Identification of Optimal Frequencies to Determine Alpha-Cypermethrin using Machine Learning Feature Selection Techniques

Patryk Kot, Magomed Muradov, Samuel Ryecroft, Montserrat Ortoneda Pedrola, Andy Shaw
BEST Research Institute
Liverpool John Moores University
Liverpool, United Kingdom
p.kot@ljmu.ac.uk

Janet Hemingway, Rinki Deb, Micheal Coleman
Liverpool School of Tropical Medicine (LSTM)
Liverpool, United Kingdom

Abstract— Machine learning feature space reduction techniques allow for vast feature spaces to be reduced with little loss or even significant improvements in the reliability of predictions of models. Microwave spectroscopy with feature spaces of over 8000 are not uncommon when considering magnitude and phase. Applying Machine learning techniques to this type of feature space allows for a quicker reduction and helps to identify the most suitable predictive features. The control of insect vectors that transmit diseases including malaria, visceral leishmaniasis and zika rely on the use of insecticide. These diseases affect millions, malaria alone accounted for 214 million new cases resulting in 438,000 deaths in 2015. One method used in controlling the vectors is through indoor residual spraying, applying insecticide to the wall surface inside houses. Alpha-cypermethrin is one of the insecticides that is currently sprayed in several countries for vector control. Quality assurance and monitoring of the control activities is challenging relying on the use of laboratory-reared insects. This was improved with a chemical based Insecticide Quantification Kit, but these assays have been challenging to operationalise. An electromagnetic sensor is being developed to investigate the potential to detect alpha-cypermethrin. Preliminary experiments were carried out to differentiate tiles sprayed with Technical Grade alpha-cypermethrin, wettable powder containing 5% alpha-cypermethrin and wettable powder with over dose of alpha-cypermethrin using a horn antenna at a frequency range between 1 GHz to 6 GHz. The experimental results indicated the potential use of electromagnetic waves to determine alpha-cypermethrin in a non-destructive manner.

Keywords— Alpha-cypermethrin; Electromagnetic waves; Malaria; Non-destructive; Sensor; Visceral Lieshmaniasis

I. INTRODUCTION

Tropical infectious diseases, such as malaria, dengue, Zika and leishmaniasis are of global public health importance. Spread by the infectious bite of arthropod insect vectors, namely mosquitoes and sand flies, the main strategies to control and eliminate such diseases are insecticide based strategies: indoor residual spraying (IRS) and long-lasting insecticide nets (LLINs)[1]. Such interventions aim to kill the vector and therefore interrupt the disease transmission cycle; however, resistance to the insecticides available for public health use is of growing concern[2].

The World Health Organisation (WHO) have set process, performance and impact indicators, which should be monitored during each round of a vector control intervention[3]. Quality assurance (QA) forms an essential component of performance monitoring, including validation that the correct dose of insecticide is delivered in any vector control intervention. However, for many programmes performing QA is costly and logistically challenging and the approach is not used.

For IRS, the current gold standard method to quantify insecticide deposited is carrying out high performance liquid chromatography (HPLC) on filter papers that were affixed to walls prior to IRS. Together with surveys monitoring insecticide efficacy and the residual decay rate, this provides a comprehensive understanding of the operational impact of IRS[3]. However, these methods can be time consuming and require insectaries with susceptible colonies, laboratories, expensive equipment and skilled technicians. Furthermore, IRS operators can become sensitised to QA methods used and focus spray efforts on filter papers, providing artificial results. Alternative methods such as the Insecticide Quantification Kit (IQK) have been developed and used in a few field surveys[4].

Electromagnetic wave technologies have been successfully applied in several industries, such as monitoring environmental pollution[5], wastewater quality[6], [7], moisture content of building materials[8], carbon emissions levels as well as for monitoring human health[9], [10]. The advantage of this method is real-time monitoring based on the interaction of electromagnetic waves with the material under test. Signal velocity changes indicated by the application of the microwave wave illustrate the composition and texture of the object[11].

There are different types of sensor structures that can be used to transmit and receive microwave signals namely Wire Antennas, Aperture Antennas, Cavity and Microstrip Antennas. In this paper aperture antenna will be used. Aperture antennas are most common at microwave frequencies. There are different types of aperture antennas such as pyramidal horn, conical horn or rectangular waveguide[12]. Horn antennas provide high gain, wide bandwidth and low Voltage Standing Wave Ratio (VSWR). There are three types of rectangular horn antennas namely H-plane sectoral horn, E-plane sectoral horn and Pyramidal horn. A microwave horn antenna produces a uniform phase front with larger aperture than the waveguide, hence greater directivity. The antenna consists of a rectangular metal tube closed at one end, flaring into an open-ended pyramidal shaped horn on the other side[13]. The microwaves are introduced into the waveguide by a coaxial cable attached to the horn antenna[14]. Aperture Antennas are often used for higher frequency applications than wire-type antennas[15]. Pyramidal horns are ideally suited for rectangular waveguide feeders. The head of a horn antenna, acts as a gradual transition from a waveguide mode to a free-space mode of the EM wave[16].

Microwave sensing systems provide the opportunity for real-time and non-destructive methods to monitor and analyse varied materials and their properties. The technique can also take on many physical forms, namely planar structures [17] and cavities[18], which have been demonstrated by the authors. In addition, the technology to generate and detect microwave signals is inexpensive; it is featured in the many millions of smartphone devices, tablets, and portable computers. Furthermore, the technique is non-ionizing, utilizing less than 10 mW of power, significantly less than modern wireless communication devices. Therefore, it is safe to use without fear of harming the residents.

The principle of microwave sensing in this study is based on the interaction of electromagnetic waves with the material under test (MUT). When a material is placed under the sensing system, i.e. EM horn antenna and exposed to the microwave radiation, it alters the velocity of the signal, attenuates, or reflects it. These changes can be captured and analysed to determine various chemical and physical properties/features of materials.

Machine Learning (ML) has been applied with exciting results to variety of data sets, providing new insights into large data sets and proving and effective technique for data processing across a range a range of data sets across a variety of disciplines including within the area of microwave spectroscopy[19]. However, other methods such as Principal Component Analysis (PCA), Partial Least Squares regression (PLS) as well as linear regression models are more common place.

In order to fully develop a portable low cost sensor it is important to identify frequencies that are sensitive to alpha- cypermethrin to allow a targeted cost effective device to be developed

This paper will investigate potential use of electromagnetic waves as a novel method to determine alpha-cypermethrin in a non-destructive manner and apply Machine Learning (ML) feature selection techniques to identify frequencies most sensitive to alpha-cypermethrin.

II. METHODOLOGY

The purpose of the experimental work outlined in this paper was to determine alpha-cypermethrin on the back of ceramic tiles. The porous surface of the back of the tiles was used as a standardised wall surface.

A commercial formulation of 5% alpha-cypermethrin in wettable powder (Tagros Chemicals India Ltd.) and technical grade alpha-cypermethrin (Sigma) were used.

A total of 15 tiles were divided into 3 groups and treated with:

1) alpha-cypermethrin wettable powder (standard mixture according to the requirements of WHO specification 454/TC (April 2006)) to a final concentration of alpha-cypermethrin of 25 mg/m^2,

2) standard mixture with added extra technical grade alpha-cypermethrin to a final total concentration of 425 mg/m^2, and

3) technical grade alpha-cypermethrin to a final total concentration of 400 mg/m^2.

The preparation of the samples were carried out in Liverpool John Moores University Industrial Chemistry laboratory following the guidance from Liverpool School of Tropical Medicine. The tiles were prepared by pouring 22.5 ml of the working suspension of alpha-cypermethrin over the surface-using a pipette; making sure the pipetting was uniform over the tile (Figure 1).

The tiles were then allowed to dry completely and stored in a cool dark place until analysis.

The surface texture of the composites/material is one of the factors that can influence the measurements. Thus, the measurements will be provided on four different positions of the tiles, i.e. I-0º, II-90º, III- 180º and IV-270º. Experimental results are presented in Figure 3.
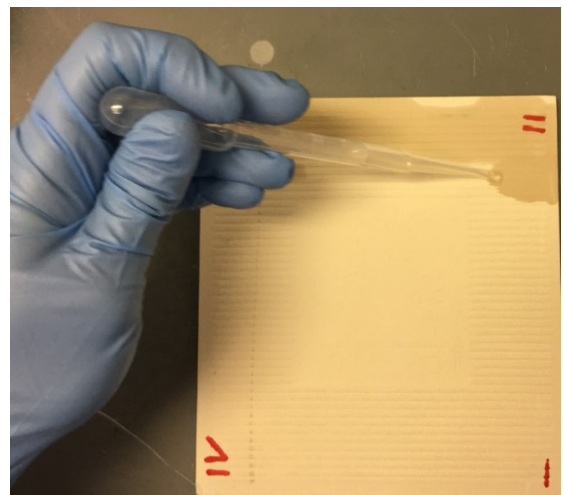

Fig. 1. Sample preparation, i.e. applying solution of alpha cypermethrin on tile samples.

The electromagnetic horn antenna was used immediately after sample preparation with the sample being placed under the horn with a 2 mm air gap as shown in Figure 2. The measurement equipment was configured on the laboratory bench with a metal plate placed on the bottom to reflect microwave signal penetrated through measured ceramic tiles. The reflected (S_{11}) power measurements were recorded for 12 hours per sample; the microwave spectrum was captured between 1 and 6 GHz every 5 minutes. A Rohde and Schwarz ZVL13 VNA was used for capturing spectral data, which was automated via bespoke National Instruments LabVIEW interface as shown in Figure 2.
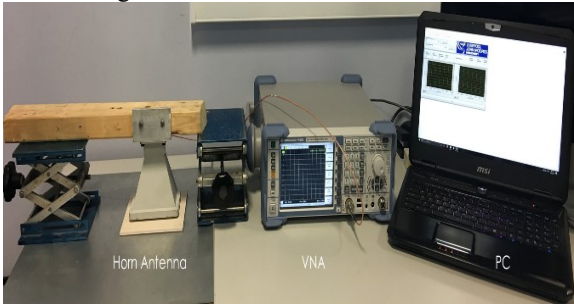
Fig. 2. Experimental Setup showing horn antenna connected to a Rohde and Schwarz Vector Network Analyzer and a bespoke LabVIEW interface capturing spectral data.

III. RESULTS AND DISCUSSION

Initial data from a fourth untreated tile placed in three separate positions was averaged and correlated against each other in order to determine a linear relationship between the readings. The results presented in Figure 3 show a strong linear agreement ($R^2 = 0.99$) between the measurements. Providing evidence that a surface texture of these particular tiles does not affect the sensor readings and will not be taken into consideration.
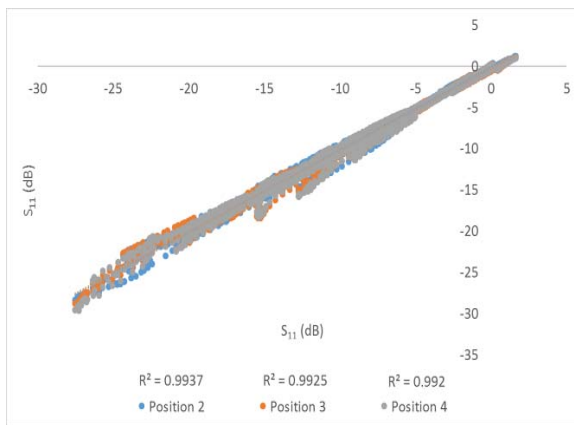

Fig. 3. R-Squared between Averaged Data for ceramic tiles positions for investigation of the effect of the texture on EM spectroscopy.

Figure 4 shows the S_{11} measurements, reflected signal of three sample groups. The measurements shown in this figure were taken every 5 minutes over a period of 12 hours. The change of the antenna response is thought to be caused by the different concentration of alpha-cypermethrin insecticide owing to the chemical properties of each sample group.
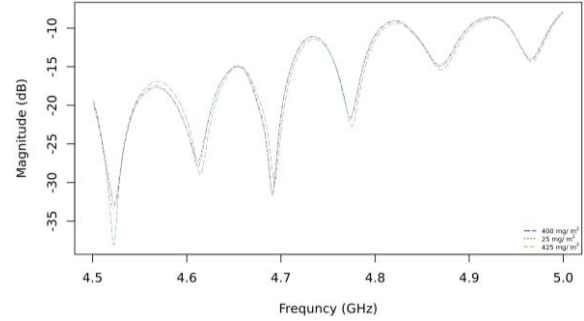

Fig. 4. Readings from the microwave horn antenna; measurements were taken for the frequency range 1-6 GHz, but for the clarity, data from 4.5 to 5 GHz is presented.

To enable the development of a low-cost miniature sensor device it is important to identify the key indicating frequencies that can be using to identify the presence of alpha cypermethrin using wrapper feature selection. It is possible to identify frequencies that show promise in the detection of alpha cypermethrin allowing more detailed data sets to be collected, focusing on the key identified frequencies.

The data set consists of three distinct classes: 25 mg/m^2, 400 mg/m^2 and 425 mg/m^2, with a feature space of 4000 magnitude readings from 1 GHz through to 6 GHz with 50 observations per class.

The data firstly underwent a normalisation process giving the data a mean of zero and a variance of one in order to reduce the sensitivity of the subsequent processes to the magnitude of the features.

To identify those features considered most important to classification process four Machine Learning (ML) classification algorithms were applied to the reduced feature space in wrapper feature selection approach. The algorithms used were Random Forest [20] (RF), Support Vector Machines [21] (SVM), K Nearest Neighbour (KNN) and Gradient Boosted Models (GBM) [22]–[24].

To initially reduce the dimensionality of the feature space an information gain filter was applied and the top 500 features carried forward to the wrapper selection process reducing the feature space to one eight of the initial size of the space. This significant reduction in the feature space was selected based through application of domain specific knowledge that a large amount of the spectrum would not be sensitive to the substances involved, allowing for a large feature space reduction without concern of influencing the overall results. The threshold selected was also based on investigations carried out at thresholds of 1000 and 250 features in order to identify the suitability of the threshold.

The application of a higher threshold of 1000 lead to the same features being identified through the GBM approach while the use clear, but more dispersed clusters being formed with less space between the 400 mg/m^2 class and the 25 mg/m^2 class. The KNN wrapper approach was significantly impacted, with merging of the 400

mg/m² and 25 mg/m² classes with a reduction of separation between the three classes. The SVM approach performed slightly worse with the application of a higher threshold, with the 400 mg/m² and 25 mg/m² classes merging together, the cluster for the 400 mg/m² was also dispersed.

Reduction of the feature space to 250 features significantly impacted on the clustered formed after application of wrapper machine learning methods. Clusters formed through the KNN wrapper methods shows considerable merging of the 400 mg/m² and 25 mg/m² classes, with little class separation between these two classes and the 425 mg/m² class. Features selected through the SVM approach showed stronger clusters formed for the 25 mg/m² and 425 mg/m² with the 400 mg/m² class showing significant dispersion. Features identified through the application of a GBM wrapper feature selection identified the same features as identified when 1000 and 500 features identified through the information gain filter were used. RF wrapper feature selection formed a strong cluster for the 425 mg/m² class that was well separated from the two other classes, the 400 mg/m² and 25 mg/m² classes, had some interference with each other, with the 400 mg/m² forming a weaker cluster than the other classes.
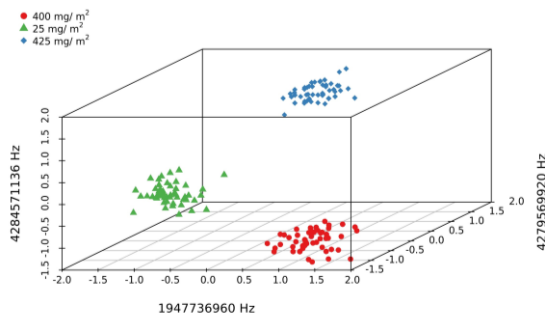

Fig. 5. Data analysis of sample measurements taken plotting the normalized magnitude of the 1947736960 Hz, 4279569920 Hz and 4284571136 Hz frequencies as identified through the information gain filter.

Table 1 shows the 10 highest-ranking frequencies in Hz identified through wrapper feature selection using the specified machine learning algorithms on the reduced feature space. The reduced feature space contains a number of recurring frequency bands, including the 1.60 GHz – 2.00 GHz, 3.60 GHz – 3.80 GHz, 4.20 GHz – 4.40 GHz and 4.88 GHz – 5.00 GHz all being highlighted four times or more by two or more algorithms. The identification of these bands across multiple algorithms gives support to the use of these frequencies as a method of detecting alpha- cypermethrin.

	TABLE I		
RF	**SVM**	**KNN**	**GBM**
1900225024	5969992704	5608652288	1947736960
1960240000	4477119488	4687171584	4279569920
4895973888	1707676928	3438109440	5353588224
1947736960	1896474112	4982245376	5997499392
4284571136	4917229056	4619655168	1701425408
1952738176	4883470848	3734433536	1960240000
1928982272	5838709760	4969742336	4289572352
1920230016	4412103168	4568392192	1692673152
1703926016	3646911744	3641910528	1950237568
4304576000	3668167168	5383596032	4358339584

Top 10 ranked frequencies and associated algorithms identified through wrapper feature selection methodology and information gain filter ordered by most important to least important
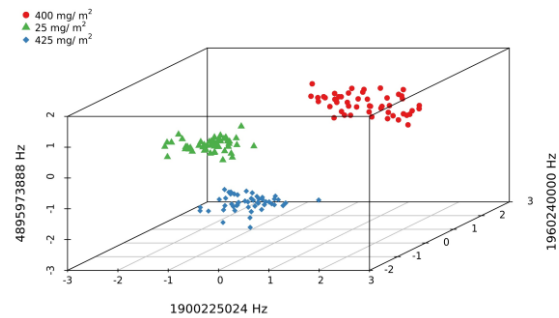

Fig. 6. Data analysis of sample measurements taken plotting the normalized magnitude of the 1900225024 Hz, 1960240000 Hz and 4895973888 Hz frequencies as identified through the RF wrapper feature selection.
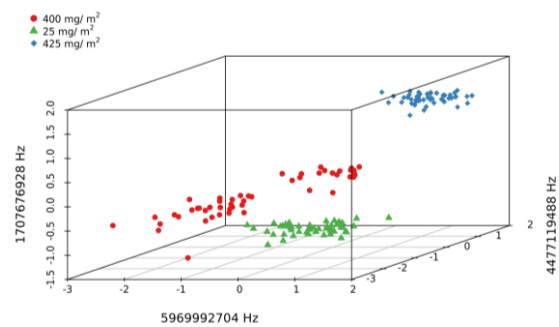

Fig. 7. Data analysis of sample measurements taken plotting the normalized magnitude of the 5969992704 Hz, 4477119488 Hz and 1707676928 Hz frequencies as identified through the SVM wrapper feature selection.
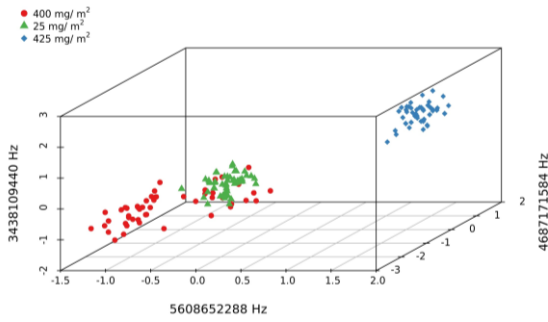
Fig. 8. Data analysis of sample measurements taken plotting the normalized magnitude of the 5608652288 Hz, 4687171584 Hz and 3438109440 Hz frequencies as identified through the KNN wrapper feature selection.
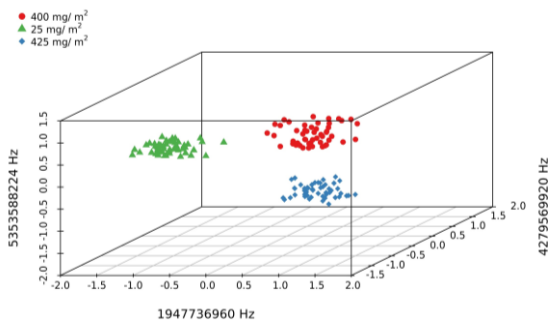

Fig. 9. Data analysis of sample measurements taken plotting the normalized magnitude of the 1947736960 Hz, 4279569920 Hz and 5353588224 Hz frequencies as identified through the GBM wrapper feature selection.

Figure 5 through to Figure 9 plot the three highest-ranking frequencies magnitude as identified through the defined feature ranking approach against each other in a three-dimensional space. The clusters formed from the RF and GBM wrapper feature selection as well as the clusters formed through the information gain filter all show strong and significant clustering using three features. The clusters formed from the three top ranking SVM features in figure 7, show significant clustering. However, the clustering of technical grade class is significantly more dispersed than other cluster formations. Clustering from the KNN wrapper feature selection shows a significant and isolated cluster formed for the 425 mg/m^2 class the standard solution forms a strong cluster. However, the 400 mg/m^2 is more dispersed and overlaps with the cluster formed for the 25 mg/m^2.

The two ensemble approaches performed well, with similar levels of class separations although, slightly improved clustering was achieved through GBM compared to RF. The KNN approach gave strong clusters for two of the classes. However, the third class was less well defined, all three features identified through the KNN approach significantly separate 425 mg/m^2 from the other classes. The SVM plot shows a large variance in the frequencies identified through the approach, though only

one instance of misclassification occurs as shown within Figure 10 where the decision boundaries are shown for the two highest ranking features. The decision boundaries show a good separation between the 425 mg/m^2 and the other two classes, the separation between the 400 mg/m^2 and 25 mg/m^2 is also well formed, through the 400 mg/m^2 class is considerably more distributed, especially in the 4477119488 Hz frequency feature.
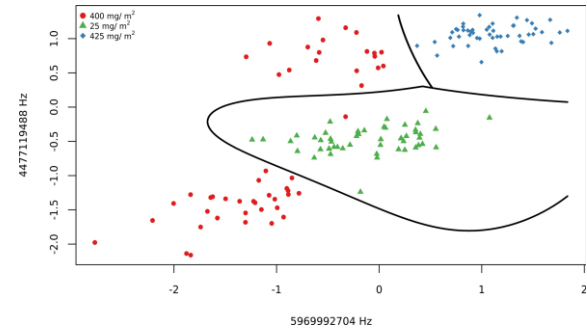

Fig. 10. Plot of the decision boundaries for the SVM for the 5969992704 Hz and 4477119488 Hz frequencies.
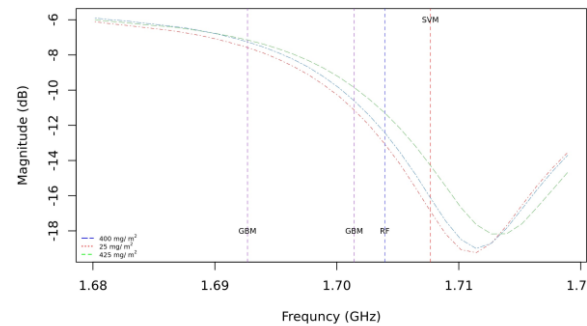

Fig. 11. Plot of the average magnitude of each class with features identified by the four applied wrapper feature selection methods highlighted for the frequency spectrum 1.68 GHz to 1.7 GHz
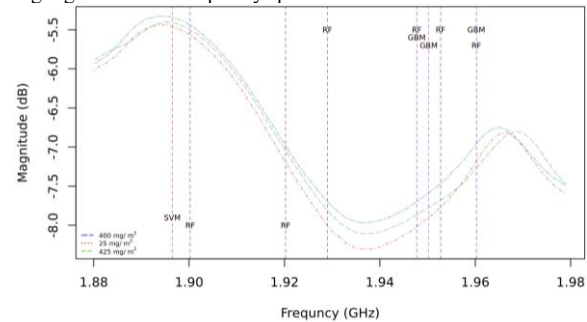

Fig. 12. Plot of the average magnitude of each class with features identified by the four applied wrapper feature selection methods highlighted for the frequency spectrum 1.88 GHz to 1.98 GHz

Fig. 13. Plot of the average magnitude of each class with features identified by the four applied wrapper feature selection methods highlighted for the frequency spectrum 3.60 GHz to 3.75 GHz
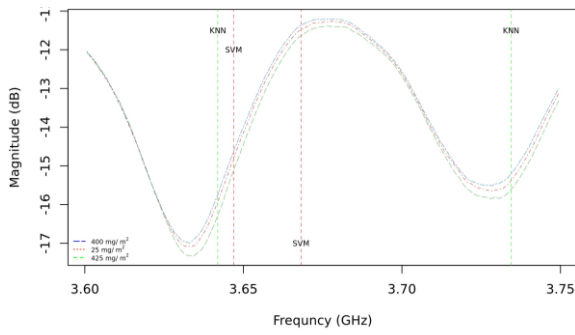

Fig. 14. Plot of the average magnitude of each class with features identified by the four applied wrapper feature selection methods highlighted for the frequency spectrum 4.26 GHz to 4.38 GHz
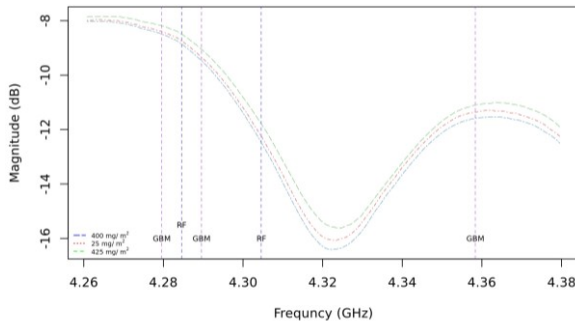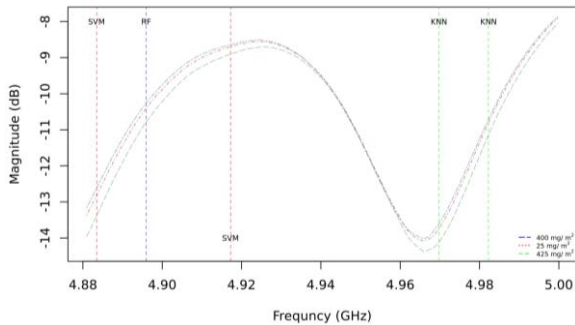

Fig. 15. Plot of the average magnitude of each class with features identified by the four applied wrapper feature selection methods highlighted for the frequency spectrum 4.88 GHz to 5.00 GHz

Figure 11 through to Figure 15 highlight areas that multiple machine learning algorithms identified important frequencies with frequencies plotted against average magnitude of each of the three classes. There is clear separation between at least one class and in many cases a clear separation between all three classes.

All of the algorithms identified changes near peaks or troughs in the spectrum though none of the methods identified frequencies at the peak or trough. None of the wrapper methods identified frequencies within the area of the peak shift such as the one that can be observed in figure 12 between the frequencies of 1.96 GHz and 1.98 GHz. Features such as peaks and troughs as well as the shift in these peaks and troughs caused by the change of dielectric constant between substance provides a reliable method of distinguishing substances in the field of microwave spectroscopy. Features selected through the

wrapper feature selection approaches highlight clear separations between the classes within the spectrum plots.

The initial data analytics carried out shows the possibility of applying wrapper machine learning feature selection methodologies to microwave spectroscopy data to refine frequency ranges to investigate, allowing further, fine-grained data to be collected in area surrounding the frequencies identified through the wrapper methodology. The clustered groups identified using the three sub 6 GHz frequencies identified through the wrapper selection methodology lends support to the feasibility of the classification of concentration though sub 6 GHz frequencies across a wider range of concentrations.

IV. CONCLUSION

Initial experiments using the horn antenna were conducted under controlled settings to investigate the potential of EM waves to detect technical grade alpha-cypermethrin, wettable powder containing 5% alpha-cypermethrin and wettable powder with over dose of alpha-cypermethrin, from ceramic tiles treated with the solution. The resultant measurements were provided by the S-parameters i.e. reflected power from the horn antenna, at a frequency range between 2 GHz to 6 GHz: suggesting that EM waves could be used to detect alpha-cypermethrin in a non-destructive manner. A unique signature was detected for each of the three solutions, which could be distinguished by analysing frequency shift and amplitude change. The frequency range where the most significant changes were observed will be selected to develop a portable sensor to determine alpha-cypermethrin insecticide concentrations deposited onto the internal wall of houses sprayed as part of an IRS programme.

The data analysis shows that ML feature space reduction techniques such as filter methods and wrapper methods can be used to reduce feature spaces within EM data sets, with the highest ranking features forming clear clusters when plotted against each other indicating the possibility of classification of the three distinct classes investigated. The process has identified several frequency ranges that could be applied to identify and detect three concentrations of alpha-cypermethrin through simple clustering approaches such as k-means or KNN. The addition of more classes as well as other variable factors such as humidity, temperature and building material may require a more complex approach in order to allow for all of these factors.

V. FURTHER WORK

Field based testing is required to validate that the unique signature detected within laboratory testing can also be seen on surfaces sprayed with alpha-cypermethrin within routine IRS activities. Baseline pre-IRS and post-IRS readings will be taken from the most common surfaces found in areas targeted by the IRS programme. Unique signatures for unsprayed surfaces and water used by IRS operators will be determined to further refine the

EM characterisation of alpha-cypermethrin. Human involvement is expected to be an important causative factor for spray variation, therefore a range of alpha-cypermethrin concentrations are expected and the sensor sensitivity will be measured against this. All activities will be done in conjunction with the IRS programme.

Further work is also required across a wider range of concentrations to investigate the possibility of applying a linear modelling approach in the detection of concentrations, alternatively additional data in a wider variety of concentrations could be used to increase the number of classes within a non-linear model allowing for the detection of concentration bands.

Further investigation should be undertaken into other possible features within this type of data set including looking at phase data as well as peak shift, which could both aid in a ML model with more complex and closer together classes. It is also important to consider environmental factors such as humidity and temperature variables, which can have a significant impact on readings from this type of data set. Finally, it is important to investigate the impact of different building surface materials which may have an impact on readings either due to the EM properties of the materials or the texture of the reading surface.

REFERENCES

[1] G. F. Killeen, A. Tatarsky, A. Diabate, C. J. Chaccour, J. M. Marshall, F. O. Okumu, G. Brunner, G. Newby, Y. A. Williams, D. Malone, L. S. Tusting, and R. D. Gosling, "Developing an expanded vector control toolbox for malaria elimination," *BMJ Glob. Heal.*, vol. 2, no. 2, pp. 1–8, 2017.

[2] W. H. Organization and Others, "Global plan for insecticide resistance management in malaria vectors," 2012.

[3] W. H. Organization and Others, "Test procedures for insecticide resistance monitoring in malaria vector mosquitoes," 2016.

[4] T. L. Russell, J. C. Morgan, H. Ismail, H. Kaur, T. Eggelte, F. Oladepo, J. Amon, J. Hemingway, H. Iata, and M. J. I. Paine, "Evaluating the feasibility of using insecticide quantification kits (IQK) for estimating cyanopyrethroid levels for indoor residual spraying in Vanuatu," *Malar. J.*, vol. 13, p. 178, 2014.

[5] O. Korostynska, A. Mason, M. Ortoneda-Pedrola, and A. Al-Shamma'a, "Electromagnetic wave sensing of NO3 and COD concentrations for real-time environmental and industrial monitoring," *Sens. Actuators B Chem.*, vol. 198, pp. 49–54, 2014.

[6] J. H. Goh, A. Shaw, J. D. Cullen, A. I. Al-Shamma'a, M. Oliver, M. Vines, and M. Brockhurst, "Water pipe leak detection using electromagnetic wave sensor for the water industry," *2011 IEEE Symposium on Computers Informatics*. pp. 290–295, 2011.

[7] M. A. Diallo, A. Shaw, J. H. Goh, S. R. Wylie, A. Mason, and A. Al-Shamma'a, "Electromagnetic sensor for leak detection and asset management for the water industry," *2012 Sixth International Conference on Sensing Technology (ICST)*. pp. 618–622, 2012.

[8] M. Adous, P. Quéffélec, and L. Laguerre, "Coaxial/cylindrical transition line for broadband permittivity measurement of civil engineering materials," *Meas. Sci. Technol.*, vol. 17, no. 8, p. 2241, Jan. 2006.

[9] O. Korostynska, A. Mason, and A. Al-Shamma'a, "Microwave sensors for the non-invasive monitoring of industrial and medical applications," *Sens. Rev.*, vol. 34, no. 2, pp. 182–191, 2014.

[10] A. Mason, O. Korostynska, M. Ortoneda-Pedrola, A. Shaw, and A. Al-Shamma'A, "A resonant co-planar sensor at microwave frequencies for biomedical applications," *Sensors Actuators, A Phys.*, vol. 202, pp. 170–175, 2013.

[11] P. Kot, A. Shaw, M. Riley, A. S. Ali, and A. Cotgrave, "The Feasibility of Using Electromagnetic Waves in Determining Membrane Failure Through Concrete," *Int. J. Civ. Eng.*, vol. 15, no. 2, 2017.

[12] C. Balanis, *Antenna Theory. Analysis and design*, Third. New Jersey: John Wiley & Sons, 2005.

[13] K. Vijayakumar, S. R. Wylie, J. D. Cullen, C. C. Wright, and A. I. Ai-Shamma'a, "Non invasive rail track detection system using microwave sensor," *J. Phys. Conf. Ser.*, vol. 178, no. 1, p. 12033, Jan. 2009.

[14] P. Kot, A. S. Ali, A. Shaw, M. Riley, and A. Alias, "The application of electromagnetic waves in monitoring water infiltration on concrete flat roof: The case of Malaysia," *Constr. Build. Mater.*, vol. 122, 2016.

[15] Y. Huang and K. Boyle, *Antennas: From Theory to Practice*. Chichester, UK: John Wiley & Sons, Ltd, 2008.

[16] A. Roy and I. Puri, "Design and Analysis of X band Pyramidal Horn Antenna Using HFSS .," *Int. J. Adv. Res. Eletronics Commun. Eng.*, vol. 4, no. 3, pp. 488–493, 2015.

[17] M. Muradov, J. Cullen, and A. Mason, "Real-Time Monitoring of Meat Drying Process Using Electromagnetic Wave Sensors," in *Next Generation Sensors and Systems*, C. S. Mukhopadhyay, Ed. Cham: Springer International Publishing, 2016, pp. 221–233.

[18] K. H. Teng, M. Ateeq, A. Shaw, A. Al-Shamma'a, S. N. Kazi, B. T. Chew, and P. Kot, "Numerical Investigation on using an Electromagnetic Wave Sensor to Detect Water Hardness in Water Cooling System Industry," *Int. J. Electromagn. Appl.*, vol. 7, no. 2, pp. 38–47, Jan. 2017.

[19] S. Cashman, O. Korostynska, A. Shaw, P. Lisboa, and L. Conroy, "Detecting the Presence and Concentration of Nitrate in Water Using Microwave Spectroscopy," *IEEE Sens. J.*, vol. 17, no. 13, pp. 4092–4099, 2017.

[20] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Jan. 2001.

[21] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Jan. 1995.

[22] L. Breiman, "Arcing the edge," Technical Report 486, Statistics Department, University of California at Berkeley, 1997.

[23] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[24] L. Mason, J. Baxter, P. L. Bartlett, and M. R. Frean, "Boosting Algorithms as Gradient Descent," in *Advances in Neural Information Processing Systems 12*, S. A. Solla, T. K. Leen, and K. Müller, Eds. MIT Press, 2000, pp. 512–518.