

## Research Article

# Ultraviolet-Visible Spectroscopy and Chemometric Strategy Enable the Classification and Detection of Expired Antimalarial Herbal Medicinal Product in Ghana

Jacob N. Mensah,<sup>1</sup> Abena A. Brobbey,<sup>1</sup> John N. Addotey <sup>1</sup>, Isaac Ayensu,<sup>1</sup>  
Samuel Asare-Nkansah,<sup>1</sup> Kwabena F. M. Opuni <sup>2</sup>, and Lawrence A. Adutwum <sup>2</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry, Faculty of Pharmacy and Pharmaceutical Sciences, College of Health Sciences, KNUST, Kumasi, Ghana

<sup>2</sup>Department of Pharmaceutical Chemistry, School of Pharmacy, College of Health Sciences, University of Ghana, Accra, Ghana

Correspondence should be addressed to Lawrence A. Adutwum; [ladutwum@ug.edu.gh](mailto:ladutwum@ug.edu.gh)

Received 2 February 2021; Accepted 18 June 2021; Published 25 June 2021

Academic Editor: Spas D. Kolev

Copyright © 2021 Jacob N. Mensah et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To meet the growing demand for complementary and alternative treatment for malaria, manufacturers produce several antimalarial herbal medicinal products. Herbal medicinal products regulation is difficult due to their complex chemical nature, requiring cumbersome, expensive, and time-consuming methods of analysis. The aim of this study was to develop a simple spectroscopic method together with a chemometric model for the classification and the identification of expired liquid antimalarial herbal medicinal products. Principal component analysis model was successfully used to distinguish between different herbal medicinal products and identify expired products. Principal component analysis showed a clear class separation between all five herbal medicinal products (HMP) studied, with explained variance for first and second principal components as 37.51% and 26.38%, respectively, while the third principal component had 18.74%. Support vector machine classification gave specificity and accuracy of 1.00 (100%) for training set data for all the products. The validation set HMP1, HMP2, and HMP3 had sensitivity, specificity, and accuracy of 1.00. HMP4 and HMP5 had sensitivity and specificity of 0.90 and 1.00, respectively, and an accuracy of 0.98. The support vector machine classification and principal component analysis models were successfully used to identify expired herbal medicinal products. This strategy can be used for rapid field detection of expired liquid antimalarial herbal medicinal products.

## 1. Introduction

The worldwide mortality attributed to malaria in 2019 was 409,000 out of which 384,000 occurred in Africa, with most of them being children and pregnant women [1]. These mortality cases in Africa represent over 93% of all malaria-related deaths worldwide. Of the 87 malaria endemic regions in the world, 28 African countries and India accounted for 95% of malaria cases reported globally [1]. Over the years, concerns have risen about the effectiveness and safety of orthodox antimalarial drugs, coupled with the development of resistance to drugs used in the treatment of malaria [2]. Artemisinin combination therapy was introduced to reduce

the rate of development of resistance. However, the associated side effects and adverse drug reactions have made it unattractive for some patients [3], as well as cost of treatment, treatment failures, and accessibility.

In such situations, plant medicines have provided a viable alternative to orthodox medicines for the treatment of malaria. In Ghana, plant medicines remain a major source of antimalaria therapy, as a large number of Ghanaians are observed to patronize herbal antimalaria remedies. Such preference stems from the lower costs of products, perceived better efficacy and reduced side effects, and acceptability based on peer recommendation. Treatment outcomes from the use of these herbal remedies have been positive in a

number of cases because medicinal plants have been sources of many bioactive compounds including natural scaffolds for antimalarial drugs such as quinine and artemisinin [4, 5]. These plant medicines may circumvent the challenges of parasite resistance and toxicity by the synergistic activity of the several constituent secondary metabolites [6].

Plant medicines are normally formulated as herbal medicinal products (HMPs) with optimum pH and minimal toxicity due to internal buffering effect [5] and relatively low concentration of constituent phytochemicals, respectively. Currently, the prevalence of use of HMPs in Ghana is about 76% [7]. Considering the debilitating nature of malaria and the potential for development of organ and neurological complications in severe cases, there is the need for surveillance and continuous quality monitoring of HMPs to safeguard the integrity of the products in order to safeguard the general public. However, the efficient quality monitoring of HMPs still remains a major challenge. In the case of orthodox medicines, an assay can be performed to determine the levels of active ingredients and other impurities that may be present. On the other hand, HMPs usually have several phytochemical constituents that nearly make it impossible to identify all the bioactive compounds and accordingly have them quantified, more especially when the products are polyherbal. In spite of this, variations in the levels of phytoconstituents in plant materials collected from different sources and at different times are a major concern for the monitoring of HMPs for content uniformity, especially in situations where the products are used for the treatment of infectious diseases such as malaria. It is already known that the levels of plant secondary metabolites are influenced by factors such as growth conditions, time and method of harvesting, and storage condition, as well as the geographical location [8–11]. Since at the moment a lot of these factors influence the levels of phytoconstituents in medicinal plants, and the manufacturing practices are not standardized, it becomes important in the interest of the HMP patrons to develop simple, cost-effective, and efficient system to assure the quality of HMPs.

In view of the difficulty associated with the development of assay methods for HMPs because of the myriad of secondary metabolites, the approach by the Food and Drugs Authority, Ghana, to approve and register HMPs has mainly focused on microbial load, toxicity, pH, and some physicochemical parameters such as ash value and acid value. It is however necessary to have a technique where the protocol involves measurement of some phytochemical components of the HMPs. Some of the challenges associated with such desired methods in resource constrained environments are the cost of equipment, accessories, and maintenance. In addition, the methods would be expected to target detection and quantification of almost all the phytochemical components of the polyherbal HMPs or selected markers [12]. These methods usually involve techniques such as high-performance liquid chromatography (HPLC), gas chromatography-mass spectrometry (GC-MS), and capillary electrophoresis (CE) which are difficult to find in countries such as Ghana for routine chemical quality monitoring of HMPs. Besides the instrumental challenges are tedious sample

preparations with high volume of solvent and solvent-wasting separation methods [12, 13].

The ultraviolet-visible (UV-Vis) spectroscopy as a simple, cost-effective, and nondestructive technique has found applications in environmental, pharmaceutical, and other related fields. For example, the British and United States Pharmacopoeias employ UV-Vis-based methods for the assay of some pharmaceutical products, as well as adjunct method for the identification of certain active pharmaceutical ingredients. The technique has also been reported to be useful for the analysis of liquid HMPs [14], and the inherent advantages are that the methods are easy to use, enabling laboratories to adopt and effectively implement the protocols without any extensive training of technical staff. In UV-Vis spectroscopy, it has been observed that spectra obtained from complex mixtures such as liquid HMPs are usually highly convoluted. This is because the absorption spectra of several components within the complex samples are superimposed on each other. The authors, however, envisaged that the application of chemometric techniques to the complex UV-Vis data can lead to useful analyses with relevant conclusions on the quality of HMPs.

Chemometric methods involve the use of mathematical and statistical tools to extract useful information from complex data [15]. These methods have been around for a while but have become popular lately due to the availability of easy-to-use software and statistical packages [16, 17]. Chemometric methods are generally used for machine learning and optimization of experiments (i.e., design of experiment) [18, 19]. This work will focus on the former as the latter is beyond the scope of this paper. Machine learning methods are usually termed as supervised or unsupervised [20]. Unsupervised learning methods such as principal component analysis (PCA) are arguably the most used chemometric method [20, 21]. PCA is a dimensionality reduction technique which reveals inherently hidden patterns in data. In a PCA model, samples of the same class, which have similar attributes, are clustered closer to each other. Thus, a confidence ellipse generated around a cluster of samples would show high variability in the score space. Samples deviating from the standard product will be seen projecting further away from the center of that cluster. In the context of our study, this can provide guidance in the detection of unwholesome or expired HMPs. Supervised learning methods, such as support vector machines (SVM), can also be used to classify liquid HMPs in order to distinguish them from other products. Thus, two different antimalarial HMPs are expected to be classified into two different groups. In addition, variations in the chemical composition due to expiration of the product can also be determined as the UV-Vis chemical fingerprint of the product becomes altered. Such alteration, detectable with PCA or SVM, can be employed to identify expired HMPs.

Therefore, in this study, we propose a routine for the classification of liquid antimalarial HMPs using SVM and UV-Vis spectroscopy. We further demonstrate the use of PCA and SVM models to monitor the variations in liquid HMPs and to identify expired products. This approach has

the potential to be expanded to the detection of adulterants in liquid antimalarial HMPs and by extension, other HMPs.

## 2. Experimental

**2.1. Sample Preparation.** Five liquid antimalarial HMPs were obtained from pharmacies and herbal shops in Accra and Kumasi, Ghana. The antimalarials used for this study were packaged in 500 mL amber colored plastic bottles. To ensure anonymity, the samples were coded HMP1, HMP2, HMP3, HMP4, and HMP5. Three different batches each were obtained for each sample. A fourth batch consisting of expired HMP4 labelled HMP4X was also obtained from a herbal shop.

A 5 mL portion of each sample was pipetted into a 50 mL volumetric flask. Distilled water was added to the volumetric flask to make up to the 50 mL mark. This yielded 5% (v/v) concentration of the liquid herbal antimalarial in distilled water.

**2.2. Data Collection.** UV-Vis spectra were collected using a JENWAY 7315 UV-Vis spectrophotometer (Jenway, UK) equipped with a Perkin Elmer Spectrum (Spectrum Two, Version 10.03.09, Serial Number 94133, Waltham, USA). The samples were analyzed using a 1 mL fused silica cuvette and at a wavelength range of 200 to 700 nm using distilled water as blank/reference. At least ten repeated scans were made for each sample. The raw data from the spectrum was imported into MATLAB R2020b (The MathWorks®, Natick, MA, USA). PCA and SVM models were generated using PLS Toolbox 8.9 (Eigenvector Research Inc., Manson, WA, USA).

**2.3. Data Processing and Analysis.** The data was organized into a matrix of samples in rows and wavelength in columns. The dataset matrix consisted of  $167 \times 501$  (samples  $\times$  wavelengths). Spectral data was smoothed with a moving average filter using a window of five. The spectral data was subsequently decluttered using generalized least squares- (GLS-) based weighting strategy using an alpha value of 0.02 [22]. Each column of the data matrix was mean centered while the rows were normalized to 1.

The data was further split into two main groups: 2/3 for model training and optimization and a third for external validation sets. The training and validation set data consisted of 111 and 56 samples, respectively.

Discriminant variable (DIVA) test was performed on the training set data to identify regions of the spectra providing information about the important regions of the data.

The PCA model was generated with the training and optimizing set data. In the PCA model, the variance explained by the model by each component is represented as a percentage. It is used to demonstrate measure of discrepancy between the model and the data. Thus, a higher explained variance is desired. The validation data were subsequently projected into the PCA model.

The training data was used to generate SVM classification models for the 5 classes of samples. The validation data

was then projected into the model to evaluate the performance on an external validation set. The evaluation was based on the model's sensitivity, specificity, and accuracy [23]. A model's ability to predict positive samples is true positive rate/sensitivity (sensitivity = true positives (TP)/number of positives (NP)). Specificity measures the model's ability to correctly identify negative samples, also known as true negative rate (specificity = true negatives (TN)/number of negatives (NN)). Accuracy measures the overall true predicting power (accuracy = (TP + TN)/(NP + NN)). These metrics are scaled from zero to 1, with 0 and 1 being the worst and best model, respectively.

The SVM model was further tested on the HP4X, which are the expired HP4 samples. PCA models were also generated and evaluated with the training and validation sets, respectively.

## 3. Results and Discussion

The assessment of quality of HMPs in Ghana has largely been based on some organoleptic and physicochemical parameters including but not limited to color, pH, and microbial load. However, it is necessary to develop advanced but simple strategies that target the phytoconstituents of HMPs in the evaluation of chemical quality. It is known that variation in these secondary metabolites supposed to be responsible for the biological activities of HMPs exists as a result of factors including environment, harvesting, manufacturing, storage, and stability. Due to the presence of myriad of chemical compounds in polyherbal products and the lack of adequate robust analytical methods to check batch-to-batch consistency in levels of phytoconstituents and identify expired or decomposed HMPs that may not show perceptible physical changes, unscrupulous persons could rebottle and sell substandard and unwholesome products to the general public. Therefore, this study has explored the application of UV-Vis spectroscopy and chemometric analysis in dealing with the problem. Antimalarial HMPs were chosen as a test case due to their high demand and over-the-counter usage.

Generally, all the UV-Vis spectra obtained from the analyses of the liquid antimalarial HMPs showed maximum absorbance around 230 nm, 280 nm, and 375 nm (Figure 1(a)), which suggests the presence of compounds with conjugated systems or chromophores and the suitability of our choice of technique. This agrees with other findings that show the presence of chromophoric compounds in plant products [12].

Due to noise in the dataset, a moving window smoothing algorithm was implemented (Figure 1(b)), which showed similar spectra characteristics relative to the raw spectra. In order to identify the more informative regions of the spectra, discriminant variable analysis was performed as previously described [24] to generate a variable selectivity ratio (SR) plot (Figure 2). In this analysis, an SR value less than 10 was considered as one with low discrimination power and, thus, was eliminated from the data. The threshold shows the variables which were above the set limit and reduced the number of variables from 501 to 324. It must be emphasized

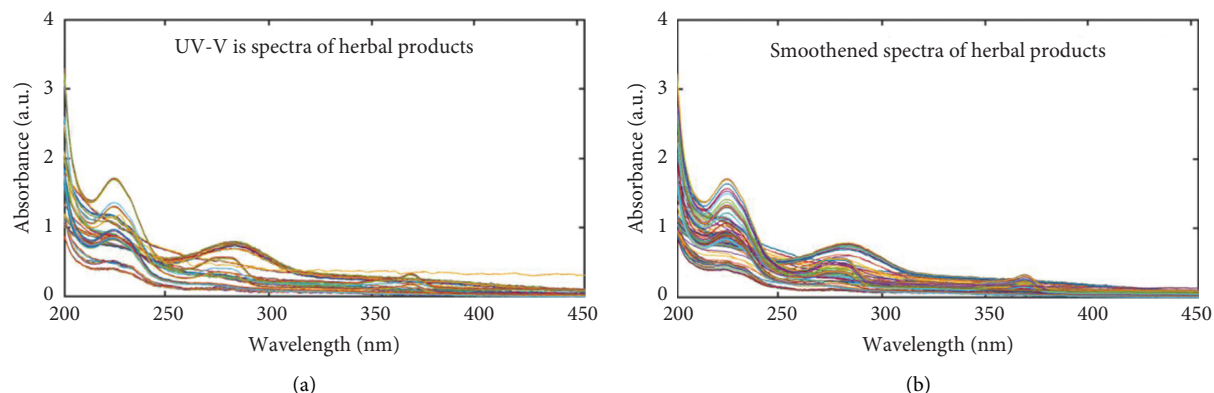


FIGURE 1: UV-Vis spectra of liquid herbal medicinal products. (a) Raw data. (b) Spectra data smoothed with a moving average filter using a window of 5.

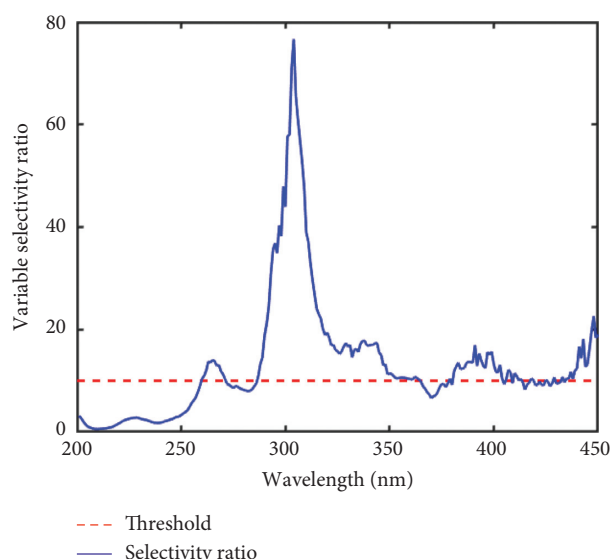


FIGURE 2: Selectivity ratio plot of UV-Vis spectra of liquid HMP obtained from DIVA test showing feature importance to discrimination between classes ( $y$ -axis) and wavelength ( $x$ -axis). Wavelengths, at which the SR is less than the threshold (red line), were eliminated from the data.

that a fewer number of descriptors with better discriminating power are much more desirable as they lead to simpler models [25, 26].

Next, principal component analysis was performed using the training set data (only the 324 variables). The results show a PCA score plot of PC1 vs. PC2 vs. PC3 (Figure 3). The explained variances for PC1 and PC2 were 37.51% and 26.38%, respectively, while the third principal component had 18.74%. Thus, a total explained variance of 82.63% was observed. It can be seen in this three-dimensional score space that the various antimalarial HMPs are all clustered in separate groups. A 95% confidence ellipse generated around each cluster shows quite a few of the training set samples straying out of the cluster. This demonstrates that UV-Vis spectra and PCA models can be used to distinguish various

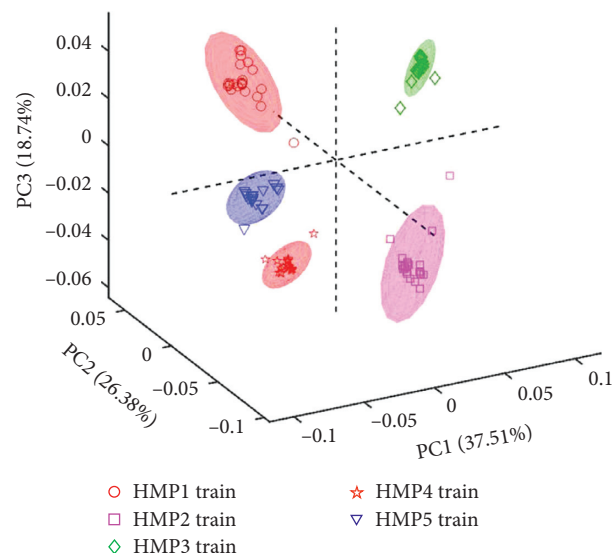


FIGURE 3: PCA score plot for training set data. A score plot for PC1 (37.51%) vs. PC2 (26.38%) vs. PC3 (18.74%). HMP1—red circles, HMP2—purple squares, HMP3—green diamonds, HMP4—red pentagrams, and HMP5—blue triangles. Training and validation data are represented by hollow and filled markers, respectively.

liquid antimalarial HMPs. Subsequently, the external validation set of each HMP class was projected in the model (Figure 4, where HMP1 are red circles, HMP2 are purple squares, HMP3 are green diamonds, HMP4 are red pentagrams, and HMP5 are blue triangles). In Figure 4, samples used for the training and validation sets are represented by hollow and filled markers, respectively. Here it is also apparent that those samples that were not used in training the model are also projected into the correct subgroups.

The ability of the PCA model generated to identify expired products was also evaluated. Expired products from HMP4, labelled as HMP4X, were also projected into the model as shown in Figure 4. These expired products were clustered into a different score space (represented as black triangles). This further demonstrates that, using the UV-Vis

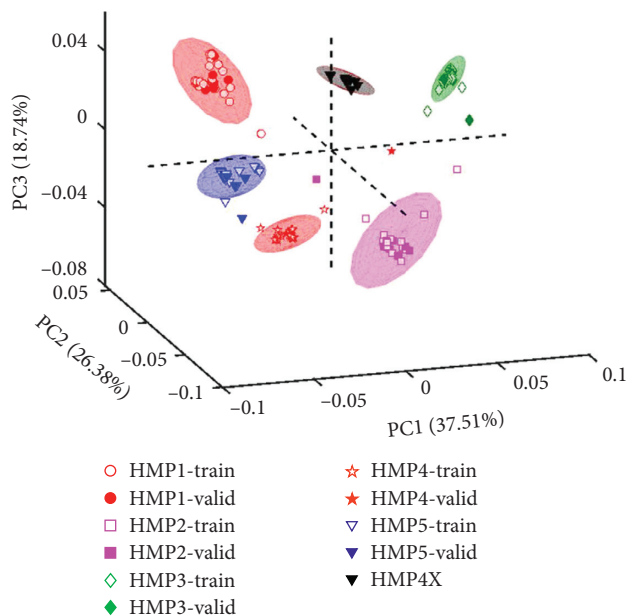


FIGURE 4: PCA plot for PC1 (37.51%) vs. PC2 (26.38%) vs. PC3 (18.74%) for training set, validation set, and expired products HMP4X. HMP1—red circles, HMP2—purple squares, HMP3—green diamonds, HMP4—red pentagrams, and HMP5—blue triangles. Black triangles: HMP4X. Training and validation data are represented by hollow and filled markers, respectively.

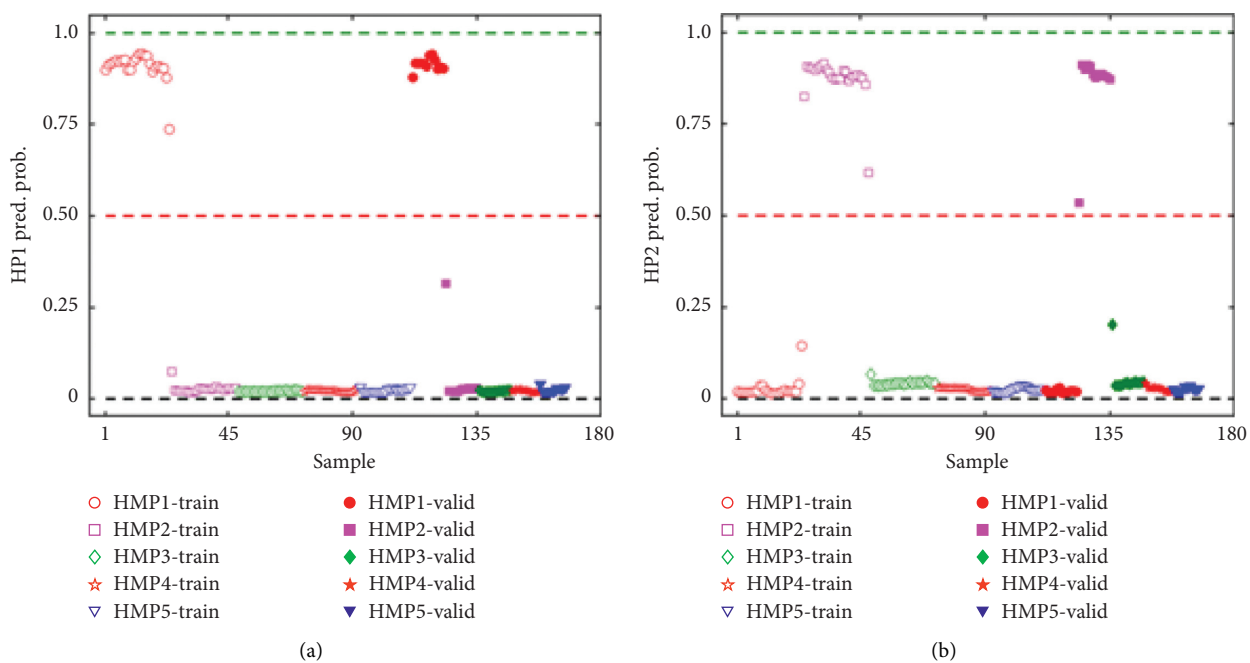


FIGURE 5: Continued.

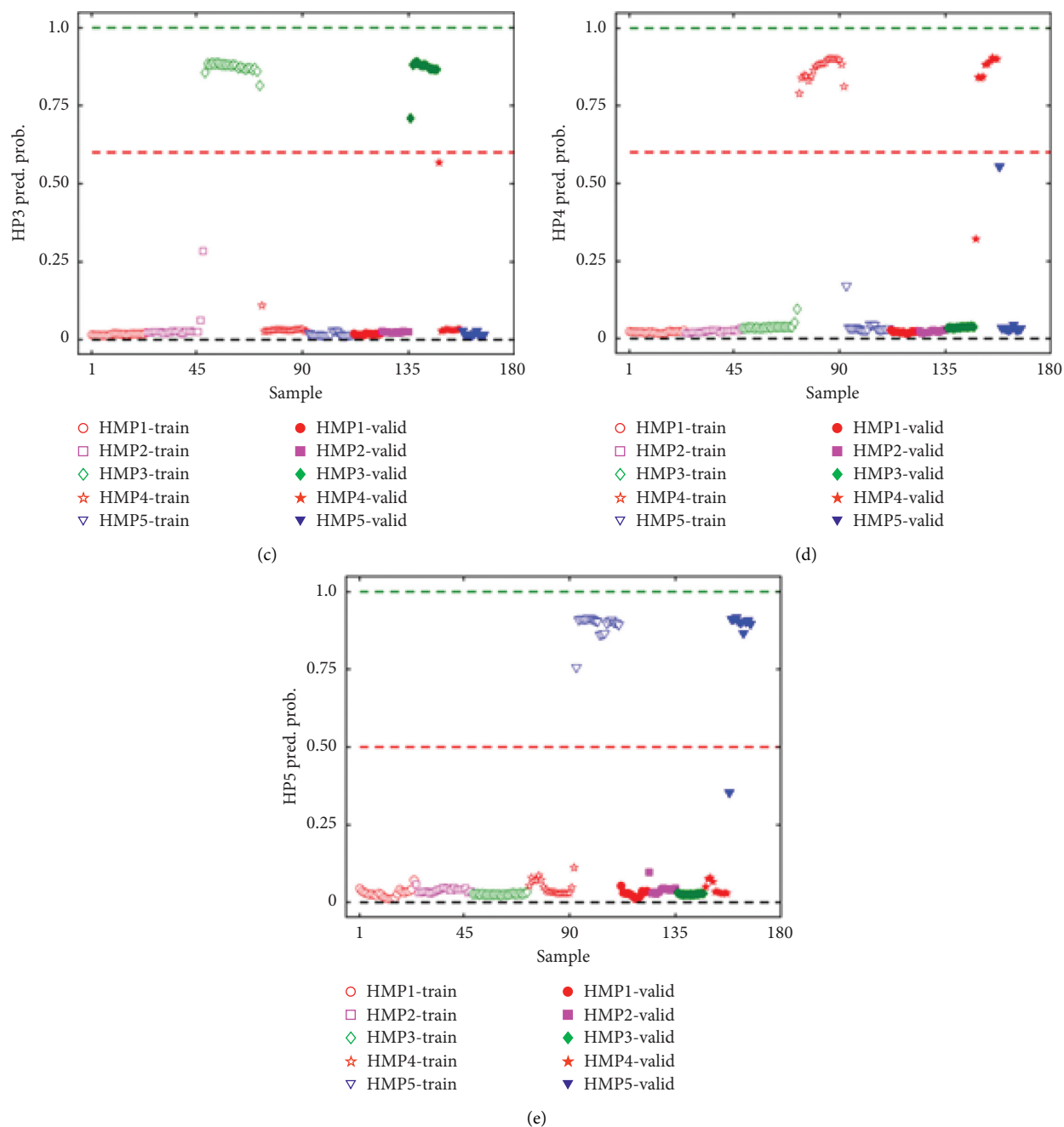


FIGURE 5: SVM class predicted probability for herbal medicinal products HMP1 (a), HMP2 (b), HMP3 (c), HMP4 (d), and HMP5 (e). HMP1—red circles, HMP2—purple squares, HMP3—green diamonds, HMP4—red pentagrams, and HMP5—blue triangles. Hollow and filled markers represent training and validation sets, respectively. The red dashed line represents the discrimination barrier above which samples are positively predicted as designated by the  $y$ -axis label.

spectra and PCA, products that are expired can be easily detected. This is of high importance due to the fact that liquid HMPs could go bad without showing perceptible variations in their physical appearance and taste.

SVM classification models were generated using the 324 features obtained from the DIVA test. The model was generated using a radial basis function kernel with cost and gamma values of 100 and 0.1, respectively. A venetian blind cross validation was employed due to structure of the data

and it gave consistent results. The model was generated with the training set data and validated using an external validation set. The class predicted probability plots for all the 5 products (HMP1, HMP2, HMP3, HMP4, and HMP5) are shown in Figures 5(a)–5(e), where HMP1 are red circles, HMP2 are purple squares, HMP3 are green diamonds, HMP4 are red pentagrams, and HMP5 are blue triangles, red dash lines are discrimination barrier, and training and validation sets are represented by hollow and filled markers,

TABLE 1: Table of results for SVM classification for training and validation sets for all herbal products.

Product ID	True positive	False negative	True negative	False positive	Sensitivity	Specificity	Accuracy
<i>Training set</i>							
HMP1	24	0	87	0	1.00	1.00	1.00
HMP2	24	0	87	0	1.00	1.00	1.00
HMP3	24	0	87	0	1.00	1.00	1.00
HMP4	20	0	91	0	1.00	1.00	1.00
HMP5	19	0	92	0	1.00	1.00	1.00
<i>Validation set</i>							
HMP1	12	0	44	0	1.00	1.00	1.00
HMP2	12	0	44	0	1.00	1.00	1.00
HMP3	12	0	44	0	1.00	1.00	1.00
HMP4	9	0	46	1	0.90	1.00	0.98
HMP5	9	0	46	1	0.90	1.00	0.98

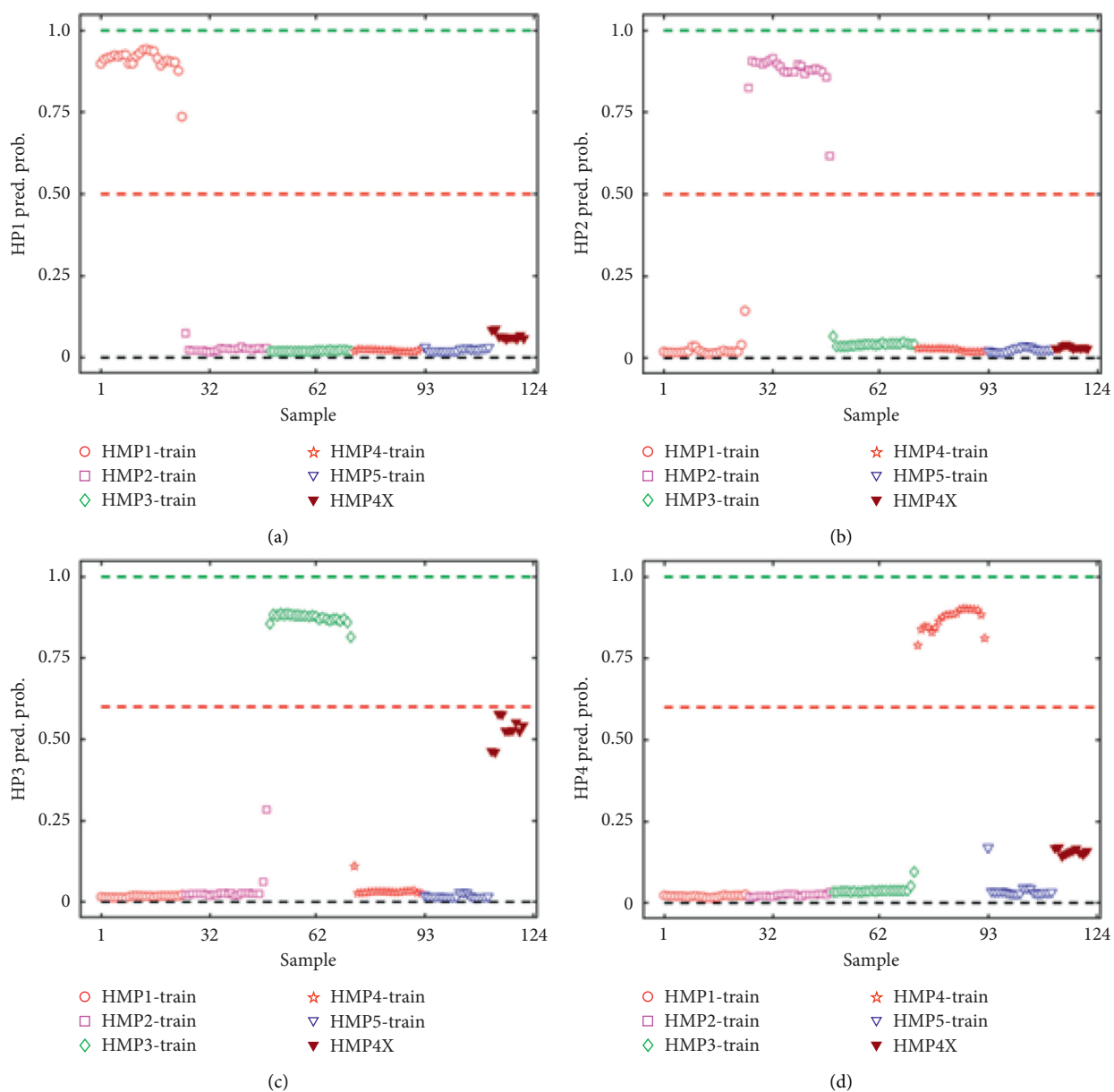


FIGURE 6: Continued.

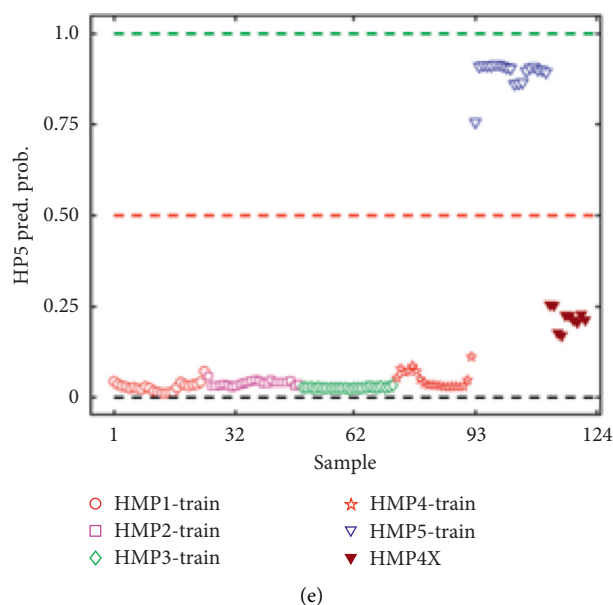


FIGURE 6: SVM class predicted probability for herbal medicinal products for training data HMP1 (a), HMP2 (b), HMP3 (c), HMP4 (d), and HMP5 (e) showing expired samples HMP4X. HMP1—red circles, HMP2—purple squares, HMP3—green diamonds, HMP4—red pentagrams, and HMP5—blue triangles. HMP4X—black triangles. Training and validation data are represented by hollow and filled markers, respectively. The red dashed line represents the discrimination barrier above which samples are positively predicted as designated by the y-axis label.

respectively. In the SVM of Figure 5, class predicted probability closer to zero (black dash lines) indicated less likelihood of samples belonging to the class being predicted. On the other hand, a class predicted probability close to 1.00 (green dash line) indicates that the sample may belong to the class being predicted. A class discrimination boundary is represented by a red dashed line. The numerical results from Figure 5 are shown in Table 1. There were no false negative or false positives leading to classification sensitivity, specificity, and accuracy of 1.00 (denoting 100%), which demonstrate the power of combination of UV-Vis spectrum and SVM for HMPs classification. However, in the validation set, two false positives were identified from HMP4 and HMP5. It is normal for a classification model to perform better on a training set data than a validation set.

The ability of the SVM model to detect expired products was evaluated using the spectra of HMP4X. The samples in HMP4X were projected into the model to check, if indeed, the model will predict it as HMP4 or others. It can be seen that HMP4X samples are not predicted to be in any of the 5 HMP groups (Figure 6). Thus, the class predicted probability is below 0.5 for all five classes of HMPs. This is because these products have undergone some chemical changes even though there is no perceptible physical change which has altered their UV-VIS spectra. This demonstrates that, using UV-VIS spectra and SVM classification, expired herbal products can be detected. The study scope would be expanded to include batch-to-batch consistency and stability of HMPs.

#### 4. Conclusion

A simple spectroscopic method (UV-Vis spectroscopy) together with chemometric models was successfully

developed and applied to the assessment of liquid antimalarial HMPs. This method demonstrates the ability of PCA to distinguish between different HMPs. In addition, we applied SVM models to UV-VIS spectra of liquid HMPs to classify different antimalarials. Prediction sensitivity, specificity, and accuracy of 1.00 (100%) were observed for training set data for all the products. With respect to the validation set, sensitivity, specificity, and accuracy of prediction were 1.00 in all cases for HMP1, HMP2, and HMP3. However, sensitivity and accuracy for HMP4 and HMP5 were 0.90 and 0.98, respectively. The SVM method also demonstrated its ability to distinguish between wholesome and expired products.

#### Data Availability

The data for this project are available at [lawrenceadutwum/herbalproducts \(github.com\)](https://github.com/lawrenceadutwum/herbalproducts).

#### Conflicts of Interest

The authors declare that they have no conflicts of interest.

#### References

- [1] World Health Organization, *World Malaria Report 2020*, WHO, Geneva, Switzerland, 2021.
- [2] F. Nosten and N. J. White, "Artemisinin-based combination treatment of *falciparum* Malaria," *American Society of Tropical Medicine and Hygiene*, vol. 77, no. 6, pp. 181–192, 2007.
- [3] R. Adisa, T. O. Fakeye, and D. Dike, "Evaluation of adverse drug reactions to artemisinin-based combination therapy in a



- Nigeria University community," *Tropical Journal of Pharmaceutical Research*, vol. 7, no. 2, pp. 937–944, 2008.
- [4] H. A. Pawar, "Natural product as a source of lead to the design of new drugs," *Natural Products Chemistry and Research*, vol. 2, no. 6, pp. 1–3, 2014, <http://paperpile.com/b/rSJ4hz/Tasf10.4172/2329-6836.1000156>.
- [5] E. G. Azeh, "Malaria and treatment: herbal antimalarials as alternative to conventional medicine," *Research Journal of Phytomedicine*, vol. 1, p. 1, 2015.
- [6] S. M. Abdel-Aziz, A. Aeron, and T. A. Kahil, "Health benefits and possible risks of herbal medicine," *Microbes in Food and Health*, vol. 36, pp. 97–116, 2016, [http://paperpile.com/b/rSJ4hz/cZL710.1007/978-3-319-25277-3\\_6](http://paperpile.com/b/rSJ4hz/cZL710.1007/978-3-319-25277-3_6).
- [7] I. A. Kretchy, A. Koduah, K. F. M. Opuni et al., "Prevalence, patterns and beliefs about the use of herbal medicinal products in Ghana: a multi-centre community-based cross-sectional study," *Tropical Medicine & International Health*, vol. 26, no. 4, pp. 410–420, 2021.
- [8] D. Donno, R. Boggia, P. Zunin et al., "Phytochemical fingerprint and chemometrics for natural food preparation pattern recognition: an innovative technique in food supplement quality control," *Journal of Food Science and Technology*, vol. 53, no. 2, pp. 1071–1083, 2016.
- [9] H. C. Andola, K. S. Gaira, R. S. Rawal, M. S. M. Rawat, and I. D. Bhatt, "Habitat-dependent variations in berberine content of berberis asiatica Roxb. ex. DC. in Kumaon, Western Himalaya," *Chemistry & Biodiversity*, vol. 7, no. 2, pp. 415–420, 2010, <http://paperpile.com/b/rSJ4hz/31xH10.1002/cbdv.200900041>.
- [10] I. Zribi, F. Omezzine, and R. Haouala, "Variation in phytochemical constituents and allelopathic potential of *nigella sativa* with developmental stages," *South African Journal of Botany*, vol. 94, pp. 255–262, 2014.
- [11] A. Lavola, A. Salonen, V. Virjamo, and R. Julkunen-Tiitto, "Phytochemical variation in the plant-part specific phenols of wild crowberry (*Empetrum hermaphroditum* Hagerup) populations," *Phytochemistry Letters*, vol. 21, pp. 11–20, 2017.
- [12] D. Kumadoh, K. O. Kwakye, N. Kuntworbe, O. Adi-Dako, and J. A. Appenahier, "Determination of shelf life of four herbal medicinal products using high-performance liquid chromatography analyses of markers and the systat sigmaplot software," *Journal of Applied Pharmaceutical Science*, vol. 10, no. 6, pp. 72–80, 2020.
- [13] M. A. Archer, D. Kumadoh, G. N. Yeboah et al., "Formulation and evaluation of capsules containing extracts of cassia sieberiana for improved therapeutic outcome," *Scientific African*, vol. 10, 2020.
- [14] K. Kucharska-Ambrożej and J. Karpinska, "The application of spectroscopic techniques in combination with chemometrics for detecting adulteration of some herbs and spices," *Microchemical Journal*, vol. 153, pp. 104–278, 2020.
- [15] I. E. Frank and B. R. Kowalski, "Chemometrics," *Analytical Chemistry*, vol. 54, no. 5, pp. 232–243, 1982.
- [16] J. Camacho and E. Saccenti, "Chemometrics analysis of big data," *Comprehensive Chemometrics*, vol. 36, pp. 437–458, 2020.
- [17] H. Martens, "Quantitative big data: where chemometrics can contribute," *Journal of Chemometrics*, vol. 30, no. 9, p. 559, 2016.
- [18] R. G. Brereton, *Applied Chemometrics for Scientists*, John Wiley & Sons, Hoboken, NJ, USA, 2007.
- [19] M. Otto, *Chemometrics: Statistics and Computer Application in Analytical Chemistry*, John Wiley & Sons, Hoboken, NJ, USA, 2016.
- [20] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, "A systematic review on supervised and unsupervised machine learning algorithms for data science," *Unsupervised vs Semi-supervised Learning*, vol. 47, pp. 3–21, 2020.
- [21] C. Christophe, "PCA: the basic building block of chemometrics," *Analytical Chemistry*, vol. 29, p. 47, 2012.
- [22] Q. Liu, Y. Okui, and Y. Yoshimura, "Generalized least squares model averaging," *Economic Review*, vol. 35, no. 8, pp. 1692–1752, 2016.
- [23] T. W. Loong, "Understanding sensitivity and specificity with the right side of the brain," *BMJ*, vol. 327, no. 7417, pp. 716–719, 2003.
- [24] T. Rajalahti, R. Arneberg, A. C. Kroksveen, M. Berle, K. M. Myhr, and O. M. Kvalheim, "Discriminating variable test and selectivity ratio plot: quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles," *Analytical Chemistry*, vol. 81, no. 7, pp. 2581–2590, 2009.
- [25] A. Amirteimoori, S. Kordrostami, A. Masoumzadeh, and M. Maghbouli, "Increasing the discrimination power of data envelopment analysis," *International Journal of Operational Research*, vol. 19, no. 2, p. 198, 2014.
- [26] L. Angulo-Meza and M. P. E. Linz, "Review of methods for increasing discrimination in data envelopment analysis," *Annals of Operations Research*, vol. 116, no. 1–4, pp. 225–242, 2002.