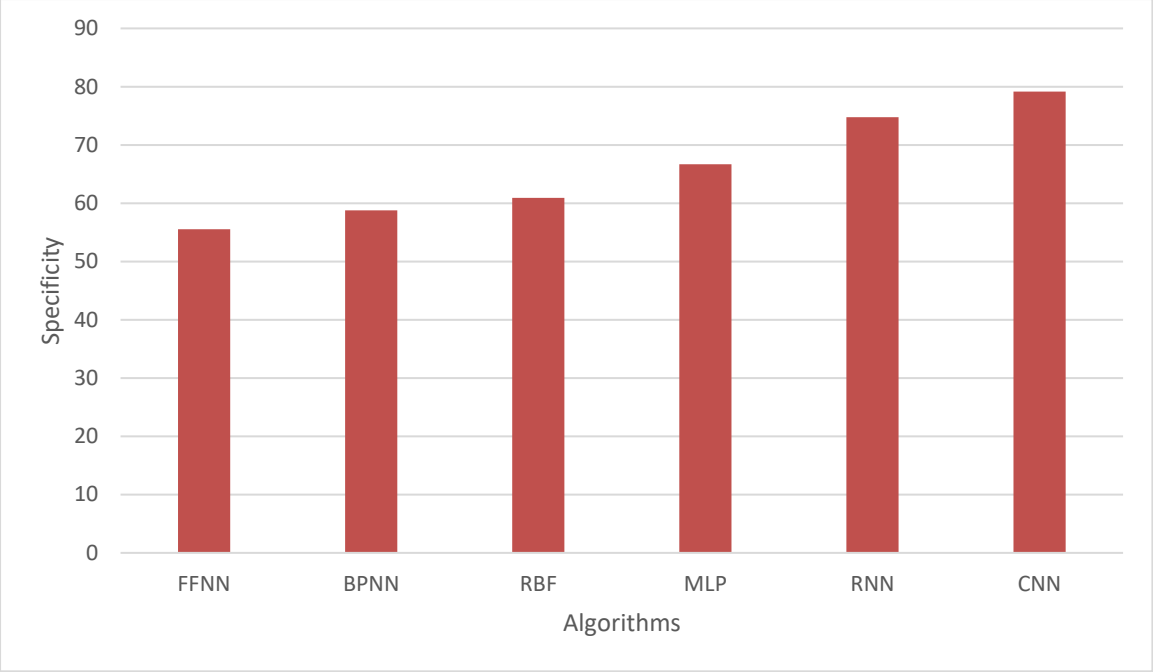

Figure 7: F-measure

Figure 5 shows the results of sensitivity, where the proposed method achieves higher rate of sensitivity than other method. Figure 6 shows the results of specificity, where the proposed method achieves higher rate of specificity. Figure 7 shows the results of F-measure, where the CNN achieves higher rate of F-measure than other methods

5. Conclusions

In this paper, CNN detects the RI of GC while processing the essential oils. In GC, the CNN involves in predicting the retention index of the GC on polar and mid-polar phases. The mean square error is measured over the stationary phases on the test datasets to validate the prediction accuracy of the model. The simulation result shows that the accuracy of the proposed CNN model is higher when compared with other methods via structure-retention relationship in linear quantitative manner.

References

[1] Sharma, S., Barkauskaite, S., Jaiswal, A. K., & Jaiswal, S. (2021). Essential oils as additives in active food packaging. Food Chemistry, 343, 128403.

[2] Falleh, H., Jemaa, M. B., Saada, M., & Ksouri, R. (2020). Essential oils: A promising eco-friendly food preservative. Food chemistry, 330, 127268.

[3] Rehman, A., Jafari, S. M., Aadil, R. M., Assadpour, E., Randhawa, M. A., & Mahmood, S. (2020). Development of active food packaging via incorporation of biopolymeric nanocarriers containing essential oils. Trends in Food Science & Technology, 101, 106-121.

[4] Jugreet, B. S., Suroowan, S., Rengasamy, R. K., & Mahomoodally, M. F. (2020). Chemistry, bioactivities, mode of action and industrial applications of essential oils. Trends in Food Science & Technology, 101, 89-105.

[5] Lammari, N., Louaer, O., Meniai, A. H., & Elaissari, A. (2020). Encapsulation of essential oils via nanoprecipitation process: Overview, progress, challenges and prospects. Pharmaceutics, 12(5), 431.

[6] Mishra, A. P., Devkota, H. P., Nigam, M., Adetunji, C. O., Srivastava, N., Saklani, S., ... & Khaneghah, A. M. (2020). Combination of essential oils in dairy products: A review of their functions and potential benefits. Lwt, 133, 110116.

[7] Syafiq, R., Sapuan, S. M., Zuhri, M. Y. M., Ilyas, R. A., Nazrin, A., Sherwani, S. F. K., & Khalina, A. (2020). Antimicrobial activities of starch-based biopolymers and biocomposites incorporated with plant essential oils: A review. Polymers, 12(10), 2403.

[8] Ruszkiewicz, D. M., Sanders, D., O'Brien, R., Hempel, F., Reed, M. J., Riepe, A. C., ... & Eddleston, M. (2020). Diagnosis of COVID-19 by analysis of breath with gas chromatography-ion mobility spectrometry-a feasibility study. EClinicalMedicine, 29, 100609.

[9] Aksenov, A. A., Laponogov, I., Zhang, Z., Doran, S. L., Belluomo, I., Veselkov, D., ... & Veselkov, K. (2021). Auto-deconvolution and molecular networking of gas chromatography–mass spectrometry data. Nature biotechnology, 39(2), 169-173.

[10] Randazzo, G. M., Bileck, A., Danani, A., Vogt, B., & Groessl, M. (2020). Steroid identification via deep learning retention time predictions and two-dimensional gas chromatography-high resolution mass spectrometry. Journal of Chromatography A, 1612, 460661.

[11] Dossin, E., Martin, E., Diana, P., Castellon, A., Monge, A., Pospisil, P., ... & Guy, P. A. (2016). Prediction models of retention indices for increased confidence in structural elucidation during complex matrix analysis: application to gas chromatography coupled with high-resolution mass spectrometry. Analytical chemistry, 88(15), 7539-7547.

[12] Matsuo, T., Tsugawa, H., Miyagawa, H., & Fukusaki, E. (2017). Integrated strategy for unknown EI–MS identification using quality control calibration curve, multivariate analysis, EI–MS spectral database, and retention index prediction. Analytical chemistry, 89(12), 6766-6773.

[13] Kumari, S., Stevens, D., Kind, T., Denkert, C., & Fiehn, O. (2011). Applying in-silico retention index and mass spectra matching for identification of unknown metabolites in accurate mass GC-TOF mass spectrometry. Analytical chemistry, 83(15), 5895-5902.

[14] Héberger, K. (2007). Quantitative structure–(chromatographic) retention relationships. Journal of chromatography A, 1158(1-2), 273-305.

[15] Kaliszan, R. (2007). QSRR: quantitative structure-(chromatographic) retention relationships. Chemical reviews, 107(7), 3212-3246.

[16] Zhokhov, A. K., Loskutov, A. Y., & Rybal'chenko, I. V. (2018). Methodological approaches to the calculation and prediction of retention indices in capillary gas chromatography. J. Anal. Chem, 73(3), 207-220.

[17] Matyushin, D. D., & Buryak, A. K. (2020). Gas Chromatographic Retention Index Prediction Using Multimodal Machine Learning. Ieee Access, 8, 223140-223155.

[18] Vrzal, T., Malečková, M., & Olšovská, J. (2021). DeepReI: Deep learning-based gas chromatographic retention index predictor. Analytica Chimica Acta, 1147, 64-71.

[19] Qu, C., Schneider, B. I., Kearsley, A. J., Keyrouz, W., & Allison, T. C. (2021). Predicting Kováts Retention Indices Using Graph Neural Networks. Journal of Chromatography A, 1646, 462100.

[20] Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. IEEE Access, 7, 53040-53065.

[21] Matyushin, D. D., Sholokhova, A. Y., & Buryak, A. K. (2019). A deep convolutional neural network for the estimation of gas chromatographic retention indices. Journal of Chromatography A, 1607, 460395.

[22] Randazzo, G. M., Bileck, A., Danani, A., Vogt, B., & Groessl, M. (2020). Steroid identification via deep learning retention time predictions and two-dimensional gas chromatography-high resolution mass spectrometry. Journal of Chromatography A, 1612, 460661.