MEKATRONIKA
e-ISSN: 2637-0883
VOL. 3, ISSUE 2, 70 – 76
DOI: https://doi.org/10.15282/mekatronika.v3i2.7390

ORIGINAL ARTICLE

The Classification of FTIR Plastic Bag Spectra via Label Spreading and Stacking

Omair Rashed Abdulwareth Almanifi[1], Ng Jee Kwan[2] and Anwar P.P. Abdul Majeed[1,3,4,5*]

[1]Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang (UMP), 26600 Pekan, Pahang Darul Makmur, Malaysia.
[2]IDIR Solutions, Jalan Kpk 1/2, Kawasan Perindustrian Kundang, 48020 Rawang, Selangor
[3]Centre for Software Development & Integrated Computing, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia
[4]Faculty of Engineering, Technology and Built Environment, UCSI University. Kuala Lumpur Campus. 56000 Cheras. Kuala Lumpur. Malaysia
[5]EUREKA Robotics Centre, Cardiff School of Technologies, Cardiff Metropolitan University, CF5 2YB. Cardiff. UK

ABSTRACT – Whereas plastics are a group of the most useful materials, widely used in all walks of life, the plastic waste that is produced daily poses a great threat towards wildlife and the planet as a whole. The use of biodegradable plastics is an important step in combating the plastic crisis. FTIR spectroscopy is a non-destructive method used for identifying different types of materials, however interpreting spectra produced by such spectrometers is both susceptible to human error, and time-consuming, not to mention that the industry suffers from a great of specialists, in the field of spectroscopy. Utilising machine learning as a method of filling the mentioned issue is suggested by this paper. Four pipelines were investigated, consisting of two machine learning algorithms, a stacked model that stacks the KNN, SVM and RF algorithms together, and Label spreading, as well as two different dimensionality reduction methods namely; SVD and UMAP. The pipelines studied seemed to show great predictivity at 100% classification accuracy acquired by the SVD-Stacked pipeline when data was sampled using an Agilent Cary 660 FTIR Spectrometer, and 99.18% by the same model when IDIR BP10 spectrometer was employed for sampling instead. The semi-supervised learning model (Label Spreading) seemed to achieve close enough accuracy at 99.82% in the case of the former dataset, and 97.54% for the latter, at a labelling rate of only 10% of the full datasets.

ARTICLE HISTORY
Received: 8th June 2021
Revised: 7th July 2021
Accepted: 31st July 2021

KEYWORDS
FTIR
Spectroscopy
Machine learning
Semisupervised learning
Plastic

INTRODUCTION

Plastic is a wide group of polymeric materials that is characterised by being light, durable, and cheap, enjoying a high rate of plasticity, as the word itself might suggest, allowing it to be moulded and reshaped with ease. These characteristics made plastics very useful in all walks of life, prompting manufacturers to produce very large numbers of plastic-based products yearly. In 2019 alone about 368 million metric tons were produced, expected to only rise to 600 million tons by 2025 [1]. Overall, the current estimated accumulated total of produced plastic since 1950 is more than 8.3 billion metric tons [2]. This high supply and demand of plastic resulted in a very high increase in plastic waste as well, with about 60% of the aforementioned accumulated considered waste. This waste is often disposed of in an unsustainable manner, with about a truck-load of plastic disposed into the oceans every minute[3]. Plastic waste provides to be a great source of pollution and a hazard to both humans and wildlife, with at least 86% of sea turtles, 23% of marine mammals and 36% of seabirds being impacted by plastic forming the majority of marine debris [4].

The devastating effects of plastic waste prompt for sustainable solutions, one of which is the use of biodegradable plastic, that is to say, plastics that decompose by natural means, typically by means of microbial living organisms. Restrictions and procedures implemented with the purpose of transitioning from the use of unsustainable plastic to biodegradable plastic is often met with difficulty especially during the process of identifying biodegradable plastics. Fourier Transform Infrared (FTIR) Spectroscopy is a popular non-destructive method of analysing organic materials, utilised worldwide in many applications, one of which is the identification of different plastics. FTIR Spectroscopy relies on recognizing patterns in wavelength spectra analysed in the spacial frequency domain, the complexity of such analysis, however, makes it highly inaccessible to non-experts. The scarcity of experts equipped with the required skills to understand and analyse FTIR Spectra poses a real challenge, not to mention that the reliance on human expertise could prove to be time consuming and susceptible to analysis errors. Hence, this paper suggests the use of machine learning as a reliable and accessible alternative for the classification of FTIR spectra of plastic bags.

Machine learning is a large field of study that investigates the concept of self-learning algorithms, that is to say, algorithms that make autonomous inferences based on available data, and very specified restrictions. This type of algorithm is particularly useful in solving mathematical problems with high complexity, where human instructed rules could prove to be very difficult. Over the years, machine learning grew to be utilised in a myriad of applications in all types of fields, becoming even essential to the bone structure to most of the recent technological advancements. In the

*CORRESPONDING AUTHOR | Anwar P.P. Abdul Majeed | ✉ amajeed@ump.edu.my

field of spectroscopy, machine learning has seen a surge in usage, lately, with a multitude of recent research work demonstrating a huge deal of potential. In this study, two machine learning techniques were employed to solve the issue of identifying the FTIR spectra of different plastic bag materials, namely a supervised model produced by stacking a K-Nearest Neighbours, Support Vector Machine, and Random Forest models, and a semi-supervised model referred to as Label Spreading. Furthermore, the work explores the use of two different dimensionality reduction algorithms, used for the purpose of feature extraction; those are Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP). The work, as well, investigates the use of FTIR spectra produced by a portable FTIR spectrometer developed by IDIR Solutions based in Malaysia, to make the process of identification even more accessible.

RELATED WORK

Although, there is little to no research specifically on the use of machine learning for the classification of FTIR plastic bag spectra, a decent body of research work could be found on the employment of machine learning classification methods for the identification of FTIR spectra of various materials in many applications, ranging from Pharmacology to Forensic science An instance of such work was authored by Gao et al, for the classification of multicategory edible fungi [5]. The work investigated three types of fungi namely, Lentinus edodes, club fungi, and Bachu mushrooms, via a dataset of 121 fungi samples, obtained using a BRUKER's VERTEX 70 infrared spectrometer. To ease the process of classification and extract identifying features, two-dimensionality reduction algorithms were used; Principle Component Analysis (PCA) and partial least squares (PLS). The number of components produced by the algorithms was set to range from 5 to 50 with an increment of 5, to study the performance of the classification algorithms under the effect of a different number of decomposed components. Three supervised learning algorithms were used in the study, Support Vector Machines (SVM), K-nearest Neighbours, and a Backpropagation Neural Network, with the highest performance achieved by SVM at 100% when 5 & 10 features are produced by PLS. At 15 features the highest accuracy was achieved by PLS-KNN, at 99.06%, and the high-end of the batch sizes, the PLS-BPNN pipeline achieved the best results at 99.07%.

Another research work that involved the use of machine learning for the classification of FTIR spectra was conducted by De Luca et al, utilizing Linear Discriminant Analysis (LDA) to classify Moroccan olive cultivars [6]. A 100 samples dataset of hand-picked – from a farm in Beni Mellal area, in central Morocco - olive endocarps was obtained via a Vector twenty-two Bruker Fourier transforms infrared spectrometer. The dataset was divided into five classes identified by the variant of olive cultivars. Similar to the aforementioned work, Principle Component Analysis was used for dimensionality reduction, before training the model for classification, obtaining a classification accuracy of about 92%.

In 2021, Tan et al authored a research paper on the use of machine learning and FTIR spectroscopy to develop an efficient method of automated differentiation of the Chinese Herb Chuan-Mutong, widely known for its diuretic effects, from its look-alike herb Guan-Mutong [7]. The samples that composed the dataset in this study were acquired from 13 different states in both Malaysia and China, with a total of 164 samples of Mutong. The FTIR spectra were then analysed and produced using a Perkin Elmer spectrometer with a Deuterated Tri-Glycate Sulfate (DTGS) detector. Classification of the Mutong herb was to be performed by four different machine learning algorithms: Partial least squares-discriminant analysis (PLS-DA), J48 (Decision Tree), Random Forest (RF), and radial basis function network (RBF), with the best results recorded for the PLS-DA model producing 100% classification accuracy and Area Under the Curve (AUC), followed by the J48 model at an accuracy of 99.32% and AUC of 99.38%.

MATERIALS AND METHODOLOGY

Data acquisition and Preparation

The investigation carried out in this study was conducted with a sample size of 3,323 FTIR spectra obtained from 1015 plastic bags, where three samples were taken per bag for repeatability. The samples were separated into 4 classes in accordance with their composing material, the classes being; Polylactic Acid (PLA), Polypropylene (PP), Polyethylene (PE), and Starch. The samples were organized and prepared for FTIR scanning, which was conducted on two devices, depicted in figure 1, the first being an Agilent Cary 660 FTIR Spectrometer installed with a PIKE MIRacle Single Reflection Horizontal ATR Accessory. The spectrometer was set to have a resolution of 4cm-1, cover a scan range from 600cm-1 to 4000cm-1 and average 16 number of scans. As for the second spectrometer, it is often referred to as BP10; a Portable FTIR Spectrometer produced by IDIR Solutions, a Malaysian company. Throughout the rest of this study, the former dataset will be referred to as the "FTIR Benchtop dataset", while the latter will be called the "Portable Device dataset" for distinction and convenience. Furthermore, both datasets were split at a ratio of 9:1 for training and testing respectively, this was conducted to avoid any biases that result from model training.

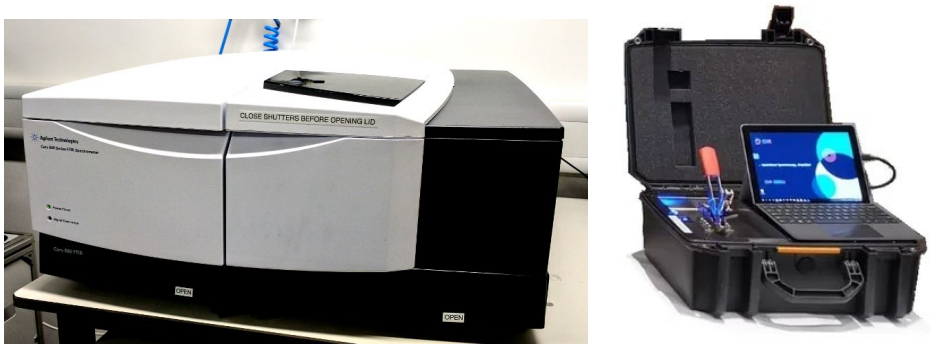

Figure 1. Spectrometers utilized in this study: Agilent Cary 660 (left) and BP-10 (Right) .

Naturally after the data was acquired, it was pre-processed in preparation for the process of machine leanring model training and classification. Primarily the data samples in this study went through two stages of pre-processing, the first being normalisation and the second being smoothing. Normalisation in this context refers to a mathematical operation in which a set of numerical values is remapped from a variant scale to a common scale, normally ranging from 0 to 1, as is the case in the study. It is important to note that normalisation in this study was not performed along the features, but along the data sample itself, where the absorbance of the data sample was adjusted to a scale of 0 to 1, as illustrated in figure 2. As for smoothening, the work made use of the Savitzky-Golay filter, a digital smoothening filter that removes noise from a digital signal by finding the least-square of a polynomial fit with an arbitrary order in a moving window with a specified size. In this study, the polynomial order was chosen to be 8 and the window frame was chosen as 11, as supported by past literature [8], [9].

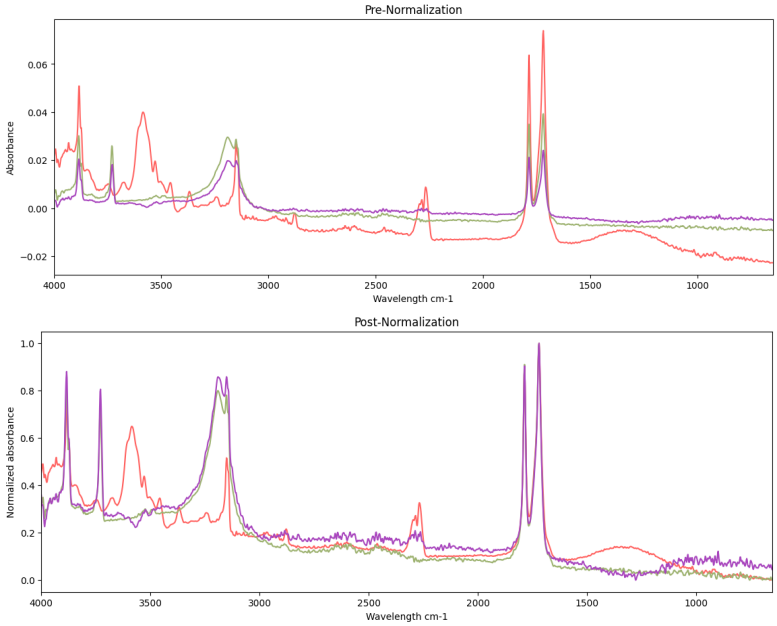

Figure 2. An illustration of the impact of normalisation on the spectra: Pre-normalised spectra (Top) and Post-normalised spectra (Bottom).

Feature extraction through dimensionality reduction

While high dimensionality in data could be advantageous - particularly in algorithms that utilize hyperspace construction, where low dimensionality could result in underfitting - a very large number of dimensions could be computationally expensive and could result in a myriad of issues, a phenomenon referred to as “The curse of dimensionality” [10]. Therefore, dimensionality reduction was carried out in this study, since both datasets enjoy a very high dimensionality rate. Primarily, two algorithms were used, Singular Value Decomposition (SVD) and Uniform Manifold Approximation and Projection (UMAP).

SVD is a decomposition algorithm that reduces dimensions in data, by factorizing it in the form of a matrix, reducing it to three constituent matrices, the use of which could produce decomposed representations of the matrix, normally referred to by Singular Values [11]. Unlike SVD, however, UMAP does not rely on linear decomposition to reduce dimensions, instead, it relies on the concept of manifold learning, that is embedding data in high-dimensional spaces into the manifold with lower dimensionality by means of trained algorithms. In the case of UMAP, this is achieved by generating fuzzy topological representations of the data and then producing close resemblance to it in lower dimensions by optimising the cross-entropy error between the data in the two-dimensional spaces [12]. The algorithms were compared in this research work for the purpose of investigating the impact of different dimensionality reduction techniques on the performance of machine learning algorithms. Note that 15 components were extracted from both algorithms as the main features for classification.

Classification of FTIR spectra using the stacked model

In supervised learning, classification often refers to a class of problems that deal with systematic categorisation of hitherto uncategorised data based on shared patterns produced through statistical learning. There exists a myriad of machine learning algorithms and techniques proposed to solve problems of classification, one of which is the technique of stacking. Stacking is an ensembling technique that simply relies on the stacking of multiple classification models and algorithms into what is called a base layer, before passing the classification output into a final classifier often referred to as a Meta-Classifier[13] . The purpose of using this technique is the combination of taking advantage of multiple machine learning algorithms to produce high performing fits. In this study, a stacked model was produced, stacking a K-Nearest Neighbours, Support Vector Machine, and Random Forest models, with a simple Logistic Regression as the Meta-Classifier.

Label Spreading a semi-supervised approach to classification

Semi-supervised learning is a subclass of machine learning that compensates between two of its other subclasses, namely supervised learning that relies on human labelling to train pattern recognising models, and unsupervised learning which produces models without the reliance on labels. Semi-supervised learning aims at combining the high performance of supervised learning during classification, and the lack of need for labels that unsupervised algorithms normally enjoy, simply by having only a few instances of labelled data observations [14], [15]. The main advantage is reducing the need for human labelling with is time-consuming and prone to errors. This study looked to utilise Label Spreading in comparison with the stacked model, for the classification of FTIR spectra of plastic bags. Label Spreading is a supervised learning algorithm that functions on the principles of graph theory, where data observations are thought of as nodes, with weighted edges. Labels propagate or spread from labelled nodes to unlabelled ones, based on the euclidean-distance-based weights of each edge, and the affinity between nodes according to a previously computed affinity matrix that represents the exchange of useful information between nodes. Figure 3, illustrates roughly the mechanism of label propagation described above.

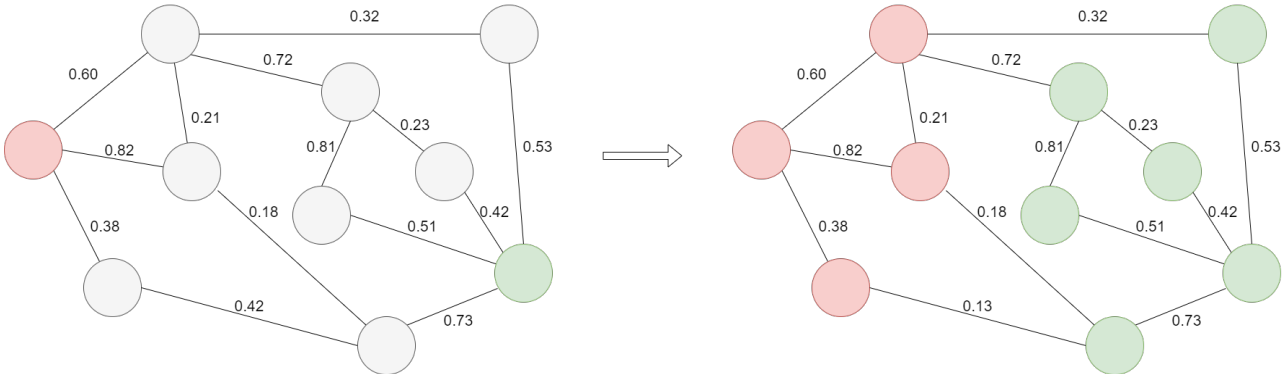

Figure 3. A depiction of the propagation of labelled in Label-Spreading based on calculated weights.

Experimental setup

As noted above, both datasets host 3,323 FTIR spectra, each spectrum covers a special frequency range from 4000 cm^{-1} to 650 cm^{-1}, at 4 cm^{-1} sampling resolution. The data were normalised between 0 and 1, and smoothened using a Savitzky-Golay filter, with a window frame of 11, and 8 orders of polynomial. Both datasets were then separated into two subclasses at 90% for training and 10% for testing, to avoid biases in results. Moreover, Feature extraction was performed using dimensionality reduction algorithms, producing a total number of 15 components, before passing the data for machine learning models training. The parameters of all models and submodels used in this study are displayed in Table 1.

Table 1. Parameters of machine learning algorithms employed in the study

Model	Submodel	Parameter 1	Parameter 2	Parameter 3
Stacked model	KNN	No. of neighbors = **5**	Weights = **equal distribution**	-
	RF	No. of trees = **100**	Min. No. of splits = **2**	-
	SVM	C = **1.0**	Type of Kernel = **Linear**	-
Label Spreading	-	No. of neighbours = **7**	Gamma = **20**	Alpha = **0.2**

Preparation of the experiment discussed in the study was conducted using the High-level, Multi-purpose programming language Python, with the aid of multiple packages to ease the process of coding and standardise it; Numpy a vector analysis package with accelerated performance and multidimensional capabilities [16], Pandas a package for data

manipulation [17], Matplotlib a visualisation tool [18], SciPy a scientific computation package [19], and finally Scikit-learn a machine learning specialised package for rapid development [20].

EXPERIMENTAL RESULTS

For the convenience of the reader, this section was separated into two subsections in which the results for both the supervised and semi-supervised algorithms are discussed, in order. The comparison should be on the basis of the closeness in performance of the semi-supervised learning pipelines to the supervised ones, as it indicates a possibility for replaceability in the event of label scarcity.

Classification performance results acquired from the stacked model

As one might observe in Table 2, the performance of the stacked model seems to be at its best when used on the FTIR Benchtop dataset, providing a classification accuracy and F1 score of 100%, This is due to the clear distinguishability that the dataset possesses in comparison with the Portable Device dataset where it lagged behind at 99.18% Accuracy and F1 score when SVD was used and at even lower scores of 97.54% accuracy and 97.34% F1 score when UMAP was the dimensionality reduction algorithm in use.

Table 2. The stacked model performance on both datasets

Feature extraction method	FTIR Benchtop		Portable Device	
	Accuracy (%)	F1(%)	Accuracy (%)	F1(%)
SVD	100	100	99.18	99.18
UMAP	100	100	97.87	97.94

The same results are reflected on the confusion matrices demonstrated below in figures 4 and 5. In the case of the FTIR Benchtop dataset, the diagonal line is most defined, as all data samples in each class were classified correctly. On the other hand, it is observed that the same diagonal line is less defined when the Portable Device dataset was utilised, especially when it comes to starch materials as some of them were detected to be PLA instead.

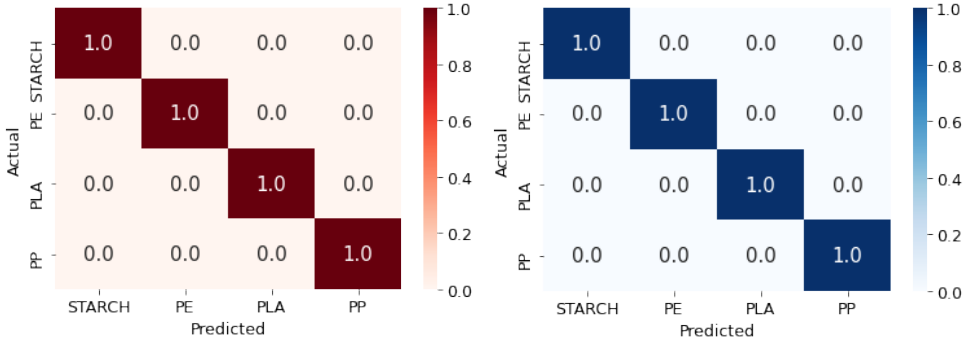

Figure 4. Confusion matrix for the stacked model when SVD (Left) and UMAP (Right) were used, trained on the FTIR Benchtop dataset.

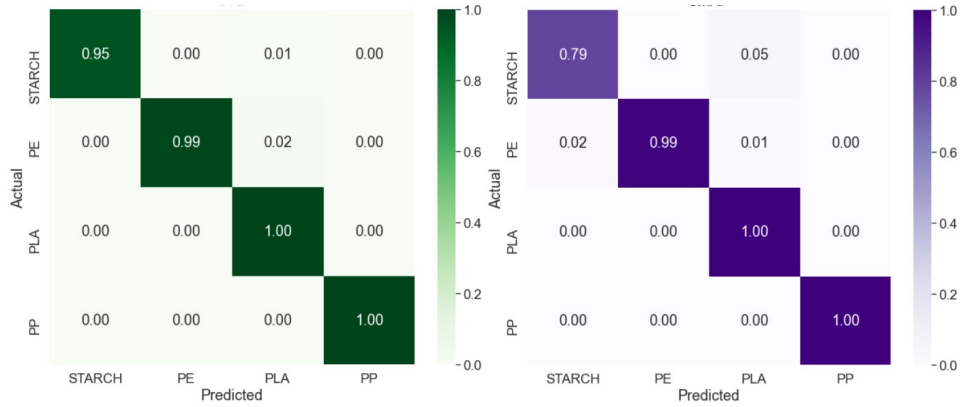

Figure 5. Confusion matrix for the stacked model when SVD (Left) and UMAP (Right) were used, trained on the Portable Device dataset.

Performance results of Label Spreading

In this subsection, the results obtained from testing the aforementioned semi-supervised learning algorithm; Label spreading were acquired and exhibited in Table 3. The accuracy metric produced by the model is displayed when both datasets were used at only a 10% labelling rate. At that rate, the model seemed to perform greatly in the case of the FTIR Benchtop data at 99.82 accuracy and 99.74% F1 score for both feature extraction methods. On the other hand, the model seemed to lag slightly behind when the Portable Device dataset was used, giving the best performance when UMAP was employed at an accuracy of 97.54% and an F1 score of 97.43%.

Table 3. Performance of Label-Spreading on both datasets when 10% of the labels were used only.

Feature extraction method	FTIR Benchtop		Portable Device	
	Accuracy (%)	F1(%)	Accuracy (%)	F1(%)
SVD	99.82	99.74	95.40	95.32
UMAP	99.82	99.74	97.54	97.43

Whereas the table above should the accuracy score of the pipelines at a labelling rate of 10%, figure 6 illustrates the performance of the models in terms of accuracy against the labelling rate from 10% to 100%. It can be observed that the two feature extraction methods in the case of the FTIR Benchtop dataset seem to perform identically, reaching a predictive accuracy of 100% when 60% of the labels were used. On the other side, It could be noted that for the Portable Device dataset the best-performing feature extraction method seems to be UMAP providing an almost consistent accuracy at around 97.5%.

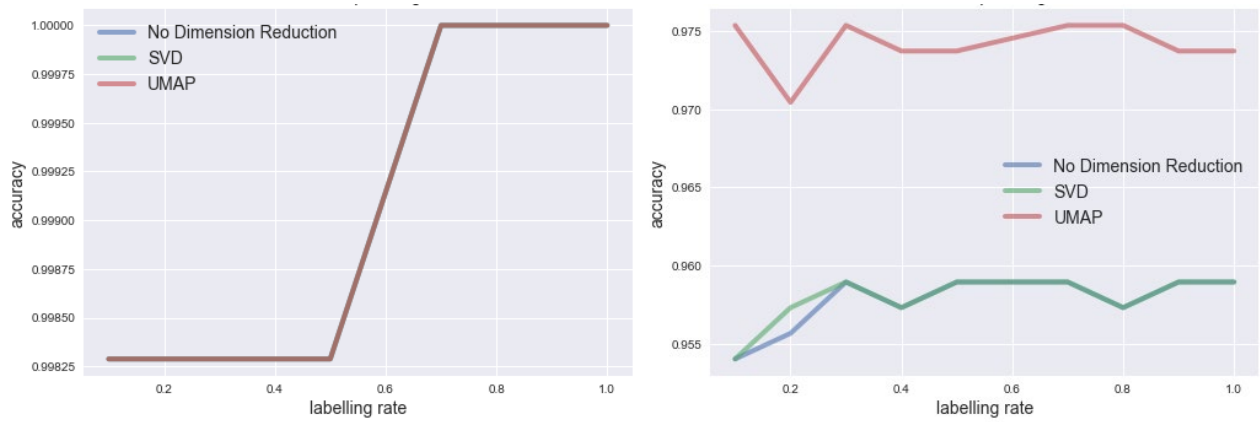

Figure 6. Plots of classification accuracy per labelling rate for the FTIR Benchtop dataset (Left) and the Portable Device dataset (Right).

It could be inferred from the afore-presented recorded results that when data labels exist in abundance, the Stacked model tested in this paper seems to perform the best when SVD is used for feature extraction. Nonetheless, when data labels are scarce Label Spreading could best with UMAP as the main feature extraction method. The high accuracies achieved by models investigated indicate that they could be used as great tools for the classification of plastic bags using various FTIR spectra. The authors of this research, however, suggest the testing of models on different types and brands of FTIR spectrometers, to test the robustness of the methods. Furthermore, different methods of feature extraction and machine learning could prove to be great material for research and investigation.

CONCLUSION

Automating mundane and time-consuming processes has witnessed, for the past years, extensive progress in terms of technological advancements thanks to artificial intelligence and machine learning in specific. This paper looks to utilize these advantages to test the capability of using machine learning algorithms for accurate and robust classification of plastic bags represented in the form of Fourier Transform Infrared spectra. First and Formost, data was acquired from plastic bag samples using two different spectrometers; a top of the line typical FTIR spectrometer, and a portable spectrometer that was developed by IDIR Solutions, based in Malaysia. The data was then normalised and smoothed in preparation for classification, and the high dimensionality of the data was reduced via two algorithms; Singular Value Decomposition and Uniform Manifold Approximation and Projection. Furthermore, Two machine learning models were targeted for study, the first being a stacked model of different supervised learning algorithms namely; K-nearest Neighbors, Support Vector Machine, and Random Forest, and the second being Label Spreading a semi-supervised learning algorithm primarily used to reduce the need for labelling in classification problems. Overall, the performance of the models seemed to be excellent with the stacked model achieving 100% classification accuracy when the benchtop FTIR spectrometer data were in use, and about 99.18% accuracy score trained on the portable device dataset when SVD was used for feature extraction. Label Spreading on the other hand seemed to provide a similar high performance trained

on the former dataset at 99.82% accuracy at 10% labels only and achieved 97.54% predictive accuracy at the same labelling rate when the portable device dataset was employed for training.

ACKNOWLEDGEMENT

The authors of this paper would like to express the utmost fo gratitude to Dr. Tay Feng Huai, Dr. Patrick Steven Wray and all of the IDIR solutions, and Elite Advanced Materials teams for the great amount of assistance they provided as this work was coming to fruition.

REFERENCES

[1] J. E. Alegado *et al.*, *PLASTIC ATLAS - Facts and figures about the world of synthetic polymers (Asia Edition)*. 2021.
[2] UNEP, *Plastics: A Roadmap for Sustainability*. 2018.
[3] "Plastic in the ocean | Magazine Articles | WWF." https://www.worldwildlife.org/magazine/issues/fall-2019/articles/plastic-in-the-ocean (accessed Apr. 14, 2021).
[4] M. Sigler, "The Effects of Plastic Pollution on Aquatic Wildlife: Current Situations and Future Solutions," *Water, Air, Soil Pollut. 2014 22511*, vol. 225, no. 11, pp. 1–9, Oct. 2014, doi: 10.1007/S11270-014-2184-6.
[5] R. Gao *et al.*, "Classification of multicategory edible fungi based on the infrared spectra of caps and stalks," *PLoS One*, vol. 15, no. 8, p. e0238149, Aug. 2020, doi: 10.1371/journal.pone.0238149.
[6] M. De Luca, W. Terouzi, F. Kzaiber, G. Ioele, A. Oussama, and G. Ragno, "Classification of moroccan olive cultivars by linear discriminant analysis applied to ATR-FTIR spectra of endocarps," *Int. J. Food Sci. Technol.*, vol. 47, no. 6, pp. 1286–1292, Jun. 2012, doi: 10.1111/j.1365-2621.2012.02972.x.
[7] C. S. Tan *et al.*, "Comparison of FTIR spectrum with chemometric and machine learning classifying analysis for differentiating guan-mutong a nephrotoxic and carcinogenic traditional chinese medicine with chuan-mutong," *Microchem. J.*, vol. 163, p. 105835, Apr. 2021, doi: 10.1016/j.microc.2020.105835.
[8] B. Zimmermann and A. Kohler, "Optimizing savitzky-golay parameters for improving spectral resolution and quantification in infrared spectroscopy," *Appl. Spectrosc.*, vol. 67, no. 8, pp. 892–902, Aug. 2013, doi: 10.1366/12-06723.
[9] A. X. Zhao, X. J. Tang, Z. H. Zhang, and J. H. Liu, "The parameters optimization selection of Savitzky-Golay filter and its application in smoothing pretreatment for FTIR spectra," *Proc. 2014 9th IEEE Conf. Ind. Electron. Appl. ICIEA 2014*, pp. 516–521, Oct. 2014, doi: 10.1109/ICIEA.2014.6931218.
[10] M. Verleysen and D. François, "The Curse of Dimensionality in Data Mining and Time Series Prediction," *Lect. Notes Comput. Sci.*, vol. 3512, pp. 758–770, 2005, doi: 10.1007/11494669_93.
[11] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular Value Decomposition and Principal Component Analysis," in *A Practical Approach to Microarray Data Analysis*, Kluwer Academic Publishers, 2005, pp. 91–109.
[12] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," *arXiv*, Feb. 2018, Accessed: May 09, 2021. [Online]. Available: http://arxiv.org/abs/1802.03426.
[13] S. Reid, "A Review of Heterogeneous Ensemble Methods," *Science (80-.).*, pp. 1–10, 2007.
[14] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with Local and Global Consistency," *Adv. Neural Inf. Process. Syst.*, vol. 16, 2003.
[15] X. Zhu, X. Zhu, and Z. Ghahramani, "Learning from Labeled and Unlabeled Data with Label Propagation," 2002, Accessed: Jan. 18, 2022. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.8280.
[16] NumPy, "NumPy — NumPy," *NumPy Website*, 2017. .
[17] W. McKinney and P. D. Team, "Pandas - Powerful Python Data Analysis Toolkit," *Pandas - Powerful Python Data Anal. Toolkit*, 2015.
[18] C. Hill and C. Hill, "Matplotlib," in *Learning Scientific Programming with Python*, 2016.
[19] P. Virtanen *et al.*, "SciPy 1.0--Fundamental Algorithms for Scientific Computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 261–272, Jul. 2019, doi: 10.1038/s41592-019-0686-2.
[20] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, 2011.