I. INTRODUCTION

Civet coffee is produced using an unconventional process called digestive bio-processing, a regular coffee bean fermentation in the civet cat’s intestine. The partially metabolized and excreted coffee beans by the Alamid were gathered, washed, and dried [1]. Civet coffee is considered highly marketable and rare [2], making it prone to fraud and scams. Civet coffee enthusiast condemns cheap imitation of Kopi luwak (civet coffee) where ordinary coffee bean varieties are mixed or claimed to be an original civet coffee (civet coffee) where ordinary coffee bean varieties are enthusiast condemns cheap imitation of Kopi luwak (civet coffee) where ordinary coffee bean varieties are mixed or claimed to be an original civet coffee [3]. For the time being, there is no globally agreed way of determining whether a bean is civet coffee [4].

Related studies regarding civet coffee have been carried by various researchers to validate civet coffee's authentic properties. Several types of research yield results that characterize civet coffee in terms of aroma and volatile [4], sensory attributes [5], metabolites [6], caffeine, α-tocopherol, and chosen elements such as carbon (C) & oxygen (O) [7], protein, sugar, fat, and acidity [8].

Some studies applied spectroscopy, that deals with the analysis of how light acts towards a given object [9], in discrimination of civet coffee. Subandy and Ulyia [10] used chosen fluorescence spectra and the SIMCA method to differentiate the two coffee varieties in terms of specificity and sensitivity values. On the other hand, another study utilized UV-Visible spectroscopy combined with various chemometrics methods [11] in which produced a 100% satisfaction in civet coffee classification. However, the major disadvantage with the mentioned studies [4]-[11] is that they require the samples to be grounded or mixed with certain chemicals, which may be time-consuming and complicated. Some samples will also be wasted because they cannot be reused for other purposes.

Arboleda [12] used Near-Infrared spectroscopy (NIRS), including Artificial Neural Network (ANN), to classify between civet coffee and Arabica coffee, which produced a classification result of 95% to 100% range. The study does not require grounded coffee beans as well as chemical mixtures. Nevertheless, the account that implements spectroscopy for discrimination of civet coffee that used whole coffee bean and did not use chemicals is very limited.

Spectroscopic methods have significant benefits, which are speed and cost per study [13]. The use of visible spectroscopy analysis is increasing and can provide an objective, repeatable, undestructive method to monitor and assess food quality and other commodities in agriculture [14]. Ultraviolet and visible spectroscopy represents the oldest and easiest process which implemented spectroscopy for discrimination of civet coffee that used whole coffee bean and did not use chemicals is very limited.

Discrimination of civet coffee using visible spectroscopy

Graciella Mae L. Adier, Charlene A. Reyes, Edwin R. Arboleda*1

Department of Computer and Electronics Engineering, Cavite State University
Indang, Cavite, Philippines 4122

How to cite: G. M. L. Adier, C. A. Reyes, and E. R. Arboleda, "Discrimination of civet coffee using visible spectroscopy," Jurnal Teknologi dan Sistem Komputer, vol. 8, no. 3, pp. 239-245, 2020. doi: 10.14710/jtsiskom.2020.13734, [Online].

Abstract - Civet coffee is considered as highly marketable and rare. This specialty coffee has a special flavor and higher price relative to regular coffee, and it is restricted in supply. Establishing a straightforward and efficient approach to distinguish civet coffee for quality; likewise, consumer protection is fundamental. This study utilized visible spectroscopy as a non-destructive and quick technique to obtain the absorbance, ranging from 450 nm to 650 nm, of the civet coffee and non-civet coffee samples. Overall, 160 samples were analyzed, and the total spectra accumulated was 960. The data gathered from the first 120 samples were fed to the classification learner application and were used as a training data set. The remaining samples were used for testing the classification algorithm. The study shows that civet coffee bean samples have lower absorbance values in visible spectra than non-civet coffee bean samples. The process yields 96.7% to 100% classification scores for quadratic discriminant analysis and logistic regression. Among the two classification algorithms, logistic regression generated the fastest training time of 14.050 seconds. The application of visible spectroscopy combined with data mining algorithms is effective in classifying between civet coffee and Arabica coffee, which produced a classification result of 95% to 100% range. The study does not require grounded coffee beans as well as chemical mixtures. Nevertheless, the account that implements spectroscopy for discrimination of civet coffee that used whole coffee bean and did not use chemicals is very limited.

Keywords – civet coffee; visible spectroscopy; classification algorithm; absorbance; classification learner application

*Corresponding author (Edwin R. Arboleda)
Email: edwin.r.arboleda@cvsu.edu.ph

DOI:10.14710/jtsiskom.2020.13734
Available at https://jtsiskom.undip.ac.id (8 June 2020)

Copyright ©2020, JTSiskom, e-ISSN: 2338-0403, p-ISSN: 2620-4002
Submitted: 1 May 2020; Revised: 26 May 2020; Accepted: 28 May 2020; Published: 31 July 2020
Coffee is restricted in supply because of special flavor and high price relative to regular coffee. The spectroscopic approach used for coffee authentication is most often paired with the strength of methods involving pattern recognition [10]. Most used in various plant studies for its two-dimensional aspect [13].

Classification aims to classify the data wherein the data included has the category type [19]. The problem occurs when a subject has to be classified into a specified group, given a set of identified characteristics aligned with that subject [20]. Pattern identification stands for the methods where the information of the sample membership type is used for classification purposes [21]. Machine learning is intended to teach systems on how to process data and gain knowledge from it [22]. A classifier is an algorithm mapping a specific category of input data using similar features [23]. The two sorts of classifiers are binary classifiers and multi-class classifiers [24]. The binary classification is concerned with two classes. On the other hand, multi-class classification deals with more than two classes for classification [25]. The classification model shall be built based on the training and data collection with defined groups [26]. The classifiers’ training gives way to identify the best possible way to separate between two data classes and ensure accuracy [27].

This study was carried out to determine the difference of civet coffee to non-civet coffee in visible spectra. Regarding the authors’ awareness, little to none evidence of discrimination regarding civet coffee using visible spectroscopy has been documented in the literature. This study also aims to identify whether data mining algorithms can distinguish between the two coffee varieties. The research concerns the efficacy of the procedure in the context of discriminating civet coffee.

II. RESEARCH METHOD

A. Research design

Civet coffee was made from the coffee berry, which the Alamid, scientifically named Paradoxorus Hermaphroditus, digested [2]. The distinct flavor of the said coffee variety was hard to verify. In cheap coffee imitations, ordinary coffee bean varieties are mixed or claimed to be an original civet coffee [3]. Due to its anticipated marketable characteristics, it became a target to fraud and imitation where ordinary coffee beans are anticipated marketable characteristics, it became a target claimed to be an original civet coffee

This study was carried out to determine the difference of civet coffee to non-civet coffee in visible spectra. Regarding the authors’ awareness, little to none evidence of discrimination regarding civet coffee using visible spectroscopy has been documented in the literature. This study also aims to identify whether data mining algorithms can distinguish between the two coffee varieties. The research concerns the efficacy of the procedure in the context of discriminating civet coffee.

B. Research procedure

The research procedure was divided into four main tasks: sample preparation, data collection and variable selection, training of classifier algorithms, and evaluation of classifier performance. Raspberry Pi Zero was used to gather spectral data from the visible sensor and stored directly on a USB flash drive for data set compilation. Classifier Learner app from MATLAB software was used for data processing. A total of 60 civet coffee beans and 60 non-civet coffee beans samples were used to train classifier algorithms, while 20 samples of both civet and non-civet coffee beans were used as a test sample to evaluate the chosen classifier algorithm if it properly classifies civet and non-civet coffee.

Sample preparation

One hundred sixty Robusta coffee bean samples were used for data gathering in this study. Samples were split equally – 80 Robusta coffee beans, which was given to a wild civet category, was considered the first class. In comparison, the succeeding class with 80 Robusta coffee beans were considered as the non-civet coffee sample for testing. Civet coffee beans excreted by a wild civet cat were collected. The samples were shown in Figure 1. Both civet and non-civet coffee beans have been washed and sun-dried. The parchment layer was removed for civet beans while the outer skin, pulp, or mesocarp layer and the parchment layer were extracted for non-civet beans. The samples were stored around 20°C-25°C or at room temperature.

Data collection

The visible sensor used in this study is AS7262. The wavelengths in the visible range detectable by the sensor are 450nm, 500nm, 550nm, 570nm, 600nm, and 650nm, each with 40nm of full-width half-max detection. This sensor gathered reflectance from the sample. The spectral data gathered was converted from .csv extension to excel data (.xls) format and converted to absorbance. The absorbance is determined by measuring the amount reflected and transmitted when there is an incident of light on the object using Eq. 1 that satisfies Beer-Lambert's law [28]. $A$ denotes absorbance, $I_0$ is the original intensity of light, and $I$ is the transmitted intensity of light.


| Classification model types | Decision Trees | Discriminant Analysis | Logistic Regression Classifier | Support Vector Machines | Nearest Neighbor Classifier | Ensemble Classifier |
|----------------------------|----------------|-----------------------|-------------------------------|-------------------------|-----------------------------|---------------------|
| Fine Tree                  | Linear Discriminant Analysis | Logistic Regression | Linear SVM                    | Fine kNN                 | Boosted Trees               |
| Medium Tree                | Quadratic Discriminant Analysis |                     | Quadratic SVM                 | Medium kNN               | Bagged Trees                |
| Coarse Tree                |                                |                       | Cubic SVM                     | Coarse kNN              | Subspace Discriminant       |

Table 2. Range of gathered absorbance values for each wavelength

| Coffee Type | 450 nm | 500 nm | 550 nm | 570 nm | 600 nm | 650 nm |
|-------------|--------|--------|--------|--------|--------|--------|
| Civet coffee | 1.5850 to 2.1308 | 1.3696 to 1.9914 | 1.3768 to 1.9872 | 1.3757 to 2.0862 | 1.5817 to 2.1739 | 1.4377 to 1.9547 |
| Non-civet coffee | 1.9586 to 2.5086 | 1.8153 to 2.4152 | 1.8500 to 2.6330 | 2.0888 to 2.4726 | 1.9547 to 2.7957 | 1.7375 to 2.7565 |

Variable selection

Data cleaning was carried out to remove unnecessary, incomplete, and fluctuating data [29]. It aimed to select data that are essential to the study for generating more accurate results. The wavelengths with large absorbance values differences have been selected as the features to be used in the model training. Upon choosing the features to be used, the principal component analysis or PCA in Classification Learner was used to minimize dimensions to avoid overfitting.

Classification

MATLAB was used for development, study, and design through simulation, application development, data processing, analysis, and mathematical computations [30], [31]. Classification Learner App from this software was used to train the data gathered from the samples. The application covers five categories of classifiers and includes a total of 23 classifiers [32]. Gathered data were analyzed, and relevant features were selected as training set for higher accuracy. This feature selection approach was used to minimize data dimensions and often includes choosing an important element of the primary data set [33].

Moreover, a validation method was also used to prevent overfitting to the training dataset. The 25-fold Cross-validation scheme was used to split the data into 25 disjoint sets. The data is grouped by k sets of data, in which the value of the fold is the k [34]. By this validation scheme, the data was set into 75% for training and 25% for testing. A total of 160 samples were used for data gathering, 75% of which is 120 samples are used for training, and 25% of which is 40 samples used for testing. This approach provides a reasonable approximation of the classification models’ accuracy that has been trained with the dataset.

Table 1 shows all the classification algorithms used to train the gathered data.

C. Evaluation

The measurement criteria used to compare the two classifiers are accuracy and precision [35], [36]. Accuracy defines the ratio of summation of true positive and true negative to the total events (Eq. 2). TP or True Positive indicates the total positive assumptions from the real positive sample, while FP or False Positive indicates the total positive assumptions from the real negative one. Moreover, TN or True Negative represents the total negative assumptions from the real negative type, while FN or False Negative indicates the total negative assumptions from the real positive set. On the other hand, precision describes the ratio of true positive to the total predicted positive observations (Eq. 3).

Accuracy\(\%\) = \(\frac{TP + TN}{TP + TN + FP + FN}\) \times 100 \hspace{1cm} (2)

Precision\(\%\) = \(\frac{TP}{TP + FP}\) \times 100 \hspace{1cm} (3)

III. RESULT AND DISCUSSION

A. The absorbance values of civet and non-civet coffee

The differences between the spectral values are 550 nm, 600 nm, and 650 nm are shown in Table 2. The graph of the absorbance values for civet and non-civet is shown in Figure 2. It can be observed that civet coffee has lower absorbance in visible spectra in comparison to non-civet coffee.

B. Training of classifier algorithms

A total of 60 civet coffee beans and 60 non-civet coffee beans were used to create a training dataset. All types of classifier models in the Classification Learner

A = 1 - \(\frac{L}{I_o}\) \hspace{1cm} (1)
The app was trained. All the accuracy of each classification algorithm was seen at the History List once they finish training. The highest overall accuracy was highlighted in a box. In choosing the most reliable and effective classifier model, such significant factors, including training period and accuracy, have to be evaluated.

Table 3 shows the highest accuracy as 98.3 %, produced by Quadratic Discriminant Analysis and Logistic Regression. Fine Tree, Medium Tree, Coarse Tree, Linear Discriminant, Linear SVM, Cubic SVM, Coarse Gaussian SVM, Cosine KNN, Cubic KNN, Weighted KNN, Bagged Trees, and Subspace Discriminant comes at second with 97.5 %. However, Coarse KNN, Boosted Trees, and RUSBoosted Trees got the lowest accuracy of 41.7 %. Among all the classification model that got the highest accuracy, Logistic Regression got the fastest training time of 14.05 seconds.

The confusion matrix was used to evaluate the accuracy percentage of Quadratic Discriminant Analysis and Logistic Regression. It shows how civet coffee beans and non-civet coffee beans are accurately classified according to their group. Table 4 indicates that Quadratic Discriminant Analysis properly classifies 100% of civet coffee beans and approximately 97 % of non-civet beans. Moreover, it was clearly shown in Table 5 that Logistic Regression correctly classifies approximately 98 % of both civet and non-civet coffee beans.

C. Testing of the final models

A selection of 20 civet coffee beans and 20 non-civet coffee beans were used to test the final models. It can be seen in Table 6 that both Quadratic Discriminant

| classifier          | classifier type | accuracy (%) | prediction speed (Object/Seconds) | training time (Seconds) |
|---------------------|-----------------|--------------|-----------------------------------|-------------------------|
| Decision Trees      | Fine Tree       | 97.5         | 160                               | 11.269                  |
|                     | Medium Tree     | 97.5         | 130                               | 11.228                  |
|                     | Coarse Tree     | 97.5         | 150                               | 13.76                   |
| Discriminant Analysis | Linear Discriminant Analysis | 97.5 | 130                               | 15.281                  |
|                     | Quadratic Discriminant Analysis | 98.3 | 130                               | 17.609                  |
| Logistic Regression Classifier | Logistic Regression | 98.3 | 130                               | 14.050                  |
| Support Vector Machines | Linear SVM      | 97.5         | 180                               | 14.432                  |
|                     | Quadratic SVM   | 96.7         | 170                               | 17.669                  |
|                     | Cubic SVM       | 97.5         | 180                               | 17.380                  |
|                     | Fine Gaussian SVM | 93.3     | 140                               | 21.332                  |
|                     | Medium Gaussian SVM | 96.7     | 140                               | 21.042                  |
|                     | Coarse Gaussian SVM | 97.5     | 160                               | 24.622                  |
| Nearest Neighbor Classifier | Fine KNN        | 96.7         | 150                               | 24.211                  |
|                     | Medium KNN      | 95.8         | 150                               | 27.673                  |
|                     | Coarse KNN      | 86.7         | 180                               | 27.376                  |
|                     | Cosine KNN      | 97.5         | 200                               | 30.415                  |
|                     | Cubic KNN       | 97.5         | 200                               | 29.990                  |
|                     | Weighted KNN    | 97.5         | 140                               | 32.272                  |
| Ensemble Classifiers | Boosted Trees   | 41.7         | 200                               | 32.506                  |
|                     | Bagged Trees    | 97.5         | 47                                | 50.965                  |
|                     | Subspace Discriminant | 97.5     | 38                                | 51.249                  |
|                     | Subspace KNN    | 96.7         | 36                                | 64.245                  |
|                     | RUSBoosted Trees| 41.7         | 250                               | 51.569                  |
Analysis and Logistic Regression accurately classify the samples according to their group with 100 % accuracy, which is slightly better than [12] with 95 % to 100 % accuracy that uses ANN classification. This study successfully implements spectroscopy for civet coffee discrimination that used whole coffee bean and did not use chemicals in contrast with [4]-[11].

It can be observed in Figure 3 that in terms of accuracy, the Quadratic Discriminant Analysis is higher. However, in terms of precision, Logistic Regression scored 98 %, which is approximately 1% higher than Quadratic Discriminant Analysis. In advantage, the Logistic Regression is simple and precise, which produces a good prediction. It is also commonly used in fraud detection in various research [37]-[39] - One of the study's main objectives.

IV. CONCLUSION

The application of visible spectroscopy in combination with a classification algorithm was effective in discriminating civet coffee from non-civet coffee. Classification models correctly define civet and non-civet by their absorbance levels. The classification models, namely quadratic discriminant analysis and logistic regression, yield 100% accuracy for both civet and non-civet classifications. Among all the classifier models, logistic regression is the most effective classifier since it has high accuracy, precision, and fastest training time of 14.050 seconds.

ACKNOWLEDGMENTS

The author(s) would like to express their sincere gratitude to the faculty members of Cavite State University-Indang for providing invaluable guidance, comments, and suggestion about the study. The authors would also like to thank Mrs. Rizalyn Gacot Latube for the samples used in the study and sample preparation guidance.

REFERENCES

[1] M. Muzaifa, D. Hasni, A. Patria, F. and A. Abubakar, "Sensory and microbial characteristics of civet coffee," International Journal on Advanced Science Engineering Information Technology, vol. 8, no. 1, pp. 165-171, 2018. doi: 10.18517/ijaseit.8.1.3092
[2] M. . F. Marcone, "Composition and properties of Indonesian palm civet coffee," Food Research International, vol. 37, pp. 901-912, 2004. doi: 10.1016/j.foodres.2004.05.008
[3] S. Gonzaga , "Authenticating the world's most expensive coffee: human world," Earthsky, 2013, [Online]. Available: https://earthsky.org/human-world/authenticating-the-worlds-most-expensive-coffee
[4] E. Ongo et al., "Chemometric discrimination of philippine civet coffee using electronic nose and gas chromatography mass," Procedia Engineering,

Table 4. Confusion matrix of Quadratic Discriminant Analysis

| Coffee Type | Classified as | Accuracy (%) |
|-------------|--------------|--------------|
| Civet       | 60           | 100          |
| Non-civet   | 2            | 96.67        |

Table 5. Confusion matrix of Logistic Regression

| Coffee Type | Classified as | Accuracy (%) |
|-------------|--------------|--------------|
| Civet       | 59           | 98.33        |
| Non-civet   | 1            | 98.33        |

Table 6. Final classification models results

| Classifier        | Coffee Type | Accuracy (%) |
|-------------------|-------------|--------------|
| Quadratic         | 20          | 100          |
| Discriminant      | 20          | 100          |
| Logistic Regression| 20          | 100          |

Figure 3. Accuracy and precision of the two chosen classifiers

vol. 47, pp. 977-980, 2012. doi: 10.1016/j.proeng.2012.09.310
[5] K. Lopetcharat, F. Kulapichitr, I. Suppavorasatit, T. Chodjarusawad, A. Phathara-aneksin, S. Pratontep, and C. Borompichaichartkul, "Relationship between overall difference decision and electronic tongue: discrimination of civet coffee," Journal of Food Engineering, vol. 180, pp. 60-68, 2016. doi: 10.1016/j.jfoodeng.2016.02.011
[6] S. P. Putri, U. Jumhawan and E. Fukusaki, "Application of GC/MS and GC/FID-based metabolomics for authentication of Asian palm civet coffee (Kopi Luwak)," Journal of Bioscience and Bioengineering, vol. 120, no. 5, pp. 33-41, 2015.
[7] S. Chan and E. Garcia, "Comparative physicochemical analyses of regular and civet coffee," The Manila Journal of Science, vol. 7, no. 1, pp. 19-23, 2011.
[8] I. Ifmalinda, I. S. Setiasih, M. Muhaemin and S. Nurjanah, "Chemical characteristics comparison of...
palm civet coffee (kopi luwak) and arabica coffee beans," *Journal of Applied Agricultural Science and Technology*, vol. 3, no. 2, pp. 280-288, 2019. doi: 10.32530/jaast.v3i2.110

[9] A. K. Pradhan and S. N. Nahar, "Atomic astrophysics and spectroscopy," in *Atomic Astrophysics and Spectroscopy*, United Kingdom, Cambridge University Press, 2011, pp. 1-2.

[10] D. Suhandy and M. Yulia, "Discrimination of several Indonesian specialty coffees using fluorescence spectroscopy combined with SIMCA method," *IOP Conference Series: Materials Science and Engineering*, vol. 334, pp. 012059, 2018. doi: 10.1088/1742-6596/334/1/012059

[11] M. Yulia and D. Suhandy, "Indonesian palm civet coffee discrimination using UV-visible spectroscopy and several chemometrics methods," in *International Symposium on Bioinformatics, Chemometrics and Metabolomics*, Bogor, Indonesia, Oct. 2016, pp. 1-6. doi: 10.1088/1742-6596/835/1/012010

[12] E. R. Arboleda, "Discrimination of civet coffee using near infrared spectroscopy and artificial neural network," *International Journal of Advanced Computer Research*, vol. 8, no. 39, pp. 324-334, 2018. doi: 10.19101/IJACR.2018.839007

[13] G. Downey, P. McIntyre, and A. N. Davies, "Detecting and quantifying sunflower oil adulteration in extra virgin olive oils from the eastern mediterranean by visible and near-infrared spectroscopy," *Journal of Agricultural and Food Chemistry*, vol. 50, no. 20, pp. 5520-5525, 2002. doi: 10.1021/jf0257188

[14] J. M. Balage, S. L. Silva, C. A. Gomide, M. N. Bonin, and A. C. Figueira, "Predicting pork quality using Vis/NIR spectroscopy," *Meat Science*, vol. 108, pp. 37-43, 2015. doi: 10.1016/j.meatsci.2015.04.018

[15] S. El-Ahmady and M. Ashour, "Advances in food authenticity testing," in *Advances in Testing for Adulteration of Food Supplements*, 2016, pp. 667-699. doi: 10.1016/B978-0-8-100220-9.00024-2

[16] G. P. Danezis, A. S. Tsagkaris, F. Camin, V. Brusic, and C. A. Georgiou, "Food authentication: Techniques, trends & emerging approaches," *Trends in Analytical Chemistry*, vol. 85, pp. 123-132, 2016. doi: 10.1016/j.trac.2016.02.026

[17] D. Pinto, I. Castro, and A. Vicente, "The use of TIC's as a managing tool for traceability in the food industry," *Food Research International*, vol. 39, pp. 772-781, 2006. doi: 10.1016/j.foodres.2006.01.015

[18] M. M. Aung and Y. S. Chang, "Traceability in a food supply chain: Safety and quality perspectives," *Food Control*, vol. 39, pp. 172-184, 2014. doi: 10.1016/j.foodcont.2013.11.007

[19] W. Wahyono, I. N. P. Trisna, S. L. Sariwening, M. Fajar, and D. Wijayanto, "Perbandingan penghitungan jarak pada k-nearest neighbour dalam klasifikasi data textual," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 1, pp. 54-58, Jan. 2020. doi: 10.14710/itsiskom.8.1.2020.54-58

[20] C. Mishra and D. L. Gupta, "Deep machine learning and neural networks: an overview," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 6, no. 2, pp. 68-73, 2017. doi: 10.11591/ijai.v6i2.pp66-73

[21] Y. Roggo, L. Duponchel, and J.-P. Huvenne, "Comparison of supervised pattern recognition methods with McNemar's statistical test Application to qualitative analysis of sugar beet by near-infrared spectroscopy," *Analytica Chimica Acta*, vol. 477, no. 2, pp. 187-200, 2003. doi: 10.1016/S0003-2670(02)01422-8

[22] A. Dey, "Machine learning algorithms: a review" *International Journal of Computer Science and Information Technologies*, vol. 7, no. 3, pp. 1174-1179, 2016.

[23] K. E. N. T. Rab, E. R. Arboleda, A. A. Andilab, and R. M. Dellosa, "Fuzzy logic based vehicular congestion estimation monitoring system using image processing and KNN classifier," *International Journal of Scientific & Technology Research*, vol. 8, no. 8, pp. 1377-1380, 24 February 2019.

[24] H. Ali, M. N. M. Salleh, R. Saududin, K. Hussain, and M. F. Mushiq, "Imbalance class problems in data mining: a review," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 14, no. 3, pp. 1560-1571, 2019. doi: 10.11591/ijeecs.v14.i3

[25] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263-1284, 2009. doi: 10.1109/TKDE.2008.239

[26] M. Sharaf, D. Illman, and B. R. Kowalski, *Chemometrics*. New York: John Wiley & Sons Inc., 1986.

[27] T. I. Netoff, "The ability to predict seizure onset," in *Engineering in Medicine: Advances and Challenges*, United Kingdom, Academic Press, 2019, pp. 365-378. doi: 10.1016/B978-0-12-813068-1.00014-2

[28] R. L. Dean, "Understanding beer's law: an interactive laboratory presentation and related exerciser," *Journal of Laboratory Chemical Education*, vol. 2, no. 3, pp. 44-49, 2014.

[29] H. Hairani, K. E. Saputro, and S. Fadli, "K-means-SMOTE untuk menangani ketidakseimbangan kelas dalam klasifikasi penyakit diabetes dengan C4.5, SVM, dan naive Bayes," *Journal Teknologi dan Sistem Komputer*, 8(3), 2020, 244.
Advanced Technology (IJEAT), vol. 8, no. 5, pp. 1563-1567, 2019.

[33] A. Rahman and R. C. Muniyandi, "An enhancement in cancer classification accuracy using a two-step feature selection method based on artificial neural networks with 15 neurons," Symmetry, vol. 12, no. 4, 271, pp. 1-21, 2020. doi: 10.3390/sym12020271

[34] K. S. Nugroho, I. Istiadi, and F. Marisa, "Optimasi naive Bayes classifier untuk klasifikasi teks pada e-government menggunakan particle swarm optimization," Jurnal Teknologi dan Sistem Komputer, vol. 8, no. 1, pp. 21-26, 2020. doi: 10.14710/jtsiskom.8.1.2020.21-26

[35] L. M. Shi, A. Mustapha, and Y. M. Mohmad Hassim, "Predicting fatalities among shark attacks: comparison of classifiers," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 8, no. 4, pp. 360-366, 2019. doi: 10.11591/ijai.v8.i4.pp360-366

[36] H. Karim, S. R. Niyakan, and R. Safdari, "Comparison of neural network training algorithms for classification of heart diseases," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 7, no. 4, p. 185-189, 2018. doi: 10.11591/ijai.v7.i4.pp185-189

[37] Y. Sahin and E. Duman, "Detecting credit card fraud by ann and logistic regression," in International Symposium on Innovations in Intelligent Systems and Applications, Istanbul, Turkey, Jun. 2011, pp. 1-5. doi: 10.1109/INISTA.2011.5946108

[38] F. Itoo, M. Meenakshi, and S. Singh, "Comparison and analysis of logistic regression, Naïve Bayes, and KNN machine learning algorithms for credit card fraud detection," International Journal of Information Technology, pp. 1-9, 2020. doi: 10.1007/s41870-020-00430-y

[39] R. Bala and D. Garg, "Credit card fraud detection using logistic regression and bayesian network," International Journal of Innovative Research in Science, Engineering and Technology, vol. 8, no. 6, pp. 7475-7480, 2019.