MJFAS
Malaysian Journal of Fundamental and Applied Sciences

Suprapto *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 19 (2023) 539-552

RESEARCH ARTICLE

Robust Regression Analysis of Full Overlapping Caffeine and Pyridoxine HCl UV-Vis Spectra in Pharmaceutical Tablet

Suprapto Suprapto*, Yatim Lailun Ni'mah, Rifda Alifah Hisana

Department of Chemistry, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Abstract The determination of caffeine and pyridoxine HCl in medicinal tablets has been successfully carried out. The mixture of caffeine and pyridoxine solution produces UV-Vis with full overlapping spectra. The full overlapping spectra can not be analyzed using conventional multicomponent analysis, as they do not have a distinct maximum wavelength. This research proposed a full overlapping spectra analysis using Robust regression. The regression models used in this research were based on Huber, RANSAC, and Theil-Sen Regression. Robust regression is a regression method that was not sensitive to the presence of outliers from the input or output data. Robust Regression models were trained using 25 standard solutions of caffeine and pyridoxine HCl at varied ratios. The models were validated using test solutions with known concentration ratios. The validated models were applied to determine the concentration of medicinal tablets. From this study, the recovery values of medicinal tablets obtained using Huber, RANSAC, and Theil-Sen Regression methods for caffeine were 96.94%, 97.19%, and 96.16% respectively, while the recoveries of pyridoxine HCl were 122.65%, 104.89%, and 107.48%.

Keywords: UV-Vis Spectrophotometry, Robust regression, Caffeine, Pyridoxine HCl.

Introduction

Robust regression is a useful technique in situations where data contain outliers, which can significantly affect the least squares regression model. Huber regression, RANSAC, and Theil-Sen are commonly used robust regression algorithms. Huber regression assigns less weight to observations identified as outliers, while RANSAC separates training data into inliers and outliers and only uses inliers for the final model. Theil-Sen regression is a nonparametric regression method that makes no assumptions about the distribution of the underlying data. Huber's regression is faster than RANSAC and Theil Sen, but RANSAC and Theil Sen regressors work well for small datasets. RANSAC regression is faster and weights data better than Theil Sen, making it a better choice for datasets with many outliers on the y-axis. Theil-Sen regression, on the other hand, works well on datasets with a moderate number of outliers along the x-axis, but it loses its effectiveness with high-dimensional data [1].

A drug refers to a substance or mixture of compounds that can aid in the treatment and recovery from illness or disease. These drugs are commonly packaged in various forms such as capsules, tablets, or syrups. Tablets, specifically, are drug forms that consist of one or more solidified substances and can include both active substances and excipients (additives). While active ingredients are biologically active and provide the main therapeutic effect, excipients serve as fillers, binders, disintegrating agents, and lubricants in drug tablets. The presence of active ingredients is crucial in medicine as they play an active role in providing benefits for human healing from illnesses or diseases. Common examples of active ingredients in medicinal tablets include caffeine and pyridoxine (B6 vitamin), which are often found in painkillers or analgesics [2], [3].

Caffeine is a substance that has psychoactive effects and acts as a stimulant for the central nervous system, as well as an analgesic and fever reducer. It has also been shown to improve a person's performance [4]. Vitamin B6, also known as Pyridoxine HCl, is a water-soluble vitamin that acts as a

***For correspondence:**
suprapto@chem.its.ac.id

Received: 26 Feb. 2023
Accepted: 27 July 2023

coenzyme in various enzymatic activities in the body [5]. The inclusion of both active ingredients in a drug tablet can enhance the benefits of the drug, as each active ingredient has a distinct function. Moreover, combining multiple active ingredients can lead to a reduction in the number of medicines a patient needs to consume compared to using medicines that contain only one active substance.

Regular quality assessment of medicinal tablets during the production process is crucial. This ensures that the process is consistent, and the material proportions are in line with the listed label. Various methods have been used to analyze the levels of caffeine and pyridoxine HCl in drugs and other pharmaceutical products. Square wave voltammetry, cyclic voltammetry [5], ion chromatography [6], Fourier transform infrared [7], and reversed-phase high-performance liquid chromatography [8] have been reported for measuring caffeine concentration. On the other hand, high-performance liquid chromatography [9], ion-selective electrode potentiometry [10], fluorimetry [11], linear voltammetry [12], square wave voltammetry [13], flow injection solid phase spectrophotometry [14], and capillary zone electrophoresis [15] have been used to determine pyridoxine HCl concentration. Although HPLC has excellent sensitivity and reproducibility, its high cost makes it less practical [16]. Thus, more cost-effective and practical methods with similar capabilities are needed. Spectrophotometry, a chemical analysis method in which the analyte interacts with electromagnetic radiation, is relatively fast and requires only a small number of samples for analysis. It is also selective and sensitive to complex mixtures, enabling accurate and reliable data on the number of impurities in a mixture [17].

Simultaneous multicomponent analysis using spectrophotometry has the advantage of not requiring sample separation prior to analysis. Machine learning algorithms are utilized in data processing as they can handle data that contain interfering components, which may not be significant but can contribute to the spectra. Unlike the usual regression method, which is sensitive to outliers and deviant data, robust regression methods such as Huber, RANSAC regression, and Theil-Sen regression are preferred when analyzing multicomponent samples since they are not significantly affected by outliers. The wavelength range selection is not critical as this method automatically extracts the required information from the main spectra. This method has been applied for the simultaneous analysis of multiple component samples with accurate and reliable results [18].

The aim of this investigation was to utilize a Robust regression approach with a UV-Vis spectrophotometer to simultaneously determine the concentrations of caffeine and pyridoxine HCl in commercial drug tablets. The study involved creating a calibration solution composed of a specific concentration of caffeine and pyridoxine HCl to serve as a model for predicting the concentration. The model's accuracy was then verified using a test solution before being applied to measure the concentration of caffeine and pyridoxine HCl in the drug tablets.

Materials and Methods

A Genesys 10S UV-Visible spectrophotometer was used for this study. The materials used in this study were caffeine (Soho Nootropics) 100%, pyridoxine HCl (Phyto Technology Laboratories) 99%, HCl (Merck), aqua DM, and samples of commercial drug tablets containing caffeine and pyridoxine HCl, 50 mg and 10 mg, respectively.

Preparation of the Caffeine and Pyridoxine HCl Solution

Caffeine powder of 99.2 mg was dissolved in 100 ml HCl 0.1 N solution using a 100 ml volumetric flask. A 100.7 mg of pyridoxine HCl powder was dissolved in 100 mL HCl 0.1 N solutions in a 100 ml volumetric flask. Caffeine stock solution, 1000 mg/L, was diluted to 10; 15; 20; 25; and 30 mg/L in a 25 mL volumetric flask. Each concentration variation was measured for its maximum absorbance using a UV-Vis spectrophotometer at a wavelength of 200-400 nm. The blank used was HCl 0.1N. Pyridoxine HCl 1000 mg/L stock solution was diluted to 4; 6; 8; 10; and 12 mg/L in a 25 mL volumetric flask. Each concentration variation was measured for its maximum absorbance using a UV-Vis spectrophotometer at a wavelength of 200-400 nm. The blank used was an HCl 0.1N solution.

The train solution was prepared by mixing 1000 mg/L caffeine stock solution and 1000 mg/L pyridoxine HCl solution. The two solutions were mixed and diluted with HCl 0.1N solution in a 25 mL volumetric flask. The concentrations variation of the two solutions were caffeine 10, 15, 20, 25, and 30 mg/L and pyridoxine HCl 4, 6, 8,10, and 12 mg/L. The combination of 25 variations in the concentration of both solutions was obtained. The absorbance of each solution was measured using a UV-Vis spectrophotometer at a wavelength of 200-400 nm. HCl 0.1 N solution was used as a blank. The test was carried out in three replications.

Preparation of Caffeine and Pyridoxine HCl Test Solutions

The test solution was prepared by mixing 1000 mg/L caffeine stock solution and 1000 mg/L pyridoxine HCl solution. The two solutions were mixed and diluted with HCl 0.1 N solution in a 25 mL volumetric flask. The concentration variations of the two solutions were caffeine: pyridoxine HCl of 12:5 and 22:9 mg, respectively. The absorbance of each test solution was then measured using a UV-Vis spectrophotometer at a wavelength of 200-400 nm. HCl 0.1 N solution was used as a blank and the measurement was carried out with three replications.

Determination of Caffeine and Pridoxine HCl Content in Drug Samples

A total of 20 drug tablets were weighed and ground into powder using a mortar. Medicinal tablet powders corresponding to the weight of 1 medicinal tablet were weighed. The obtained drug tablet powder was dissolved in a small amount of HCl 0.1 N solution. After dissolving, the drug solution was transferred to a 100 ml volumetric flask. The beaker was then rinsed three times with 10 mL of HCl 0.1 N solution. The obtained solution was sonicated for 10 minutes to accelerate the dissolution. Then, the solution was placed in a Falcon tube and centrifuged at 1000 rpm for 10 minutes to separate the precipitate and the solution. A total of 1 mL of the centrifuged supernatant was pipetted into a 25 mL volumetric flask and diluted with HCl 0.1 N solution. The absorbance of the drug solution was measured using a UV-Vis spectrophotometer at a wavelength of 200-400 nm. HCl 0.1 N solution was used as a blank. The test was carried out in three replications.

Results and Discussion

The Spectra of Caffeine and Pyridoxine HCl

The absorbances of the individual caffeine and pyridoxine HCl train solutions were measured using a Genesys 10S UV spectrophotometer at a wavelength of 200-400 nm. The absorbances of each solution were shown in Figure 1. (a) for pyridoxine HCl, (b) for caffeine, and (c) for caffeine and pyridoxine HCl mixtures. Figure 1(a) shows the spectrum of pyridoxine HCl at concentrations of 4, 6, 8, 10, and 12 mg/l. The maximum wavelength of pyridoxine HCl was 293 nm. According to Aboul-Enein, the wavelength of pure pyridoxine HCl was 291 nm in pure ethanol solvent. The maximum wavelength shift can occur due to the HCl-water used as a solvent in this research [19]. Figure 1(b) shows the single-component spectrum of caffeine at concentrations of 10, 15, 20, 25, and 30 mg/l. The maximum wavelength of caffeine was 287 nm. The wavelength of pure caffeine in water was 273 nm according to Atomssa [20]. The differences in the value of the maximum wavelength could be due to the use of HCL 0.1 N as solvents.

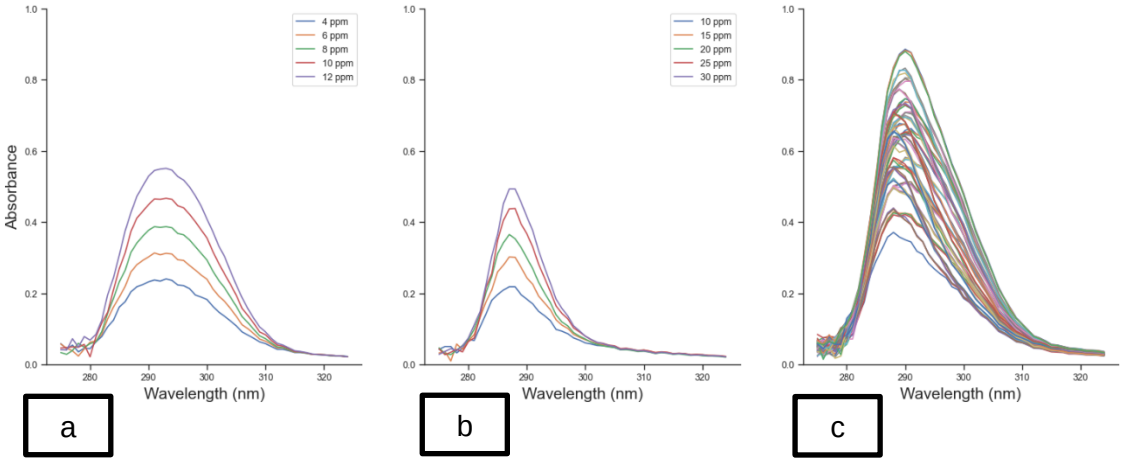


Figure 1. The individual spectrum of pyridoxine HCl (a), caffeine (b), and the concentration combination spectra of a mixture containing caffeine at concentrations of 10, 15, 20, 25, and 30 mg/L, and pyridoxine HCl at concentrations of 4, 6, 8, 10, and 12 mg/L (c).

The spectrum in Figure 1(c) illustrates the overlapping spectra of caffeine and pyridoxine HCl, which vary based on their concentrations. The fluctuations in the spectra of caffeine and pyridoxine HCl are correlated with their compositions. When combined, the absorbance of the mixed spectrum is higher

than that of the individual caffeine and pyridoxine HCl spectra. This is due to the additive nature of each component, resulting in increased absorbance at certain wavelengths. Figure 2 shows the chemical structures and topological similarity of caffeine and pyridoxine, calculated using RDKit in Python. The green substructures in Figure 2(c) indicate similarity, while the pink substructures indicate dissimilarity.

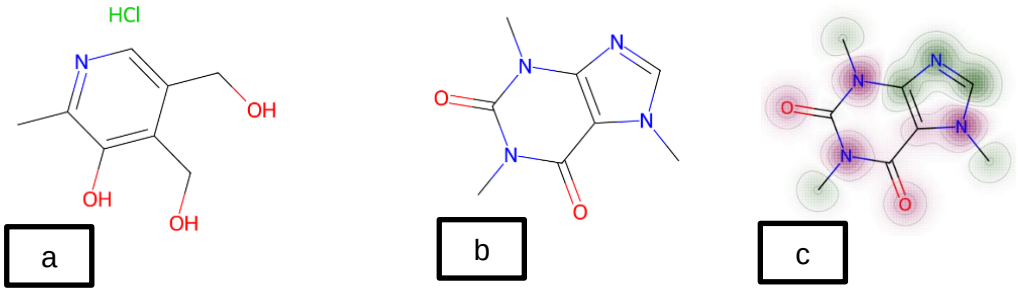


Figure 2. The chemical structure of pyridoxine HCl (a), caffeine (b), and similarity maps between pyridoxine and caffeine (c)

Figure 3 presents the UV-Vis spectra of caffeine and pyridoxine HCl in water, calculated using ORCA DFT with the B3LYP functional and def2-SVP basis set. The individual spectra of caffeine and pyridoxine are shown in Figures 3(a) and (b), respectively. Caffeine exhibits two peaks at 256 nm and 216 nm, while pyridoxine displays three peaks at 251 nm, 245 nm, and 224 nm. The overlapping spectra of caffeine and pyridoxine are presented in Figure 1(c), without peak broadening, and in Figure 1(d), with peak broadening. Although the wavelengths from the ORCA calculation were higher than those from the experimental values, the spectra in Figures 1(c) and 3(d) have a similar pattern.

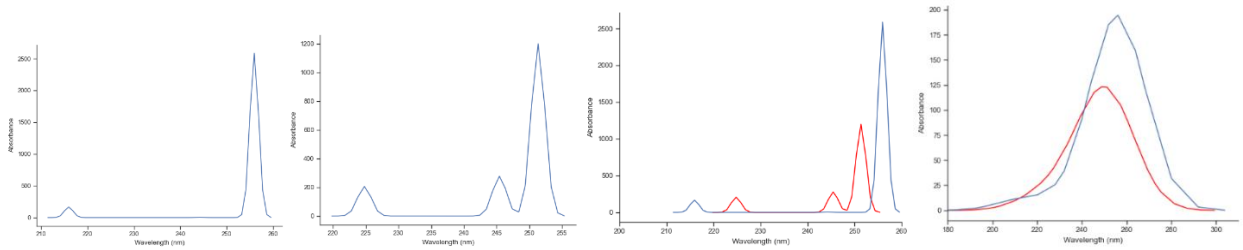


Figure 3. The UV-Vis spectra of caffeine (a), pyridoxine (b), caffeine-pyridoxine overlay (c), and caffeine-pyridoxine overlay with peak broadening (d)

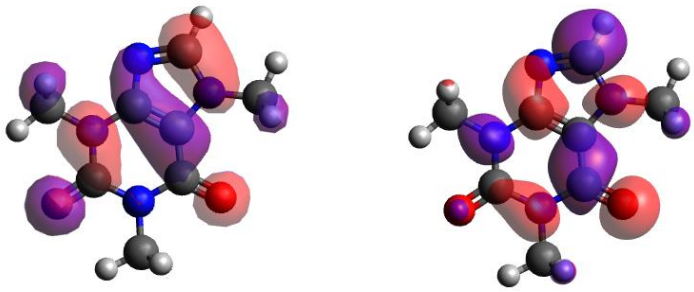


Figure 4. The caffeine electron density at HOMO (a) and LUMO (b) orbital for electronic transition at 256 nm

The caffeine peak at 256 nm comes from the electronic transition of caffeine in the ground state (Figure 4(a)) to the excited states with electron density shown in Figure 4(b). The caffeine HOMO energy level was -5.989 eV while the LUMO energy was -0.873 eV. The electron transition at 216 nm peak comes from 2 prominent HOMO and LUMO transitions as shown in Figure 5. The first transition HOMO-LUMO energy was from -7.445 eV to -0.873 nm. The second electronic transition was from HOMO energy of -5.989 eV to LUMO energy of 0.424 eV, from left (HOMO) to right (LUMO).

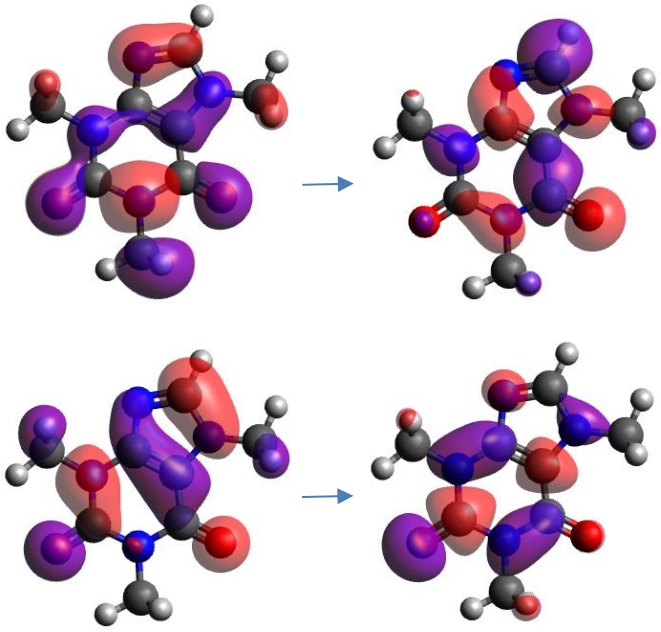


Figure 5. The caffeine electron density of the HOMO-LUMO transition at 216 nm

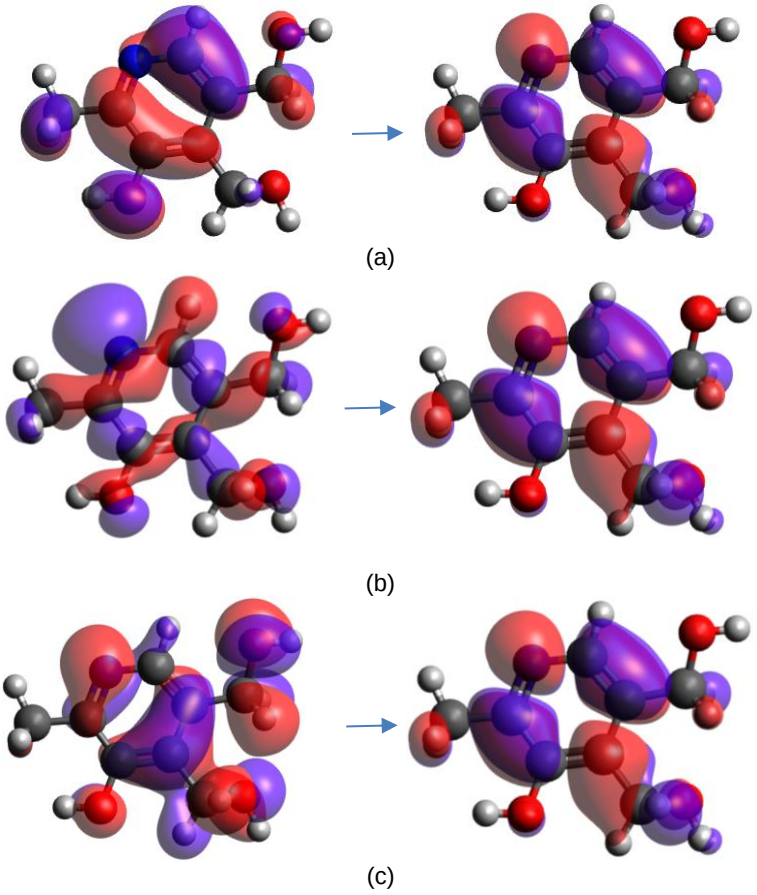
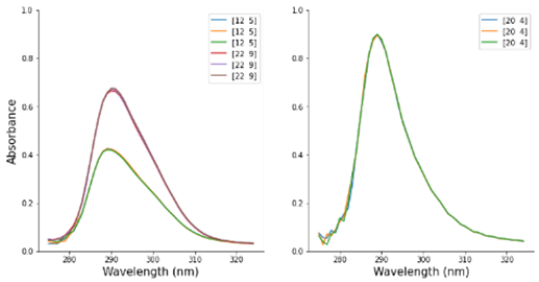


Figure 6. Pyridoxine electron transition at 251 nm (a) 245 nm (b), and 224 nm (c)

Pyridoxine peak at 251 nm comes from the electronic transition of HOMO energy of -6.238 eV to LUMO energy of -0.747 eV. The HOMO to LUMO transition from -7.034 eV to -0.747 eV produced a peak at 245 nm. The peak at 224 nm was an electronic transition from -6.858 eV to -0.747 eV. The electron density at HOMO and LUMO states were shown in Figure 6.

Multi-component Analysis of Caffeine and Pyridoxine HCl in the Test Solution and Pharmaceutical Drug

The test solution contains caffeine and pyridoxine HCl concentration combination of 12:5 and 22:9 mg/L, respectively. The absorbance of the test solution was shown in Figure 7(a). The spectrum of the pharmaceutical drug was shown in Figure 7(b). From Figure 7(b), its looks that the caffeine content in the drug was higher than in the test solution.



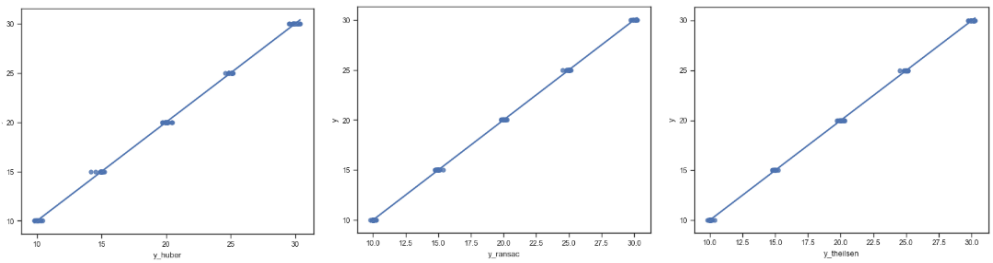
(a) (b)

Figure 7. Spectrum of pyridoxine HCl and caffeine test solutions (a) and pharmaceutical drug (b)

The maximum wavelength of the 25 concentration variations was different but was still in the range between 287-293 nm, which was the maximum wavelength range of caffeine and pyridoxine HCl. The higher the concentration of caffeine, the closer the maximum wavelength was to the 287 nm wavelength. On the other hand, the higher the concentration of pyridoxine HCl, the maximum wavelength was shifted towards the value of 293 nm.

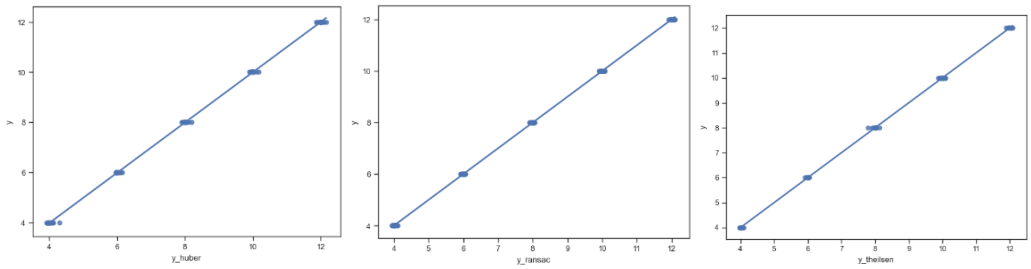
The test solution measurement was carried out to validate the regression models. The test solutions were prepared with a concentration variation that was not applied in training or calibration solutions. This was done to avoid leaking in the models. The success of the regression models was evaluated by the closeness of the predicted concentration compared to the test solution labels. These closeness features can be observed from the recovery value. The recovery value was calculated from the percentage ratio of the predicted value and experimental value.

The regression curves of each regression model for caffeine and pyridoxine were shown in Figure 8 and Figure 9. From Figure 8 and Figure 9, it was shown that all the training datasets prediction falls in the linear regression lines both for caffeine and pyridoxine. The r-values, root mean squared error (RMSE), and recoveries for caffeine and pyridoxine training data were shown in Figure 10 and Figure 11, respectively. The RMSE of Huber regression models was higher than RANSAC and Theil-Sen. No significant difference was observed for r-values and the recoveries between the regression models.



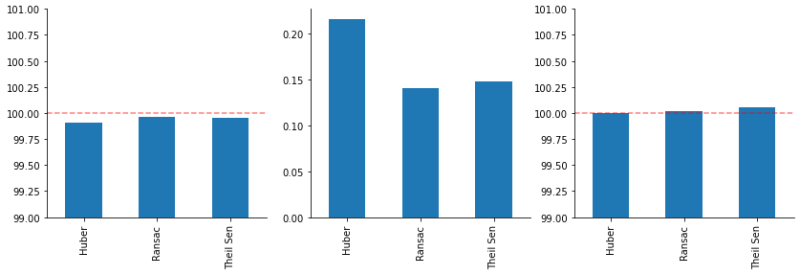
(a) (b) (c)

Figure 8. Regression curves of Huber (a), RANSAC (b), and Theil-Sen (c) Regression of caffeine training and predicted data



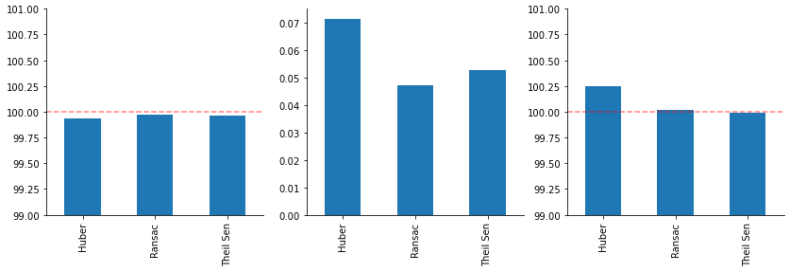
(a) (b) (c)

Figure 9. Regression coefficient of Huber (a), RANSAC (b), and Theil-Sen (c) Regression of pyridoxine training data



(a) (b) (c)

Figure 10. Correlation coefficient (a), RMSE (b), and recovery of caffeine train datasets

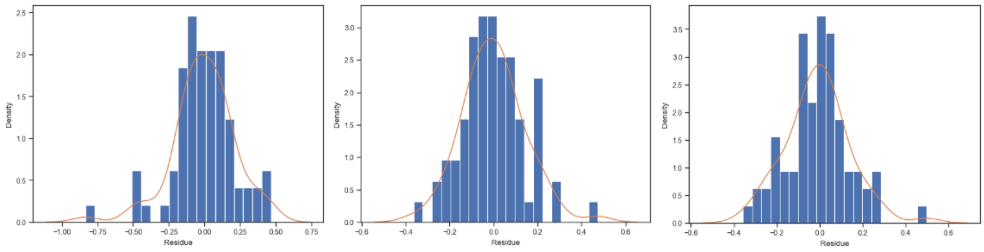


(a) (b) (c)

Figure 11. Correlation coefficient (a), RMSE (b), and recovery (c) of pyridoxine train datasets

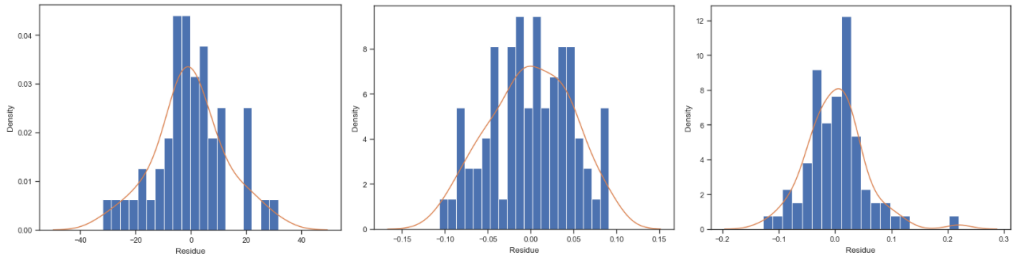
The residues of the regression models were shown in Figure 12 for caffeine and Figure 13 for pyridoxine. The p-values of the Anderson-Darling normality test for the caffeine and pyridoxine residue distributions were 0.75 for all the models with different statistical values, indicated that the residues were normally distributed.

The weight of absorbance in each wavelength was different between each regression model, but the weight of coefficients for RANSAC and Theil-Sen regression have slightly similar patterns both for caffeine and pyridoxine HCl, as shown in Figure 14 and Figure 15. RANSAC and Theil-Sen give a smaller weight to absorbance at a lower wavelength, both for caffeine and pyridoxine. Huber distributed the weight of the absorbance either for lower wavelength as well as higher wavelength.



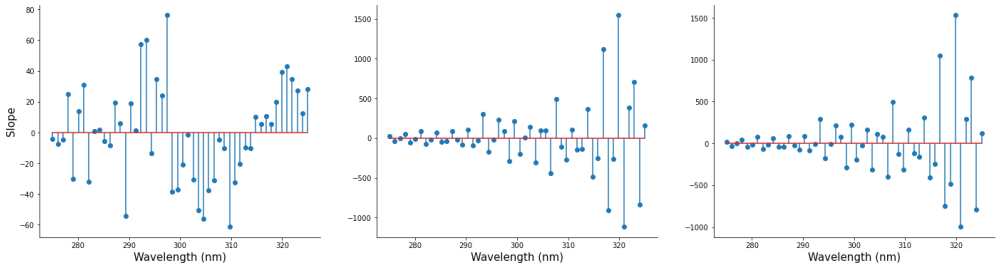
(a) (b) (c)

Figure 12. The residue of Huber (a), RANSAC (b), and Theil-Sen (c) Regression of caffeine training data



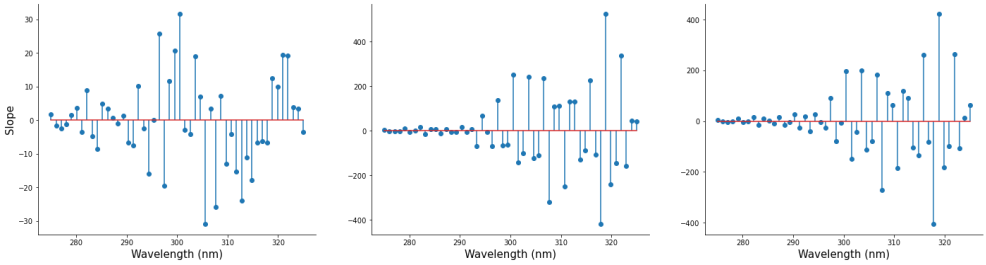
(a) **(b)** **(c)**

Figure 13. The residue of Huber (a), RANSAC (b), and Theil-Sen (c) Regression of pyridoxine training data



(a) (b) (c)

Figure 14. Regression coefficient of Huber (a), RANSAC (b), and Theil-Sen (c) Regression of caffeine training datasets



(a) (b) (c)

Figure 15. Regression coefficient of Huber (a), RANSAC (b), and Theil-Sen (c) Regression of pyridoxine training datasets

The recoveries of the datasets were calculated as the ratio between predicted and experimental values. The boxplots for both caffeine and pyridoxine HCl recoveries in the training datasets showed values close to 100%, indicating a good fit of the regression model in correlating the input variables and response variables. However, there were several outliers present in each regression model, as observed in Figure 16.

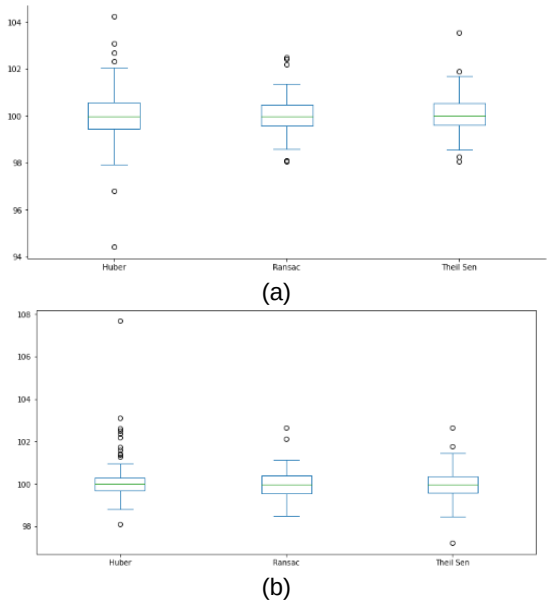


Figure 16. Box plot of model recovery for caffeine (a) and pyridoxine (b)

Analysis of the Test Solutions

The model validation using a known caffeine and pyridoxine test solution indicated that Huber regression has lower performance compared to both RANSAC and Theil-Sen, as shown in Figure 17 for caffeine and Figure 18 for pyridoxine. The bar plot of the predicted and experiment caffeine and pyridoxine ratio was shown in Figure 19 for caffeine and Figure 20 for pyridoxine.

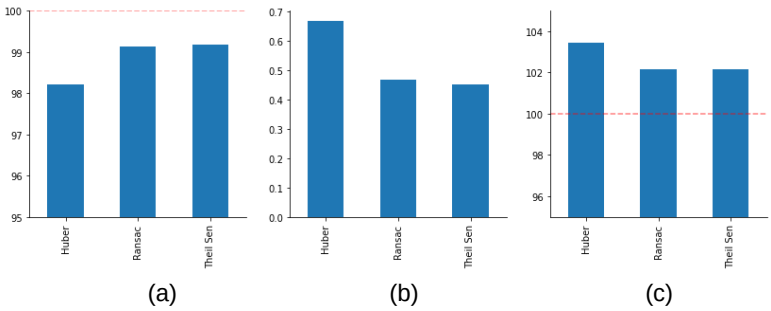


Figure 17. Correlation coefficient (a), RMSE (b), and recovery (c) of caffeine test solution

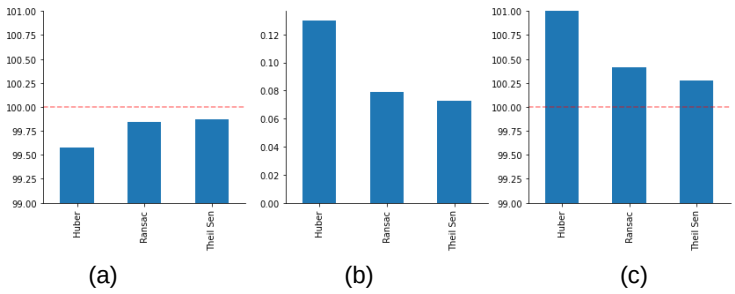
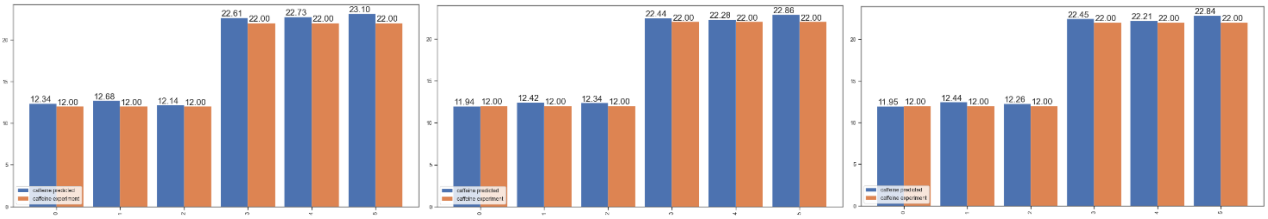
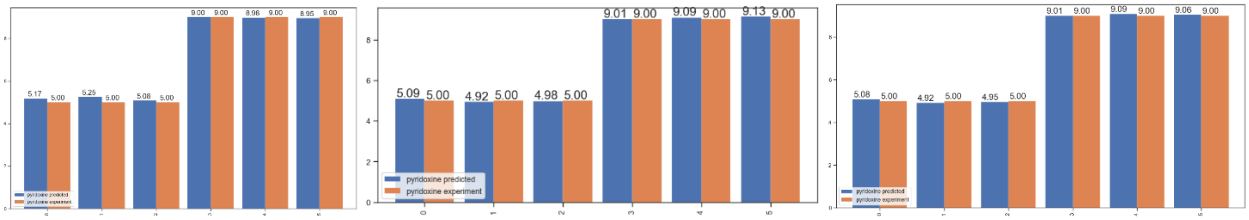


Figure 18. Correlation coefficient (a), RMSE (b), and recovery (c) of pyridoxine test solution



(a) (b) (c)

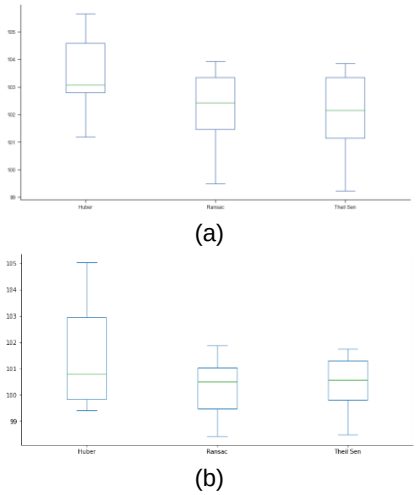
Figure 19. Prediction of caffeine test solution using Huber (a), RANSAC (b), and Theil-Sen (c)



(a) (b) (c)

Figure 20. Prediction of pyridoxine test solution using Huber (a), RANSAC (b), and Theil-Sen (c)

The box plot for caffeine recovery, as shown in Figure 21(a), revealed that the distribution of the recovery in the test solution was close to a normal distribution for RANSAC and Theil-Sen regression models. However, the Huber regression model showed a skew towards higher values. It is important to note that the box plot may not provide detailed information due to the relatively small size of the data.



(a)

(b)

Figure 21. Model recovery box plot of (a) caffeine (b) pyridoxine test solution

Individual comparison between experimental concentration and regression models' prediction was tested using the Wilcoxon rank test. The p values for predicted and experimental data for Huber were 0.03125, RANSAC was 0.0625, and Theil-Sen was 0.0625. Thus, the predicted and experimental values for Huber regression were significantly different.

The pyridoxine experimental and prediction values comparison using the Wilcoxon rank test produces p values 0.4375, 0.3125, and 0.3125 for Huber, RANSAC, and Theil-Sen, respectively. Thus, no significant difference between pyridoxine HCl experimental values and all the regression predicted values.

RANSAC regression was able to overcome large data sets that contain outliers. The correlation

coefficient of the RANSAC regression method shows a value of 0.99 in predicting the concentration of caffeine and pyridoxine HCl, so it can be said that this method has good linearity. Method validation is required to ensure the correctness of the chosen method in predicting the concentration. The validation of the RANSAC regression method was performed with two test solution with different caffeine and pyridoxine HCl ratio. In test solution 1, the mean recovery values of caffeine are known to be 101.96% and 99.98% for pyridoxine HCl. In test solution 2, the mean recovery values of caffeine are known to be 102.40 and 100.55% for pyridoxine HCl. The mean recoveries of test solution 1 and test solution 2 meet the accuracy requirements, so it can be concluded that the RANSAC regression method has good accuracy in predicting the concentration of test solution 1 and test solution 2.

The Theil-Sen regression method employs a robust line fitting approach to estimate the relationship between paired data points in a plane. It calculates the median slope from all the lines connecting these points [21]. In our study, this regression method was used to establish the correlation between the calculated concentration and the experimental concentration (obtained by predicting the concentration of the calibration solution using Theil-Sen regression).

For test solution 1, the mean recovery values for caffeine were found to be 101.86%, and for pyridoxine HCl, the mean recovery values were 99.84%. Similarly, for test solution 2, the mean recovery values for caffeine were determined to be 102.12%, while for pyridoxine HCl, the mean recovery values were 100.67%. As the mean recovery values for both test solution 1 and test solution 2 meet the required criteria, it can be concluded that the Theil-Sen regression method exhibits good accuracy in predicting the concentrations of these test solutions.

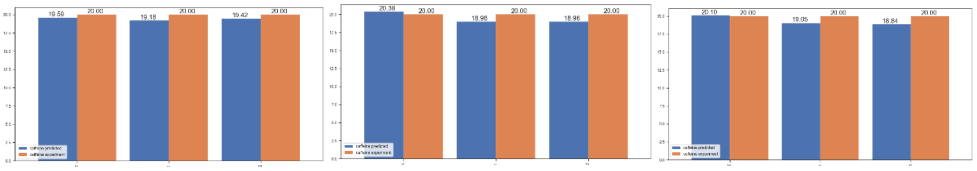
Pharmaceutical Tablets Analysis

In this study, the brand “N” drug tablet containing 50 and 10 mg of caffeine and pyridoxine HCl indicated on the label was used as a sample. A total of 20 drug tablets were weighed individually using an analytical balance. The mean mass of drug tablets was calculated as in Equation (1).

$$\text{Mean of Mass} = \frac{13{,}2445}{20} = 0{,}6622 \text{ g} \qquad (1)$$

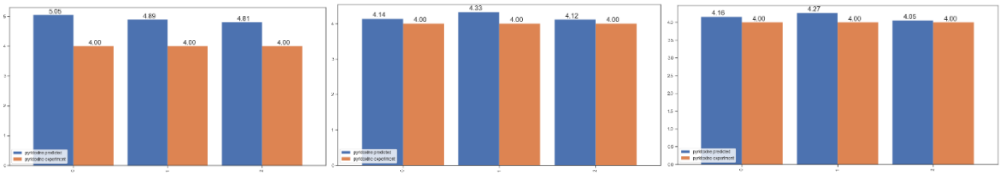
The comparison of predicted and experimental ratios of caffeine and pyridoxine is presented in Figure 22 for caffeine and Figure 23 for pyridoxine. It can be observed that the Huber regression method resulted in higher errors for pyridoxine prediction compared to the other two methods. In terms of caffeine analysis in pharmaceutical drugs, the RANSAC method exhibited higher errors compared to the other two methods. Among the three regression methods, Huber regression demonstrated the lowest error rates with a moderate recovery. On the other hand, Theil-Sen regression exhibited the lowest recovery rate and the highest root mean square error (RMSE), as depicted in Figure 24 and Figure 25.

In the case of pyridoxine measurement, Huber regression displayed higher error compared to the other two methods. Furthermore, the recovery of Huber pyridoxine prediction exceeded the permitted threshold of 90% - 110%.



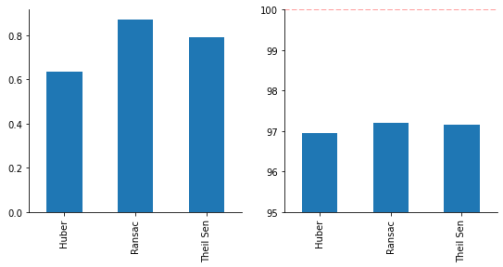
(a) (b) (c)

Figure 22. Comparison of RMSE (a) and recovery (b) for caffeine in the pharmaceutical drug



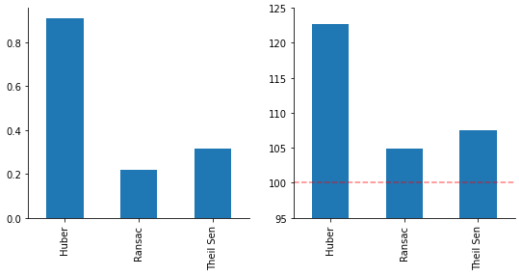
(a) (b) (c)

Figure 23. Comparison of RMSE (a) and recovery (b) for pyridoxine in the pharmaceutical drug



(a) (b)

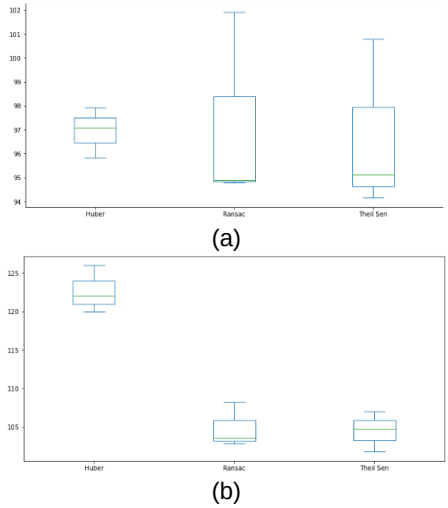
Figure 24. Comparison of RMSE (a) and recovery (b) for caffeine in the pharmaceutical drug



(a) (b)

Figure 25. Comparison of RMSE (a) and recovery (b) for pyridoxine in the pharmaceutical drug

While the Huber regression model provided caffeine predictions in pharmaceutical tablets that were close to 100%, the predictions for pyridoxine were beyond the control limits. This suggests that the Huber regression model may not be suitable for accurately predicting pyridoxine in pharmaceutical tablets. It is worth noting that the Huber regression model performs well when the number of analyzed features is lower than the number of datasets. The box plot depicting the predicted caffeine and pyridoxine values in pharmaceutical tablets can be seen in Figure 26.



(a)

(b)

Figure 26. Recovery boxplot for (a) caffeine and (b) pyridoxine in the pharmaceutical drug

The average recovery of caffeine in pharmaceutical tablets was determined to be 97.19%, while for Pyridoxine HCl it was 104.89%. The average concentration of caffeine in the drug was found to be 48.5955 mg, and for pyridoxine HCl, it was 10.5225 mg. These values closely align with the labeled concentrations of caffeine (50 mg) and pyridoxine HCl (10 mg), indicating that the RANSAC regression method is suitable for predicting drug concentrations.

Similarly, the mean recovery values for caffeine and pyridoxine HCl were determined to be 99.77% and 103.69%, respectively. The average concentration of caffeine in the drug was 49.8841 mg, and for pyridoxine HCl, it was 10.3688 mg. These values are in close agreement with the labeled concentrations of caffeine (50 mg) and pyridoxine HCl (10 mg), suggesting that the Theil-Sen regression method is effective for predicting drug concentrations.

After performing calculations using the three regression procedures, it was observed that the Theil-Sen regression method exhibited the ability to predict drug concentrations that closely matched the values reported on the label. This can be attributed to the fact that the Theil-Sen regression method generates estimation models based on the dimensional combination of the items in the dataset. Additionally, this method calculates the spatial median of all parameters within each model. As a result, the Theil-Sen regression method produces more robust data compared to other regression methods, making it particularly suitable for datasets with substantial issues. These findings align with previous research [6].

Conclusions

Based on the research conducted, it can be concluded that the UV-Vis spectrophotometric method with multivariate regression analysis i.e., Huber, RANSAC, and Theil-Sen regression can predict the concentrations of caffeine and pyridoxine HCl drug tablets simultaneously. The recovery on drug tablets using RANSAC regression, and Theil-Sen regression for caffeine were 97.19, 97.19, and 99.77%, and for pyridoxine HCl were104.89, 104, 89, and 103.69%. Prediction of caffeine and pyridoxine HCl content in drug tablets according to the levels on the label, caffeine was 48.5955 mg using RANSAC regression and 49.8841 mg using Theil-Sen regression. Pyridoxine HCl content was 10.5525 mg using RANSAC regression and 10.3688 mg using Theil-Sen regression. The predicted levels of caffeine and pyridoxine HCl were close to those reported on the drug label.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

The authors gratefully acknowledge financial support from the Institut Teknologi Sepuluh Nopember for this work, under the project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2023.

References

[1] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondelt, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*. *12*(85), 2825-2830.

[2] Leacock, R. E., Stankus, J. J., and Davis, J. M. (2011). Simultaneous determination of caffeine and vitamin B6 in energy drinks by high-performance liquid chromatography (HPLC). *J. Chem. Educ*., *88*(2), 232-234.

[3] Ruxton, C. H. S. (2008). The impact of caffeine on mood, cognitive function, performance, and hydration: A review of benefits and risks. *Nutrition Bulletin*, *33*(1), 15-25.

[4] Bender, D. A. (1999). Non-nutritional uses of vitamin B6. *Br J Nutr*., *81*(1), 7-20.

[5] Yiğit, A., Yardım, Y., Çelebi, M., Levent, A., and Şentürk, Z. (2016). Graphene/Nafion composite film modified glassy carbon electrode for simultaneous determination of paracetamol, aspirin, and caffeine in pharmaceutical formulations. *Talanta*, *158*, 21-29.

[6] Chen, Q., Mou, S., Hou, X., and Ni, Z. (1998). Simultaneous determination of caffeine, theobromine, and theophylline in foods and pharmaceutical preparations by using ion chromatography. *Analytica Chimica Acta*, *371*(2), 287-296.

[7] Nugrahani, I., Manosa, E. Y., and Chintya, L. (2019). FTIR-derivative as a green method for simultaneous content determination of caffeine, paracetamol, and acetosal in a tablet compared to HPLC. *Vibrational Spectroscopy*, *104*, 102941.

[8] Martín, M. J., Pablos, F., and González, A. G. (1999). Simultaneous determination of caffeine and non-steroidal anti-inflammatory drugs in pharmaceutical formulations and blood plasma by reversed-phase HPLC from linear gradient elution. *Talanta*, *49*(2), 453-459.

[9] Dinç, E., Kökdil, G., and Onur, F. (2000). A comparison of matrix resolution method, ratio spectra derivative spectrophotometry and HPLC method for the determination of thiamine HCl and pyridoxine HCl in pharmaceutical preparation. *Journal of Pharmaceutical and Biomedical Analysis*, *22*(6), 915-923.

[10] Vaze, V. D. and Srivastava, A. K. (2008). Determination of pyridoxine hydrochloride in pharmaceutical

preparations by calixarene based potentiometric sensor. *Journal of Pharmaceutical and Biomedical Analysis*, *47*(1), 177-182.

[11] Escandar, G. M., Bystol, A. J., and Campiglia, A. D. (2002). Spectrofluorimetric method for the determination of piroxicam and pyridoxine. *Analytica Chimica Acta*, *466*(2), 275-283.

[12] Qu, W., Wu, K., and Hu, S. (2004). Voltammetric determination of pyridoxine (Vitamin B6) by use of a chemically-modified glassy carbon electrode. *Journal of Pharmaceutical and Biomedical Analysis*, *36*(3), 631-635.

[13] Hernández, S. R., Ribero, G. G., and Goicoechea, H. C. (2003). Enhanced application of square wave voltammetry with glassy carbon electrode coupled to multivariate calibration tools for the determination of B6 and B12 vitamins in pharmaceutical preparations. *Talanta*, *61*(5), 743-753.

[14] Portela, J. G., Costa, A. C. S., and Teixeira, L. S. G. (2004). Determination of Vitamin B6 in pharmaceutical formulations by flow injection-solid phase spectrophotometry. *Journal of Pharmaceutical and Biomedical Analysis*, *34*(3), 543-549.

[15] Huopalahti R., and Sunell J. (1993). Use of capillary zone electrophoresis in the determination of B vitamins in pharmaceutical products, *Journal of Chromatography A*, *636*(1), 133-135. Doi: 10.1016/0021-9673(93)80065-G.

[16] McMaster, M. C. (1994). HPLC : a practical user's guide / Marvin C. McMaster, Universitas Indonesia Library, 1994. https://lib.ui.ac.id.

[17] Kaflé, B. P. (2020). *Chemical analysis and material characterization by spectrophotometry*. Elsevier.

[18] Lavagnini, I., Badocco, D., Pastore, P., and Magno, F. (2011). Theil–Sen nonparametric regression technique on univariate calibration, inverse regression, and detection limits. *Talanta*, *87*, 180-188.

[19] Aboul-Enein, H. Y., and Loutfy, M. A. (1984). Pyridoxine hydrochloride. *Analytical Profiles of Drug Substances*. Academic Press.

[20] Atomssa T., and Gholap A. V. (2011). Characterization of caffeine and determination of caffeine in tea leaves using UV-visible spectrometer. *African Journal of Pure and Applied Chemistry*, *5*(1), 1-8.

[21] Chervenkov, H., and Slavov, K. (2019). Theil–Sen estimator vs. Ordinary least squares — Trend analysis for selected ETCCDI climate indices. *Comptes rendus de l'Académie bulgare des sciences: sciences mathématiques et naturelles*, *72*, 47-54.

[22] Rousseeuw, P. J., and Leroy, A. M. (1987). *Robust regression and outlier detection*. John Wiley & Sons.