2019 1st International Conference on Cybernetics and Intelligent System (ICORIS)
22nd-23rd August 2019 at Institut Teknologi dan Bisnis (ITB) STIKOM Bali, Indonesia

Application of PCA-SVM and ANN Techniques for Plastic Identification by Raman Spectroscopy

Wilem Musu
Department of Biological and Environmental Chemistry
Kindai University
Iizuka, Japan
wilem.musu@dipanegara.ac.id

Hirofumi Kawazumi
Department of Biological and Environmental Chemistry
Kindai University
Iizuka, Japan
kawazumi@fuk.kindai.ac.jp

Akihiro Tsuchida
Saimu Corporation
Iizuka, Japan
akihiro@saimu-net.ne.jp

Nobuto Oka
Department of Biological and Environmental Chemistry
Kindai University
Iizuka, Japan
nobuto.oka@fuk.kindai.ac.jp

***Abstract*—The mechanical recycling of plastics is one of the most efficient approaches for reducing carbon dioxide emissions. Purification of the plastics from shredded waste materials requires versatile techniques, such as optical identification by Raman spectroscopy. The identification procedure demands the spectroscopy expertise to assign molecular structures from spectral peaks. In this study, we demonstrate applications to classify plastics using machine learning techniques under practical recycling industry conditions. Combining the techniques of principal component analysis (PCA) and support vector machine provides an accurate and robust classification of the valuable plastics of polypropylene, polystyrene, and acrylonitrile–butadiene–styrene copolymer. The identification accuracy remained above 95%, even with noise 3 times larger than the original intensity. For noise 10 times larger, the accuracy was more than 70%. Fast and simple computation is also useful for industrial applications, resulting from dimension reduction of the spectroscopic data by PCA. Furthermore, artificial neural networks showed high accuracy, close to 100%, after a few epoch calculations.**

Keywords*—*mechanical recycling, optical identification, principal component analysis, support vector machine, artificial neural networks

I. INTRODUCTION

The annual growth of solid plastic waste has created a serious problem for municipal life of global pollution, both on the land and in the oceans [1]. Accordingly, the plastic recycling industry has developed in many countries, aimed at reducing environmental damage from plastic waste. Mechanical recycling is the most useful approach to save fuel resources and to reduce carbon dioxide emissions. Major plastic sorting processes are currently based on gravity separation. However, with this method, the purity of the homogeneous plastic component is around 90%. For closed-loop recycling that produces high-grade recycled resins, the purity must be increased. Recently, optical identification techniques of plastics have been developed, such as near-infrared (NIR) absorption and Raman scattering spectroscopy. A Raman plastic identification apparatus has been implemented for the industrial and massive recycling of a Japanese company, as shown in Fig. 1 [2]; the small shredded plastic pieces on the conveyor belt (speed of 100 m/s) are classified into two categories based on their Raman spectra in less than 10 ms. The achieved recognition accuracy, of approximately 95%, is higher than that of NIR identification.

Raman spectroscopy is adequate to gain strong discrimination between samples [3,4] because spectral peaks involve information on the vibrations of chemical bonds, resulting in molecular structures, particularly for the assignment of plastic resins. Nevertheless, Raman scattering measurements produce spectra with pixel numbers of 512, 1024, and 2048 in a CCD detector. Analysis of the spectra requires chemistry and spectroscopy expertise, and the classification process may become time-consuming. Meanwhile, no empirical and automatic recognition procedures are available in the industrial field, which save human resources and expand the applicability of the optical identification of plastic recycling. For this purpose, machine-learning techniques should be useful according to the recent progress of processing power and algorism of computer science.

In this paper, two applications are demonstrated to classify three kinds of plastics: polypropylene (PP), polystyrene (PS),

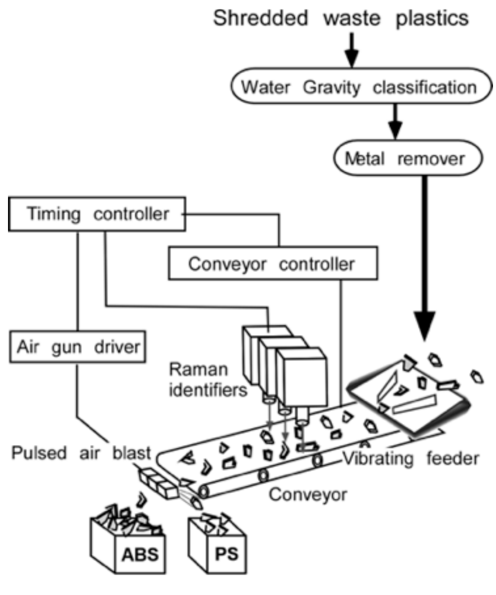

Fig. 1. Mechanical recycling with Raman spectroscopy identification.

and acrylonitrile–butadiene–styrene copolymer (ABS), which are major target components in waste home electric appliances and the end-of-life vehicle recycling industry. We adopted the machine-learning algorithms of the support vector machine (SVM) and artificial neural networks (ANNs) because this spectroscopic analysis could use supervised learning based on the correct spectra. SVM is a remarkable technique that separates two or more classes using optimum boundaries to maximize the margins between data in different categories [5]. To make the computation simple and fast, we reduced the dimensions of the essential peaks using the principal component analysis (PCA). The combination of PCA-SVM was accurate and robust under an artificially noisy condition. To compare the achievement of the combination of PCA-SVM, we also implemented an ANN algorithm as proof that machine learning can improve mechanical recycling technology.

II. Experimental

A. Measurement of Raman Spectra

To achieve the high-speed identification of plastics for industrial-scale recycling, we developed a homemade Raman apparatus [6-8]. It comprises a high-power laser diode (785 nm, 1 W), high throughput optical system, a highly sensitive back-illuminated FFT-CCD (1024 px), and an on-board digital signal processing system. In the study, the measurement time was 6 ms to consider an appropriate noise level. The plastic samples were 115 small pieces taken from used electric appliances, including 61 pieces of PP, 33 pieces of PS, and 21 pieces of ABS. Each correct plastic resin type was validated using FT-IR measurements.

B. Peak Extraction and Composition of Noisy Test Data

Fig. 2 shows an example of the Raman spectra of PP, PS, and ABS for the wavenumber 100–3300 cm-1. The horizontal axis in this figure is also displayed in pixel number of the CCD for rapid data transfer. There are many peaks in the Raman spectra, but we only chose the four peaks corresponding to the characteristic molecular vibration to distinguish each type of

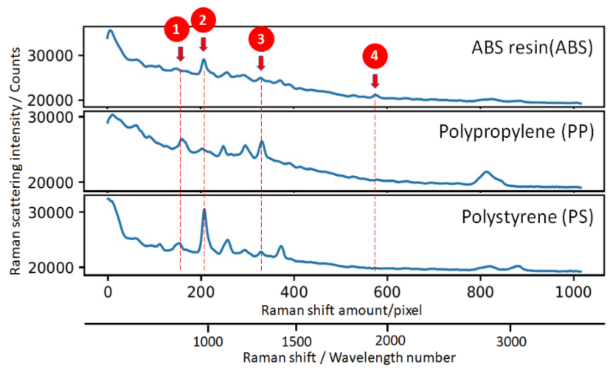

Fig. 2. Raman spectra of plastics and the four peak positions extracted for the identification.

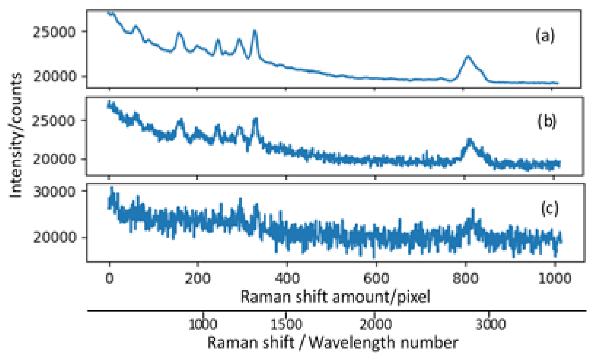

Fig. 3. Comparison of the noise levels in the Raman spectra of PP: (a) original, and with (b) 7.5-times noise addition and (c) 15-times noise addition.

plastic: peak 1 of C-H bending at 820 cm-1 / 165-th pixel, peak 2 of benzene ring breathing at 1000 cm-1 / 208-th pixel, peak 3 of C=C stretch at 1600 cm-1 / 373-rd pixel, and peak 4 of C=N stretch at 2240 cm-1 / 580-th pixel at the center of the peaks. We accumulated values of 20-point data around each peak center and subtracted the next 10-point data as a baseline from the sum to obtain the peak intensity. The four peak intensities in the spectrum were normalized as the total intensity is unity. This numerical peak extraction supplied four variables of P1, P2, P3, and P4.

A training dataset was obtained from these 115 spectra of the plastic pieces. Moreover, a test dataset was generated by the simulation to add artificial noises; we added a Gaussian noise to the spectra measured originally, as shown in Fig. 3. The noise composed of 0.75, 1.5, 2.5, 3.75, 7.5, and 15-times larger normal distribution fluctuations than the original spectrum baseline. Fig. 3 shows a comparison of the noise intensity in the Raman spectra of PP by way of example. Some weak peaks vanish by visual looking.

C. Principal Component Analysis

After extracting the values of the four essential peaks, we made a two-dimensional dataset for the SVM using the PCA algorithm for further simple and fast computation in the identification. PCA is the most popular multivariate statistical technique to reduce the dimensionality of a dataset [9].

The features of PCA extract the important information from the four peaks and express it as new orthogonal variables, called principal components, by the calculation of

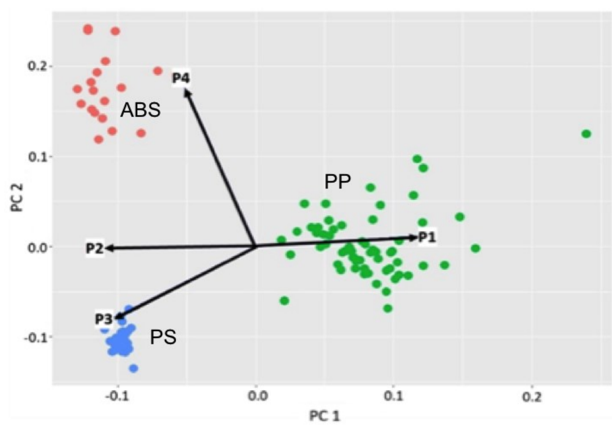

Fig. 4. Contributions of the essential peaks toward principal components.

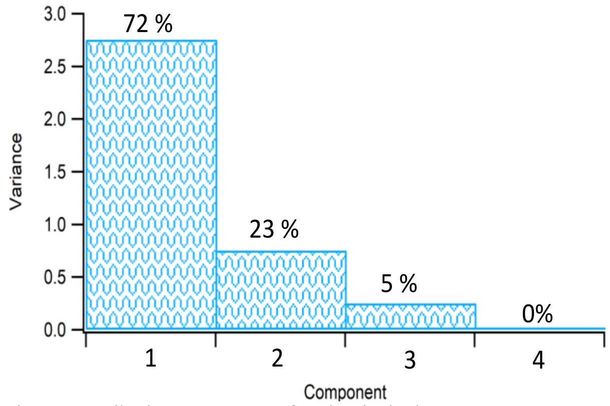

Fig. 5. Contribution percentage of each principal component.

a covariant matrix and eigenvalues. The number of variables decreased to two new variables of the first and second principal components (PC1 and PC2, respectively) used in the SVM classification.

In Fig. 4, the four arrows show the contribution of each peak toward the principal components; the arrow projection length to each principal component axis represents the contribution amount. Major information regarding PP and PS comes from P1 and P3, respectively, which both have a positive correlation, while for ABS comes strongly from P4. These results are reasonable for the origin of the peaks from the vibration of the chemical bonds. The arrow of P2 indicates a negative correlation to PP and positive to PS and ABS. The important information of 95% in the four peaks was collected from PC1 and PC2, as shown in Fig. 5. Therefore, the PCA dimension reduction may work well, as expected.

D. *Implementation of Machine-Learning Algorithm*

The analysis procedure implementing the machine-learning algorithms of SVM and ANNs is shown in Fig. 6. We used a free statistical computing environment of R version 3.4.2, installed with “e1071” library, on a Windows 10 platform for the PCA and SVM. For ANNs, the anaconda package was used for the Python-TensorFlow developing environment installed with Keras library. Python was also used in data pretreatments, such as for peak extraction, peak area calculation, normalization, and noise addition.

III. Results and Discussion

A. *Support Vector Machine Classification*

The multiclass SVM discriminates more than two classes by using the best plural hyperplanes in the feature space with a maximized margin scale. We applied the SVM technique to the datasets of PC1 and PC2. Fig. 7 shows the results of the categorization for the original and two noisy data. The scatter plots indicate the increase of the noise. The original data, without the noise addition, shows 100%-correct identification, and the hyperplanes of the original data were

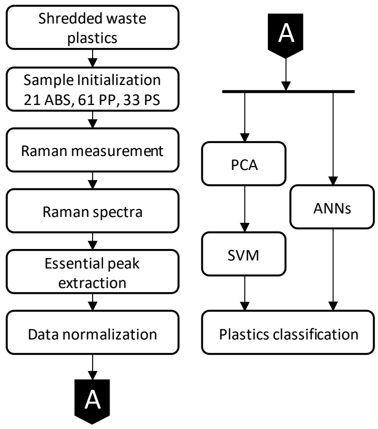

Fig. 6. Schema of implementation of machine-learning algorithms.

used as the training data. A similar approach to the PCA-SVM analysis in the NIR spectroscopy for waste plastics also provides accurate categorization, as in a recent report [10]. Such a mutilative analysis straightforwardly presents the high accuracy in the original data with less noisy condition. An evaluation of the robustness to the noise is crucial because industrial-scale plastic waste sorting is operated in larger, noisy conditions resulting from shorter measurement times and many external noise sources.

The accuracy rates of plastic identification were obtained by counting the number of plots settled in the correct categories. At first glance, the accuracy rate of ABS

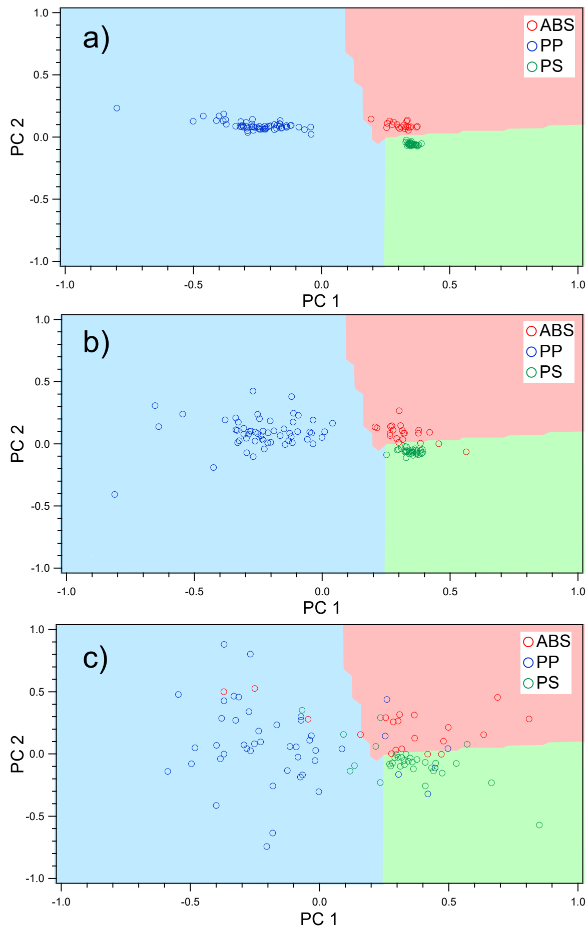

Fig. 7. Classification of combination technique of PCA-SVM: (a) original, and with (b) 7.5-times noise addition and (c) 15-times noise addition.

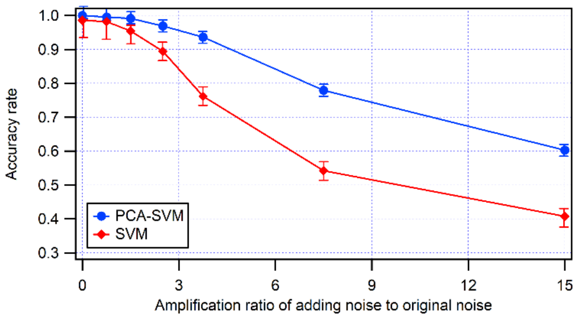

Fig. 8. Comparison of the robustness in the identification accuracy between PCA-SVM combination and simple SVM with changing noise intensity. The vertical bars indicate deviations in the five experiments.

decreases more easily than those of PP and PS because the peak intensities of the Raman spectrum of ABS, shown in Fig. 2, are small, although the noise addition effects should be larger compared with the others.

We evaluated the robustness through the combination of PCA and SVM analyses by increasing the noise intensity artificially added. The relationships between the noise intensity and the total accuracy of the three kinds of plastics are shown in Fig. 8. Two curves show the dependences on the accuracies in the PCA-SVM combination analysis and direct SVM analysis with all four-peak data. The former presents a smaller decrease of the accuracy and a stronger tolerance to the noise than the latter. The identification accuracy remained more than 95% with noise 3-times larger than the original intensity. Furthermore, the accuracy remained above 70% under a 10-times larger noisy condition in which some peaks in the spectra vanish at quick glance.

The vertical bars in Fig. 8 represent the standard deviations of the five evaluating experiments because the random Gaussian noise additions produced fluctuation of the accuracy. The PCA-SVM combination technique also presents smaller deviations than the four-peak SVM. Such robustness to the noise in the recognition may indicate that the dimension reduction by PCA is effective to remove the excess information in the noisy spectral data. Meanwhile, the four-peak SVM analysis requires larger computer resources and may induce a time-consuming calculation. Furthermore, the four-dimension data cannot draw the categorization figure visually, such as in Fig. 7, because its result is plotted in four-dimension hyperspace. The PCA-SVM combination technique developed in this paper provides intuitive identification in the simple scattered plots that is very useful in the practical field of the recycling industry.

B. *Artificial Neural Network Classification*

ANNs are computational algorithms that are accepted in many disciplines for modeling complex real-world problems [11], especially classification problems. In this paper, we demonstrate the applicability of the implementation of the ANN technique to the categorization of plastics with the four essential peaks in Raman spectra. In addition, we intend to extend the application of machine-learning techniques beyond SVM.

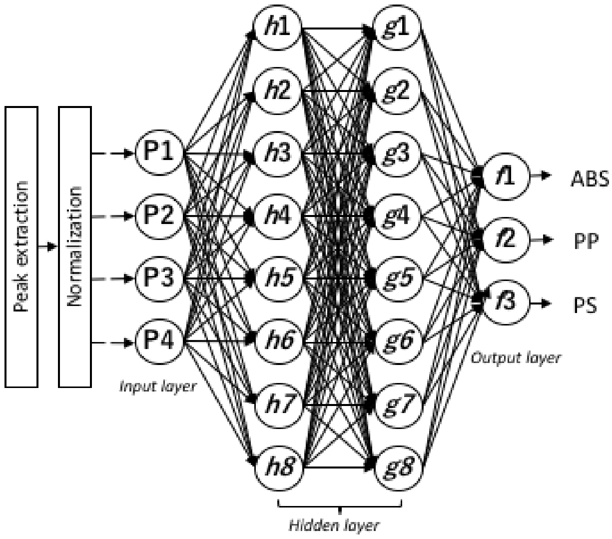

Fig. 9. Architecture of ANN classification.

The data condition is similar to that used in the PCA-SVM approach, but in this case, we did not apply PCA because ANNs can automatically improve the error value by a backpropagation mechanism depending on the amount of epoch [12]. The values in the input layer are the normalized data of the four peaks. The output layer is the categories of ABS, PP, and PS. We included two or three hidden layers to compare the results using the ReLu activation function, as shown in Fig. 9.

The spectral data of the 115 plastic pieces were divided into three types using the ANNs; training data, validation data, and testing data [13]. We used 80% of the data as training data, while 20% was used for validation. At this stage, we wanted to confirm whether ANNs are valid as a useful platform to evaluate plastic classification. Therefore, we did not add noise to the spectra as testing data. The 19-testing data, obtained randomly from the original spectra, were divided proportionally and automatically by the software.

Fig. 10 shows a comparison of the convergences of the loss values of the two and three hidden layers according to epochs up to 1000. The converging curves indicate an improvement of the loss value, which became better and shorter when the hidden layer was added because of the converging point around 600 and 300 epochs for the two and three hidden layers, respectively. The accuracy of classification became 100% after several hundred epochs, and it shows a similar behavior as the loss value curve. Increasing the hidden layer offers fast convergence, as shown in Fig. 10.

IV. Conclusion

In this research, application of machine-learning techniques for plastic identification based on Raman spectroscopy was evaluated to demonstrate the possibility of accurate and fast categorization in the recycling industry. After determining the parameters in the machine-learning algorithms, the plastic sorting system was operated without spectroscopy.

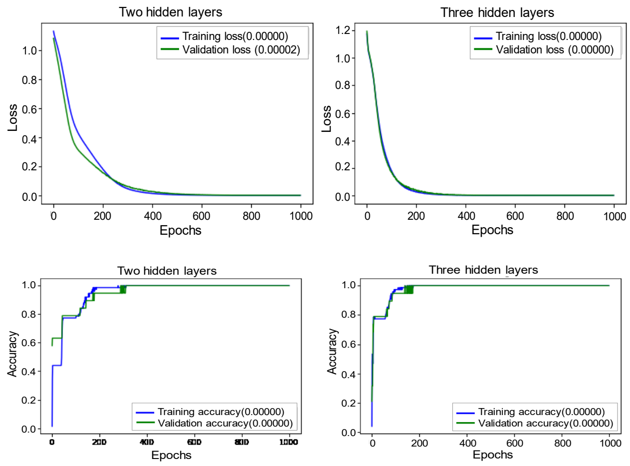

Fig. 10. Loss values and accuracies in training and validation of ANNs.

The combined PCA-SVM technique developed in this paper could recognize the kinds of plastics with purities of over 90% under the rather noisy condition. Simple and fast computation is also useful for industrial applications resulting from dimension reduction of the spectroscopic data by PCA.

The ANN technique showed a high accuracy, of close to 100%, of plastic identification after a few hundred epoch calculations. Furthermore, increasing the hidden layer improved the accuracy straightforwardly. The implementation of these machine-learning algorithms to the online apparatus, as shown in Fig. 1, is under contemplation as future works.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 15K05589.

REFERENCES

[1] R. Geyer, J. R. Jambeck, and K. L. Law, "Production, use, and fate of all plastics ever made - Supplementary Information", *Sci. Adv.*, July, pp. 25–29, 2017.

[2] K. Ragaert, L. Delva, and K. Van Geem, "Mechanical and Chemical Recycling of Solid Plastic Waste", *Waste Manag.*, vol. 69, pp. 24–58, 2017.

[3] H. Vašková, "A Powerful Tool for Material Identification: Raman Spectroscopy", *Int. J. Math. Model. Methods Appl. Sci. A*, vol. 5, pp. 1205–1212, 2011.

[4] C. Muehlethaler, G. Massonnet, and P. Esseiva, "The Application of Chemometrics on Infrared and Raman Spectra as a Tool for The Forensic Analysis of Paints", *Forensic Sci. Int.*, vol. 209, pp. 173–182, 2011.

[5] W. Schumacher, M. Kühnert, P. Rösch, and J. Popp, "Identification and Classification of Organic and Inorganic Components of Particulate Matter via Raman Spectroscopy and Chemometric Approaches", *J. Raman Spectrosc.*, vol. 42, pp. 383–392, 2011.

[6] H. Kawazumi, A. Tsuchida, T. Yoshida, and Y. Tsuchida, "High-performance Recycling System for Waste Plastics using Raman Identification", in *Progress in Sustainable Energy Technologies Vol II: Creating Sustainable Development*, 2014, pp. 519–529.

[7] A. Tsuchida, H. Kawazumi, A. Kazuyoshi, and T. Yasuo, "Identification of Shredded Plastics in Milliseconds using Raman Spectroscopy for Recycling", *Proc. IEEE Sensors*, pp. 1473–1476, 2009.

[8] A. Tsuchida, T. Yoshida, Y. Tsuchida, and H. Kawazumi, "Identification and Degradation Estimation of Waste Plastics for Recycling Using Raman Spectroscopy", *Bunseki Kagaku*, vol. 61, pp. 1027–1032, 2012.

[9] I. T. Jolliffe, "Principal Component Analysis. Second Edition", *Springer Ser. Stat.*, vol. 98, p. 487, 2002.

[10] S. Zhu, H. Chen, M. Wang, X. Guo, Y. Lei, and G. Jin, "Plastic Solid Waste Identification System based on Near Infrared Spectroscopy in Combination with Support Vector Machine", *Adv. Ind. Eng. Polym. Res.*, vol. 2, pp. 77–81, 2019.

[11] I. A. Basheer and M. Hajmeer, "Artificial Neural Networks: Fundamentals, Computing, Design, and Application", *J. Microbiol. Methods*, vol. 43, pp. 3–31, 2000.

[12] Svidhya, "Fundamentals of Deep Learning – Starting with Artificial Neural Network", pp. 1–16, 2016.

[13] F. Karray, "Introduction Features Fundamentals Madaline Case Study: Binary Classification Using Perceptron Fundamentals of Artificial Neural Networks", 2009.