SCIENCE AND TECHNOLOGY RESEARCH CENTRE
UNIVERSITAS PGRI SEMARANG - INDONESIA

ISSN: 2715-4211

Journal ASSET

Advance Sustainable Science, Engineering and Technology

Volume 3 Number 1 April 2021

URL LINK: http://journal.upgris.ac.id/index.php/asset

Advance Sustainable Science, Engineering and Technology

e-ISSN: 2715-4211

Volume 3 Number 1 April 2021

Science and Technology Research Centre, Universitas PGRI Semarang, Indonesia

Advance Sustainable Science, Engineering and Technology (ASSET) is a peer-reviewed open-access international scientific journal dedicated to the latest advancements in sciences, applied sciences and engineering, as well as relating sustainable technology. This journal aims to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas of sciences, engineering, and technology.

ASSET papers will be published twice a year by Universitas PGRI Semarang, Central Java, Indonesia.

Advance Sustainable Science, Engineering and Technology
e-ISSN: 2715-4211
Volume 3 Number 1 April 2021
Science and Technology Research Centre, Universitas PGRI Semarang, Indonesia

Editorial Preface

Advance Sustainable Science, Engineering and Technology (ASSET)
Volume 3 Number 1 April 2021

It is our great pleasure to present the Volume 3 Number 1 Advance Sustainable Science, Environmental Engineering and Technology (ASSET). This issue includes eight manuscripts. Utari's article: *Artificial Neural Network for Classifying Injected Materials under Ultrasonography*. They have constructed an artificial neural network (ANN) architecture to classify four different classes of ultrasonography recorded from a jelly box phantom that was injected by iron, glass, or plastic marble, or without any injection. Mikuriya's article: *Synthesis, Crystal Structures, Electronic Spectra, and Magnetic Properties of Thiolato-Bridged Trinuclear Cobalt(II) Complexes with N, N, S-Tridentate Thiolate Ligands*. They synthesized and characterized a series of thiolato-bridged trinuclear cobalt(II) complexes by the use of two kinds of tridentate thiolate ligands, 2-[(3-aminopropyl)amino]ethanethiol and 1-[(3-aminopropyl)amino]-2-methylpropane-2-thiol. Tijing's article: *3D-Printed Absorbers for Solar-Driven Interfacial Water Evaporation: A Mini-Review*. This study hopes to provide more insights into the use of additive manufacturing for improving the absorber design and performance of SWE. Margono's article*: Current and Voltage Imbalances at BPTIK Universitas PGRI Semarang*. This research conducted an energy audit in the BPTIK (Badan Pengembangan Teknologi Informasi dan Komunikasi) department room at the Universitas PGRI Semarang, observing the value of electrical parameters, especially unbalanced voltage and current that have the potential to cause neutral currents. Islami's article: *Application Biplot and K-Medians Clustering to Group Export Destination Countries of Indonesia's Product*. this research applies K-Medians Clustering Method using Manhattan Distance to resolve outliers while grouping the countries based on their export data. Anwar's article: *Automatic Complaints Categorization Using Random Forest and Gradient Boosting*. This project aims to utilize Data Mining to automatize complaints categorization. Sihombing's article: *Implementation of K-Means and K-Medians Clustering in Several Countries Based on Global Innovation Index (GII) 2018*. The result clustering with using k-means method and k-medians method showed that k-medians is better than k-means method because the variance value of k-medians is smaller than k-means. The final article, Raulnadi's article: *"Branket" Design as a Safe Deposit Box Security System using Arduino-Based Tap Sensor*. "Branket" is expected to be a new innovation in the locking system in a unique safe with a high level of security.

We thank all of the 26 authors affiliated from Kwansei Gakuin University, Japan; Yamagata University, Japan; Shimane University, Japan; University of Technology Sydney, Australia; Bataan Peninsula State University, Philippines; Indian Institute of Technology, India; Universitas Bali Internasional, Indonesia; Universitas Bina Insan, Indonesia; Badan Pusat Statistik, Jakarta, Indonesia; Politeknik STMI Jakarta, Indonesia; Universitas Kristen Satya Wacana, Indonesia; and Universitas PGRI Semarang, Indonesia who have contributed to this issue.

Advance Sustainable Science, Engineering and Technology
e-ISSN: 2715-4211
Volume 3 Number 1 April 2021
Science and Technology Research Centre, Universitas PGRI Semarang, Indonesia

ASSET is now indexed with GARUDA, Google Scholar, NELITI, PKP INDEX, CROSSREF, and DIMENSIONS. We expect that ASSET will soon be indexed with SINTA and SCOPUS which will provide a higher platform for the authors and the readers, with a comprehensive overview on the latest advancements in sciences, applied sciences and engineering, as well as relating sustainable technology.

April 2021
Asst. Prof. Mega Novita
Asst. Prof. Rizky Muliani Dwi Ujianti

Advance Sustainable Science, Engineering and Technology
e-ISSN: 2715-4211
Volume 3 Number 1 April 2021
Science and Technology Research Centre, Universitas PGRI Semarang, Indonesia

Table of Content

Advance Sustainable Science, Environmental Engineering and Technology (ASSET)
Volume 3 Number 1 April 2021

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210101-01 ~ 0210101-07
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8324

Artificial Neural Network for Classifying Injected Materials under Ultrasonography

Galuh Retno Utari[1], Giner Maslebu[1], Suryasatriya Trihandaru[2*]

[1] Physics Study Program, Faculty of Science and Mathematics, Universitas Kristen Satya Wacana, Jl. Diponegoro 52-60 Salatiga 50714 Jawa Tengah Indonesia

[2] Magister Data Science, Faculty of Science and Mathematics, Universitas Kristen Satya Wacana, Jl. Diponegoro 52-60 Salatiga 50714 Jawa Tengah Indonesia

*suryasatriya@uksw.edu

Abstract. We have constructed an artificial neural network (ANN) architecture to classify four different classes of ultrasonography recorded from a jelly box phantom that was injected by iron, glass, or plastic marble, or without any injection. This jelly box was made as a phantom of a human body, and the injected materials were the cancers. The small size of the injected materials caused only little disturbances those could not easily distinguished by human eyes. Therefore, ANN was used for classifying the different kind of the injected materials. The number of original imagestaken from ultrasonographs were not so many, therefore we did data augmentation for providing large enough dataset that fed into ANN. The data augmentation was constructed by pixel shifting in horizontal and vertical directions. The procedure proposed here produced 98.2% accuracy for predicting test dataset, though the result was sensitive to the choice of augmentation area.

Keywords: Artificial Neural Network, Tensorflow, Multiclass Classification

(Received 2021-03-26, Accepted 2021-04-30, Available Online by 2021-04-30)

1. Introduction

In this article, we describe the use of machine learning, especially the artificial neural network (ANN), in classifying different injected objects in a phantom by analyzing their ultrasonography images. The phantom was made from jelly, a kind of translucent food. The injected materials were iron, glass and plastic marbles. Due to the small size of the objects, the resulting images were not easily classified by human eyes, therefore any computational method could help to solve the problem. Here, we used ANN to make prediction for classifying an unknown injected material. The method has given good results, i.e. it can give more than 95% accuracy for the test datasets. It should be very useful for predicting the density class of any injected material outside the database used in this research by giving appropriate assumption.

1.1. Neural Network Research Based Review

Neural network has been used widely in different areas and it is now still growing fast as a powerful learning machine to use in artificial intelligence as reported by Lecun [1]. For some medical or related areas, here we give some examples. Moein et.al [2] reported several application of neural network in medical diagnosis, for examples are cancer, hepatitis, heart diagnosis, pattern recognition and also in drug development. Lee et.al. [3] uses ANN to predict prostate cancer using transrectal ultrasonography, using 684 patients database. Lee has compared two difference ANN methods depending on the available additional data. For more complicated medical imaging, Convolutional Neural Network (CNN) is used instead of ANN. For example, Wu et.al. [4] used CNN for analyzing brain tumors with MRI dataset. Zhang [5] also used ANN and CNN to classify breast cancers through mammograms. Dealing with unstructured radiology reports, Pandey [6] used a specific CNN method, i.e. Natural language processing (NLP) approach to extract clinical terms. Zhou [7] used CNN for identifying capsule defects. Almezhghwi [8] used a branch of CNN, i.e. GAN, to classify four difference blood cells.

Our analysis was based on the ANN given by Tensorflow module. The basic procedures are given in many papers, for example in [9]. When we did this research, we had only small size database. In order to have high accuracy, we made some experiments in creating simple data augmentation before running ANN. Actually, data augmentation procedure is also provided by Tensorflow, and often used by researchers, for example is the work of Dong et.al. [10].

1.2. Mathematical Background of ANN

ANN is a method for calculating the optimal parameters used for fitting data target into a (quasi) linear model. The simplest one of classification procedures is to connect between a single input vector, ***x***, to a output vector, ***y***, through a composition of an activation function and a linear equation:

$$\mathbf{y} = \sigma(\mathbf{x} \cdot \mathbf{w} + \mathbf{b}) \tag{1}$$

where σ is the activation function, **w** and **b** are weights and bias matrix respectively. When we add one hidden layer between input and output, we add also one more composition function as

$$\mathbf{y} = \sigma_2(\sigma_1(\mathbf{x} \cdot \mathbf{w}_1 + \mathbf{b}_1) \cdot \mathbf{w}_2 + \mathbf{b}_2) \tag{2}$$

Therefore, a long ANN (usually known as deep learning) will have long form of composition function too, because it has many hidden layers.

The activation function can be chosen differently for different layers [11]. In this research, we use only two different types, i.e.

$$\sigma_h(z) = \max(0, z) \tag{3}$$

And

$$\sigma_f(z)_i = \exp(z_i) / \sum_{k=1}^{K} \exp(z_k)) \tag{4}$$

Here, σ_h is called by ReLU (rectified linear unit) or threshold function [12] and used for all hidden layers and σ_f is called by Softmax [13] and used only for the final (or output) layer that produces *K* classes.

The goal of ANN is to calculate the best weights **w** and bias **b** (in all layers) that minimize the cross-entropy loss function defined as [14]

$$L(\mathbf{w},\mathbf{b}) = \sum_{i=1}^{K} - y_i \log \hat{y}_i(\mathbf{w},\mathbf{b}) \quad (5)$$

Here, $\hat{y}$ is the function that has to be calculated such that it becomes very close to the data target y. The minimization uses Adaptive Moment Estimation (Adam) method [15].

2. Methods

There were two steps of works. First, the preparation of sample and data acquisition using 2 dimensional USG Mindray DP-10; second, the data augmentation and ANN architecture.

2.1. Preparation and Data Acquisition

First, we made a phantom that simulate a human body using jelly. Jelly is a material for making food, it has clear or translucent color. A soft-solid box of jelly, with size 15cm × 15 cm × 10 cm, was easily created by cooking 40-gram jelly powder in 2.25 liter water. In room temperature, it became a soft-solid form. We cut the jelly box at the center in such a way that a marble can be placed into the center or taken away from it. USG Mindray DP-10 was used to scan the jelly box with a marble inside it. There were three different marbles, and each marble was scanned 16 times with variations of the position of USG convex transducer 35C50EB. The data acquisition was situated as shown in Figure 1. Without marble, the jelly box was also scanned 16 times. The physical properties of the injected marbles are given in Table 1.

Table 1. The properties of the injected marbles in the jelly box

Material	Diameter (mm)	Mass (g)	Density (g/cm^3)
Iron marble	17.34	21	7.7
Glass marble	16.30	6	2.6
Plastic marble	15.75	2	0.98
Jelly box			1.14

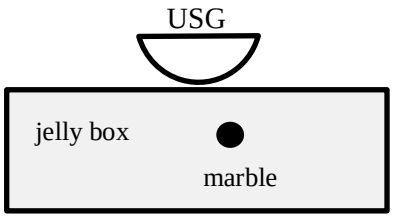


Figure 1. A marble inside the jelly box is scanned using USG convex transducer 35C50EB

2.2. Data Augmentation and ANN Architecture

Each image taken from USG was augmented using a fix size of box, i.e. 280×200×3 pixels, that shifted horizontally and vertically as shown in Figure 2. Using 5 times horizontal shifting and 5 times vertical shifting, we got 5×5=25 data for each image, therefore we had 25×16×4=1600 images as the dataset.

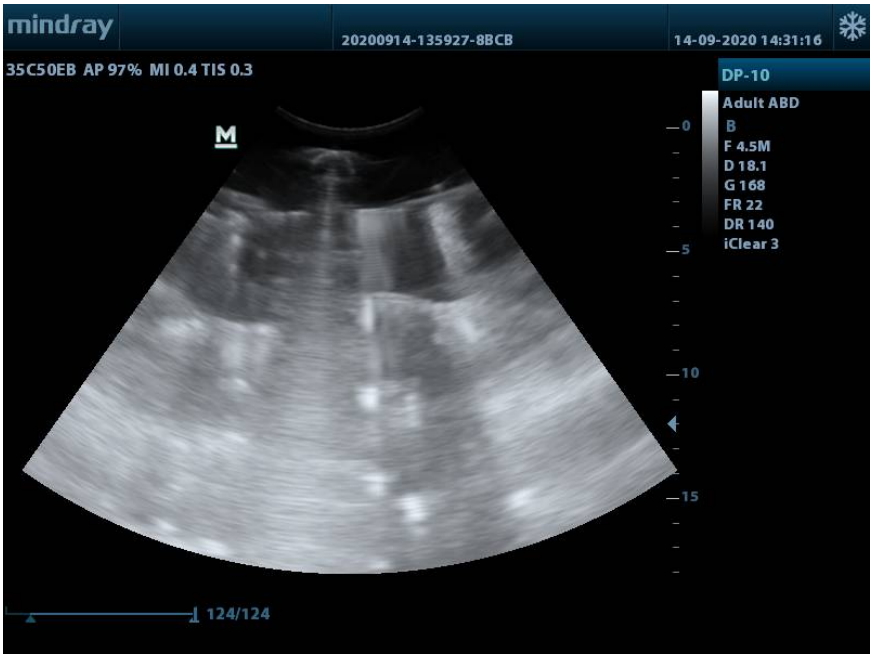


Figure 2. Data augmentation using pixel shifting.

The architect of ANN proposed here is shown in Figure 3. Each data image taken from the augmentation has to be reshaped from 3 dimensional (RGB) image of size [280, 200, 3] into 1 dimensional image with size [1, 280×200×3] before flowing it into the input layer of ANN that consists of 64 neurons. After reshaping data, all data flow to the first hidden layer with 32 neurons, to the second one with 16 neurons and finally to the final (output) layer that has to classify the image into 4 categorical classes. ANN will calculate the best weights and bias and also return the accuracy.

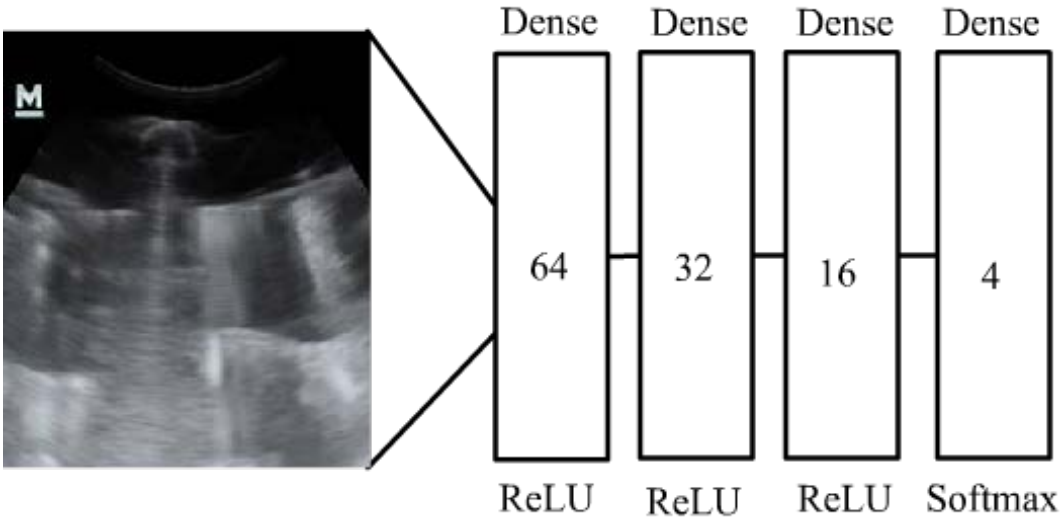


Figure 3. The architect of ANN used in this research.

The dataset that consists of 1600 data images has to be split into two parts, i.e. training dataset and testing dataset in order to make appropriate analysis. Here we used 21.9% data for testing (or 350 images from 1600).

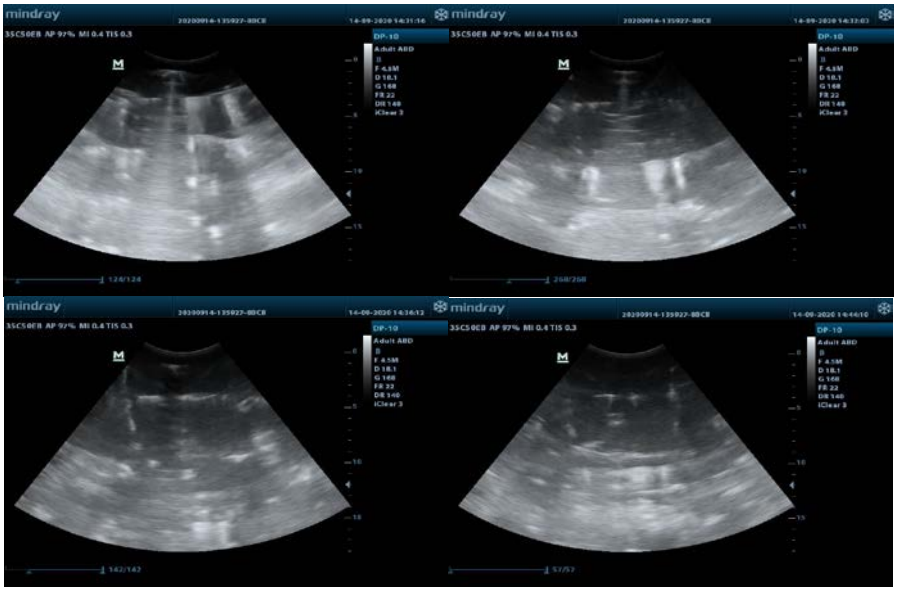


Figure 4. Four different classes: iron marble (left-top), glass marble (right-top), plastic marble (left-bottom), without marble (right-bottom).

3. Results and Discussion

3.1. Data Acquisition Results

From Figure 4, we understood that a manual attempt to classify any injected object into jelly box using human eyes was not easy, though we could differ the empty (without marble) box with the filled box. It simulates a situation for a radiologist to predict whether there is a strange object inside a known (or normal) human body or not. The small size of the injected material in the jelly box caused only little perturbation in the image, therefore we could select only a specific range in the image and did some experiments around the selected position by doing data augmentation. The accuracy depends on the choice of the selected area. Here we reported only the best we could achieve. The accuracy is defined as

$$a = \frac{N - f}{N} \tag{6}$$

In Equation (6), N is the number data and f is the number of incorrect predictions.

3.2. The ANN Result and Discussion

According to Figure 5, the algorithm converges after 35 epochs, though the profile of the loss function and the accuracy are not smooth enough. The accuracy of the testing datasets reaches 98.2%. Unfortunately, this result is sensitive to the choice of area for data augmentation. The accuracy may vary very wide if we wrongly select the area. Therefore, other robust methods in the preprocessing data, such as data augmentation, have to be investigated more if we want to use ANN, before migrating to convolution neural network.

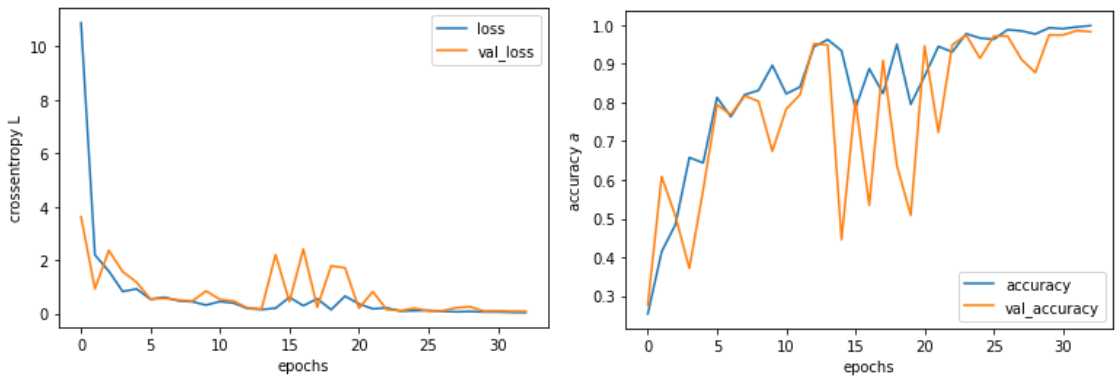


Figure 5. The profile of the loss function (left) for training dataset (loss) and testing dataset (val_loss), and the accuracy for training dataset (accuracy) and testing dataset (val_accuracy)

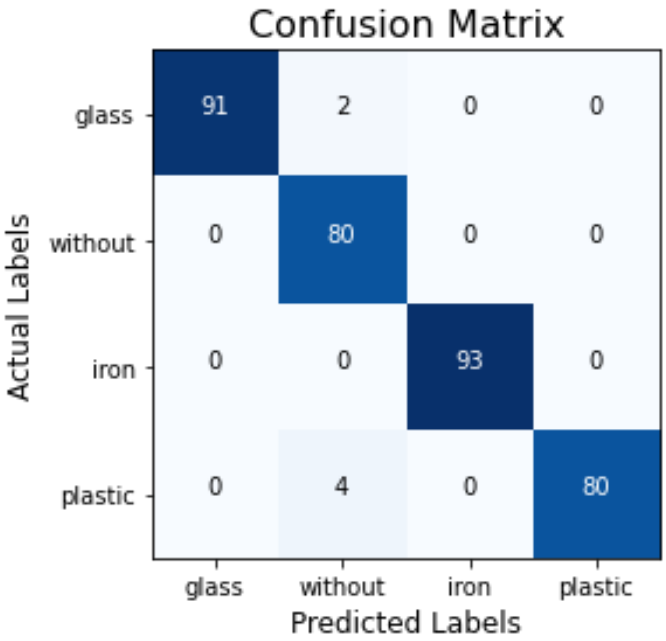


Figure 6. The confusion matrix of the prediction of the testing dataset. It shows 98.2% accuracy

The confusion matrix in Figure 6 shows that the incorrect predictions mainly come from the plastic marbles those are wrongly predicted as the class without marble. But it shows only a small fraction, i.e. 4/80, and it shows a promising method to distinguish a material that has density near the phantom (jelly box) density as given in Table 1.

4. Conclusion

We have reported that ANN can be used for classifying injected materials, such as iron marble, glass marble and plastic marble, that injected in a jelly box. Data augmentation from the ultrasonography images and ANN architecture proposed in this paper produced 98.2% accuracy of prediction, though the result was sensitive to the choice of augmentation area.

Acknowledgements

The Authors would like to thank our colleagues, dr. Jodelin Muninggar and Tafip Hariyanto who have helped us during data acquisition at Physics Laboratory of Universitas Kristen Satya Wacana.

References

[1] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, 27-May-2015, doi: 10.1038/nature14539.

[2] S. Moein, "Artificial Neural Network for Medical Diagnosis," in *Medical Diagnosis Using Artificial Neural Networks*, IGI Global, 2014, pp. 85–94.

[3] H. J. Lee *et al.*, "Role of transrectal ultrasonography in the prediction of prostate cancer: artificial neural network analysis.," *J. Ultrasound Med.*, vol. 25, no. 7, pp. 815-21–4, Jul. 2006, doi: 10.7863/jum.2006.25.7.815.

[4] W. Wu *et al.*, "An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm.," *Comput. Math. Methods Med.*, pp. 1–10, Jul. 2020.

[5] C. Zhang, J. Zhao, J. Niu, and D. Li, "New convolutional neural network model for screening and diagnosis of mammograms.," *PLoS One*, vol. 15, no. 8, pp. 1–20, Aug. 2020.

[6] M. Pandey *et al.*, "Extraction of radiographic findings from unstructured thoracoabdominal computed tomography reports using convolutional neural network based natural language processing.," *PLoS One*, vol. 15, no. 7, pp. 1–15, Jul. 2020.

[7] J. Zhou, J. He, G. Li, and Y. Liu, "Identifying Capsule Defect Based on an Improved Convolutional Neural Network.," *Shock Vib.*, pp. 1–9, Jul. 2020.

[8] K. Almezhghwi and S. Serte, "Improved Classification of White Blood Cells with the Generative Adversarial Network and Deep Convolutional Neural Network.," *Comput. Intell. Neurosci.*, pp. 1–12, Jul. 2020.

[9] R. Shanmugamani, *Deep Learning for Computer Vision : Expert Techniques to Train Advanced Neural Networks Using TensorFlow and Keras*. Birmingham, UK: Packt Publishing, 2018.

[10] M. Dong, S. Mu, A. Shi, W. Mu, and W. Sun, "Novel method for identifying wheat leaf disease images based on differential amplification convolutional neural network," *Int. J. Agric. Biol. Eng.*, vol. 13, no. 4, pp. 205–210, 2020, doi: 10.25165/j.ijabe.20201304.4826.

[11] Ö. F. Ertuğrul, "A novel type of activation function in artificial neural networks: Trained activation function," *Neural Networks*, vol. 99, pp. 148–157, Mar. 2018, doi: 10.1016/j.neunet.2018.01.007.

[12] K. Fukushima, "Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements," *IEEE Trans. Syst. Sci. Cybern.*, vol. 5, no. 4, pp. 322–333, 1969, doi: 10.1109/TSSC.1969.300225.

[13] B. Gao and L. Pavel, "On the properties of the softmax function with application in game theory and reinforcement learning," *arXiv*. arXiv, 03-Apr-2017.

[14] L. Li, M. Doroslovacki, and M. H. Loew, "Approximating the Gradient of Cross-Entropy Loss Function," *IEEE Access*, vol. 8, pp. 111626–111635, 2020, doi: 10.1109/ACCESS.2020.3001531.

[15] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Dec. 2014.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210102-01 ~ 0210102-13
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8352

Synthesis, Crystal Structures, Electronic Spectra, and Magnetic Properties of Bis(μ-Thiolato)-Bridged Trinuclear Co^{II} Complexes with Tridentate-*N*, *N*, *S*-Thiolates

Masahiro Mikuriya[1*], Atsushi Fujita[1], Takanori Kotera[1], Daisuke Yoshioka[1], Hiroshi Sakiyama[2], Makoto Handa[3], Motohiro Tsuboi[1]

[1]School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda 669-1337, Japan

[2]Faculty of Science, Yamagata University, 1-4-12 Kojirakawa, Yamagata 990-8560, Japan

[3]Graduate School of Natural Science and Technology, Shimane University, 1060 Nishikawatsu, Matsue 690-8504, Japan

*junpei@kwansei.ac.jp

Abstract. New trinuclear Co^{II} complexes, $[\{Co(apaet)_2\}_2Co]X_2$ ($apaet^-$ = 2-[(3-aminopropyl)amino]ethanethiolato; X = SCN (**1**), ClO_4 (**2**), NO_3 (**3**), Cl (**4**), Br (**5**), I (**6**)) and $[\{Co(apampt)_2\}_2Co]X_2$ ($apampt^-$ = 1-[(3-aminopropyl)amino]-2-methylpropane-2-thiolato; X = NO_3(**7**), ClO_4 (**8**), Cl (**9**), Br (**10**), I (**11**)), and mononuclear Co^{III} complexes, $[Co(apaet)_2]X$ (X = ClO_4 (**12**), NO_3 (**13**)), were synthesized. Single-crystal X-ray crystallography of **1** and **7** confirmed that the trinuclear complexes have a linear arrangement of octahedral $Co^{II}S_2N_4$-tetrahedral $Co^{II}S_4$-octahedral $Co^{II}S_2N_4$ chromophores where two thiolate ligands are coordinated to each terminal Co atom in a *mer* coordination mode and the two thiolato S atoms are further bound to the central Co atom, which is consistent with the electronic spectra and antiferromagnetic properties.

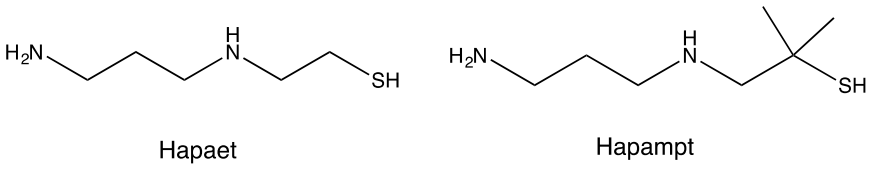
Keywords: thiolato complex, cobalt complex, mononuclear Co(III) complex, trinuclear Co(II) complex

(Received 2021-04-04, 2021-04-29, Available Online by 2021-04-30)

1. Introduction

The synthesis and crystal structures of a great number of metal thiolates have appeared in the literature over the past fifty years, since there have been their diverse coordination compounds of many kinds of metal atoms with redox reactions in metal-assembling [1-8]. The growing interests in metal thiolates

have stimulated research efforts for the efficient synthesis and synthetic methods of thiolate ligands have been developed [8,9,23]. Metal thiolates are also of interest from a point of view that their metal complexes have similar functionalities to those of active sites of metallo-enzymes containing cysteine residue in catalyzing organic oxidation or electron-transfer reactions. We are interested in using organic thiolate ligands containing nitrogen atoms besides thiolate sulfur atom, because synthesis of metal thiolates may be controlled by the formation of chelate rings by the combination of donor set [9-32]. Reactions of metal ions and tridentate-*N*, *N*, *S*-thiolates (Figure 1) such as 2-[(3-aminopropyl)amino]ethanethiol (abbreviated as Hapaet), we isolated trinuclear thiolates of homonuclear $[\{M(apaet)_2\}_2M](ClO_4)_2$ (M = Mn^{II} [17], Fe^{II} [18], Cd^{II} [24]) as well as heterometal $[\{M(apaet)_2\}_2M'](ClO_4)_2$ (M = Mn^{II}, Fe^{II}, Co^{II}, Ni^{II}; M' = Zn^{II}, Cd^{II}, Hg^{II}) [19], which have a linearly arranged trinuclear metal ions with octahedral S_2N_4, tetrahedral S_4, and octahedral S_2N_4 coordination environments. The linear trinuclear structure seems to be most favorable for metal thiolates with apaet. In addition to the trinuclear species, dinuclear nickel(II) [12], dinuclear iron(II) [18], dinuclear molybdenum(V) [30], cyclic trinuclear zinc(II) [22], mixed-valent trinuclear molybdenum(V,VI) [31], tetranuclear palladium(II) [20,21], adamantane-like teranuclear manganese(II) and iron(II) [23], mixed-valent hexanuclear and octanuclear copper(I,II) [25,26,28], infinite polynuclear manganese(II) [14,17] and zinc(II) [22] species also have been isolated in these systems. In the case of 2-[(2-pyridylmethyl)amino]ethanethiol, a mononuclear rhenium thiolate was isolated [29]. It is well known that cobalt(III) ion favours an octahedral geometry because of the crystal field stabilization energy, resulting in the mononuclear octahedral Co(III) species, $[Co(apaet)_2](ClO_4)$, and dinuclear octahedral cobalt(III) species, $[Co_2\{SCH(CH_2CH_2NH_2)_2\}_3](ClO_4)_3$, for Hapaet and Hdpet (Hdpet = 1,5-diamino-3-pentanethiol), respectively [10,27,32].



Hapaet Hapampt

Figure 1. Tridentate-*N*, *N*, *S*-thiolates

In this study, we made an effort to extend cobalt thiolate chemistry by synthesizing new cobalt thiolates in addition to the mononuclear thiolates, because the isolated compounds may have relevance to model compounds of metallo-enzymes in bacteria, Co-nitrile hydratases, which have non-corrinoid Co atom with N and S donor-atoms and catalyze the hydration of nitriles to the corresponding amides [33-35]. Crystallographic study of the Co-nitrile hydratases has revealed that the metal is coordinated by two deprotonated carboxamide-N and three cysteine-S atoms which can be converted to cysteine-sufenic and -sulfinic group as like those of Fe-nitrile hydratases [36,37]. We synthesized a series of cobalt thiolates with $apaet^-$, using a variety of cobalt salts, and further, synthesized 1-[(3-aminopropyl)amino]-2-methylpropane-2-thiol (Hapampt), which corresponds to an analogous tridentate-*N*, *N*, *S*-thiol substituted the α-methylene hydrogens by methyl groups, and performed synthesis of cobalt thiolates with $apampt^-$. We report herein the synthesis, crystal structures, electronic spectral and magnetic properties of the isolated cobalt thiolates.

2. Research Methods

2.1. Synthesis of Complexes

All commercial chemicals and solvents were used without further purification. Methanol was dried using standard laboratory techniques. The thiolate ligands, Hapaet and Hapampt, were synthesized according to literature procedure [12,25]. All manipulations were performed under an argon atmosphere using Schlenk techniques.

Safety Notes. *Warning!* Cobalt(II) perchlorate hexahydrate are potentially explosive and should only be handled in small quantities with great care.

$[\{Co(apaet)_2\}_2Co](SCN)_2\cdot CH_3OH\cdot 2H_2O$ (**1**$\cdot CH_3OH\cdot 2H_2O$)

To a mixed solution of 4 mL of methanol and 2 mL of *N*,*N*-dimethylacetamide containing Hapaet (34 mg, 0.25 mmol) and triethylamine (two drops) was added 4 mL of methanol solution of cobalt(II) thiocyanate (44 mg, 0.25 mmol). The solution was stirred for 5 min at room temperature and was left standing without disturbance for several days. The deposited crystals were filtered off, washed with methanol, and dried. Yield 26 mg (49%). Anal. Calcd for $C_{23}H_{60}N_{10}O_3S_6Co_3$: C, 30.98; H, 6.48; N, 15.67%. Found: C, 31.44; H, 6.50; N, 15.64%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3288(m), 3239(m); ν(C-H) 2930(m), 2866(m), 2840(m); ν(SCN) 2051(s). Diffused reflectance spectra λ_{max}/nm 264, 326, 636, 732, 784, 1360, 1598, 1700. μ_{eff}(300 K)/μ_B 6.95, μ_{eff}(4.5 K)/μ_B 4.58.

$[\{Co(apaet)_2\}_2Co](ClO_4)_2\cdot H_2O$ (**2**$\cdot H_2O$)

To a mixed solution of 4 mL of methanol and 2 mL of *N*,*N*-dimethylacetamide containing Hapaet (34 mg, 0.25 mmol) and triethylamine (two drops) was added 4 mL of methanol solution of cobalt(II) perchlorate hexahydrate (93 mg, 0.25 mmol). The solution was left standing without disturbance at room temperature for several days. The deposited crystals were filtered off, washed with methanol, and dried. Yield 12 mg (21%). Anal. Calcd for $C_{20}H_{54}N_8O_9S_4Cl_2Co_3$: C, 25.92; H, 5.87; N, 12.09%. Found**:** C, 25.84; H, 5.80; N, 11.97%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3327(m), 3282(m); ν(C-H) 2930(m), 2823(m); ν(Cl-O) 1145(s), 1100(s), 1080(s), 623(s). Diffused reflectance spectra λ_{max}/nm 272, 342(sh), 458(sh), 666, 730, 782, 1368, 1630. Electronic absorption spectra in dmf (*N*,*N*-dimethylformamide) λ_{max}/nm (ε/dm^{3}mol^{-1}cm^{-1}) 673 (300), 731 (370), 786 (870), 1355 (94). μ_{eff}(300 K)/μ_B 5.11, μ_{eff}(4.5 K)/μ_B 3.55.

$[\{Co(apaet)_2\}_2Co](NO_3)_2\cdot 3H_2O$ (**3**$\cdot 3H_2O$)

To a mixed solution of 4 mL of methanol and 2 mL of *N*,*N*-dimethylacetamide containing Hapaet (34 mg, 0.25 mmol) and triethylamine (two drops) was added 4 mL of methanol solution of cobalt(II) nitrate hexahydrate (73 mg, 0.25 mmol). The solution was left standing without disturbance at room temperature for several days. The deposited crystals were filtered off, washed with methanol, and dried. Yield 38 mg (69%). Anal. Calcd for $C_{20}H_{56}N_{10}O_7S_4Co_3$: C, 27.06; H, 6.58; N, 15.18%. Found**:** C, 27.37; H, 6.25; N, 15.33%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3301(m), 3265(m); ν(C-H) 2931(m), 2882(m); ν(N-O) 1382(s), 1336(s). Diffused reflectance spectra λ_{max}/nm 276, 342(sh), 486(sh), 636, 726, 782, 1358, 1642.

$[\{Co(apaet)_2\}_2Co]Cl_2\cdot 2CH_3OH\cdot 2H_2O$ (**4**$\cdot 2CH_3OH\cdot 2H_2O$)

To a mixed solution of 4 mL of methanol and 2 mL of *N*,*N*-dimethylacetamide containing Hapaet (34 mg, 0.25 mmol) and triethylamine (two drops) was added 4 mL of methanol solution of cobalt(II) chloride hexahydrate (60 mg, 0.25 mmol). The solution was left standing without disturbance at room temperature for several days. The deposited crystals were filtered off, washed with methanol, and dried. Yield 34 mg (61%). Anal. Calcd for $C_{22}H_{64}N_8O_4S_4Cl_2Co_3$: C, 30.00; H, 7.32; N, 12.72%. Found**:** C, 30.42; H, 7.44; N, 12.70%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3204(m), 3148(m); ν(C-H) 2925(m), 2875(m), 2820(m). Diffused reflectance spectra λ_{max}/nm 266, 316, 626, 724, 778, 1366, 1644. μ_{eff}(300 K)/μ_B 6.45, μ_{eff}(4.5 K)/μ_B 3.67.

$[\{Co(apaet)_2\}_2Co]Br_2\cdot 2CH_3OH\cdot 4H_2O$ (**5**$\cdot 2CH_3OH\cdot 4H_2O$)

This complex was prepared in the same manner as complex **4**$\cdot 2CH_3OH\cdot 2H_2O$ using cobalt(II) bromide hexahydrate. Yield 20 mg (32%). Anal. Calcd for $C_{22}H_{68}N_8O_6S_4Br_2Co_3$: C, 26.27; H, 6.82; N, 11.14%. Found**:** C, 26.30; H, 6.51; N, 11.06%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3194(m), 3147(m); ν(C-H) 2926(m), 2866(m), 2842(m). Diffused reflectance spectra λ_{max}/nm 274, 350, 380, 664(sh), 778, 1364, 1650.

$[\{Co(apaet)_2\}_2Co]I_2\cdot 2CH_3OH\cdot 2H_2O$ (**6**$\cdot 2CH_3OH\cdot 2H_2O$)

This complex was prepared in the same manner as complex **4**$\cdot 2CH_3OH\cdot 2H_2O$ using cobalt(II) iodide. Yield 26 mg (38%). Anal. Calcd for $C_{22}H_{64}N_8O_3S_4I_2Co_3$: C, 24.84; H, 6.06; N, 10.53%. Found**:** C, 24.57; H, 5.85; N, 10.30%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3186(m), 3129(m); ν(C-H) 2926(m), 2866(m),

2798(m). Diffused reflectance spectra λ_{max}/nm 348, 728, 664(sh), 782, 1352, 1658. μ_{eff}(300 K)/μ_B 4.86, μ_{eff}(4.5 K)/μ_B 2.60.

[{Co(apampt)$_2$}$_2$Co](NO$_3$)$_2$·2CH$_3$OH (**7**·2CH$_3$OH)

To a mixed solution of 5 mL of methanol and 2 mL of *N*,*N*-dimethylacetamide containing Hapampt (41 mg, 0.25 mmol) and triethylamine (two drops) was added 5 mL of methanol solution of cobalt(II) nitrate hexahydrate (73 mg, 0.25 mmol). The solution was stirred for 5 min at room temperature and was left standing without disturbance for several days. The deposited crystals were filtered off, washed with methanol, and dried. Yield 39 mg (64%). Anal. Calcd for $C_{30}H_{76}N_{10}O_8S_4Co_3$: C, 35.67; H, 7.58; N, 13.87%. Found**:** C, 35.84; H, 7.79; N, 14.22%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3324(m), 3238(m); ν(C-H) 2947(m), 2839(m); ν(N-O) 1402(s), 1384(s), 1325(s). Diffused reflectance spectra λ_{max}/nm 288, 336, 434(sh), 696, 754(sh), 796, 1392, 1744. μ_{eff}(300 K)/μ_B 6.60, μ_{eff}(4.5 K)/μ_B 4.52.

[{Co(apampt)$_2$}$_2$Co](ClO$_4$)$_2$·dmac·H$_2$O (dmac = *N*,*N*-dimethylacetamide) (**8**·dmac·H$_2$O)

This complex was prepared in the same manner as complex **7**·2CH$_3$OH using cobalt(II) perchlorate hexahydrate. Yield 12 mg (16%). Anal. Calcd for $C_{32}H_{79}N_9O_{10}S_4Cl_2Co_3$: C, 33.60; H, 7.14; N, 11.02%. Found**:** C, 33.75; H, 7.01; N, 10.81%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3331(m), 3285(m), 3264(m); ν(C-H) 2928(m), 2843(m); ν(Cl-O) 1106(s), 1058(s), 1014(s), 626(s). Diffused reflectance spectra λ_{max}/nm 290, 348(sh), 694, 754(sh), 798, 1384, 1730. Electronic absorption spectra in dmf λ_{max}/nm (ε/dm^{3}mol^{-1}cm^{-1}) 687 (390), 742 (540), 794 (1140), 1389 (130). μ_{eff}(300 K)/μ_B 6.84, μ_{eff}(4.5 K)/μ_B 4.57.

[{Co(apampt)$_2$}$_2$Co]Cl$_2$·CH$_3$OH·H$_2$O (**9**·CH$_3$OH·H$_2$O)

To a mixed solution of 5 mL of methanol and 2 mL of *N*,*N*-dimethylacetamide containing Hapampt (41 mg, 0.25 mmol) and triethylamine (two drops) was added 5 mL of methanol solution of cobalt(II) chloride hexahydrate (60 mg, 0.25 mmol). The solution was stirred for 5 min at room temperature and was left standing without disturbance for several days. The deposited crystals were filtered off, washed with methanol, and dried. Yield 15 mg (26%). Anal. Calcd for $C_{29}H_{74}N_8O_2S_4Cl_2Co_3$: C, 36.94; H, 7.91; N, 11.88%. Found**:** C, 36.86; H, 7.77; N, 11.79%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3244(m), 3204(m); ν(C-H) 2954(m), 2921(m), 2865(m). Diffused reflectance spectra λ_{max}/nm 216, 260, 294(sh), 340(sh), 506(sh), 694, 754(sh), 794, 1504, 1580, 1758.

[{Co(apampt)$_2$}$_2$Co]Br$_2$·CH$_3$OH (**10**·CH$_3$OH)

This complex was prepared in the same manner as complex **9**·CH$_3$OH·H$_2$O using cobalt(II) bromide hexahydrate. Yield 14 mg (22%). Anal. Calcd for $C_{29}H_{72}N_8OS_4Br_2Co_3$: C, 34.36; H, 7.16; N, 11.05%. Found**:** C, 34.28; H, 7.01; N, 10.96%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3244(m), 3204(m); ν(C-H) 2954(m), 2921(m), 2865(m). Diffused reflectance spectra λ_{max}/nm 280, 338, 442(sh), 694, 754(sh), 794, 1394, 1502(sh), 1730. μ_{eff} (300 K)/μ_B 6.02, μ_{eff}(4.5 K)/μ_B 2.93.

[{Co(apampt)$_2$}$_2$Co]I$_2$·CH$_3$OH (**11**·CH$_3$OH)

This complex was prepared in the same manner as complex **9**·CH$_3$OH·H$_2$O using cobalt(II) iodide. Yield 16 mg (22%). Anal. Calcd for $C_{29}H_{72}N_8OS_4I_2Co_3$: C, 30.96; H, 6.63; N, 9.95%. Found**:** C, 31.04; H, 6.62; N, 9.96%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3205(m), 3147(m); ν(C-H) 2945(m), 2839(m). Diffused reflectance spectra λ_{max}/nm 288, 340(sh), 696, 754(sh), 796, 1386, 1732, 1782. μ_{eff}(300 K)/μ_B 5.71, μ_{eff}(4.5 K)/μ_B 3.24.

[Co(apaet)$_2$]ClO$_4$·CH$_3$OH (**12**·CH$_3$OH)

This complex was prepared according to the method reported recently [32].

[Co(apaet)$_2$]NO$_3$·H$_2$O (**13**·H$_2$O)

To 3 mL of methanol containing Hapaet (34 mg, 0.25 mmol) and triethylamine (two drops) was added 3 mL of methanol solution of cobalt(II) nitrate hexahydrate (36 mg, 0.18 mmol). The solution was left standing without disturbance at room temperature for several days. The deposited crystals were filtered off, washed with methanol, and dried.-Yield 45 mg (70%). Anal. Calcd for $C_{10}H_{54}N_5O_4S_2Co$: C, 29.63; H, 6.96; N, 17.27%. Found**:** C, 29.66; H, 6.56; N, 17.40%. IR (KBr, pellet) ν/cm^{-1} ν(N-H) 3252(m), 3213(m), 3142(m); ν(C-H) 2969(m), 2928(m); ν(N-O) 1372(s), 1348(s). Diffused reflectance spectra λ_{max}/nm 272, 348(sh), 492, 838, 916, 1370, 1584, 1712, 1748, 1782, 1928.

2.2 Measurements

C, H, and N elemental analyses were carried out by the use of a Perkin-Elmer 2400 Series II CHNS/O Analyzer. IR spectra (4000—600 cm^{-1}) were recorded on a JASCO MFT-2000 FT-IR Spectrometer with the samples prepared as KBr pellets. Electronic spectra (200—2000 nm) were recorded on a Shimadzu UV-3100 Spectrophotometer. Magnetic susceptibility measurements in the temperature range 4.5—300 K were performed with a Quantum Design MPMS-5S SQUID magnetometer at an applied field of 0.5 T. The data were corrected for diamagnetic contributions calculated from Pascal's constants for all constituent atoms [38].

The X-ray data were collected at room temperature with a Bruker CCD diffractometer (SMART APEX) with Mo-$K\alpha$ radiation. The data collection and refinement processes are summarized in Table 1. The structures were solved by direct methods and refined by full-matrix least-squares method, using the SHELXTL program package [39]. All non-hydrogen atoms were refined anisotropically. The hydrogen atoms were located on calculated positions.

Table 1. Crystal data and structure refinement details for **1** and **7**.

Compounds	**1**	**7**
Empirical Formula	$C_{24}H_{60}Co_3N_{10}O_2S_6$	$C_{30}H_{60}Co_3N_{10}O_7S_4$
Formula weight	889.97	977.91
Temperature / K	293	293
Crystal dimensions / mm	0.31×0.30×0.14	0.40×0.39×0.20
Crystal system	Monoclinic	Tetragonal
Space group	*C*2/*c*	*I*4/*m*
a / Å	24.728(6)	12.9269(11)
b / Å	9.415(2)	
c / Å	18.336(5)	28.238(4)
β / °	104.178(3)	
V / $Å^3$	4138.8(18)	4718.7(8)
Z	4	4
d_{calcd}	1.428	1.377
μ / mm^{-1}	1.527	1.268
F(000)	1868	2044
Reflections collected	10932	10567
Independent reflections	4533	1751
θ range for data collection	1.70 to 28.27°	1.73 to 23.23°
Data / Restraints / Parameters	4533/0/206	1751/0/129
*R*1, *wR*2[$I>2\sigma(I)$][a)]	0.1008, 0.2823	0.0387, 0.1340
*R*1, *wR*2[all data][a)]	0.1182, 0.3160	0.0503, 0.1387
Goodness-of-fit on F^2	1.218	1.122
CCDC number	1953679	2073611

[a] $R1 = \Sigma||F_o|-|F_c||/\Sigma|F_o|$; $wR2 = [\Sigma w(F_o^2-F_c^2)^2/\Sigma w(F_o^2)^2]^{1/2}$

3. Results and Discussion

3.1 Synthesis and Structural Characterization of Co Thiolates

Our previous studies showed that the reactions of Hapaet with Mn(II) [17], Fe(II) [18], and Cd(II) [24] afforded trinuclear thiolate complexes, $[\{M(apaet)_2\}_2M]X_2$, where the three metal atoms are arranged linearly with *Oh-Td-Oh* geometries. In this study, we isolated analogous trinuclear Co(II) complexes with $apaet^-$ and $apampt^-$, $[\{Co(apaet)_2\}Co]X_2$ (X = SCN (**1**), ClO_4 (**2**), NO_3 (**3**), Cl (**4**), Br (**5**), I (**6**)) and $[\{Co(apampt)_2\}_2Co]X_2$ (X = NO_3(**7**), ClO_4 (**8**), Cl (**9**), Br (**10**), I (**11**), under the similar condition with equivalent molar ratio of the thiolate ligand : metal salt.

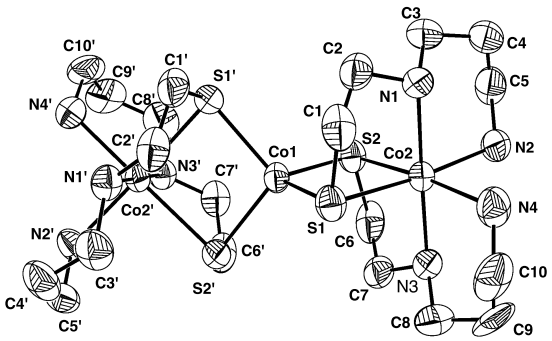


Figure 2. ORTEP Diagram of Trinuclear Complex Cation in **1**

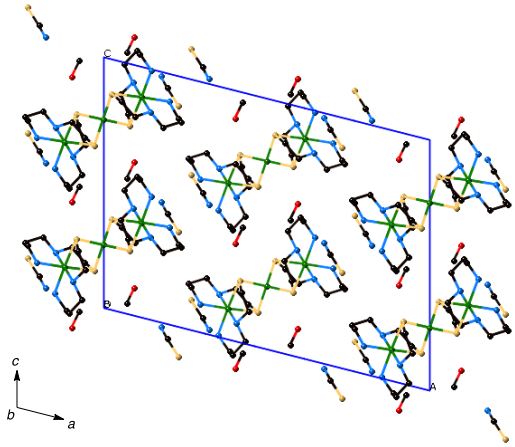


Figure 3. Packing Diagram of **1**

The results of the elemental analyses suggest that the cations of these complexes should consist of three cobalt atoms and four thiolato ligands, apaet$^-$ or apampt$^-$. The infrared spectra of these complexes are essentially the same except for the bands due to the counter anion X, showing the characteristic bands of apaet$^-$ and apampt$^-$ deprotonated ligands with the ν(N-H) bands at 3327—3129 cm^{-1} and the ν(C-H) bands at 2969–2798 cm^{-1} and without the ν(S-H) band around 2500 cm^{-1} [40,41]. If we use an excess amount of thiolate ligand to metal salt with 1.4:1 molar ratio, mononuclear cobalt(III) thiolates, $[Co(apaet)_2]X_2$ (X = ClO_4 (**12**) and NO_3 (**13**)) were isolated as described in the previous paper [32]. X-ray structure analysis was performed for the single crystals of **1** and **7**. The crystal structure of **1** contains trinuclear complexes, $[Co\{Co(apaet)_2\}_2]^{2+}$, thiocyanate anions, and methanol molecules in a 1:2:2 molar ratio. An ORTEP diagram of $[Co\{Co(apaet)_2\}_2]^{2+}$ is depicted in Figure 2. The three cobalt atoms are arranged almost linearly (Co2-Co1-Co2' = 179.57(3)°). The cation has a C_2-symmetry, where the crystallographic C_2 axis passes through the Co1 atom and the midpoint of S1 and S1' atoms. The Co1 atom is bonded to four thiolato-sulfur atoms of apaet$^-$ in a distorted tetrahedral geometry, while each of the Co2 and Co2' atoms is coordinated by two thiolato-sulfur atoms and four amino-nitrogen atoms of apaet$^-$ in a distorted octahedral geometry. Each apaet$^-$ ligand is coordinated to the terminal cobalt atom in a meridional mode to form a fused chelate with six- and five-membered rings. The bond distances of the terminal octahedral cobalt atoms [Co2-S 2.524(2)—2.526(2) Å] are longer than those of the central tetrahedral cobalt atom [Co1-S 2.331(2)—2.340(2) Å]. The bond distances of Co2-S and Co1-S are similar to the Co-S bonds of octahedral and tetrahedral thiolato Co(II) complexes, respectively [19,27]. The Co2-N distances [2.203(5)—2.224(5)Å] are also similar to the Co-N bonds of octahedral Co(II) complexes [19,27]. The Co2-S1-C1-C2-N1 and Co2-S2-C6-C7-N3 chelate rings are in a gauche form which is common for five-membered chelate ring, whereas the Co2-N1-C3-C4-C5-N2 and Co2-N3-C8-C9-C10-N4 chelate rings are in a chair form. The complex cation structure is similar to that of

2·2dmac·2CH_3OH (dmac = *N*,*N*-dimethylacetamide) [27]. In the crystal, thiocyanate ions and methanol molecules are arranged among the trinuclear cations as like separating these molecules (Figure 3). The crystal structure of **7** contains trinuclear complexes, $[Co\{Co(apampt)_2\}_2]^{2+}$, nitrate anions, methanol molecules, and water molecules in a 1:2:0.5:0.5 molar ratio. An ORTEP diagram of $[Co\{Co(apampt)_2\}_2]^{2+}$ is depicted in Figure 4. The trinuclear structure of the complex cation is similar to those for **1**·2CH_3OH and **2**·2dmac·2CH_3OH [27]. However, the three cobalt atoms are arranged in a crystallographical line (Co2-Co1-Co2" = 180°). The cation has a C_2-symmetry, where the crystallographic C_2 axis passes through the Co1 atom and the midpoint of S1 and S1' atoms. The Co1 atom is bonded to four thiolato-sulfur atoms of apampt$^-$ in a distorted tetrahedral geometry, while each of the Co2 and Co2" atoms is coordinated by two thiolato-sulfur atoms and four amino-nitrogen atoms of apampt$^-$ in a distorted octahedral geometry. Each apampt$^-$ ligand is coordinated to the terminal cobalt atom in a meridional mode to form a fused chelate with six- and five-membered rings. The bond distances of the terminal octahedral Co2-S [2.477(1) Å] are significantly shorter than those of **1**·2CH_3OH and **2**·2dmac·2CH_3OH [2.509(4)-2.526(2) Å]. As the electron density of the sulfur atom is increased, compared with that of apaet$^-$, by the methyl groups which are attached at the alpha position of the sulfur atom, bond lengths between the octahedral Co and the sulfur atoms would be appreciably shortened. The bond distances of the Co2-N [2.204(3)-2.232(3) Å] are comparable to those of **1**·2CH_3OH and **2**·2dmac·2CH_3OH [27], which are normal as the Co-N bond lengths of octahedral Co(II) complexes [42]. On the other hand, the bond distances of the central Co1-S [2.334(1) Å] are significantly longer than those of **1**·2CH_3OH and **2**·2dmac·2CH_3OH [Co1-S 2.305(4)—2.340(2) Å] [27], reflecting the steric hindrance with the methyl groups attached to the alpha carbon atom of the

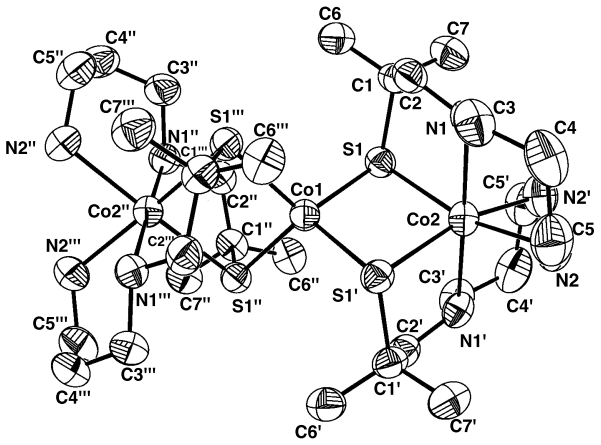


Figure 4. ORTEP Diagram of Trinuclear Complex Cation in **7.**

thiolato sulfur atom.

The Co2-S1-C1-C2-N1 five-membered chelate ring is in a gauche form, whereas the Co2-N1-C3-C4-C5-N2 six-membered chelate ring is in a chair form. The nitrate anion is considered to be hydrogen-bonded to the nitrogen atoms of the thiolate ligand as indicated by the distances O1(NO_3)···N2 3.052(6) Å, O1(NO_3)···N2' 3.057(6) Å, and O1(NO_3)···N1' 3.171(6) Å (Figure 5).

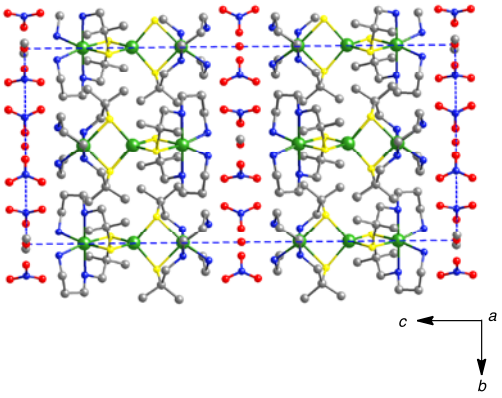


Figure 5. Packing Diagram of **7**

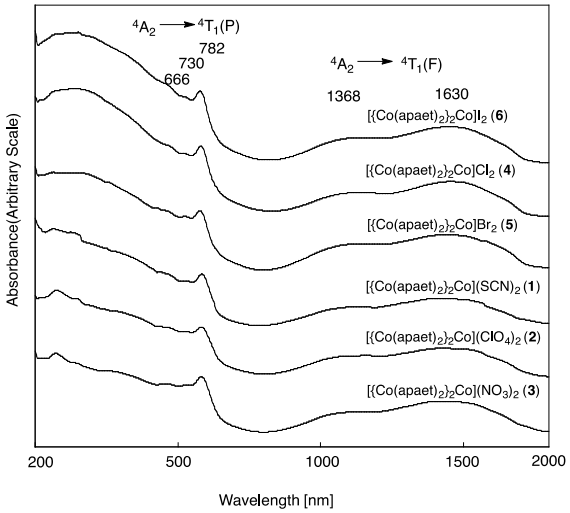


Figure 6. Diffuse Reflectance Spectra of Trinuclear Co(II) Complexes with apaet⁻

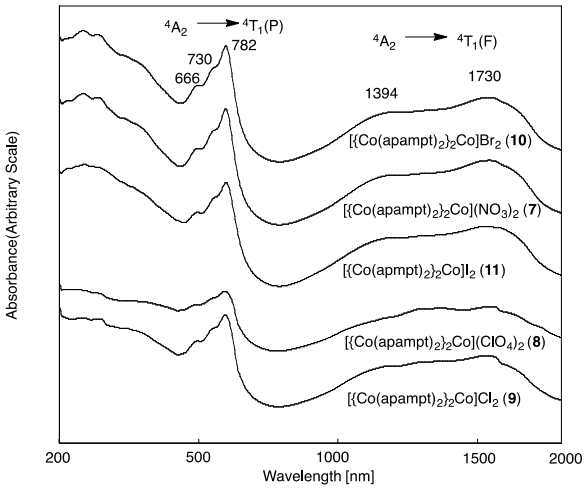


Figure 7. Diffuse Reflectance Spectra of Trinuclear Co(II) Complexes with apampt⁻

3.2 Electronic Spectra of Co Thiolates

Diffuse reflectance spectra of the trinuclear cobalt complexes with apaet⁻ **1**—**6** are shown in Figure 6. These spectra show essentially the same feature, having a broad band attributable to thiolato-sulfur-to-cobalt(II) charge transfer transition at around 350 nm, three bands at around 670, 730, and 780 nm, and two broad bands at around 1370 and 1630 nm, which may be attributed to d-d transition bands of high-spin tetrahedral cobalt(II) [43]. D-d transitions of octahedral cobalt(II) can be considered to be obscured by the stronger tetrahedral d-d transition bands [43]. As shown in Figure 7, diffuse reflectance spectra of the trinuclear cobalt(II) complexes with apampt⁻ **7**—**11** are similar to those of **1**—**6**, exhibiting a broad band attributable to thiolato-sulfur-to-cobalt(II) charge transfer transition at around 350 nm, three bands at around 690, 750, and 790 nm, and two broad bands at around 1390 and 1730 nm, which may be attributed to d-d transition bands of high-spin tetrahedral cobalt(II) species ($^4A_2 \rightarrow {}^4T_1$(P) and $^4A_2 \rightarrow {}^4T_1$(F), respectively) [43]. These d-d transition bands are located at a little lower frequency side compared with those of the unmethylated thiolato complexes **1**—**6**. We can ascribe this weak coordination field of the methylated thiolato complexes **7**—**11** to the steric crowdedness with the methyl groups as indicated in the bond lengthening of the central Co-S bonds. The electronic absorption spectra of **2** and **8** in dmf are essentially the same as those in the solid state (Figure 8). The three absorption band at visible and near infra-red region can be attributed to ligand field absorptions for high-spin tetrahedral Co(II). The ligand field absorptions for **8** show red-shift features and molar absorption coefficients are slightly large compared with those of **2**. This result suggests that the crystal field of the central cobalt atom is weaker than that of **2** and the deviation from the tetrahedral arrangement may be smaller by adding methyl groups to the alpha-carbon atom of the thiolato sulfur atom. The differences between the bond distances and angles around the central cobalt atoms of **1**, **2**, and **7** [Co1-S 2.305(4)—2.322(2) Å, S-Co1-S 100.60(5)—119.9(2)° for **1** and **2**; Co1-S 2.334(1) Å, S-Co1-S 99.35(5), 114.76(3)° for **7**] are consistent with this result. The diffuse reflectance spectra of the mononuclear cobalt (III) complexes with apaet⁻ are shown in Figure 9. The three ligand field absorptions at around 480—930 nm can be ascribed as octahedral d-d transitions of low-spin cobalt (III) $^1A_{1g} \rightarrow {}^1T_{2g}$, $^1T_{1g}$, $^3T_{2g}$) [43].

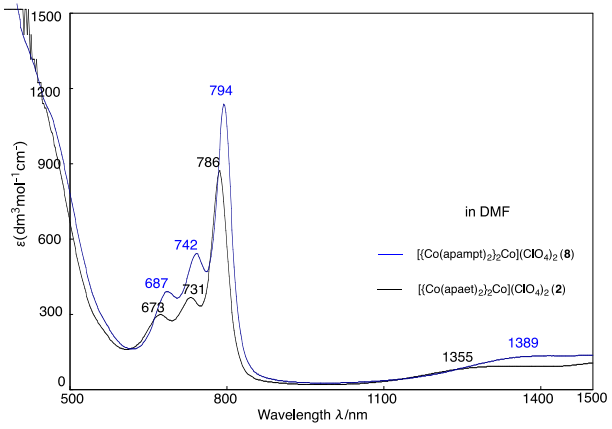


Figure 8. Electronic Absorption Spectra of **2** and **8**

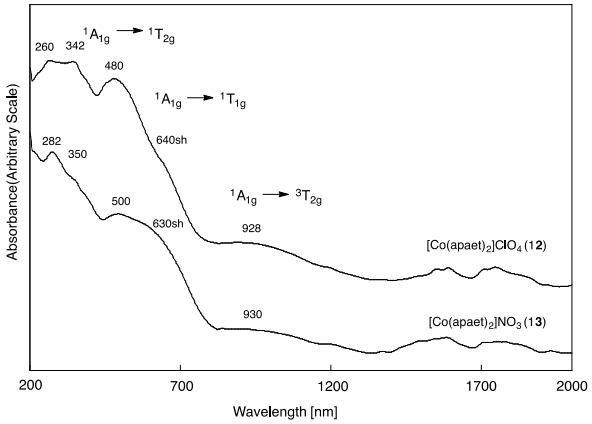


Figure 9. Diffuse Reflectance Spectra of Mononuclear Co(III) complexes **12** and **13**

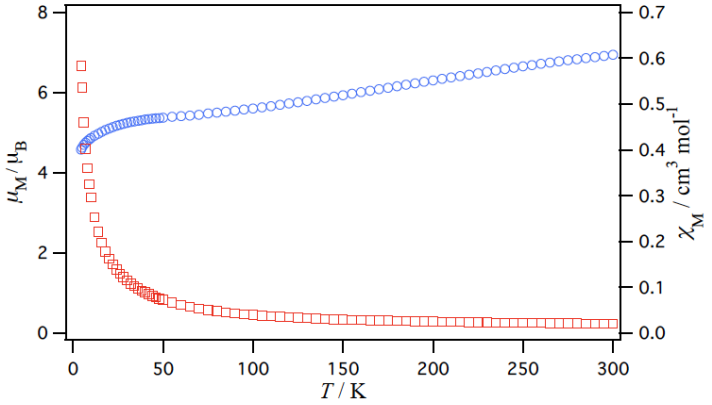


Figure 10. Magnetic Data of **1** (Blue Circle: Magnetic Moment; Red Square: Magnetic Susceptibility)

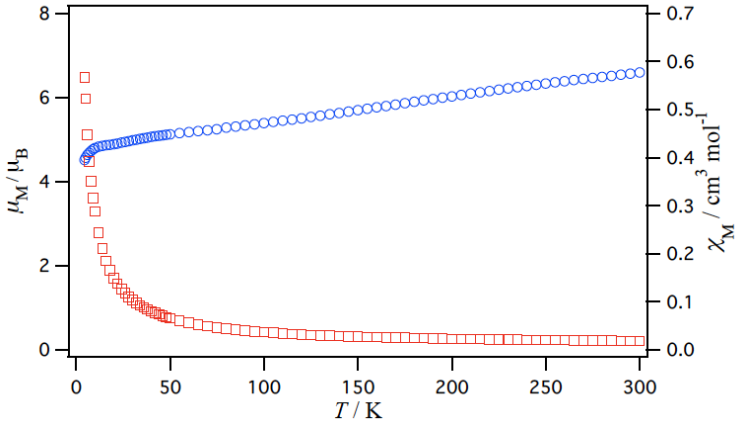


Figure 11. Magnetic Data of **7** (Blue Circle: Magnetic Moment; Red Square: Magnetic Susceptibility)

3.3 Magnetic Properties of Co Thiolates

Variable-temperature magnetic data were measured for the complexes **1**, **2**, **4**, **6**, **7**, **8**, **10**, and **11** in the temperature range of 4.5—300 K (Figures 10 and 11). The effective magnetic moments at 300 K of **1**, **2**, **4**, **6**, **7**, **8**, **10**, and **11** are 6.95, 5.11, 6.45, 4.86, 6.60, 6.84, 6.02, and 5.71 μ_B, per trinuclear molecule,

respectively, showing generally lower values, compared with the spin-only magnetic moment (6.71 μ_B) for a non-interacted spin system of high-spin octahedral d^7-tetrahedral d^7-octahedral d^7. The magnetic moments decrease upon cooling from room temperature to 4.58, 3.55, 3.67, 2.60, 4.52, 4.57, 2.93, and 3.24 μ_B, respectively, at 4.5 K, suggesting an overall antiferromagnetic interaction in these systems.

4. Conclusions

In this study, we synthesized and characterized a series of thiolato-bridged trinuclear cobalt (II) complexes by the use of two kinds of tridentate thiolate ligands, 2-[(3-aminopropyl) amino] ethanethiol and 1-[(3-aminopropyl) amino]-2-methylpropane-2-thiol. Although no large structural, spectral, and magnetochemical changes were found in this series of trinuclear complexes, whether or not for the existence of the added methyl groups at the alpha position of the sulfur atom, some minor differences were observed. We also isolated mononuclear cobalt (III) complexes for the former thiolate ligand by changing the reaction condition, but not for the latter thiolate ligand. The mononuclear thiolate complexes may react with hydrogen peroxide, resulting in the sulfinato species [32].

Acknowledgements

This work was partially supported by the MEXT (the Grants-in-Aid for Scientific Research No. 17K05820).

References

[1] I. G. Dance, "The Structural Chemistry of Metal Thiolate Complexes," *Polyhedron*, vol. 5, pp. 1037-1104, 1986.

[2] P. J. Blower and J. R. Dilworth, "Thiolato-complexes of the Transition Metals," *Coord. Chem. Rev.*, vol. 76, pp. 769-788, 1991.

[3] B. Krebs and G. Henkel, "Transition-Metal Thiolates: From Molecular Fragments of Sulfidic Solids to Models for Active Centers in Biomolecules," *Angew. Chem. Int. Ed. Engl.*, vol. 30, pp.769-788, 1991.

[4] J. R. Dilworth and J. Hu, "Complexes of Sterically Hindered Thiolate Ligands," *Adv. Inorg. Chem.*, vol. 40, pp. 411–459, 1993.

[5] A. C. Marr, D. J. E. Spencer, and M. Schröder, "Structural mimics for the active site of [NiFe] hydrogenase," *Coord. Chem. Rev.*, vol. 219-221, pp. 1055–1074, 2001.

[6] T. Konno, "Aggregation of Octahedral Thiolato Complexes by Forming Sulfur-Bridged Structures with Transition Metal Ions," *Bull. Chem. Soc. Jpn.*, vol. 77, pp. 627-649, 2004.

[7] H. Fleischer, "Structural chemistry of complexes of $(n - 1)d^{10} ns^m$ metal ions with β-N-donor substituted thiolate ligands (m = 0, 2)," *Coord. Chem. Rev.*, vol. 249, pp. 799–827, 2005.

[8] M. H. Asim, M. Ijaz, A. C. Rosch, and A. Bernkop-Schnurch,"Thiolated cyclodextrins: New perspectives for old excipients," *Coord. Chem. Rev.*, vol. 420, p. 213433, 2020.

[9] M. Mikuriya, K. Miyoshi, S. Kurano, J. Taguchi, T. Kotera, and D. Yoshioka, "Synthesis and Crystal Structure of 1,4,7-Tris(2-mercaptoethyl)-1,4,7-triazacyclononane," *X-ray Struct. Anal. Online*, vol. 31, pp. 59-60, 2015.

[10] M. Mikuriya, S. Kida, S. Ueno, and I. Murase, "Crystal and Molecular Structure of a Triply-thiolate-bridged Binuclear Cobalt(III) Complex, $[Co_2\{SCH(CH_2CH_2NH_2)_2\}_3](ClO_4)_2 \cdot H_2O$," *Bull. Chem. Soc. Jpn.*, vol. 58, pp. 1857-1858, 1985.

[11] M. Mikuriya, S. Kida, and I. Murase, "Crystal and Molecular Structure of a Thiolate-bridged Binuclear Nickel(II) Complex, $[Ni_2\{SCH(CH_2CH_2NH_2)_2\}_2]Br_2$," *Bull. Chem. Soc. Jpn.*, vol. 60, pp. 1180-1182, 1987.

[12] M. Handa, M. Mikuriya, J. Z. Zhuang, H. Okawa, and S. Kida, "Synthesis, Properties, and Crystal Structure of Binuclear Nickel(II) and Palladium(II) Complexes of Tridentate Thiolic Ligands with NNS-Donor Set," *Bull. Chem. Soc. Jpn.*, vol. 61, pp. 3883-3887, 1988.

[13] M. Mikuriya, S. Shigematsu, M. Handa, and T. Kohzuma, "Redox properties of thiolate-bridged

binuclear nickel(II) complexes and correlation of their electronic spectra, *Transition Met. Chem.*, vol. 16, pp. 532-534, 1991.
[14] M. Mikuriya, F. Adachi, H. Iwasawa, M. Handa, M. Koikawa, and H. Okawa, "A novel thiolate-bridged polynuclear manganese(II) complex with 2-[2-(2-pyridyl)ethylamino]ethanethiol," *Inorg. Chim. Acta*,, vol. 179, pp. 3-5, 1991.
[15] M. Mikuriya, M. Handa, S. Shigematsu, S. Funaki, F. Adachi, and H. Okawa, "Thiolate-Bridged Binuclear Nickel(II) Complexes with a Mixed-Spin State," *Bull. Chem. Soc. Jpn.*, vol. 65, pp. 512-516, 1992.
[16] M. Mikuriya, M. Handa, S. Shigematsu, S. Funaki, T. Fujii, H. Okawa, K. Toriumi, T. Koshiba, and H. Terauchi, "Synthesis and Characterization of Thiolate-Bridged Dinuclear Nickel(II) Complexes with Thioether Pendant Arms," *Bull. Chem. Soc. Jpn.*, vol. 66, pp. 1104-1110, 1993.
[17] M. Mikuriya, F. Adachi, H. Iwasawa, M. Handa, M. Koikawa, and H. Okawa, "Syntheses and Characterization of Thiolate-Bridged Manganese(II) Complexes with NNS-Tridentate Thiolic Ligands," *Bull. Chem. Soc. Jpn.*, vol.67, pp. 3263-3270, 1994.
[18] M. Mikuriya, T. Kotera, F. Adachi, M. Handa, M. Koikawa, and H. Okawa, "Synthesis and Structural Characterization of Dinuclear and Trinuclear Iron(II) Complexes of Tridentate Thiolic Ligands with an NNS Donor Set," *Bull. Chem. Soc. Jpn.*, vol. 68, pp. 574-580, 1995.
[19] M. Mikuriya, H. Tsutsumi, R. Nukada, M. Handa, and Y. Sayama, "Synthesis and Characterization of Thiolato-Bridged Trinuclear Heterometal Complexes with 2-[(3-Aminopropyl)amino]ethanethiol," *Bull. Chem. Soc. Jpn.*, vol. 69, pp. 3489-3498, 1996.
[20] T. Kawahashi, H. Tsutsumi, and M. Mikuriya, "A Novel Tetranuclear Palladium(II) Complex with 2-[(3-Aminopropyl)amino]ethanethiol," *Polyhedron*, vol. 15, pp. 169-171, 1996.
[21] T. Kawahashi, M. Mikuriya, R. Nukada, and J.–W. Lim, "Synthesis and Structural Characterization of Thiolato-Bridged Tetranuclear Palladium(II) Complexes with *N*,*N*,*S*-Tridentate Ligands," *Bull. Chem. Soc. Jpn.*, vol. 74, pp. 323-329, 2001.
[22] M. Mikuriya, J. Xiao, S. Ikemi, T. Kawahashi, and H. Tsutsumi, "Synthesis and Structural Characterization of Thiolato-Bridged Zinc(II) Complexes with NNS-Tridentate Ligands," *Bull. Chem. Soc. Jpn.*, vol. 71, pp. 2161-2168, 1998.
[23] M. Mikuriya and T. Kotera, "Novel Thiolato-Bridged Tetranuclear Manganese(II) and Iron(II) Complexes with Adamantane-like Cores," *Chem. Lett.*, pp. 971-972, 1998.
[24] M. Mikuriya, J. Xiao, S. Ikemi, T. Kawahashi, H. Tsutsumi, A. Nakasone, and J. -W. Lim, "Synthesis and characterization of thiolato-bridged cadmium(II) complexes with NNS-chelating thiolic ligands," *Inorg. Chim. Acta*, vol. 312, pp. 183-187, 2001.
[25] T. Kotera and M. Mikuriya, "Thiolato-Bridged Hexanuclear $Cu^{I}Cu^{II}$ Mixed-Valence Complex," *Chem. Lett.* vol. 31, pp. 654-655, 2002.
[26] T. Kotera, A. Fujita, M. Mikuriya, and M. Handa, "Thiolao-bridged copper complexes with N,N,S-tridentate ligands," *Materials Science (Poland)*, vol. 21, pp. 171-179, 2003.
[27] T. Kotera, A. Fujita, M. Mikuriya, H. Tsutsumi, and M. Handa, "Cobalt complexes with 2-[(3-aminopropyl)amino]ethanethiol," *Inorg. Chem. Commun.*, vol. 6, pp. 322-324, 2003.
[28] T. Kotera, T. Sugimoto, and M. Mikuriya, "A thiolato-bridged Octanuclear Copper(I,II) Mixed-Valence Complex with N,N,S-Tridentate Ligand," *Chem. J. Moldova. General, Industrial and Ecological Chemistry*, vol. 2(1), pp. 102-107, 2007.
[29] M. Mikuriya, T. Watanabe, A. Suyama, and D. Yoshioka, "Rhenium Complex with 2-[(2-Pyridylmethyl)amino]ethanethiol," *X-ray Struct. Anal. Online*, vol. 31, pp. 3-4, 2015.
[30] M. Mikuriya, K. Kusunoki, T. Kotera, D. Yoshioka, and K. Ogasawara, "Synthesis, Crystal Str ucture, and DFT Calculation of a Dioxido-bridged Dinuclear Oxidomolybdenum(V) Complex with 2-(2-Aminoethyl)aminoethanethiol," *X-ray Struct. Anal. Online*, vol. 33, pp. 37-39, 2017.
[31] M. Mikuriya, K. Kusunoki, T. Kotera, D. Yoshioka, S. Takemura, and K. Ogasawara, "Synthesis, Crystal Structure, and Relativistic DV-Xα Calculation of a μ-Oxido-μ-molybdato(VI)-bridged Dinuclear Oxidomolybdenum(V) Complex with 2-(3-Aminopropyl)aminoethanethiol," *X-ray Struct. Anal. Online*, vol. 34, pp. 19-21, 2018.

[32] T. Kotera, A. Fujita, M. Mikuriya, D. Yoshioka, T. Shibahara, K. Nishimoto, and M. Handa, “Cobalt Thiolates with Organic *N*,*N*,*S*-Tridentate Ligand: Conversion of Thiolate Complex to Sulfinate Complex via Sulfenate,” in *IOP Conference Series: Materials Science and Engineering*, vol. 835, p. 012024, 2020.
[33] P. K. Mascharak, "Structural and functional models of nitrile hydratase," *Coord. Chem. Rev*., vol. 225, pp. 201-214, 2002.
[34] T. C. Harrop and P. K. Mascharak, "Fe(III) and Co(III) Centers with Carboxamide Nitrogen and Modified Sulfur Coordination: Lessons Learned from Nitrile Hydratase," *Acc. Chem. Res*., vol. 37, pp. 253-260, 2004.
[35] J. A. Kovacs, "Synthetic Analogues of Cysteinate-Ligated Non-Hem Iron and Non-Corrinoid Cobalt Enzymes," *Chem. Rev*., vol. 104, pp.825-848, 2004.
[36] W. Huang, J. Jia, J. Commings, M. Nelson, G. Schneider, and Y. Lindqvist, "Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold," *Structure***,** vol. 5, pp. 691-699, 1997.
[37] S. Nagashima, M. Nakasato, N. Dohmae, M. Tsujimura, K. Takio, M. Okada, M. Yoshida, N. Kamiya, and I. Endo, "Novel non-heme iron center of nitrile hydratase with a claw setting of oxygen atoms," *Nature Struct. Biol*., vol. 5, pp. 347-351, 1998.
[38] O. Kahn, *Molecular Magnetism*, VCH Publications, New York, U.S.A. 1993, pp.3-4, ISBN 1-56081-566-3.
[39] G. M. Sheldrick, "A short history of SHELX," *Acta Crystallogr. Sect A***,** vol. 64, pp.112-122, 2008.
[40] K. Nakanishi, P. H. Solomon, and N. Furutachi, *Infrared Absorption Spectroscopy* (in Japanese), Nankodo, Tokyo, Japan, 1978.
[41] K. Nakamoto, *Infrared and Raman Spectra of Inorganic and Coordination Compounds, Part B*, 6th ed., John Wiley & Sons, Hoboken, U.S.A., 2009, pp. 64-67, ISBN 978-0-471-74493-1.
[42] R. Grobelny, M. Melnik, and J. Mrozinski, *Cobalt Coordination Compounds: Clasification and Analysis of Crystallographic and Structural Data*, Dolnoslaskie Wydawnictwo Edukacyjne, Wroclaw, Poland, 1996.
[43] Y. Murakami and K. Sakata, "Electronic Spectra of Metal Complexes," in *Kireto Kagaku* (in Japanese), K. Ueno, Ed., Nankodo, Tokyo, Japan, 1976, vol.1, pp. 91-396. (in Japanese)

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210103-1 ~ 0210103-9
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8367

3D-Printed Absorbers for Solar-Driven Interfacial Water Evaporation: A Mini-Review

Leonard D. Tijing[1*], John Ryan C. Dizon[2,3], Gil G. Cruz Jr.[4]

[1]Centre for Technology in Water and Wastewater, School of Civil and Environmental Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney, 15 Broadway, Ultimo 2007 New South Wales, Australia

[2]Design, Research, Extension in Additive Manufacturing, Advanced Materials and Advanced Manufacturing (DR3AM) Center, Bataan Peninsula State University (Main Campus), City of Balanga, Bataan, 2100, Philippines

[3]Department of Industrial Engineering, College of Engineering and Architecture, Bataan Peninsula State University (Main Campus), City of Balanga, Bataan, 2100, Philippines

[4]Department of Civil Engineering, College of Engineering and Architecture, Bataan Peninsula State University (Main Campus), City of Balanga, Bataan, 2100, Philippines

*leonard.tijing@uts.edu.au

Abstract. Solar-driven interfacial water evaporation (SWE) is considered as a promising sustainable solution for clean water production especially for remote and off-grid communities. Various approaches have been developed in the last decade to improve the evaporation and thermal efficiency of the system, and to make it more robust for long-term operation. In recent years, 3D printing has emerged as an attractive method to fabricate simple and complex absorber geometries for SWE. In this mini-review, we present the new developments of 3D-printed solar absorbers including the various designs, fabrication strategies, challenges and opportunities. This study hopes to provide more insights into the use of additive manufacturing for improving the absorber design and performance of SWE.

Keywords: Solar water evaporation, 3D printing, desalination, 3D absorber

(Received 2021-04-09, Accepted 2021-04-30, Available Online by 2021-04-30)

1. Introduction

Population growth and rapid urbanization has resulted in increasing water scarcity issues around the

world. This has driven the rise of desalination technologies for the production of freshwater from unconventional yet abundant sources such as seawater and wastewater [1]. However, the current state-of-the-art desalination technologies such as reverse osmosis and multi-stage flash desalination are mostly centralized and are energy-intensive which are heavily based on fossil fuel use. In recent years, solar-driven interfacial water evaporation (SWE) has attracted increasing interest as a decentralized desalination technology that can produce freshwater, especially in remote areas, utilizing solar energy making it more sustainable [2]. SWE works via the process of evaporation-distillation, wherein a photothermal material floating on the source water is used that can absorb solar light and convert it into heat to evaporate the water at the absorber-water interface (see Figure 1 for schematic) [3]. The water vapor is then condensed back by a condenser and collected as clean water. This desalination technology is mainly passive, wherein no pumps or moving parts are needed, and is attractive for application especially for remote areas where sunlight is abundant and connection to electricity grid is a challenge [4].

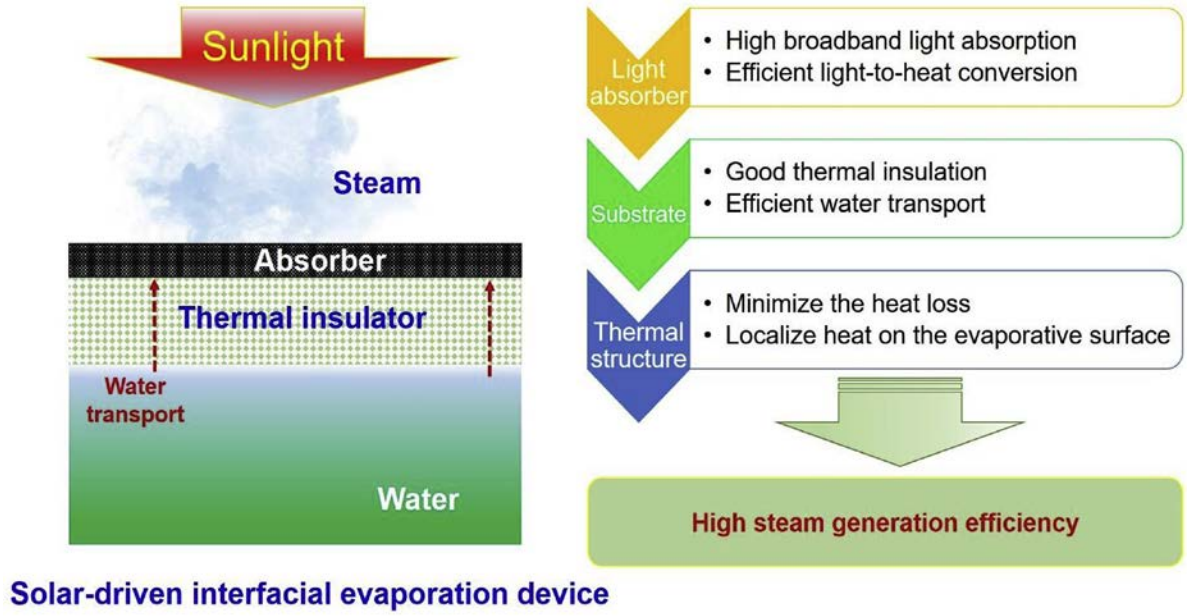


Figure 1. Schematic representation of the common solar-driven interfacial water evaporation system based on 2D-plane solar absorber and the effect of key components [3]

However, SWE still suffers from low water production and stability issues such as salt crystallization and heat losses for long-term use. The absorbers are either made of plasmonic nanoparticles [5], carbon-based materials [6], semiconductors [7], biomass [8], and polymeric materials. The most common design is 2D solar absorbers, where it is floated horizontally on top of the source water to enable evaporation, but still, they have low evaporation rate due to low overall effective surface area and high heat losses. To improve the performance, 3D absorber design has been considered which increases the overall surface area, and enables internal solar light reflection that can enhance the overall solar-thermal conversion efficiency [9]. In recent years, the 3D printing approach for the fabrication of 3D solar absorbers has attracted increased interest due to its ease of fabrication and the ability to make complex shapes with little geometrical constraints [10]. A number of new studies show good SWE performance of various 3D-printed absorber designs from cone-shapes, to hydratable plastics, to all-in-one evaporators, etc. Several reviews related to SWE have been published but none so far focusing on the potential of 3D printing for SWE. Thus, this mini-review will present the recent progress of the use and potential of 3D printing for the fabrication of absorbers for solar-driven interfacial water evaporation, and discuss its various challenges and opportunities.

2. 3D Printing Techniques for Solar-Driven Water Evaporation

3D printing is increasingly explored in the field of solar-driven interfacial water evaporation especially with the advent of new fabrication techniques and printable materials that are continuously developed. 3D printing offers a lot of potentials to ensure rapid prototyping of these materials [11]. In brief, 3D printing is a layer-by-layer fabrication technique that enables production of 3D structures according to a computer-aided design (CAD) model. Figure 2 shows the various steps in 3D printing process. Since it is based on a CAD model, it can basically fabricate any designs from simple to very complicated geometries [12]. 3D printing can be done in various ways such as by material extrusion, laser sintering, vat polymerization, direct-ink writing (DIW), and jetting [13]. The accuracy, resolution and productivity vary depending on the type of printer and virgin material (e.g., thermoplastics, metals, ceramics, etc).

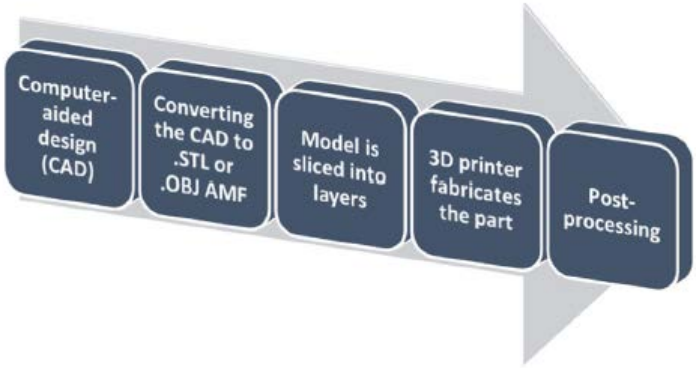


Figure 2. The five general steps in the 3D printing process [10].

In SWE, the absorbers are usually in the macro-scale and can also be designed to have microscale structures especially for water pathway channels. Material extrusion approach or fused deposition modeling (FDM) is one of the most common techniques used for SWE absorber preparation due to its fast processing and relatively cheap cost. A thermoplastic is usually used, and is melted and extruded in the printer following the 3D model design in the computer. There is an opportunity to combine the virgin material with photothermal nanoparticles before printing. This printing technique commonly requires post-processing. Another 3D printing method is digital light processing (DLP) (vat polymerization), wherein a thermoset photopolymer resin is cured upon exposure to an arc lamp. The light follows curing of pre-determined CAD model design at a single projection, making it fast and with good resolution. One study [14] employed DLP continuous 3D printing system to fabricate 3D absorber structures with surface distributed micropores for SWE. The base material was a polyacrylate-based resin that was incorporated with carbon nanotubes and a pore-former material, signifying the technique's versatility for use with composite resin material. In a recent study, DIW technique, a form of extrusion method, was utilized for the fabrication of a 3D hydratable plastic solar absorber. In DIW, the key consideration is to have a printable ink with the right rheological properties, i.e., is able to be extruded easily via the pneumatic air pressure from the 3D-printing system, and retain its shape immediately after extrusion without defects. Therefore, there is a need to play with the various fabrication parameters like ratio of polymer, solvent, and nanoparticles to achieve the right viscosity and consistency of the solution for writing. The size of the printed inks will also depend on the size of the nozzles. There is a great opportunity for this technique to use multi-materials by utilizing multi-nozzles with different solutions. The use of 3D printing in SWE is still at an early stage, but it looks very promising with still a lot of opportunities for further optimization and improvements.

3. Design and Performance of 3D-Printed Solar Absorbers

Recent studies have shown the positive enhancement in solar water evaporation performance when a 3D design is used for absorbers. Thus, 3D printing has come in handy for this application as it can

produce complex shapes with high accuracy and scalability. Among the various 3D printing techniques, vertical printing, stereolithography, digital light processing (DLP), and extrusion methods have been so far used for SWE. This section will discuss some of the latest developments in terms of design and SWE performance of 3D-printed solar absorbers in literature.

One of the main issues for 2D flat solar absorbers is the direct contact of the absorber to the water surface, which can lead to high conduction heat loss and salt accumulation. Thus, some groups have tried isolating the absorber from water, which resulted to improvements in efficiency, however, it still suffers from reflective light losses due to the 2D-plane geometry of the absorber [3]. To address this issue, 3D absorber designs have been tested with isolated absorber configuration. However, most of the reported 3D designs are just simple structures made from folding of materials or just simple fabrication methods, which has also limitations on their stability and the possible designs that can be prepared. Hence, 3D printing poses a great promise in the development of novel absorbers without any geometrical design limitations.

This was the approach taken by Cao et al. [15] when they fabricated a tree-inspired, cone-shape Janus solar absorber via 3D printing of acrylonitrile-styrene-acrylate (ASA) copolymer for SWE application. The 3D-printed cones were designed with various apex angles from 30-90^{o}, which could easily be fabricated by 3D printing. The hydrophobic ASA itself acted as the solar absorber, which had a solar absorptivity of ~99.80% across a wide solar wavelength (200-2,200 nm). A hydrophilic air-laid paper was wrapped around the outside of the cone, forming a dual-wetting layer or a Janus design, i.e., an outer hydrophilic part, and an inner hydrophobic part, essentially isolating the evaporating surface from the water wicking surface. The lower narrow end of the cone was immersed in the source water, with an expanded polystyrene (EPS) foam wrapped around as a floater. This configuration enables the continuous wicking of water to the upper part of the cone via the air-laid paper, leading to water evaporation upon light-to-heat conversion at the absorber part. Under 1 sun illumination, the solar evaporation rate reached 1.713 kg/m^{2}h which was stable for 14 days utilizing 3.5wt% NaCl solution. This high performance was attributed to the excellent heat localization effect, constant water supply to the absorber, and self-cleaning ability from salt crystallization.

Figure 3. Solar desalination and durability of the biomimetic 3D evaporator. **a.** Scheme of the batch purification prototype which simulates the practical solar water purification apparatus using 3D evaporator. **b.** Measured concentrations of four primary ions in the actual seawater sample from Jiaozhou Bay, the Yellow Sea before and after desalination. Orange and green columns represent the metal ion concentrations before and after purification, respectively (images adapted from [14]).

Another study [14] utilized a biomimetic 3D cone-shape structure inspired from bird beak and pitcher plant peristome surface (see Figure 3). This cone-shape absorber was fabricated by size-refilling dependent DLP 3D printing of a polyacrylate-based UV curable composite resin containing carbon

nanotubes (CNTs) and citrate sodium powder. The incorporated CNTs act as the photothermal material, while citrate sodium powder acts as pore inducer. The 3D printed parts with surface distributed micropores were then subjected to plasma post-treatment to endow hydrophilic property to the surface. Asymmetric groves were provided along the cone that allow a continuous water suction pathway. The presence of the surface micropores together with the gradient microgravity array was found to help in the fast suction of water, even faster than a porous filter paper as a comparison. The 3D cone absorber with absorptivity of >90% of input light was floated with the wide base touching the water, and SWE experiments were conducted at 1 sun irradiation. Results showed high evaporation rate of 1.72 kg/m^{2}h in a closed system with 3.5 wt% NaCl solution for the biomimetic 3D cone absorber compared to 1.07 kg/m^{2}h for a 2D plane absorber fabricated with the same materials. In an open system and at 25 wt% NaCl solution, the evaporation rate was found to be 2.63 kg/m^{2}h with an energy efficiency >96%. Interestingly at this high NaCl solution, salt crystallization only happened near the apex of the cone, and was easily detached. This high performance was attributed to the uneven heating and evaporation due to the cone structure, where near the apex part showed higher temperature due to thinner water film and faster evaporation, compared to the bottom wider part of the cone. In addition, the gradient liquid film thickness along the grooves also enabled constant water pumping and evaporation that helps in the continuous SWE process. The 3D cone design was found superior in SWE performance when compared to a 3D columnar structure with similar grooves and micropores as a fabricated control. The precise fabrication via 3D printing of the 3D cone absorber and its unique characteristics with grooves and micropores has definitely helped in the excellent performance of the SWE.

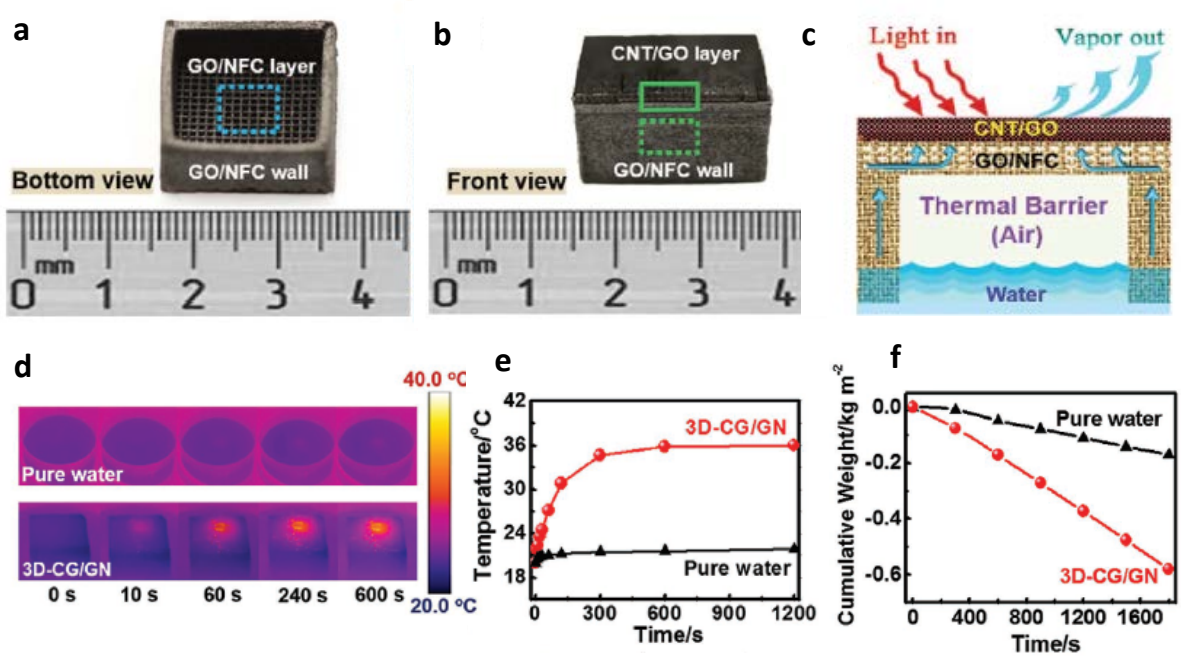


Figure 4. Photographic images of the 3D-printed evaporators: (a) bottom view, (b) front view; (c) schematic illustration of the SWE system; (d) infrared images of pure water only and that with the 3D all-in-one evaporator upon 1 sun illumination; (e) graph showing changes in surface temperature of pure water only and with the 3D evaporator with respect to time; (f) Mass change of water over time under solar illumination for pure water only and for the 3D evaporator (images are adapted from [16]).

Li et al [16] utilized an extrusion 3D printing approach to fabricate an all-in-one solar evaporator for SWE. The design was a box-like evaporator with one open side (see Figure 4a and 4b). The upper part was made of CNT/graphene oxide (GO) layer that serves as the photothermal layer. Then below it is a GO/nanofibrillated cellulose (GO/NFC) layer with mesh-like structure that acts as a support and a water-

wicking channel. In addition, four walls of GO/NFC were also 3D-printed, which contributes to the suction of the water towards the top photothermal material. All the parts were 3D-printed and formed as one structure (see Figure 4). The CNT/GO photothermal layer obtained high solar absorptivity (>97%), and together with the high porosity of the absorber (~97.3%), this resulted to an evaporation rate of 1.25 kg/m^{2}h under 1 sun illumination (see Figure 4f). This good performance was attributed to the integrated structure of the CNT/GO layer and the GO/NFC layer, ensuring a continuous supply of water to the CNT/GO layer, while facilitating heat localization and effective solar light-thermal conversion.

Another study [17] by the same group designed a jellyfish-like solar absorber fabricated by vertical 3D printing technique. Carbon black/GO (CB/GO) and highly-concentrated GO (HC-GO) solutions were prepared and extruded using a 3D printer. First, the CB/GO layer was printed into a mosquito coil-like shape, and then HC-GO pillars at specified distances were vertically printed. After printing, the materials were placed in a freeze dryer and then annealed at 150^{o}C to obtain the stable printed solar absorber structure. The integrated CB-GO with GO pillars structure was then embedded into an EPS matrix with matched holes for the GO pillars. The CB-GO layer serves as the photothermal material (high absorption of 99% at wide wavelength), while the GO pillars act as directed water transport channel (as they are hydrophilic due to functional groups), and the EPS foam is the insulating support and floater. Due to the directed water channel, there was less contact area between the absorber and the bulk water, thus decreasing the heat losses. This unique 3D-printed design displayed good energy conversion efficiency of 87.5% under one sun irradiation, and obtained an evaporation rate of 1.27 kg/m^{2}h. The jellyfish-like design is interesting but only achieved lower than the theoretical evaporation rate limit perhaps due to the lower water pumping capacity from the few GO pillars.

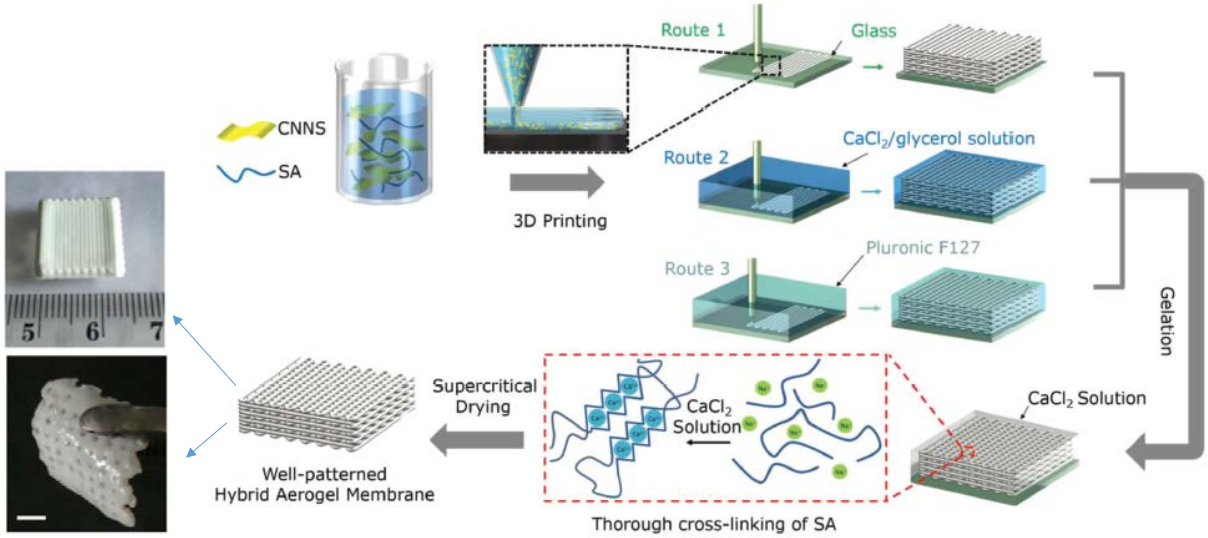


Figure 5. Schematic of the preparation and fabrication of the 3D-printed hybrid aerogel solar absorber. Photographic images of the fabricated materials can be seen on the two lower left images (adapted from [18]).

Due to the challenge of controlled assembly of 2D building blocks, He et al. [18] utilized DIW technique to fabricate a carbon nitride-based hybrid aerogel with patterned microscopic structures as solar absorber for SWE. Various printing approaches were tested to check the viability of the DIW technique (see Figure 5). The resultant hybrid aerogel showed high visible light absorption and obtained 2.5 times greater activity compared to the control sample. Another interesting recent study by Koh et al. [19] also utilized DIW 3D printing technique to fabricate hydratable plastics made from amorphous regenerated cellulose with various designs. The hydratable plastics have two main components, the water transporting substrate (cellulose acetate), and the light absorbing surface (cellulose acetate with

carbon black nanoparticles), both of which were 3D printed. The substrate was designed to have woodpile structure, formed by alternating perpendicular stacked layers with square pores to allow water wicking channels. The light absorber layer was also designed into two ways, one as a plate-like solid material, and the other one is a criss-cross woodpile structure. Results revealed that using 3D-printing technique has significantly improved the rehydration rate and exhibited long-term stability and anti-salt-fouling ability in evaporating saline water. The evaporation rate was 3.01 kg/m^{2}h at 1 sun irradiation, which was way higher than the theoretical limit. This high performance was attributed to the ability of the hydratable plastic for vaporization enthalpy reduction, mainly due to the weaker cellulose-water hydrogen bonds and more deviation from non-linearity. At much concentrated solar irradiation (3 sun), it achieved >7 kg/m^{2}h evaporation rate, which is the highest so far reported at such light intensity. This study signifies that with proper architectural design via 3D printing, rehydration rate can be improved which can lead to enhanced SWE performance.

Based from literature, only few studies have yet used 3D printing for solar absorber fabrication and SWE application, but the trend is starting to increase. Compared with membrane fabrication for other desalination and water treatment processes, the structures for SWE absorber do not generally need very high printing resolution (micro to mm scale is enough), thus it presents good potential as a fabrication technique. Besides, there is good accuracy with 3D printing, which is very beneficial for the proper design and manufacture of absorbers leading to high evaporation rate performance as indicated in the studies mentioned above.

4. Challenges and Opportunities

Up until now, water shortage is a serious global issue that urgently needs to be addressed [16]. Hence, a technology that allows freshwater generation and offers reliability and efficiency are of great interest in dealing with this global water crisis, especially for people living in off-grid areas [20]. While there are emerging techniques being employed, there remain several challenges that require further study to fully understand the complexity of solar-driven interfacial water evaporation using 3D printing technology. 3D printing promotes sustainable manufacturing as it has many applications related to water and the environment. Generally, 3D printing has lesser waste compared with other manufacturing methods and highly complex/intricate designs may be produced. Though 3D printing looks promising for application in SWE, there are still some challenges that need to be addressed.

As absorbers are constantly subject to light irradiation, there is a need to further develop new materials that are chemically, thermally and mechanically robust and will not degrade easily for long-term operation and that can be 3D printed. Examples of such are high performance polymers to increase strength of materials for both solar absorption and water transport. The main consideration is the strength, weight, surface finish, etc. Strength is needed when cleaning the part/material. 3D printing is still limited in its capability to print at ultra-high resolution, though much more advanced printers are now being introduced. The capability to accurately print micro to nano roughness and pores would be very promising for SWE especially on providing precise wicking channels or surface roughness/grooves. Generally, 3D printing is still relatively more costly compared to other conventional fabrication techniques primarily due to slower fabrication times and the need for specific materials for printing [21]. This is also related to the upscaling capacity when printing large amounts or big size solar evaporators. The issue of safety during 3D printing due to emission of particulate matters and volatile organic compounds is an issue that still needs to be addressed as its environmental and health impact is still not yet fully understood at this time.

Aside from these challenges, there are great opportunities for further development and study of additive manufacturing for SWE. One is the potential of 4D-printed materials, where the solar absorber can be designed with stimuli-responsive materials [22] enabling the opening or closing of water wicking channels when exposed to stimuli such as temperature, pH, etc. For example, it would be great to see a solar absorber with pores that open wide at lower temperature especially at night time when sun is not shining, which can lead to re-migration of precipitated salts to the bulk water for dissolution and maintain a clean surface. Another opportunity for research is the use of composite solutions for 3D

printing. Many of the 3D printing works are based on single-material fabrication, but the versatility of 3D printing allows combination of materials [23] especially for SWE, where photothermal nanoparticles are usually needed in the polymeric matrix. 3D printing is a promising approach not only for the production of solar absorbers but also as a way to fabricate all-in-one evaporators or parts of SWE in an integrated design. Its potential is enormous especially if newer high-resolution 3D printers will become available, and increased accuracy and versatility in base material are used.

5. Conclusion

This study presents the various progress and developments of the use of 3D printing for absorber preparation in solar-driven interfacial water evaporation applications. So far, only few studies are available in literature utilizing 3D printing for SWE, but it presents an enormous potential. It can be used for the efficient and fast fabrication of solar absorbers, by the direct printing of polymeric absorbers, or printing of resins incorporated with photothermal materials. Additionally, all-in-one evaporators can also be directly prepared by 3D printing, with advantage of robust integrated design, and capability to fabricate complex structures. Results in literature show enhanced water evaporation performance of properly-designed 3D-printed absorbers, even surpassing the theoretical limit. However, it is also acknowledged that there are various challenges that still need to be addressed such as resolution limitation, overall cost, and health and environmental safety. Overall, 3D printing presents itself as a very exciting technique for rapid fabrication of efficient solar absorbers/evaporators for SWE, and further studies are ought to be done in this area.

References

[1] M. A. Shannon, P. W. Bohn, M. Elimelech, J. G. Georgiadis, B. J. Mariñas, and A. M. Mayes, "Science and technology for water purification in the coming decades," in *Nanoscience and Technology*, pp. 337-346.

[2] I. Ibrahim, D. H. Seo, A. M. McDonagh, H. K. Shon, and L. Tijing, "Semiconductor photothermal materials enabling efficient solar steam generation toward desalination and wastewater treatment," *Desalination*, vol. 500, p. 114853, 2021.

[3] V.-D. Dao, N. H. Vu, and S. Yun, "Recent advances and challenges for solar-driven water evaporation system toward applications," *Nano Energy*, vol. 68, p. 104324, 2020.

[4] L. Zhu, M. Gao, C. K. N. Peh, and G. W. Ho, "Recent progress in solar-driven interfacial water evaporation: Advanced designs and applications," *Nano Energy*, vol. 57, pp. 507-518, 2019.

[5] Y. Shi, C. Zhang, Y. Wang, Y. Cui, Q. Wang, G. Liu, S. Gao, and Y. Yuan, "Plasmonic silver nanoparticles embedded in flexible three-dimensional carbonized melamine foam with enhanced solar-driven water evaporation," *Desalination*, vol. 507, p. 115038, 2021.

[6] T. Yang, H. Lin, K.-T. Lin, and B. Jia, "Carbon-based absorbers for solar evaporation: Steam generation and beyond," *Sustainable Materials and Technologies*, vol. 25, p. e00182, 2020.

[7] S. Zuo, D. Xia, Z. Guan, F. Yang, S. Cheng, H. Xu, R. Wan, D. Li, and M. Liu, "Dual-functional CuO/CN for highly efficient solar evaporation and water purification," *Separation and Purification Technology*, vol. 254, p. 117611, 2021.

[8] I. Ibrahim, V. Bhoopal, D. H. Seo, M. Afsari, H. K. Shon, and L. D. Tijing, "Biomass-based photothermal materials for interfacial solar steam generation: A review," *Materials Today Energy*, p. 100716, 2021.

[9] D. N. Thoai, Q. T. Hoai Ta, T. T. Truong, H. Van Nam, and G. Van Vo, "Review on the recent development and applications of three dimensional (3D) photothermal materials for solar evaporators," *Journal of Cleaner Production*, vol. 293, p. 126122, 2021.

[10] L. D. Tijing, J. R. C. Dizon, I. Ibrahim, A. R. N. Nisay, H. K. Shon, and R. C. Advincula, "3D printing for membrane separation, desalination and water treatment," *Applied Materials Today*, vol. 18, p. 100486, 2020.

[11] H. A. Balogun, R. Sulaiman, S. S. Marzouk, A. Giwa, and S. W. Hasan, "3D printing and surface imprinting technologies for water treatment: A review," *Journal of Water Process Engineering*, vol. 31, p. 100786, 2019.
[12] J. R. C. Dizon, A. H. Espera, Q. Chen, and R. C. Advincula, "Mechanical characterization of 3D-printed polymers," *Additive Manufacturing*, vol. 20, pp. 44-67, 2018.
[13] A. D. Valino, J. R. C. Dizon, A. H. Espera, Q. Chen, J. Messman, and R. C. Advincula, "Advances in 3D printing of thermoplastic polymer composites and nanocomposites," *Progress in Polymer Science*, vol. 98, p. 101162, 2019.
[14] L. Wu, Z. Dong, Z. Cai, T. Ganapathy, N. X. Fang, C. Li, C. Yu, Y. Zhang, and Y. Song, "Highly efficient three-dimensional solar evaporator for high salinity desalination by localized crystallization," *Nature Communications*, vol. 11, no. 1, p. 521, 2020.
[15] N. Cao, S. Lu, R. Yao, C. Liu, Q. Xiong, W. Qin, and X. Wu, "A self-regenerating air-laid paper wrapped ASA 3D cone-shaped Janus evaporator for efficient and stable solar desalination," *Chemical Engineering Journal*, vol. 397, p. 125522, 2020.
[16] Y. Li, T. Gao, Z. Yang, C. Chen, W. Luo, J. Song, E. Hitz, C. Jia, Y. Zhou, B. Liu, B. Yang, and L. Hu, "3D-Printed, All-in-One Evaporator for High-Efficiency Solar Steam Generation under 1 Sun Illumination," *Advanced Materials*, vol. 29, no. 26, p. 1700981, 2017.
[17] Y. Li, T. Gao, Z. Yang, C. Chen, Y. Kuang, J. Song, C. Jia, E. M. Hitz, B. Yang, and L. Hu, "Graphene oxide-based evaporator with one-dimensional water transport enabling high-efficiency solar desalination," *Nano Energy*, vol. 41, pp. 201-209, 2017.
[18] P. He, X. Tang, L. Chen, P. Xie, L. He, H. Zhou, D. Zhang, and T. Fan, "Patterned Carbon Nitride–Based Hybrid Aerogel Membranes via 3D Printing for Broadband Solar Wastewater Remediation," *Advanced Functional Materials*, vol. 28, no. 29, p. 1801121, 2018.
[19] J. J. Koh, G. J. H. Lim, S. Chakraborty, Y. Zhang, S. Liu, X. Zhang, S. C. Tan, Z. Lyu, J. Ding, and C. He, "Robust, 3D-printed hydratable plastics for effective solar desalination," *Nano Energy*, vol. 79, p. 105436, 2021.
[20] C. Chen, Y. Kuang, and L. Hu, "Challenges and Opportunities for Solar Evaporation," *Joule*, vol. 3, no. 3, pp. 683-718, 2019.
[21] M. N. Nadagouda, M. Ginn, and V. Rastogi, "A review of 3D printing techniques for environmental applications," *Current Opinion in Chemical Engineering*, vol. 28, pp. 173-178, 2020.
[22] S. Ma, Y. Zhang, M. Wang, Y. Liang, L. Ren, and L. Ren, "Recent progress in 4D printing of stimuli-responsive polymeric materials," *Science China Technological Sciences*, vol. 63, no. 4, pp. 532-544, 2020.
[23] M. A. Skylar-Scott, J. Mueller, C. W. Visser, and J. A. Lewis, "Voxelated soft matter via multimaterial multinozzle 3D printing," *Nature*, vol. 575, no. 7782, pp. 330-335, 2019.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210104-01 ~ 0210104-10
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8385

Current and Voltage Imbalances at BPTIK Universitas PGRI Semarang

Margono[1*], Muhammad Amiruddin[1], Irna Farikhah[2]

[1]Department of Electrical Engineering, Faculty of Engineering and Informatics, Universitas PGRI Semarang, Jl. Jalan Sidodadi Timur No.24 Semarang, Central Java Indonesia

[2]Department of Mechanical Engineering, Faculty of Engineering and Informatics, Universitas PGRI Semarang, Jl. Jalan Sidodadi Timur No.24 Semarang, Central Java Indonesia

*margono.27@gmail.com

Abstract. Energy conservation is an effort to increase energy efficiency. This research conducted an energy audit in the BPTIK (Badan Pengembangan Teknologi Informasi dan Komunikasi) department room at the Universitas PGRI Semarang, observing the value of electrical parameters, especially unbalanced voltage and current that have the potential to cause neutral currents. In this study the calculated unbalanced voltage value is 0.64% and unbalanced current is 20.25%. From the current development, the neutral current produces an average of 1.72 Amperes. From the value of the neutral current can be calculated the energy lost in the conductor neutral wire per day is about 0.01% of energy consumption by the load per day.

Keywords: current, voltage, imbalance

(Received 2021-04-10, Accepted 2021-04-30, Available Online by 2021-04-30)

1. Introduction

Energy conservation is an effort to optimize energy use (increase energy efficiency) in a company with the ultimate goal of increasing economic, social and environmental added value. Basically, energy audit is one of the efforts in energy conservation and the environment that focuses on survey activities to obtain data and information that explains the portrait of energy use, and whether there are opportunities to improve energy quality.

The dominant factors in the consideration of energy efficiency and electrical power quality in the

distribution of electricity in buildings are harmonic distortion and imbalance distortion [3]. According to Hossein [2], his research suggests current imbalance and voltage imbalance is a serious problem in the distribution of low-voltage level three-phase systems. In a three-phase system the cause of current imbalance is due to load imbalance, and current imbalance is the main cause of voltage imbalance. The task of balancing currents and loads is the responsibility of both parties, namely electricity providers and consumers. Therefore, it is necessary to calculate special rates based on imbalance.

The importance of the quality of electric power, seen from several studies that have been done before, is the investigation of the quality of electric power in the PGRI University of Semarang. For this research, a study of the electricity network will be carried out at the Main Building of the PGRI University in Semarang, especially on the 3rd floor in The Unit of Technical Information and Communication (BPTIK), due to observations, there are many 1-phase loads such as PC (Personal Computer), servers, lighting and AC (Air Conditioner) installed on network with 3 phase sources. In addition, in the BTIK department, there is a UPS (Uninteruptible Power Supply) which is risky against neutral voltage to ground. To maintain the continuity of the performance of the BPTIK department of the University of PGRI Semarang, it is necessary to conduct an energy audit on the electricity network. In a previous study by Wanimbo [5] located in the same place, the BPTIK room IKE (Energy Consumption Intensity) was 8.11 kWh per month / m2 and calculated daily energy consumption in the measurement time range of 43.29 kWh. This paper compares the amount of energy consumption by the load in the room with the energy lost in the neutral conductor as a result of the imbalance.

2. Methods

This research was conducted in several stages, starting with the observation of the electrical panel in the field, followed by measuring and recording data. After the data recording process is continued with data processing related to current and voltage imbalance parameters. After processing the data, it is continued with drawing conclusions for recommendations for improving the electricity network.

2.1. Measurement and Data Record

Measurements were made on the BPTIK power panel connected to a 3-phase UPS (Uniteruptible Power Supply) that provides energy for the BPTIK room using the Scheneider PM5350. The measured parameters are the current and voltage per phase and recorded for approximately 24 hours with data retrieval every 5 minutes. The results of the current measurement data record per phase are as shown in Figure 1.

As can be seen in Figure1 The highest rated phase current T and the lowest rated phase current R throughout the measurement time range. From this data, the current values in the three phases are obtained as in Table I below. For the results of recording the voltage measurement data per phase as shown in Figure 2. For the results of recording the voltage measurement data per phase as shown in Figure 3. From the voltage per phase data shown in Figure 2, the stress values in the three phases are obtained as in Table 2.

Table 1. Flow rate per phase

Value	Voltage per phase (A)		
	R	**S**	**T**
Maximum	4.96	5.55	7.48
Minimum	3.22	4.10	4.40
Average	3.85	4.52	5.58

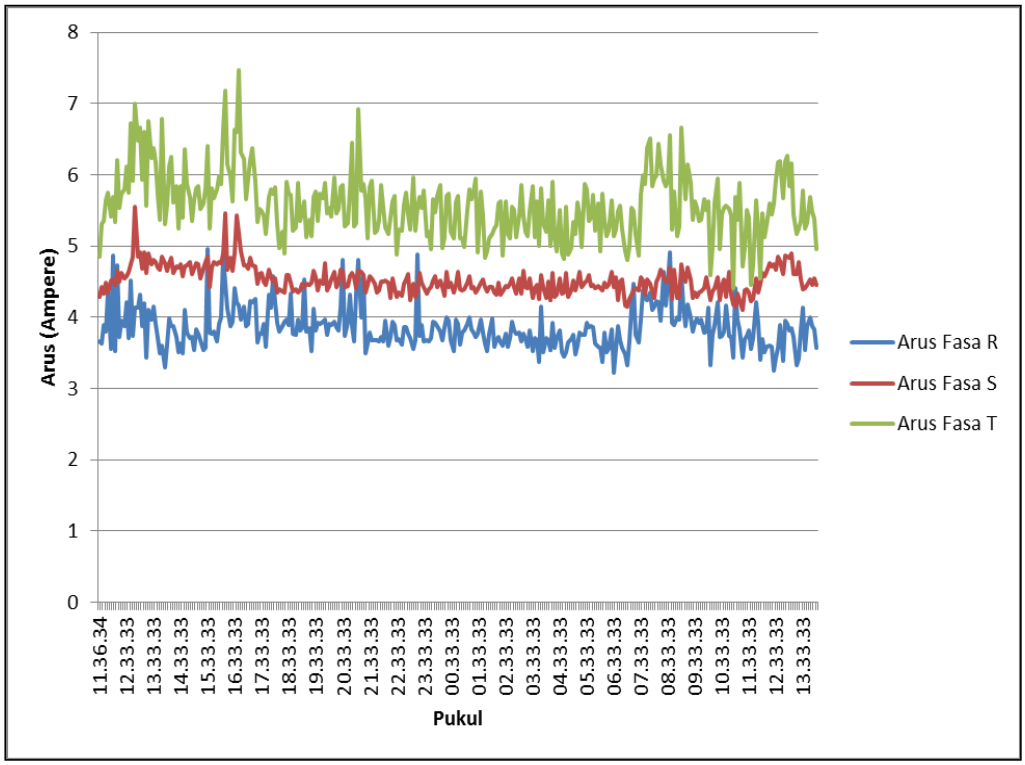


Figure1. Graph of current measurement data per phase

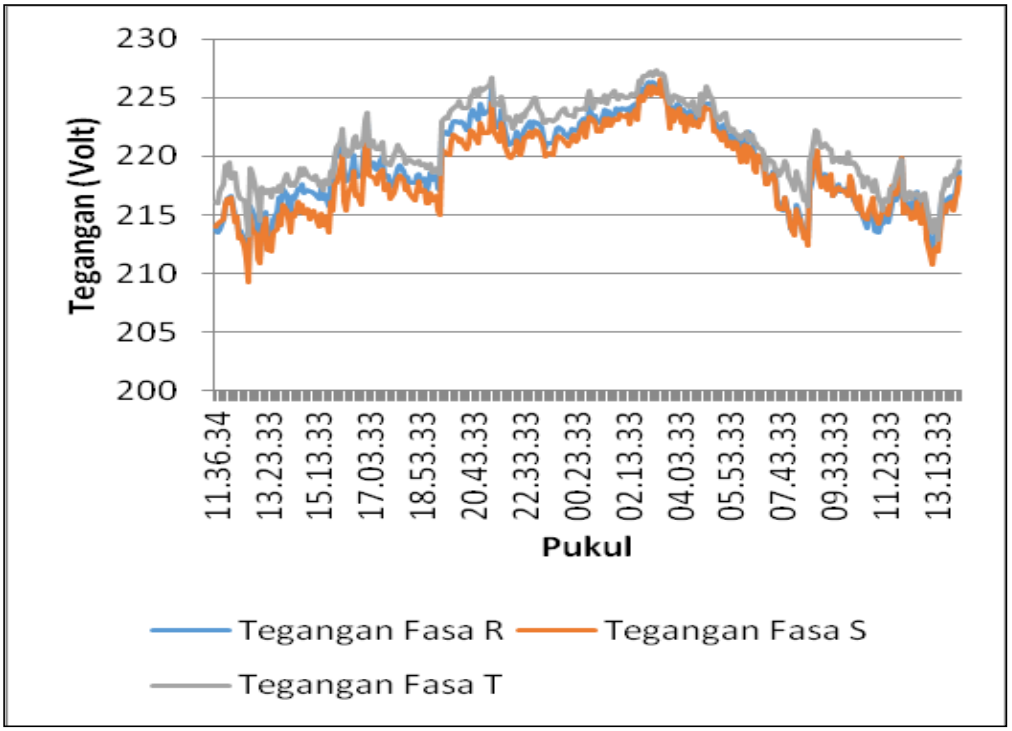


Figure 2. Graph of voltage measurement data per phase

Table 2. Voltage per phase against neutral

Value	Voltage per phase (V)		
	R	**S**	**T**
Maximum	226.4	226.59	227.35
Minimum	210.38	209.31	212.94
Average	219.33	218.58	220.97

2.2. Data Processing

2.2.1. Voltage Imbalance

At 11:38:33 WIB the value of the voltage parameter in each phase towards neutral is as shown in Table 3. At the data processing stage, the three parameters are calculated, the value of the voltage imbalance, current imbalance and neutral current. Therefore, the calculated average voltage for each phase to neutral is 214.58 Volts. The largest difference between the value of the voltage per phase and the average voltage per phase at that time is 1.38 volts occurs in phase T. By calculating the imbalance with the IEEE Std 112 standard in equation 2.6, the voltage imbalance is calculated as equation 1.

$$\text{Voltage Imbalance} = \frac{1.38\ Volt}{214.58\ Volt} \times 100\% = 0.64\ \% \quad (1)$$

From all measurement points are calculated, then the result of the stress imbalance calculation is plotted in a graph as shown in the graph in Figure 3 as follows. From the calculation of the voltage imbalance, it is found that the average voltage imbalance is 0.64%.

Table 3. Value of voltage per phase to neutral

Time	Voltage per phase for neutral (V)			
	R	S	T	rata-rata
11:38:33	213.5	214.28	215.96	214.58

2.2.2. Current Imbalance

At 11:38:33 WIB the value of the voltage parameter in each phase to neutral is shown in Table 4. So that the average current for each phase to neutral is calculated at 4.45 Amperes. The largest difference between the current value per phase and the average current per phase at that time is 0.85 Ampere occurs in phase T. By calculating the imbalance with the IEEE Std 112 standard in equation 2.6, the current imbalance is calculated as,

$$\text{Current Imbalance} = \frac{0.85\ ampere}{4.45\ ampere} \times 100\% = 19.09\% \quad (2)$$

From all measurement points are calculated, then the calculation results of the current imbalance are plotted in a graph as shown in the graph in Figure 4 as follows: From the current imbalance calculation data, the results of the calculation of the average current imbalance are 20.25%.

Tabel 4. Flow value per phase against neutral

Time	Voltage per phase for neutral (V)			
	R	S	T	rata-rata
11:38:33	3.63	4.42	5.3	4.45

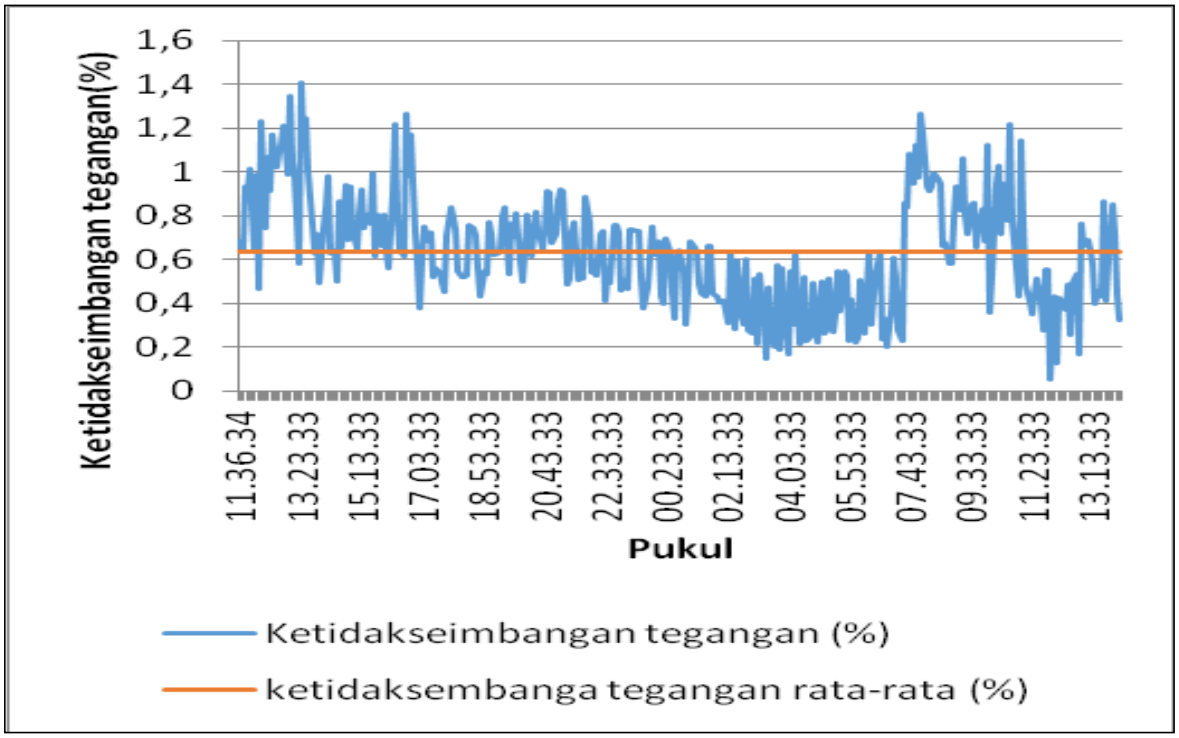


Figure 3. Graph of voltage imbalance between phases

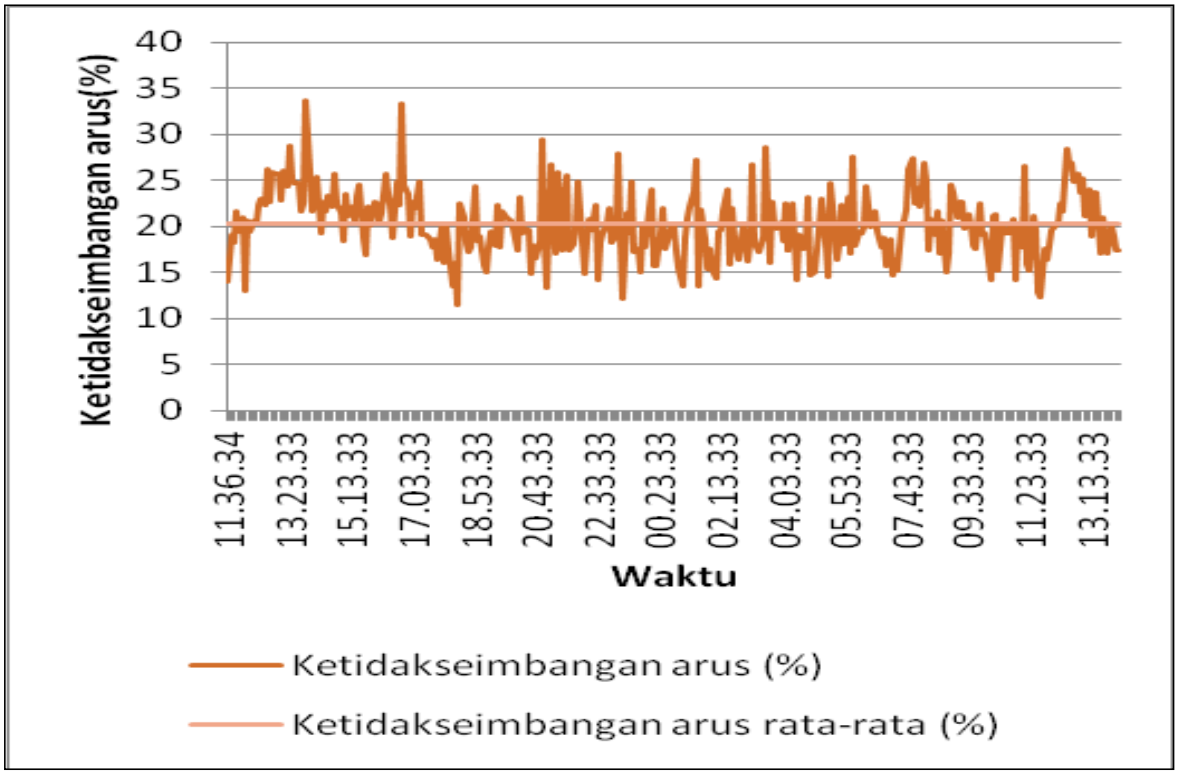


Figure 4. Graph of current imbalance between phases

2.2.3 Neutral Current

From the current imbalance that occurs between the phases, it causes a neutral current which is illustrated in the graph in Figure 5, which is the resultant current vector between the three phases as in formula 2.2 above, each of which has a direction and angle as shown in Figure 6. We get the direction and angle from the calculation of the power factor value. It is known in the 11:38:33 measurement that the values are obtained in Table 5.

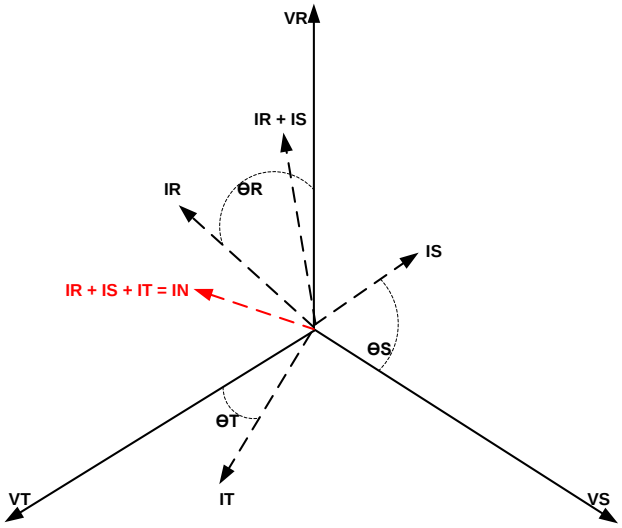


Figure 5. Neutral current vector calculation

Table 5. Flow value and power factors per phase

Time	Voltage per phase (A)			Power factors		
	R	S	T	R	S	T
11:38:33	3.63	4.42	5.30	0.53	0.43	0.64

From this value it can be calculated the value of the angle of the current and the voltage per phase, that is $\theta_R = cos^{-1}\ 0.53$, or the phase angle R $\theta_R = 57.85°$, while the phase angle S is obtained from $\theta_S = cos^{-1}\ 0.43$ or $\theta_S = 64.75°$. In the same way, to find the value of the phase angle T is $\theta_T = cos^{-1}\ 0.64\ or\ \theta_T = 50.33°$. Then it can be calculated the sum of the current vector of phase R and current of phase S as follows

$$\left|\overrightarrow{I_R} + \overrightarrow{I_S}\right| = \sqrt{{I_R}^2 + {I_S}^2 + 2I_R I_s \cos(120 + \theta_R - \theta_S)} \quad (3)$$

$$\left|\overrightarrow{I_R} + \overrightarrow{I_S}\right| = \sqrt{3.63^2 + 4.42^2 + 2 \times 3.63 \times 4.42\ \times \cos(113.09)} \quad \left|\overrightarrow{I_R} + \overrightarrow{I_S}\right| = 4.48\ ampere$$

From this value, the value of the current angle and the voltage per phase can be calculated, that is $\theta_R = cos^{-1}\ 0.53$, or phase angle R $\theta_R = 57.85°$, while the phase angle S is obtained from $\theta_S = cos^{-1}\ 0.43$ or $\theta_S = 64.75°$. In the same way, to find the value of the phase angle T is $\theta_T = cos^{-1} 0.64$ or $\theta_T = 50.33°$. Then it can be calculated the sum of the current vector of phase R and current of phase S as follows

$$\left|\overrightarrow{I_R} + \overrightarrow{I_S}\right| = \sqrt{{I_R}^2 + {I_S}^2 + 2I_R I_s \cos\,(120 + \theta_R - \theta_S)} \quad (4)$$

$$|\vec{I_R}+\vec{I_S}| = \sqrt{3.63^2 + 4.42^2 + 2 \times 3.63 \times 4.42 \times \cos(113.09)} \qquad |\vec{I_R}+\vec{I_S}| = 4.48\ ampere$$

Calculated the amount of current R and S equal to 4.48 ampere and form an angle θ_{RS} namely the intermediate angle $\vec{I_R}+\vec{I_S}$ to $\vec{I_S}$.

$$\frac{|\vec{I_R}+\vec{I_S}|}{\sin(120+\theta_R-\theta_S)} = \frac{|\vec{I_R}|}{sin\theta_{RS}} \tag{5}$$

$$sin\theta_{RS} = \frac{|\vec{I_R}|\sin(120+\theta_R-\theta_S)}{|\vec{I_R}+\vec{I_S}|}$$

$$sin\theta_{RS} = \frac{3.63\sin(113.09)}{4.48}$$

$$sin\theta_{RS} = 0.905$$

$$\theta_{RS} = 64.93°$$

Then proceed with calculating the value of the neutral current I_N.

$$|\vec{I_N}| = \sqrt{|\vec{I_R}+\vec{I_S}|^2 + {I_T}^2 + 2|\vec{I_R}+\vec{I_S}|I_T\cos(120+\theta_{RS}+\theta_S-\theta_T)} \tag{6}$$

$$|\vec{I_N}| = \sqrt{4.48^2 + 5.30^2 + 2 \times 4.48 \times 5.30 \times \cos(199.36)}$$

$$|\vec{I_N}| = 1.85\ ampere$$

All current and power factor measurement data points during the measurement timeframe are used for the calculation of neutral currents and are plotted in a graph in Figure 6.

.

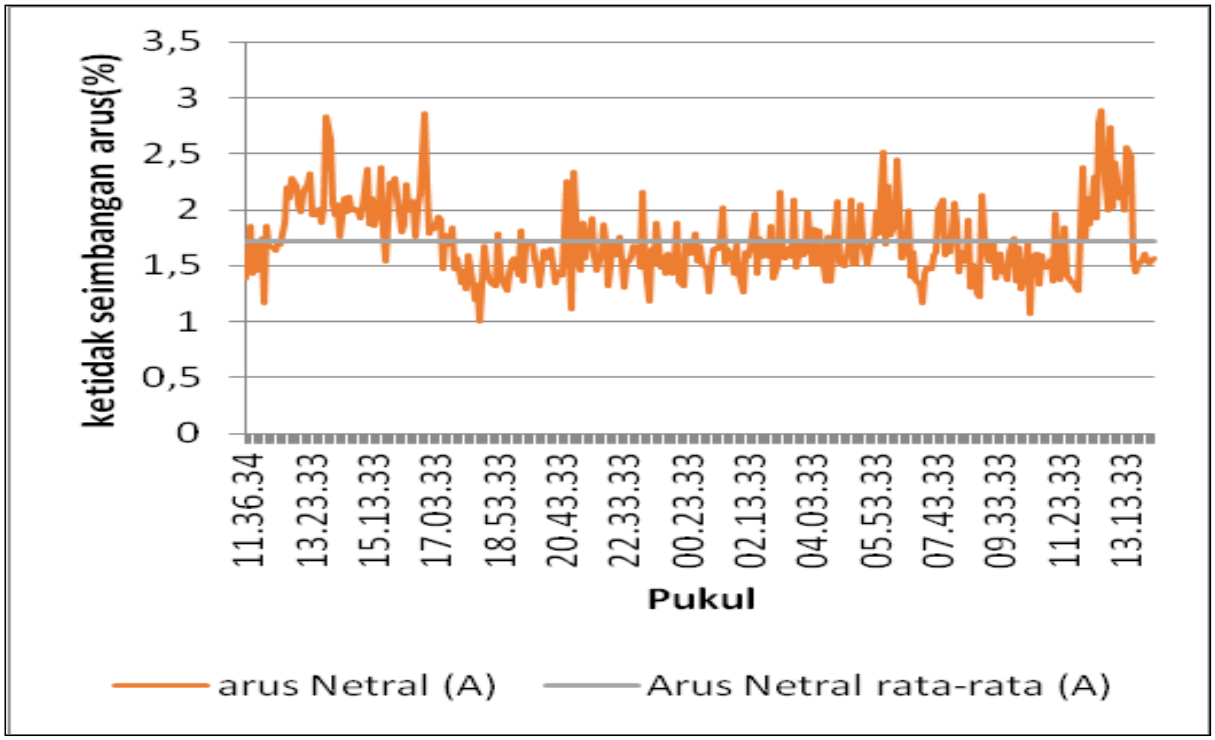


Figure 6. Graph of neutral currents due to current imbalance between phases

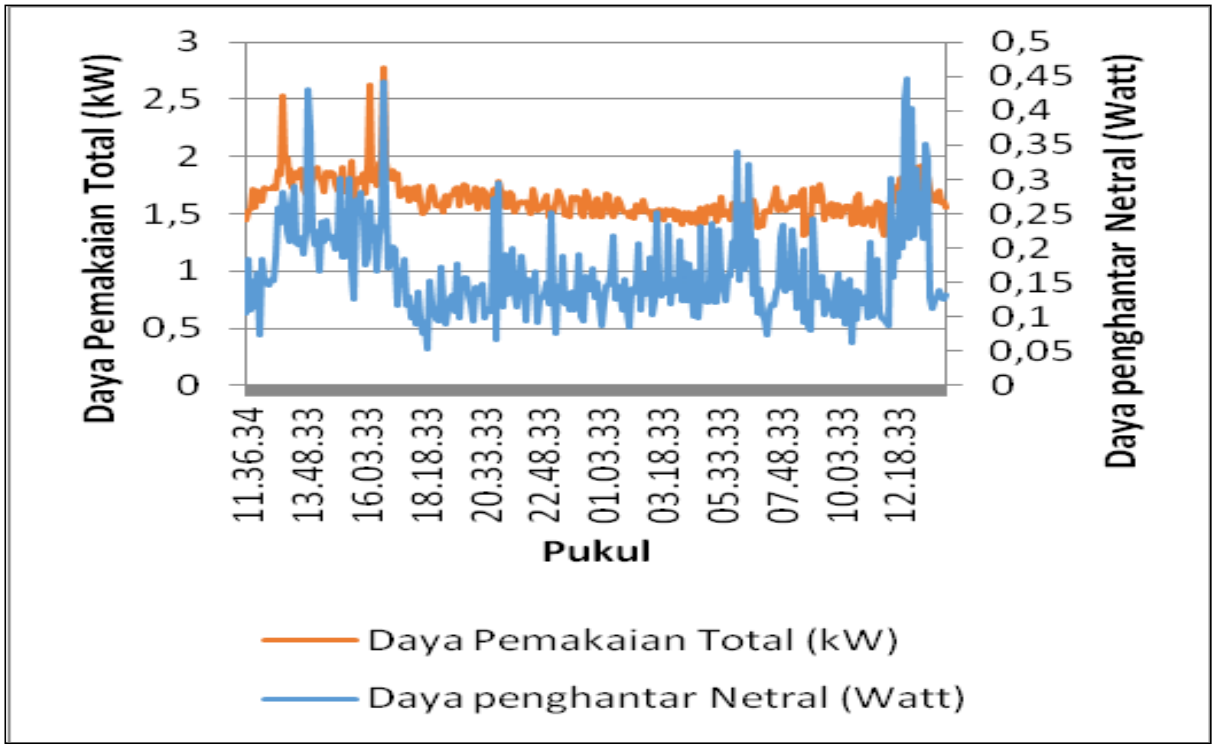


Figure 7. Graph of the total consumption power and neutral power

From the graph in Figure 7 it can be seen that the value of the neutral current is influenced by the current imbalance between the three phases. From the calculation, the neutral current value is obtained as in Table 6.

Table 6. Neutral Current Value

Value	Neutral Current (A)
Max	2.89
Min	1.00

In the electrical panel of this research room, a neutral conducting wire is installed + 50 meters long, with a cross-sectional area of 16 mm2, made of copper, calculated as resistance according to equation 2.4 as follows

$$R = \frac{1{,}72 \times 10^{-8} \times 50}{16 \times 10^{-6}} = 0{,}05375\ \Omega \quad (7)$$

In the previous study [11], energy consumption by load per day was calculated, which was 43.29 kWh, with the same calculation method in the previous study, namely the numerical integration method, which calculated energy in the neutral conducting wire per day of 4.37 Wh, which is when compared to consumption. energy by load per day is about 0.01%.

3. Conclusion

During the measurement period, the voltage imbalance value is 0.64% and the current imbalance is 20.25%. From this current imbalance, it has the potential to produce an average neutral current of 1.72 Ampere, so that the energy flowing in the neutral wire conductor per day is 4.37 Wh, about 0.01% of the energy consumption by the load per day.

Acknowledgements

This work was supported by Research Institute and Services (LPPM) Universitas PGRI Semarang

References

[1] Bina,M.T ,A. Kashefi,*Three Phase Unbalance of Distribution Systems: Complementary Analysis And Experimental Case Study*, Elsevier, Februari 2011.

[2] Arghavani H., Mitra P., *Unbalanced current based tariff,* 24th International Conference and Exhibition on Electricity Distribution (CIRED), 12 -15 Juni 2017

[3] W.U. Kwoktin, Martin, *"Standards Of Power Quality With Reference To The Code Of Practice For Energy Efficiency Of Electrical installations*,Energy Efficiency Office, Electrical And Mechanical Services Departement, September 2003

[4] Kurnianto, D., Hadi, A.M. dan Wahyudi, E., Perancangan Sistem Kendali Otomatis Pada Smart Home Menggunakan Modul Arduino Uno, Jurnal Nasional Teknik Elektro, vol. 5, no.2, hal.260-270, (2016)

[5] PT. PLN (Persero) P3B SUMATERA, Pedoman O dan M Transformator Tenaga, (2014).

[6] Salustiano, R., Neto, E., and Martinez, M., The Unbalanced Load Cost on Transformer Losses at A Distribution System, 22nd International Conference on Electricity Distribution, pp.1-3, (2013).

[7] Motors and Generators, Nema Standards Publications No.MG-1, (1993).

[8] Von Jouanne, A., and Banerjee, B., Assessment of Voltage Unbalance, IEEE Transaction on Power Delivery, vol. 16, no.4, pp.782-790, (2015)

[9] Wang, Y.J., and Yang, M.J., Probabilistic Modeling of Three-Phase Voltage Unbalanced Caused by Load Fluctuations, IEEE PES Winter Meeting, pp.2588-2593, (2000)

[10] Sentosa J,*Pengaruh Ketidakseimbangan Beban terhadap Arus Netral dan Losses pada Trafo Distribusi* Researchgate, Januari 2008.

[11] Wanimbo E.,Muhammad A.,*Perhitungan Nilai Intensitas Konsumsi Energi (ike) Ruangan BPTIK Universitas PGRI Semarang*, Science And Engineering National Seminar 4 (SENS 4), Semarang 7 Desember 2020

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

Application Biplot and K-Medians Clustering to Group Export Destination Countries in Asia, Africa, America, and Europe of Indonesia's Product

Rahmi Lathifah Islami[1], Pardomuan Robinson Sihombing[2]

[1]Universitas Padjadjaran, Jl. Raya Bandung Sumedang KM.21, Hegarmanah, Jatinangor, Kabupaten Sumedang, Jawa Barat 45363 Indonesia

[2*]Badan Pusat Statistik, Jl. Dr. Sutomo 6-8 Jakarta 10710 Indonesia

*robinson@bps.go.id

Abstract. A good increasing export will yield foreign exchange to a country, and subsequently funding its country growth. In Indonesia, export is one of the biggest foreign contributors. As we can see that the countries Indonesia export to are more than 100, it is a must to group them based on their similarity. Biplot and cluster analysis are statistic methods which are used as tool to classify data based on variable explanatory. There are outliers in data acquired. Outliers are observation data which is appeared to be extremely different to the other data. Those data are identified by leverage method. in summary, this research applies K-Medians Clustering Method using Manhattan Distance to resolve outliers while grouping the countries based on their export data. The data contains export data of 182 countries in the year of 2017. R 3.5.1 software was used to calculate in this analysis. The clustering shows us that each continent form difference clusters. Asia has 4 clusters while the rest each has 3 clusters. In addition, we can conclude that several clusters have high value export of Indonesia for certain variables.

Keywords: Biplot, Clustering, K-medians, Silhouette Coefficient, Export

(Received 2021-04-06, Accepted 2021-04-30, Available Online by 2021-04-30)

1. Introduction

Business Development data is really important to a country. It will affect country's annual income or economic growth. One of these business development data is country's export data. A good increasing export will yield foreign exchange to a country, and subsequently funding its country growth. In Indonesia, export is one of the biggest foreign contributors. Based on Badan Pusat Statistik [1], cumulatively Indonesia export value from January to November 2017 reached US$153.90 Billion. It is 17.16% increase year on year (*y-on-y*). Meanwhile, in January-November 2018 its value reached US$150.88 Billion or 8.84% increase in yoy. As we can see that the countries Indonesia export to are more than 100, it is a must to group them based on their similarity. Furthermore, the groups can be used

as consideration to take decision and policy as Indonesia keep export goods into those countries.

Biplot and cluster analysys are statistic methods which are used as tool to classify data based on variable explanatory. Biplot analysys purpose is to group data based upon their similarity which have been mapped on 2 dimensions diagram. so, biplot analysys is easier to be understood and analyzed compared to other classification method. Meanwhile clustering analysys purpose is to identify a group of objects which have certain character similarity which can be separated from the other group of objects. It caused the object in the same group tend to be homogen compared to other group of objects. The objects of this research are export activities of countries in 4 continents: Asia, Africa, America, and Europe.

There are outliers in data acquired. Outliers are observation data which is appeared to be extremely different to the other data. Those data are identified by leverage method. in summary, this research applies K-Medians Clustering Method using Manhattan Distance to resolve outliers while grouping the countries based on their export data.

2. Methods

2.1 *Research Variables*

The Data that is used in this research is secondary data obtained from http://trademap.org which only the export data was being taken. The data contains export data of 182 countries which are consisted of 48 Asia countries, 52 Africa countries, 42 America countries and 40 Europe countries in the year of 2017. There are 15 variables which are the most export in that data:

Table 1. Research Variable

X1: Animal or vegetable fats and oils and their cleavage products; prepared edible fats; animal or vegetable waxes (code 15)
X2: Ores, slag and ash (code 26)
X3: Mineral fuels, mineral oils and products of their distillation; bituminous substances; mineral wax (code 27)
X4: Miscellaneous chemical products (code 38)
X5: Rubber and articles thereof (code 40)
X6: Wood and articles of wood; wood charcoal (code 44)
X7: Paper and paperboard; articles of paper pulp, of paper or of paperboard (code 4
X8: Articles of apparel and clothing accessories, knitted or crocheted (code 61)
X9: Articles of apparel and clothing accessories, not knitted or crocheted (code 62)
X10: Footwear, gaiters and the like; parts of such articles (code 64)
X11: Natural or cultured pearls, precious or semi-precious stones, precious metals, metals clad with precious metal, and articles thereof; imitation jewellery; coin (code71)
X12: Iron and steel (code72)
X13: Machinery, mechanical appliances, nuclear reactors, boilers; parts thereof (code 84)
X14: Electrical machinery and equipment and parts thereof; sound recorders and reproducers, television image and sound recorders and reproducers, and parts and accessories of such articles (code 85)
X15: Vehicles other than railway or tramway rolling stock, and parts and accessories (code 87)

2.2. Biplot Analysis

Biplot analysis is introduced for the first time by [2] which has been developed with Principle Component Analysis (PCA). The biplot display is a graph of row and column markers obtained from data that forms a two-way table. The markers are calculated from the singular value decomposition of the data matrix.

2.3 K-medians analysis

In cluster analysis there is a need of distance type which is suitable to the analysis method so there will be optimal analysis results. Based on [3], it says the clustering using Manhattan Distance results more optimal if applied to k-medians clustering technique, because the distortion or deviation resulted is smaller than other distance. The same report was found in [4]. They found the use of median to measure center point is same as objective function of manhattan distance. K-medians algorithm needs 3 components:

1. **Defined the number of K cluster**

   The number of K must be identified firstly because K-medians is one of non-hierarchy method. There are several ways to define the number of cluster and find the quality and the power of a clustering, one of those ways is Silhouette Coefficient. Silhouette Coefficient is a combination of 2 methods which are cohesion method and separation method. Cohesion method is used to measure how close the relation of objects in a cluster, meanwhile separation method is used to measure how far a cluster separated from the others is [5]. The steps to measure Silhouette Coefficient is shown in the following:

   a) Calculate the average distance between objects within the cluster:

   $$a(i) = \frac{1}{[\mathrm{A}]-1} \sum j \in_{\mathrm{A}, j \neq 1} d(i,j) \quad (1)$$

   Where,

   a(i): the average distance between cluster members

   i: an object in A cluster

   j: others object in A cluster

   d (i, j): distance between *i* and *j*

   b) Calculate the average distance between objects within A and objects within other cluster (C), then define the minimum value:

   $$d(i, \mathrm{C}) = \frac{1}{[\mathrm{A}]} \sum j \in \mathrm{C}\, d(i,j) \quad (2)$$

   Where,

   *d(i,*C*)*: the average distance between *I* and all objects within other cluster (C), where A≠ C

   c) Calculate Silhouette Coefficient:

   $$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (3)$$

   Where,

   b(i): the smallest distance between cluster member with the closest cluster member. Subjective criteria of clustering quality based on silhouette coefficient that was made by [6] is shown in Table 2.

Table 2. Subjective criteria of clustering quality based on silhouette coefficient

Silhouette Coefficient	Kauffman Interpretation
0.71 – 1.00	Strong
0.51 – 0.70	Good
0.26 – 0.50	Weak
0 – 0.25	Bad

2. **Define centroid**

   Based on Teknomo [7], centroid could be picked randomly from all of observation.

3. **Define the distance of each observation unit to each centroid**

   Distance calculation is used to locate the observation into the cluster based on the nearest centroid. Distance calculation which is being used in k-medians method is manhattan [8]. The Manhattan Distance is a distance of a point to the other point in the Cartesian. Following the vertical and

horizontal axes, with no comeback. As instance, if known 2 points, which are p_1 and p_2 in two dimensions' plane (x_1, y_1) and (x_2, y_2), then the manhattan distance between p_1 and p_2 is $|x_1 - x_2| + |y_1 - y_2|$. Each distance observation is calculated to each centroid using Manhattan Distance with this following equation

$$d(\vec{x}, \vec{y}) = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_p - y_p| \quad (4)$$

where $d(\vec{x}, \vec{y})$ is the distance between data and each centroid, x is export proportion data, y is *centroid* and p is number of variable. Each observation is grouped based on the closest distance between the data and the centroid. Update the centroid of each cluster using median calculation from observations in the same centroid. Repeat the steps until no member cluster movement left.

3. Results and Discussion

3.1 Asia Continent

This research uses 48 countries in Asia Continent. The average value and the standard deviation from each variable is calculated before clustering the countries of Asia using k-medians clustering. The calculation is shown in Table 3. As we can see in the table, there is little difference of average value and standard deviation which shown the variance in the data. X_{14} is the most popular variable in Asia meanwhile X_2 is the least. The highest variance was found in X_1 meanwhile the least is X_8. The higher variance the more distributed the variable.

Table 3. Variable description of Asia

Variable	Standard Deviation	Mean
X_1	0.267767	0.22378
X_2	0.028901	0.008757
X_3	0.207446	0.119571
X_4	0.029024	0.018748
X_5	0.06051	0.04949
X_6	0.187337	0.092494
X_7	0.106071	0.08379
X_8	0.014934	0.009312
X_9	0.072084	0.020405
X_{10}	0.023495	0.011209
X_{11}	0.15334	0.037698
X_{12}	0.019599	0.009542
X_{13}	0.163303	0.875247
X_{14}	0.208736	0.94454
X_{15}	0.183955	0.705525

Based on R, the outliers were detected as the leverage method was used. The outliers were found in Georgia, North Korea, Nepal, State of Palestine, Philippine, and Timor Leste. Further treatment is needed to cluster these countries. The biplot analysis of Asia is shown in Figure 1. Figure 1 is showing biplot graph of Asia countries. In the graph we can see the similarity relation between countries and similarity of countries to variables. Furthermore, cluster analysis is conducted so the data in the biplot analysis would not be overlapped.

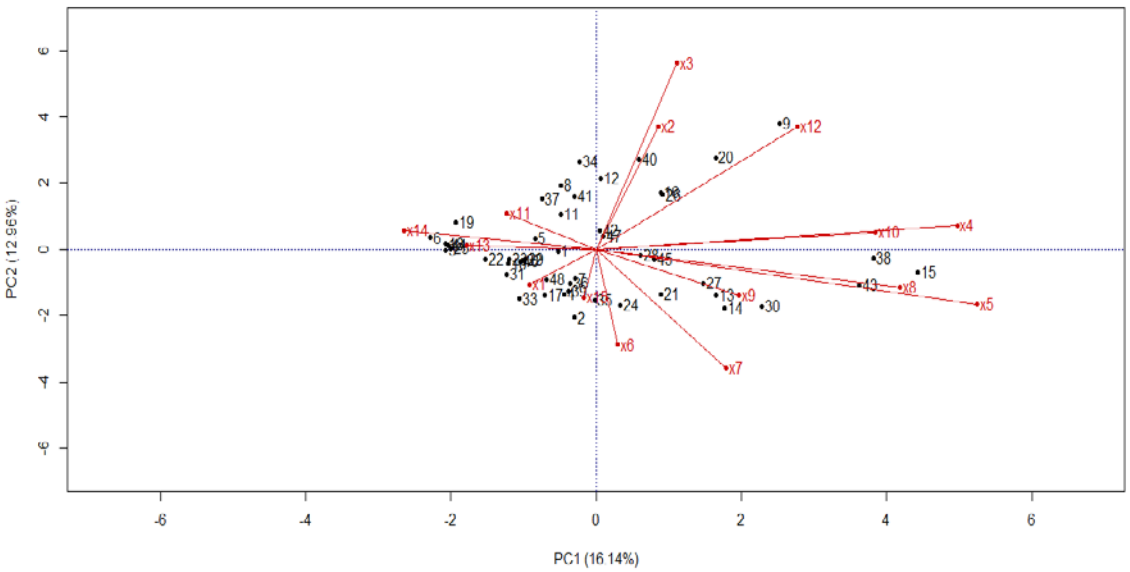


Figure 1. Biplot graph of Asia countries

The first step to do cluster analysis is to define the number of clusters. When the first k cluster is formed, calculate at once the silhouette coefficient until the good criteria coefficient is obtained. The coefficient is calculated using R 3.5.1 software and is shown in Table 3. this number is showing how well the clustering process and the cluster quality that is formed. As we can see in the Table 4, the highest value belongs to K = 4. Therefore, the writer chose 4 clusters as the number of clusters in this continent. The value of silhouette coefficient, 0.060142649 is categorized as 'Good' based on Kauffman.

Tabel 4. Silhouette coefficient value of Asia

K	Silhouette Coefficient
3	0.1346155
4	0.60142649
5	0.10396336

After cluster analysis was conducted using k-medians through R 3.5.1 software, we can see the names of countries in Asia for each of 4 clusters the Table 5. To measure the classification of sum of export proportion in Asia for each cluster, the following can be interpretated:

a. Good Criteria if the value is far above the average of overall, and marked as '++'
b. '+' is given for the value below the '++' as long as it is still above overall average.
c. For the value under the average of overall, is given '-'mark
d. And for the value that is far under the overall average is given '—' mark

Tabel 5. Cluster analysis of Asia countries-based export proportion using k-medians clustering

Cluster	Number of Countries	Name of Country
1	14	Bahrain, Brunei Darussalam, Kuwait, Laos, Lebanon, Mongolia, Oman, Philippines, Qatar, Saudi Arabia, Thailand, Timor-Leste, Vietnam
2	26	Armenia, Azerbaijan, Bangladesh, Cambodia, China, Georgia, Hong Kong, India, Iran, Israel, Japan, North Korea, SouthKorea, Kyrgyzstan, Malaysia, Myanmar, Nepal, Pakistan, Singapore,

		Sri Lanka, Syria, Chinese Taipei, Turkey, United Arab Emirates, Uzbekistan, Yemen
3	3	Jordan, Maldives, State of Palestine
4	5	Afghanistan, Bhutan, Kazakhstan, Macao China, Turkmenistan

Table 6 is showing us the level of export proportion classification between the clusters. As we can see in the Table 4, the characteristic of export proportion data of Asia is a focus of the main cluster which is cluster 2. Meanwhile, the rest 3 clusters tend to far from the average. In addition, there are two variables that is undesirable in the cluster 3 and 4, they are X_2 and X_3.

Table 6. The level of export proportion classification between the clusters of Asia

Variable	Mean	Cluster				Classification			
		C1	C2	C3	C4	C1	C2	C3	C4
X_1	0.22378	0.10020	0.30346	0.11449	0.18037	--	++	--	-
X_2	0.00875	0.01296	0.00932	0	0	++	+	--	--
X_3	0.11957	0.07227	0.17777	0	0	-	++	--	--
X_4	0.01874	0.00729	0.02560	0.02294	0.00899	--	++	+	-
X_5	0.04949	0.04347	0.06083	0.02794	0.01679	-	++	--	--
X_6	0.09249	0.05842	0.06088	0.66796	0.00649	-	-	++	--
X_7	0.08379	0.10550	0.08836	0.06686	0.01277	++	+	-	--
X_8	0.00931	0.00918	0.01127	0.00653	0.00072	-	++	--	--
X_9	0.02040	0.00971	0.02823	0.01784	0.00743	--	++	-	--
X_{10}	0.01120	0.01164	0.01325	0.00831	0.00078	+	++	-	--
X_{11}	0.03769	0.00070	0.06651	0.00146	0	--	++	-	--
X_{12}	0.00954	0.00953	0.01206	0.00277	0	-	++	--	--
X_{13}	0.07595	0.09313	0.0748	0.01965	0.07126	++	-	--	-
X_{14}	0.11115	0.06646	0.0405	0.00957	0.66960	-	--	--	++
X_{15}	0.128103	0.399476	0.02708	0.033608	0.02475	++	--	-	--

In summary, the main export countries of Indonesia is the countries in cluster 2 with the main variables of X_1, X_3, X_4, X_5, X_8, X_9, X_{10}, X_{11} dan X_{12}. The second export countries of Indonesia is cluster 1 with main variables of X_2, X_7, X_{13}, dan X_{15}. The rest are the third and fourth export countries, which are cluster 3 and 4, with the main carriable is X_6 and X_{14}, respectively.

3.2 Africa Continent

This research uses 52 countries in the Africa Continent. The average value and standard deviation of each variable are calculated before the clustering (with k-medians clustering). The results are shown in Table 7. The highest mean is found on X_1 as it is the most popular variable in Africa, meanwhile the least is shown by X_2. The distribution of variables in Africa is the most varied compared to 3 other continents. X_1 shows the most varied variables among the others, meanwhile X_2 is the least. Based on R, outliers were found using leverage method. They are Botswana, Burkina Faso, Cabo Verde, Egypt, Malawi, Mali, Mauritania, Seychelles, Somalia, and South Africa. They need further treatment to be clustered using k-medians.

Table 7. Africa continent research variable description

Variable	Standard Deviation	Mean
X_1	0.311546	0.6525
X_2	5.67E-05	7.86E-06

X_3	0.005032	0.00111
X_4	0.080258	0.023741
X_5	0.027638	0.014147
X_6	0.070435	0.020112
X_7	0.126892	0.087727
X_8	0.064997	0.01627
X_9	0.024258	0.012914
X_{10}	0.026484	0.009541
X_{11}	0.007522	0.001475
X_{12}	0.037929	0.006506
X_{13}	0.066613	0.04142
X_{14}	0.22428	0.087251
X_{15}	0.06685	0.025312

The biplot analysis of Africa was shown in Figure 2. The graph shows us the relation of similarity of each countries and similarity of countries with variables. The cluster analysis was conducted to map the data clearly as the data in biplot analysis tend to be overlapped one to another.

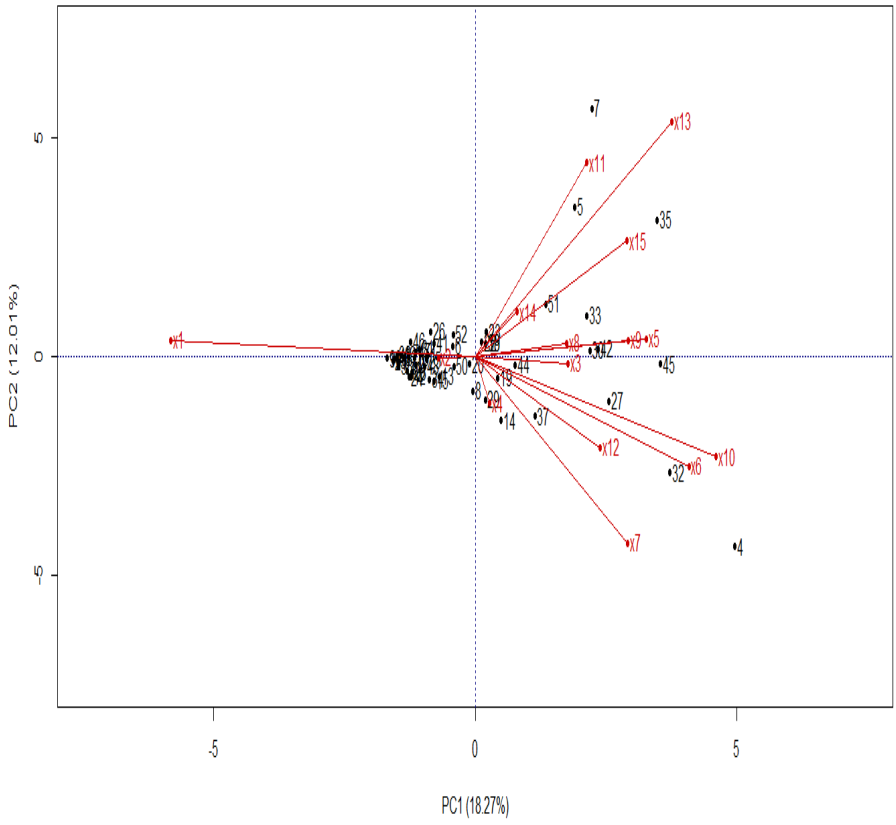


Figure 2. Biplot analysis graph of Africa

The first step to do cluster analysis is to define the number of clusters. When the first k cluster was formed, calculate at once the silhouette coefficient until the good criteria coefficient was found. Silhouette coefficients were obtained using R 3.5.1 and were shown in Table 8.

Table 8. Silhouette coefficient value of Africa

K	Silhouette Coefficient
3	0.88830048
4	0.1059707
5	0.3543382

As we can see in the Table 8, the highest value belongs to K = 3. Therefore, the writer chose 3 clusters as the number of clusters in this continent. The value of silhouette coefficient, 0.88830048 is categorized as 'Strong' based on Kauffman. After cluster analysis was conducted using k-medians through R 3.5.1 software, we can see the names of countries in Africa for each of 3 clusters the Table 9.

Table 9. Cluster analysis of Africa countries-based export proportion using k-medians clustering

Cluster	Number of Countries	Name of Countries
1	32	Algeria, Angola, Benin, Burundi, Central African Republic, Chad, Comoros, Congo, Democratic Republic of the Congo, Djibouti, Egypt, Equatorial Guinea, Gabon, Gambia, Guinea-Bissau, Kenya, Liberia, Madagascar, Mauritania, Mozambique, Niger, Rwanda, Sao Tome and Principe, Senegal, Sierra Leone, Somalia, Swaziland, Tanzania, Togo, Tunisia, Uganda, Zimbabwe
2	17	Botswana, Burkina Faso, Cabo Verde, Cameroon, Côte d'Ivoire, Ethiopia, Ghana, Guinea, Libya, Malawi, Mali, Mauritius, Morocco, Namibia, Nigeria, Seychelles, South Africa
3	3	Eritrea, Saint Helena, Zambia

To measure the classification of sum of export proportion in Africa for each cluster, the following can be interpretated:

a. If the value is above the average of overall, and marked as '+'
b. For the value under the average of overall, is given '-' mark
c. And for the value that is far under the overall average is given '—' mark

The level of export proportion classification between the clusters of Africa was shown in Table 10. As we can see from the table, the characteristic of export proportion data in Africa focused on the main cluster, which is cluster 2. The cluster 2 is above the average, meanwhile the rest show the contrary. In addition, we can find a lot of variables in cluster 3 that are least desired. They are X_1, X_2, X_3, X_6, X_7, X_9, X_{11}, X_{12} dan X_{15}.

In summary, the main cluster of export countries of Indonesia in Africa is the cluster 2 with the main variables of X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13} dan X_{15}. Cluster 1 become the second export target with the main variables of X_1 and X_2, while the cluster 3 become the third with the main variables of X_{13} and X_{14}.

Table 10. The level of export proportion classification between the clusters of Africa

Variable	Mean	Cluster			Classification		
		C1	C2	C3	C1	C2	C3
X_1	0.6525	0.8653	0.3668	0	+	-	--
X_2	7.86E-06	1.28E-05	0	0	+	--	--
X_3	0.0011	0.0007	0.0019	0	-	+	--
X_4	0.0237	0.0132	0.0475	0.0002	-	+	--
X_5	0.0141	0.0076	0.0288	0.0002	-	+	--
X_6	0.0201	0.0039	0.0540	0	-	+	--

X_7	0.0877	0.0455	0.1809	0.0091	-	+	--
X_8	0.0162	0.0030	0.0432	0.0042	--	+	-
X_9	0.0129	0.0095	0.0215	0	-	+	--
X_{10}	0.0095	0.0017	0.0244	0.0085	--	+	-
X_{11}	0.001475	1.79E-06	0.0045	0	-	+	--
X_{12}	0.006506	0.000923	0.018162	0	-	+	--
X_{13}	0.0414	0.0163	0.0867	0.052288	-	+	+
X_{14}	0.0872	0.0277	0.0513	0.9252	--	-	+
X_{15}	0.0253	0.0040	0.0697	0	-	+	--

3.3 America Continent

This research uses 42 countries in the America Continent. The average value and standard deviation of each variable are calculated before the clustering (with k-medians clustering). The results are shown in Table 11. The highest mean is found on X_{15} as it is the most popular variable in America, meanwhile the least is shown by X_2. The most varied variables in America was found in X_1, meanwhile X_2 is the least. Based on R, outliers were found using leverage method. They are British Virgin Islands, Canada, Cuba, Mexico, Saint Lucia, and United States of America. They need further treatment to be clustered using k-medians.

Table 11. America continent research variable description

Variable	Standard Deviation	Mean
X_1	0.232036	0.133325
X_2	0.000232	3.68E-05
X_3	0.011614	0.002628
X_4	0.014868	0.010278
X_5	0.212741	0.11665
X_6	0.260717	0.11811
X_7	0.189049	0.103628
X_8	0.05618	0.023826
X_9	0.040393	0.023006
X_{10}	0.139164	0.074092
X_{11}	0.007889	0.00259
X_{12}	0.009216	0.003147
X_{13}	0.069327	0.05992
X_{14}	0.108067	0.07306
X_{15}	0.321278	0.2557

The biplot analysis of America was shown in Figure 3. The graph shows us the relation of similarity of each countries and similarity of countries with variables. The cluster analysis was conducted to map the data clearly as the data in biplot analysis tend to be overlapped one to another.

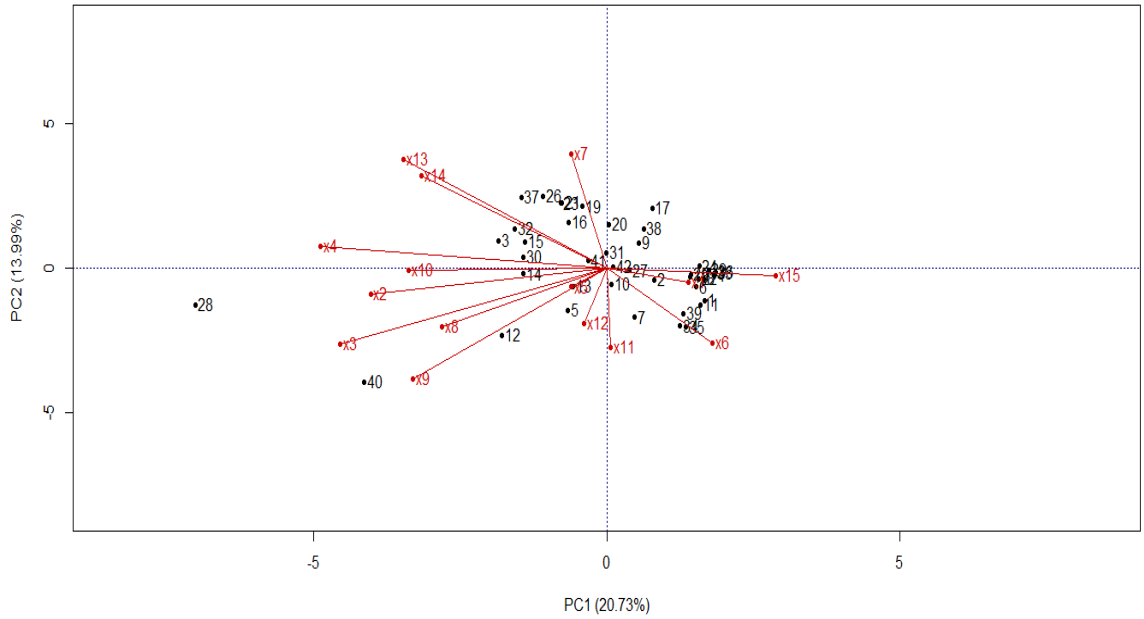


Figure 3. Biplot analysis graph of America

The first step to do cluster analysis is to define the number of clusters. When the first k cluster was formed, calculate at once the silhouette coefficient until the good criteria coefficient was found. Silhouette coefficients were obtained using R 3.5.1 and were shown in Table 12.

Table 12. Silhouette coefficient value of America

K	Silhouette Coefficient
3	0.7140413
4	0.34643963
5	0.3640942

As we can see in the Table 12, the highest value belongs to K = 3. Therefore, the writer chose 3 clusters as the number of clusters in this continent. The value of silhouette coefficient, 0.7140413 is categorized as 'Strong' based on Kauffman. After cluster analysis was conducted using k-medians through R 3.5.1 software, we can see the names of countries in America for each of 3 clusters the Table 13.

To measure the classification of sum of export proportion in America for each cluster, the following can be interpretated:

a. If the value is above the average of overall, and marked as '+'
b. For the value under the average of overall, is given '-' mark
c. And for the value that is far under the overall average is given '—' mark

The level of export proportion classification between the clusters of America were shown in Table 14. As we can see from the table, the characteristic of export proportion data in America focused on the main cluster, which is cluster 2. The variables of cluster 2 almost all are above the average. Meanwhile the rest, cluster 1 and 3, show the contrary. In addition, we can find a lot of variables in cluster 1 and 3 that are least desired. They are X_2, X_3, X_{13}, X_{14} and X_2, X_3, X_{12}, respectively.

In summary, the main cluster of export countries of Indonesia in America is the cluster 2 with the main variables of X_1, X_2, X_3, X_4, X_5, X_7, X_8, X_9, X_{10}, X_{13} dan X_{14}. Cluster 1 become the second export

target with the main variables of X_6, X_9 dan X_{12}. while the cluster 3 become the third with the main variables of X_{11} dan X_{15}.

Table 13. Cluster analysis of America countries-based Export Proportion using k-medians clustering

Cluster	Number of Countries	Name of Countries
1	6	Anguilla, Bahamas, Barbados, British Virgin Islands, Saint Vincent and the Grenadines, Turks and Caicos Islands
2	28	Antigua and Barbuda,Argentina, Belize, Brazil, Canada,Cayman Islands, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, Grenada, Guatemala, Haiti, Honduras, Jamaica, Mexico, Panama, Paraguay, Peru, Suriname, Trinidad and Tobago , United States of America , Uruguay, Venezuela
3	8	Aruba, Bermuda, Bolivia, Guyana, Nicaragua, Saint Kitts and Nevis, Saint Lucia, St Pierre and Miquelon

Table 14. The level of export proportion classification between the clusters of America

Variable	Mean	Cluster			Classification		
		C1	C2	C3	C1	C2	C3
X_1	0.1333	0.0816	0.1664	0.056	-	+	--
X_2	3.68E-05	0	5.51E-05	0	--	+	--
X_3	0.0026	0	0.0039	0	--	+	--
X_4	0.0102	0.0063	0.0137	0.001	-	+	--
X_5	0.1166	0.0056	0.1709	0.009	--	+	-
X_6	0.1181	0.7017	0.0206	0.021	+	-	-
X_7	0.1036	0.0049	0.1520	0.008	--	+	-
X_8	0.0238	0.0024	0.0350	0.0004	-	+	--
X_9	0.0230	0.0280	0.0264	0.007	+	+	-
X_{10}	0.0740	0.0332	0.1037	0.0009	-	+	--
X_{11}	0.0025	0.0021	0.0018	0.005	-	--	+
X_{12}	0.0031	0.0076	0.0030	0	+	-	--
X_{13}	0.0599	0	0.0889	0.003	--	+	-
X_{14}	0.0730	0	0.0958	0.048	--	+	-
X_{15}	0.2557	0.125	0.1171	0.837	-	-	+

3.4 Europe Continent

This research uses 40 countries in the Europe Continent. The average value and standard deviation of each variable are calculated before the clustering (with k-medians clustering). The results are shown in Table 15.

The highest mean is found on X_1 as it is the most popular variable in Europe, meanwhile the least is shown by X_2. The most varied variables in Europe was found in X_5, meanwhile X_2 is the least. Based on R, outliers were found using leverage method. They are Belarus, Ireland, Slovenia, Switzerland, and Ukraine. They need further treatment to be clustered using k-medians.

The biplot analysis of Europe was shown in Figure 4. The graph shows us the relation of similarity of each countries and similarity of countries with variables. The cluster analysis was conducted to map the data clearly as the data in biplot analysis tend to be overlapped one to another.

The first step to do cluster analysis is to define the number of clusters. When the first k cluster was formed, calculate at once the silhouette coefficient until the good criteria coefficient was found. Silhouette coefficients were obtained using R 3.5.1 and were shown in Table 16.

Table 15. Europe continent research variable description

Variable	Standard Deviation	Mean
X_1	0.259281	0.200674
X_2	0.002923	0.000479
X_3	0.042051	0.01128
X_4	0.046908	0.02561
X_5	0.251331	0.20975
X_6	0.038053	0.025969
X_7	0.135202	0.07002
X_8	0.05489	0.04398
X_9	0.053991	0.038759
X_{10}	0.161764	0.117964
X_{11}	0.154227	0.027231
X_{12}	0.00647	0.003387
X_{13}	0.111575	0.05473
X_{14}	0.190139	0.138673
X_{15}	0.069105	0.03149

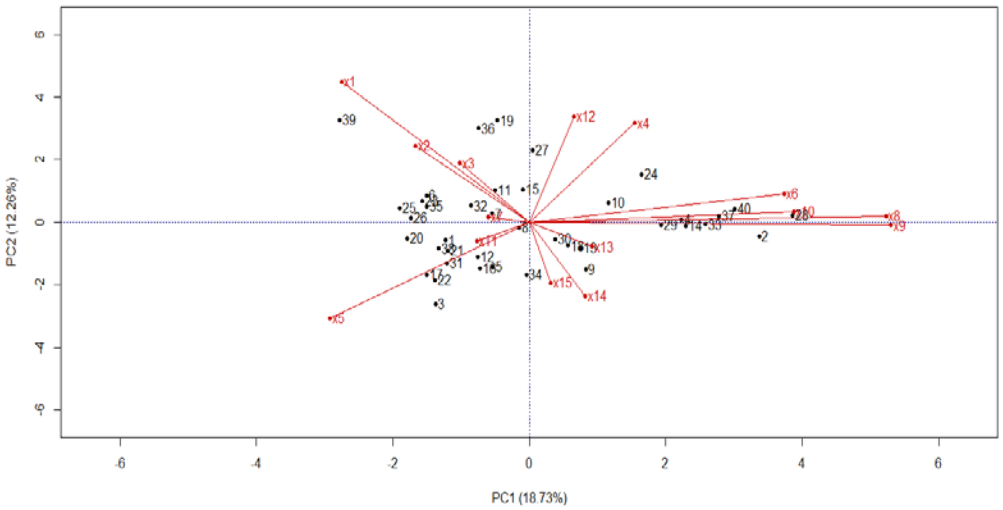


Figure 3. Biplot analysis graph of Europe

Table 16. Silhouette coefficient value of Europe

K	Silhouette Coefficient
3	0.3688017
4	0.3599565
5	0.1291649

As we can see in the Table 16, the highest value is belong to K = 3. Therefore, the writer chose 3 clusters as the number of clusters in this continent. The value of silhouette coefficient, 0.3688017 is categorized as 'Weak' based on Kauffman. After cluster analysis was conducted using k-medians

through R 3.5.1 software, we can see the names of countries in Europe for each of 3 clusters the Table 17.

Tabel 17. Cluster analysis of Europe countries-based export proportion using k-medians clustering

Cluster	Number of Countries	Name of Countries
1	12	Albania, Croatia, Estonia, Greece, Italy, Macedonia, Moldova, Montenegro, Netherlands, Russian Federation, Spain, Ukraine
2	16	Austria, Belgium, Cyprus, Czech Republic, Denmark, France, Germany , Ireland, Malta, Norway, Poland, Serbia, Slovakia, Sweden, Switzerland, United Kingdom
3	12	Belarus, Bosnia and Herzegovina, Bulgaria, Finland , Hungary, Iceland , Latvia, Lithuania, Luxembourg, Portugal, Romania, Slovenia

To measure the classification of sum of export proportion in Europe for each cluster, the following can be interpretated:

a. If the value is above the average of overall, and marked as '+'
b. For the value under the average of overall, is given '-' mark
c. And for the value that is far under the overall average is given '—' mark

The level of export proportion classification between the clusters of Europe were shown in Table 18. As we can see from the table, the characteristic of export proportion data in America focused on the main cluster, which is cluster 2. The variables of cluster 2 tend to be above the average. Meanwhile the rest, cluster 1 and 3, show the contrary. In addition, we can find a variable in cluster 3 which is undesired. The variable is X_2.

In summary, the main cluster of export countries of Indonesia in Europe is the cluster 2 with the main variables of X_4, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}, dan X_{15}. Cluster 1 become the second export target with the main variables of X_1, X_2, X_3, X_4, X_7 dan X_{12}, while the cluster 3 become the third with the main variables of X_3, X_5 dan X_{15}.

Table 18. The level of export proportion classification between the clusters of Europe

Variable	Mean	Cluster			Classification		
		C1	C2	C3	C1	C2	C3
X_1	0.200674	0.536578	0.045614	0.071516	+	--	-
X_2	0.0004	0.0015	2.9E-05	0	+	-	--
X_3	0.01128	0.015845	0.001177	0.020197	+	-	+
X_4	0.0256	0.0342	0.0315	0.0090	+	+	--
X_5	0.2097	0.0877	0.0523	0.5416	--	--	+
X_6	0.0259	0.0185	0.0438	0.0095	-	+	--
X_7	0.0700	0.0946	0.0702	0.0450	+	+	-
X_8	0.0439	0.0118	0.0884	0.0167	-	+	-
X_9	0.0387	0.0156	0.0814	0.0049	-	+	--
X_{10}	0.1179	0.0379	0.2033	0.0841	--	+	-
X_{11}	0.0272	0.0009	0.0666	0.0009	--	+	--
X_{12}	0.0033	0.0051	0.0030	0.002	+	-	--
X_{13}	0.0547	0.0375	0.0837	0.0332	--	+	--
X_{14}	0.1386	0.0874	0.1931	0.1172	--	+	-
X_{15}	0.0314	0.0141	0.0351	0.0438	-	+	+

4. Conclusion

The clustering shows us that each continent form difference clusters. Asia has 4 clusters while the rest each has 3 clusters. In addition, we can conclude that several clusters have high value export of Indonesia for certain variables. So, this research could be utilized as consideration to make policies related to export.

Reference

[1] Badan Pusat Statistik.(2017).Perkembangan Ekspor Impor Indonsia.

[2] Gabriel, K, R. (1971), "The Biplot Graphic Display of Matrices with Aplication to Principal Component Analysis," Journal of Biometrica. 58, 453-467.

[3] Anderson, Gross, Musicant, Ritz, Smith, dan Steinberg. (2006), "Adapting K-Medians to Generate Normalized Cluster Centers."

[4] Wierzchon and Klopotek. (2017), "Modern Algorithms of Cluster Analysis," Switzerland: Springer.

[5] Pramesti, Dyang Falila, M. Tanzil Furqon, and Candra Dewi. (2017), "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X Vol. 1, No. 9, 723-732.

[6] Kaufman, L., Rousseeuw, PJ. (1990), "Finding Groups in Data: An Introduction to Gerombol Analysis," New York: John Wiley.

[7] Teknomo, Kardi. (2007), "K-Means Clustering Tutorials."

[8] Ritz, A., Anderson B., Nelson K. (2006), "Clustering Algorithms: A New Look at Mass Spectra," Computer Science Carleton College.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210106-01 ~ 021016-07
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8460

Automatic Complaints Categorization Using Random Forest and Gradient Boosting

Muchamad Taufiq Anwar[1*], Anggy Eka Pratiwi[2], Khadijah Febriana Rukhmanti Udhayana[3]

[1]Politeknik STMI Jakarta, Jl. Letjend Suprapto No.26, Central Jakarta 10510, Special Capital Region of Jakarta, Indonesia

[2]Department of Computer Science and Engineering, Indian Institute of Technology Jodhpur, Jodhpur, India

[3]Department of Informatics, Universitas Bali Internasional, Gg. Jeruk, Tonja, Kec. Denpasar Timur, Denpasar, Bali 80234, Indonesia

*taufiq@stmi.ac.id

Abstract. Capturing and responding to complaints from the public is an important effort to develop a good city/country. This project aims to utilize Data Mining to automatize complaints categorization. More than 35,000 complaints in Bangalore city, India, were retrieved from the "I Change My City" website (https://www.ichangemycity.com). The vector space of the complaints was created using Term Frequency–Inverse Document Frequency (TF-IDF) and the multi-class text classifications were done using Random Forest (RF) and Gradient Boosting (GB). Results showed that both RF and GB have similar performance with an accuracy of 73% on the 10-classes multi-class classification task. Result also showed that the model is highly dependent on the word usage in the complaint's description. Future research directions to increase task performance are also suggested.

Keywords: automatic complaints categorization, multi-class classification, data mining, random forest, gradient boosting

(Received 2021-03-26, Accepted 2021-04-30, Available Online by 2021-04-30)

1. Introduction

One way of giving a good service to the public is by giving room for the public to post a complaint to the government so that necessary actions can be taken as soon as possible. This is especially important for a developing country such as India to make the cities better. One of the initiatives that already taken is the "I Change My City" website which first debuted in Bangalore City, India, in 2012. This platform is managed by the Swachh Bharat Mission of the Ministry of Urban Development of the Government

of India. By using this platform, citizens can post complaints they face in their city and the government will resolve the issue. To be resolved, the issues must be categorized and forwarded to the relevant agencies. Until now, this platform already has nearly 20 million users and 53 million complaints have been submitted, of which 94% percent have been successfully resolved. This research aims to create a model so that when a new complaint is registered, it will get automatically classified. The classification task can be done by using Data Mining techniques. The automatic categorization of the complaints will help both the user and the government in reducing the overall complaints resolution time.

2. Methods

The research methods are shown in Figure 1. Complaint data are scraped from the "I Change My City" website (https://www.ichangemycity.com). The website itself was started in October 2019. The original scraped data from the website are in JSON format. The JSON file was then transformed into two-dimensional data using Python. The data consists of 44961 data which represent complaints from October 2019 to December 2020. The data are spread across 14 different classes. Each object is having 8 features that include id, category, sub-category, my category (specifically mentioned by the person lodging the problem), description of the complaint, and an associated image to describe the problem. Of the eight features, only two features are used in this research, i.e the complaint description and the complaint category. For this study, we focused only on Bangalore city and its top-10 complaint categories which happened to represent 79% of the total data. The sample data is shown in Table 1.

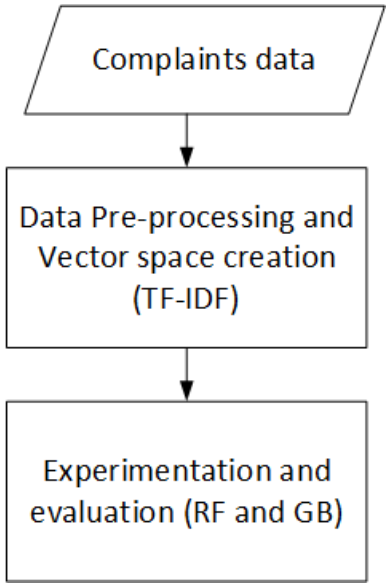


Figure 1. Research methods

Table 1. Sample data

Complaint_description	Complaint_category
Existing road inside GSS layout near Kudlu VGP...	Provide good driveable Roads
Street lights not working in this are 2 to 3 l...	Maintenance/Repair Of Streetlights
47th Cross, 8th Block Jayanagar Road Maintenan...	Fixing/Reparing Potholes

This research aims to categorize data into 10 broad classes. The data is also not perfectly balanced among classes as shown in Table 2. To handle this class imbalance, we utilize ensemble methods such as Random Forest (RF) and Gradient Boosting (GB) since the ensemble method takes care of the class imbalance. The data has majorly two components, a) free-flowing text (the text to be predicted), and b) complaint categories (target classes). We are using Term Frequency–Inverse Document Frequency (TF-

IDF) to give weights for every word to create a vector space for the text. TF-IDF is commonly used in text classification tasks such as in [1]. This research is essentially a multi-class classification problem. The implementation of these classification tasks was done using Python alongside its machine learning libraries i.e the ScikitLearn library.

RF and GB are newer-generation DM techniques that utilize bagging / boosting which improve the performance of the previous "traditional" techniques. These methods are also known as ensemble learning methods since they are using multiple learners (models) for the prediction task. Both of these methods are based on tree structure such as used in C4.5. C4.5 uses a single tree to make a prediction. C4.5 itself is a simple yet versatile classification method and had been used in wildfire modeling [2], rain prediction[3], [4], drug resistance prediction[5], etc. Bagging and boosting methods are a class of classification techniques that increase (boost) its performance by adding bags of models or by iteratively improve the model.

RF is a tree-based classifier that utilizes multiple (a bag of) trees where each tree has its own prediction ability. The creation of each tree is based on a random subset of the entire dataset and also a random subset of the entire attributes. In RF, each tree will give its prediction and the final prediction is taken from the voting from all trees. RF was introduced in 2014 by Breiman [6]. RF had been recently used in various fields including rain prediction[4], drug resistance prediction[5], geospatial pattern alaysis[7], prediction of electricity production [8], battery modeling [9], and landslide susceptibility mapping [10].

GB is another tree-based classifier that improves (boosts) its prediction ability by iteratively building a better tree from the previous iterations. Each iteration will reduce the error of the previous tree. GB was introduced in 2014 and is widely available on multiple platforms including Python, R [11], etc. GB is capable of both classification and regression tasks. GB had been recently used in various fields including rain prediction[4], rainfall prediction[12], transpiration estimation [13], undrained shear strength prediction[14], concrete strength prediction[15], groundwater level prediction[16], and solar irradiation forecasting[17].

3. Results and Discussion

The complaints from Bangalore City consist of 35,375 rows of data. The topmost recurrent complaints are related to road infrastructure, garbage/waste management, stray dogs, and floods as shown in Table 2. The word cloud (top 50 words) for complaints description is shown in Figure 2. This confirms the fact that most complaints are related to road infrastructure and waste management.

Table 2. The top recurring complaints in Bangalore

Rank	Category	Count	Percentage
1	Fixing/Reparing Potholes	10185	29
2	Maintenance/Repair Of Streetlights	6744	19
3	Clearance Of Garbage Dump Or Black Spot	5151	15
4	Repair of Existing Footpaths	3464	10
5	Maintenance Of Dry Waste Collection Centre	2638	7
6	Stray Dog Sterilisation/Animal Birth Control (ABC)	2022	6
7	Garbage Dumping In Vacant Lot/Land	1512	4
8	Provide good driveable Roads	1287	4
9	Desilting of storm water drains	1236	3
10	Collection Of Door-to-door Garbage	1136	3
	TOTAL	**35375**	**100**

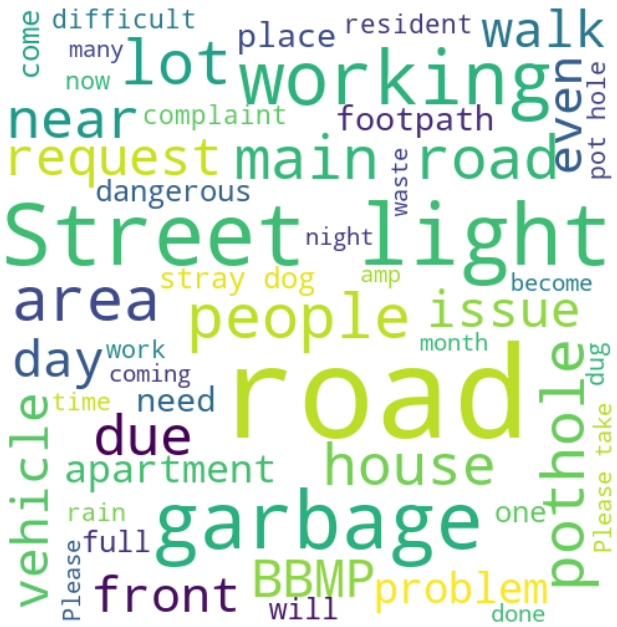


Figure 2. The wordcloud for the complaints

Data were split into 80% training data and 20% testing data. Our result showed that RF (n estimator = 100, using all features) gave better overall accuracy of 0.73 with the best f-1 score in the "Maintenance/Repair Of Streetlights" and "Stray Dog Sterilisation/Animal Birth Control (ABC)" categories. The worst classification performance is on the "Provide good driveable Roads" and "Maintenance Of Dry Waste Collection Centre" categories. This might be caused by similar words being used in several other categories, eg: "waste"/"garbage", and "road"/"street". Whereas the good performing categories have a distinguished use of words such as "streetlights", "dogs", "water", and "potholes". The result of performance metrics of the RF model is shown in Table 3 and the confusion matrix is in Table 4.

Table 3. Performance metrics for RF model

Categories	Precision	Recall	F1-score	Support
Clearance Of Garbage Dump Or Black Spot	0.523224	0.796258	0.631492	962
Collection Of Door-to-door Garbage	0.664634	0.488789	0.563307	223
Desilting of storm water drains	0.804878	0.788845	0.796781	251
Fixing/Reparing Potholes	0.715209	0.901294	0.797541	2087
Garbage Dumping In Vacant Lot/Land	0.764151	0.517572	0.617143	313
Maintenance Of Dry Waste Collection Centre	0.454054	0.159696	0.236287	526
Maintenance/Repair Of Streetlights	0.964784	0.941977	**0.953244**	1396
Provide good driveable Roads	0.588235	0.039216	0.073529	255
Repair of Existing Footpaths	0.615591	0.349618	0.445959	655
Stray Dog Sterilisation/Animal Birth Control (ABC)	0.900474	0.933661	**0.916767**	407

Table 4. Confusion matrix for RF model

Cat	0	1	2	3	4	5	6	7	8	9
0	766	30	7	48	30	63	4	1	5	8
1	99	109	0	2	1	10	1	0	0	1

2	9	0	198	33	0	4	1	0	2	4
3	44	0	17	1881	1	4	26	3	106	5
4	121	3	1	6	162	15	1	0	2	2
5	349	22	7	35	16	84	2	0	4	7
6	16	0	2	42	2	3	1315	0	4	12
7	14	0	3	205	0	0	3	10	20	0
8	31	0	10	370	0	1	8	3	229	3
9	15	0	1	8	0	1	2	0	0	380

The GB method produced a slightly worse accuracy of 0.72 (parameter setting: n estimator = 100, Learning Rate (LR) = 0.1, max features=20, max depth=20). The performance metrics of the GB model are shown in Table 4 and the confusion matrix is in Table 5. The confusion matrix shows that most of the data are misclassified to the “Fixing/Repairing Potholes” category. This might be caused by the words used in that category are also present in the other categories e.g: “street”/”road”. The best performance is, again, on the ”Maintenance/Repair Of Streetlights” and “Stray Dog Sterilisation/Animal Birth Control (ABC)” category which uses a distinctive keyword such as ”streetlight” and “dogs”.

Table 5. Performance metrics for GB model

Category	Precision	Recall	F1-score	Support
Clearance Of Garbage Dump Or Black Spot	0.532641	0.746362	0.621645	962
Collection Of Door-to-door Garbage	0.648000	0.363229	0.465517	223
Desilting of storm water drains	0.848039	0.689243	0.760440	251
Fixing/Reparing Potholes	0.666091	0.923335	0.773896	2087
Garbage Dumping In Vacant Lot/Land	0.725888	0.456869	0.560784	313
Maintenance Of Dry Waste Collection Centre	0.406977	0.13308	0.200573	526
Maintenance/Repair Of Streetlights	0.976709	0.931232	**0.953429**	1396
Provide good driveable Roads	0.394737	0.117647	0.181269	255
Repair of Existing Footpaths	0.617391	0.325191	0.426000	655
Stray Dog Sterilisation/Animal Birth Control (ABC)	0.953125	0.899263	**0.925411**	407

Table 6. Confusion matrix for GB model

Cat	0	1	2	3	4	5	6	7	8	9
0	718	19	6	106	36	66	6	2	3	0
1	115	81	0	8	0	17	2	0	0	0
2	6	2	173	63	0	1	2	0	2	2
3	14	3	11	1927	1	2	13	27	88	1
4	123	1	0	30	143	11	1	0	2	2
5	351	17	7	53	13	70	2	1	6	6
6	4	1	0	77	2	2	1300	1	4	5
7	3	0	1	193	1	0	0	30	27	0
8	10	0	6	403	1	2	3	15	213	2
9	4	1	0	33	0	1	2	0	0	366

Table 7 shows the accuracy result of the RF and GB model on the same parameter setting. From the experiments, it is concluded that RF benefitted more from additional tree depth. Whereas n_estimator and max feature only slightly improve the accuracy. It is also worth noting that GB is slower in training. For future research in multi-class classification tasks, it is recommended to use a large value on the max_depth parameter. Regarding the insight from the word usage and categorization accuracy, it is recommended to create a better categorization / sub-categorization of the complaints by grouping similar categories and splitting distinct categories. For future research, one can explore different methods such as Word2vec, Paragraph2vec, Word embeddings, and other state-of-the-art Natural Language Processing (NLP) techniques to create the vector and use other text classification approaches such as BERT, CNN, etc.

Table 7. Performance comparison of RF and GB model using the same parameter setting

Method	n_estimator	LR	Max_depth	Max features	Accuracy
RF	50	N/A	10	10	0.48
GB	50	0.1	10	10	0.69
RF	100	N/A	10	10	0.49
GB	100	0.1	10	10	0.71
RF	100	N/A	20	10	0.60
GB	100	0.1	20	10	0.72
RF	100	N/A	20	20	0.66
GB	100	0.1	20	20	0.72

4. Conclusion

Capturing and responding to complaints from the public is an important effort to develop a good city/country. This project aims to utilize Data Mining to automatize complaints categorization. More than 35,000 complaints data in Bangalore City, India, were retrieved from https://www.ichangemycity.com. The vector space of the complaints was created using TF-IDF and the multi-class text classifications were done using Random Forest and Extreme Gradient Boosting. Results showed that both RF and GB have a similar performance with an accuracy of 73%, with RF benefitted from larger max_dept whereas GB is slower in training. Result also showed that the model is highly dependent on the word usage in the complaint's description. Future research suggestions to increase the task performance include a) using bagging methods instead of boosting methods, b) creating a better categorization / sub-categorization of the complaints by grouping similar categories and splitting distinct categories. Future research may also experiment using Word2vec, Paragraph2vec, Word embeddings, and other state-of-the-art NLP techniques for the vector space creation and using other text classification approaches like BERT, CNN, etc.

References

[1] E. Zuliarso, M. T. Anwar, K. Hadiono, and I. Chasanah, "Detecting Hoaxes in Indonesian News Using TF/TDM and K Nearest Neighbor," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 835, no. 1, p. 12036.

[2] M. T. Anwar, H. D. Pumomo, S. Y. J. Prasetyo, and K. D. Hartomo, "Decision Tree Learning Approach To Wildfire Modeling on Peat and Non-Peat Land in Riau Province," in *2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2018, pp. 409–415.

[3] M. T. Anwar, S. Nugrohadi, V. Tantriyati, and V. A. Windarni, "Rain Prediction Using Rule-Based Machine Learning Approach," *Adv. Sustain. Sci. Eng. Technol.*, vol. 2, no. 1, 2020.

[4] M. T. Anwar, W. Hadikurniawati, E. Winarno, and W. Widiyatmoko, "Performance Comparison of Data Mining Techniques for Rain Prediction Models in Indonesia," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2020, pp. 83–

88.
[5] W. Hadikurniawati, M. T. Anwar, D. Marlina, and H. Kusumo, "Predicting tuberculosis drug resistance using machine learning based on DNA sequencing data," in *Journal of Physics: Conference Series*, 2021, vol. 1869, no. 1, p. 12093.
[6] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[7] Z. Xia, K. Stewart, and J. Fan, "Incorporating space and time into random forest models for analyzing geospatial patterns of drug-related crime incidents in a major us metropolitan area," *Comput. Environ. Urban Syst.*, vol. 87, p. 101599, 2021.
[8] M. Zolfaghari and M. R. Golabi, "Modeling and predicting the electricity production in hydropower using conjunction of wavelet transform, long short-term memory and random forest models," *Renew. Energy*, vol. 170, pp. 1367–1381, 2021.
[9] K. Liu, X. Hu, H. Zhou, L. Tong, D. Widanalage, and J. Macro, "Feature analyses and modelling of lithium-ion batteries manufacturing based on random forest classification," *IEEE/ASME Trans. Mechatronics*, 2021.
[10] D. Sun, J. Xu, H. Wen, and D. Wang, "Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest," *Eng. Geol.*, vol. 281, p. 105972, 2021.
[11] T. Chen, T. He, M. Benesty, V. Khotilovich, and Y. Tang, "Xgboost: extreme gradient boosting," *R Packag. version 0.4-2*, pp. 1–4, 2015.
[12] M. T. Anwar, E. Winarno, W. Hadikurniawati, and M. Novita, "Rainfall prediction using Extreme Gradient Boosting," in *Journal of Physics: Conference Series*, 2021, vol. 1869, no. 1, p. 12078.
[13] J. Fan, J. Zheng, L. Wu, and F. Zhang, "Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models," *Agric. Water Manag.*, vol. 245, p. 106547, 2021.
[14] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," *Geosci. Front.*, vol. 12, no. 1, pp. 469–477, 2021.
[15] L. Cui, P. Chen, L. Wang, J. Li, and H. Ling, "Application of Extreme Gradient Boosting Based on Grey Relation Analysis for Prediction of Compressive Strength of Concrete," *Adv. Civ. Eng.*, vol. 2021, 2021.
[16] A. I. A. Osman, A. N. Ahmed, M. F. Chow, Y. F. Huang, and A. El-Shafie, "Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia," *Ain Shams Eng. J.*, 2021.
[17] P. Kumari and D. Toshniwal, "Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance," *J. Clean. Prod.*, vol. 279, p. 123285, 2021.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210107-01 ~ 0210107-10
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8461

Implementation of *k*-means and K-Medians Clustering in Several Countries Based on Global Innovation Index (GII) 2018

Ade Famalika[1], Pardomuan Robinson Sihombing[2]

[1]Universitas Bina Insan, Jl. Raya Siliwangi No.6, RT.001/RW.004, Sepanjang Jaya, Kec. Rawalumbu, Kota Bekasi, Jawa Barat 17114 Indonesia

[2*]Badan Pusat Statistik, Jl. Dr. Sutomo 6-8 Jakarta 10710 Indonesia

*robinson@bps.go.id

Abstract. The Global Innovation Index (GII) is an instrument to assess the ranking of innovation capabilities of all countries. The sub-index of the GII has seven enabler pillars: Institutions, Human Capital and Research, Infrastructure, Market sophistication, Business Sophistication, Knowledge and Technology Outputs, and Creative Outputs. The *k*-means method and *k*-medians method are methods for cluster countries based on GII. Cluster 1 in *k*-means method consists of 48 Countries, Cluster 2 consists of 45 Countries and Cluster 3 consists of 33 Countries and has the average value of seven variables are the highest. Cluster 1 in *k*-medians method consists of 33 Countries and has the average value of seven variables are the highest., Cluster 2 consists of 53 Countries and Cluster 3 consists of 40 Countries. The result clustering with using *k*-means method and *k*-medians method showed that *k*-medians is better than *k*-means method because the variance value of *k*-medians is smaller than *k*-means.

Keywords: GII, *k*-means Cluster, K-Medians Cluster, Variance

(Received 2021-04-28, Accepted 2021-04-30, Available Online by 2021-04-30)

1. Introduction

The Global Innovation Index (GII) is an instrument to assess the ranking of innovation capabilities of 126 countries carried out by the World Intellectual Property Organization (meaning in the English World Intellectual Property Organization-WIPO) in coordination with INSEAD Institute (France) and Cornell University (United States). The Global Innovation Index (GII) project was launched by Professor Dutta at INSEAD in 2007 with the simple goal of determining how to find metrics and approaches that better capture the richness of innovation in society and go beyond such traditional measures of innovation as the number of research articles and the level of research and development (R&D) [1].

There were several motivations for setting this goal. First, innovation is important for driving economic progress and competitiveness-both for developed and developing economies. Many governments are putting innovation at the centre of their growth strategies. Second, the definition of

innovation has broadened, it is no longer restricted to R&D laboratories and to published scientific papers. Innovation could be and is more general and horizontal in nature, and includes social innovations and business model innovations as well as technical ones. Last but not least, recognizing and celebrating innovation in emerging markets is seen as critical for inspiring people, especially the next generation of entrepreneurs and innovators.

The sub-index of the GII has seven enabler pillars: Institutions, Human Capital and Research, Infrastructure, Market sophistication, Business Sophistication, Knowledge and Technology Outputs, and Creative Outputs. Seeing the importeance of the Global Innovation Index, this study aimed at finding out the clustering of countries based on GII data in 2018 with using *k*-means method and *k*-medians method, then compare the two methods [1].

2. Methods

2.1. Research Variables

This study used secondary data sourced from World Intellectual Property Organization-WIPO coordinating with INSEAD and Cornell University. The three institutions measured a country's level of global innovation based on seven variables, including [1]:

- Institutions Variable (X_1)
  The Institutions pillar captures the institutional framework of a country. Nurturing an institutional framework that attracts business and fosters growth by providing good governance and the correct levels of protection and incentives is essential to innovation.
- Human Capital and Research Variable (X_2)
  The level and standard of education and research are activities in a country are prime determinants of the innovation capacity of a nation. This pillar tries to gauge the human capital of countries.
- Infrastructure Variable (X_3)
  The infrastructure includes three sub-pillars: Information and communication technologies (ICTs), General infrastructure, and Ecological sustainability.
- Market Sophistication Variable (X_4)
  The availability of credit and an environment that supports investment, access to the international market, competition, and market scale are all critical for businesses to prosper and for innovation to occur.
- Business Sophistication Variable (X_5)
  The business sophistication tries to capture the level of business sophistication to assess how conducive firms are to innovation activity.
- Knowledge and Technology Outputs Variable (X_6)
  This pillar covers all those variables that are traditionally thought to be the fruits of inventions and/or innovations. The first subpillar refers to the creation of knowledge. The second sub-pillar, on knowledge impact, includes statistics representing the impact of innovation activities at the micro- and macroeconomic. The third sub-pillar, on knowledge diffusion.
- Creative OutputsVariable (X_7)
  The last pillar, on creative outputs, has three sub-pillars The first sub-pillar on intangible assets includes statistics on trademark applications by residents at the national office. The second sub-pillar on creative goods and services and the third sub-pillar on online creativity.

2.2. Stage of Research

2.2.1. Cluster Analysis

Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters [2].

Data clustering algorithms can be hierarchical or partitional. Hierarchical algorithms find successive clusters using previously established clusters, whereas partitional algorithms determine all

clusters at time. Hierarchical algorithms can be agglomerative (bottom-up) or divisive (top-down). Agglomerative algorithms begin with each element as a separate cluster and merge them in successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters.

There are two assumptions that must be fulfilled in cluster analysis, namely samples that are representative (representing the population) and there are no cases of multicollinearity between variables[3]. A representative sample can be seen from the Kaiser-Meyer-Olkin (KMO) value that is greater than 0.5[4].

$$KMO_j = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} u_{ij}^2} \quad (1)$$

The presence or absence of multicollinearity between variables can be seen from the value of Variance Inflation Factor (VIF) which is greater than 10 [5].

$$VIF_k = \frac{1}{1 - {R_k}^2} \quad (2)$$

Here ${R_k}^2$ is the $R^2 = value$ obtained by regressing the k^{th} predictor on the remaining predictors.

2.2.2. K-Means Method

Clustering is a classification of similar objects into several different groups, it is usually applied in the analysis of statistical data which can be utilized in various fields, for example, machine learning, data mining, pattern recognition, image analysis and bioinformatics [6]. In general, partitioning algorithms such as *k*-means and EM highly recommended for use in large-size data. This is different from a hierarchical clustering algorithm that has good performance when they are used in small size data [7]. The method of K-means algorithm as follows [8]:

1) Determine the number of clusters *k* as in shape. To determine the number of clusters K was done with some consideration as theoretical and conceptual considerations that may be proposed to determine how many clusters.
2) Generate K centroid (the center point of the cluster) beginning at random. Determination of initial centroid done at random from objects provided as K cluster, then to calculate the *i*-cluster centroid next, use the following formula:

$$v = \sum_{i=1}^{n} \frac{x_i \quad ; i = 1,2, \dots, n}{n} \quad (3)$$

where $\boldsymbol{v}$ is cluster centroid, $\boldsymbol{n}$ is the number of objects to members of the cluster and $\boldsymbol{x_i}$ is the object to-*i*.

3) Calculate the distance of each object to each centroid of each cluster. To calculate the distance between the object with the centroid author using Euclidian Distance.

$$\boldsymbol{d(x,y) = \|x - y\| = \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2}} \quad (4)$$

4) Allocate each object into the nearest centroid. To perform the allocation of objects into each cluster during the iteration can generally be done in two ways, with a hard K-means, where it is explicitly

every object is declared as a member of the cluster by measuring the distance of the proximity of nature towards the center point of the cluster, another way to do with fuzzy *k*-means.

5) Do iteration, then specify a new centroid position using equation in step 2.
6) Repeat step 3 if the new centroid position is not the same.

2.2.3. K-Medians Method

The *k*-medians method is the development of the *k*-means method. Both produce *k*-cluster formed by measuring the distance between the center point and each object, then each object is grouped according to the nearest center point. Both of these methods have differences, one of which is at the center of the cluster. As the name implies, *k*-means uses the mean (mean) and *k*-medians using the median. Furthermore, the median is descriptive statistics which tend to be more resistant to outliers. Therefore, the use of the K-Medians method will minimize errors in the cluster [9]. The method of *k*-medians algorithm as follows:

1) Determine the number of clusters
   In the *k*-median method the number *k* must be determined in advance and there is no specific rule in determining the number of clusters k, because sometimes the determination of the number of clusters is based on the subjectivity of the researcher. In this study, cluster number *k* was determined using Silhouette Coefficient. Stages of silhouette coefficient calculation[10]:
   - Calculate the average distance of objects with other objects in the cluster with the equation:

$$\boldsymbol{a(i) = \frac{1}{[A] - 1} \sum j \in_{A, j \neq 1} d(i, j)} \quad (5)$$

     Where *a(i)* is average distance between group components, *I* is an object in *cluster* A
     J is other objects in *cluster* A, *d(i,j)* is distance betweenobject *i* and *j*.
   - Calculate the average distance of objects with all other objects in another cluster, then take the minimum value with the equation:

$$d(i, C) = \frac{1}{[A]} \sum j \in C \, d(i, j) \quad (6)$$

     Where, *d(i,*C*)* is the average distance between objects *i* with all objects in another cluster (C), where A ≠ C
   - Calculate the value of silhouette coefficient with the equation:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (7)$$

2) Determine the center point (centroid)
   Some opinions on choosing centroids for the *k*-medians method are as follows:
   - Based on Hartigan (1975), the selection of centroids can determined based on the interval of the number of each observation [11].
   - Based on Rencher (2002), the selection of centroids can be determined through the approach of one of the hierarchical methods [12].
   - Based on Teknomo (2007), the selection of centroids can be randomized from all observations [13].

   In this study, the centroid was chosen based on Teknomo's opinion in determining the centroid, which is to choose centroids randomly from all observation units.
3) Determine the distance of each observation unit to each centroid
   In this case, distance measurements are used to place observations into clusters based on the nearest centroid. The measure of distance used in the *k*-medians method is Manhattan's distance [14].

Manhattan distance is a measurement based on a grid system in which the points in question are placed. The concept is that in order to move from start to end point, one of four directions must be chosen for the point to advance: up, down, left, or right. Each decision will move the start point one unit in the chosen direction. The Manhattan distance is determined by the number of decisions required for the start point to reach the end point [15]. Manhattan distance can be written as follows:

$$\boldsymbol{d(x_{ij}, c_{ij}) = \sum_{j=1}^{p} |x_{ij} - c_{ij}| \; ; i = 1, 2, \dots, k} \tag{8}$$

2.3. Determining the Goodness of the Clustering Method with Standard Deviation

Varians can be calculated by

$$varians = \frac{S_w}{S_B} \tag{9}$$

To find out which method has the best result, we can use the standard deviation in the cluster (S_w) and the standard deviation between clusters (S_B)[16]. The average standard deviation formula in the cluster (S_w):

$$S_w = K^{-1} \sum_{k=1}^{K} S_k \tag{10}$$

Here K is number of clusters formed and S_k is standard cluster k^{th}. Standard deviation formula between clusters (S_B):

$$S_B = \left[(K-1)^{-1} \sum_{k=1}^{K} (\bar{X}_k - \bar{X})^2 \right]^{1/2} \tag{11}$$

Where $\bar{X}_k$ is kth cluster average and X is average of overall clusters.

3. Results and Discussion

3.1. Variables Description

Before clusters of countries using *k*-medians clustering, the average, median and standard deviations of each variable are calculated first. This calculation is done to calculate the confidence interval that will be used in classifying clusters. The calculation results can be seen in Table 1.

Table 1. Descriptive of variables

Variables	Mean	Median	Standard deviation
Institutions	64.07	62.2	15.11
Human Capital and Research	32.57	30.45	15.62
Infrastructure	45.19	45.15	12.66
Market sophistication	48.04	46.8	10.73
Business Sophistication	33.89	30.25	12.34
Knowledge and Technology Outputs	26.58	23.1	13.88
Creative Outputs	30.41	28.35	13.15

3.2. Silhouette coefficient value

The value of the silhouette coefficient is obtained by using software R and shown in Table 3.2 These values show how good the grouping process and the quality of the group formed.

Table 2. Silhouette coefficient value

K	Silhouette Coefficient (*k*-means)	Silhouette Coefficient (*k*-medians)
3	0,4485	0,4730

4	0,2859	0,2581
5	0,2806	0,2237
6	0,3087	0,3032

Based on Table 2 it can see that the highest value of silhouette coefficient on each cluster is K = 3. Therefore, the study uses 3 clusters

3.3. Cluster Analysis

3.3.1. Outlier Detection and Sample Representing the Population

Using R application obtained, the data has an outlier. The value of Kaiser-Meyer-Olkin Measure of Sampling Adequacy is 0.919. The KMO value of 0.919 ranges from 0.5 to 1, it can be concluded that the sample can represent the population and variables can be used for further analysis [17].

3.3.2. Multicollinearity Assumption

All values of VIF are less than five. Based on the results show that VIF in variables $\boldsymbol{X_1, X_2, X_3, X_4, X_5, X_6}$, and $\boldsymbol{X_7}$, there is no multicollinearity.

3.3.3. Cluster Results using K-Means Clustering

After cluster analysis using the *k*-means method, obtained 3 clusters of countries based on the Global Innovation Index.

Table 3. Cluster results using *k*-means clustering

Cluster 1	Albania, Argentina, Bosnia and Herzegovina, Bulgaria, Bahrain, Belarus, Brunei Darussalam, Brazil, Chile, Colombia, Costa Rica, Croatia, Georgia, Greece, Hungary, India, Iran, Jordan, Kazakhstan, Kuwait, Latvia, Lithuania, Moldova, Montenegro, Mongolia, Mauritius, Malaysia, Mexico, Morocco, Namibia, Oman, Panama, Peru, Poland, Qatar, Romania, Russia, Saudi Arabia, Serbia, Slovakia, South Africa, Thailand, Turkey, The former Yugoslav Republic of Macedonia, Tunisia, Ukraine, Uruguay, Vietnam
Cluster 2	Algeria, Armenia, Azerbaijan, Bangladesh, Benin, Bolivia, Botswana, Burkina Faso, Cambodia, Cameroon, Côte d'Ivoire, Dominican Republic, Ecuador, Egypt, El Savador, Ghana, Guinea, Guatemala, Honduras, Indonesia, Jamaica, Kenya, Kyrgyzstan, Lebanon, Madagascar, Malawi, Mali, Mozambique, Niger, Nigeria, Nepal, Philippines, Pakistan, Paraguay, Rwanda, Senegal, Sri Lanka, Tajikistan, Tanzania, Togo, Trinidad and Tobago, Uganda, Yemen, Zambia, Zimbabwe
Cluster 3	AE United Arab Emirates, Australia, Austria, Belgium, Canada, China, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hong Kong (China), Ireland, Israel, Iceland, Italy, Japan, Korea, Luxembourg, Malta, Netherlands, New Zealand, Norway, Portugal, Singapore, Slovenia, Spain, Sweden, Switzerland, United Kingdom, United States of America

Based on Table 3, we can find out the results of grouping using the *k*-means algorithm using Euclidean distance, which is in Cluster 1 consists of 48 Countries, Cluster 2 consists of 45 Countries and Cluster 3 consists of 33 Countries. Then to differentiate the cluster results that is formed, it is necessary to do profilization by calculating the average value of each variable on Table 4. The result as follows:

Ta**ble 4.** Characteristic cluster *k*-means

Variable	Cluster 1	Cluster 2	Cluster 3
Institutions	62.960	50.551	84.112
Human Capital and Research	33.421	17.791	51.485
Infrastructure	46.385	32.296	61.030

Market sophistication	47.083	40.413	59.842
Business Sophistication	30.156	25.029	51.412
Knowledge and Technology Outputs	24.290	15.613	44.879
Creative Outputs	29.740	18.727	47.330

Based on Table 4, it can be known the characteristics of each cluster. Cluster 1 has the average value of seven variables are quite high. Cluster 2 has the average value of seven variables are low, whereas Cluster 3 has the average value of seven variables are the highest.

3.3.4. Cluster Results using K-Medians Clustering

Using R application to find *k*-medians cluster, obtained 3 clusters of countries based on the Global Innovation Index as shown in Table 5.

Table 5. Cluster results using K-Medians Clustering

Cluster 1	AE United Arab Emirates, Australia, Austria, Belgium, Canada, China, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hong Kong (China), Ireland, Israel, Iceland, Italy, Japan, Korea, Luxembourg, Malta, Netherlands, New Zealand, Norway, Portugal, Singapore, Slovenia, Spain, Sweden, Switzerland, United Kingdom, United States Of America
Cluster 2	Albania, Argentina, Armenia, Azerbaijan, Bosnia And Herzegovina, Bulgaria, Bahrain, Belarus, Brunei Darussalam, Botswana, Brazil, Chile, Colombia, Costa Rica, Croatia, Georgia, Greece, Hungary, India, Iran, Jamaica, Jordan, Kazakhstan, Kuwait, Latvia, Lithuania, Moldova, Montenegro, Mongolia, Mauritius, Malaysia, Mexico, Morocco, Namibia, Oman, Philippines, Panama, Peru, Poland, Qatar, Romania, Russia, Saudi Arabia, Serbia, Slovakia, South Africa, Thailand, Turkey, The Former Yugoslav Republic Of Macedonia, Tunisia, Ukraine, Uruguay, Vietnam
Cluster 3	Algeria, Bangladesh, Benin, Bolivia, Burkina Faso, Cambodia, Cameroon, Côte D'ivoire, Dominican Republic, Ecuador, Egypt, El-Savador, Ghana, Guinea, Guatemala, Honduras, Indonesia, Kenya, Kyrgyzstan, Lebanon, Madagascar, Malawi, Mali, Mozambique, Niger, Nigeria, Nepal, Pakistan, Paraguay, Rwanda, Senegal, Sri Lanka, Tajikistan, Tanzania, Togo, Trinidad and Tobago, Uganda, Yemen, Zambia, Zimbabwe

Based on Table 5 we can find out the results of grouping using the *k*-means algorithm using Euclidean distance, which is in Cluster 1 consists of 33 Countries, Cluster 2 consists of 53 Countries and Cluster 3 consists of 40 Countries. Then to differentiate the cluster results that is formed, it is necessary to do profilization by calculating the average value of each variable on Table 6. The result as follows:

Table 6. Characteristic Cluster K-Median

Variable	cluster 1	cluster 2	cluster 3
Institutions	84.112	62.851	49.533
Human Capital and Research	51.485	32.364	17.218
Infrastructure	61.030	45.649	31.745
Market sophistication	59.842	46.789	40.060
Business Sophistication	51.412	30.008	24.590
Knowledge and Technology Outputs	44.879	23.874	15.090
Creative Outputs	47.330	29.249	18.098

Based on Table 6, it can be known the characteristics of each cluster. Cluster 1 has the average value of seven variables are the highest. Cluster 2 has the average value of seven variables are quite high, whereas Cluster 3 has the average value of seven variables are low.

3.3.5. Determining the Goodness of The Clustering Method with Standard Deviation on K-Means
Standard deviation of cluster 1, for the variable mean in each country, where the value is $\bar{x}_I$ = 39.15.

$$S_1 = \sqrt{\frac{(\bar{x}_1 - \bar{x}_I)^2 + \ldots + (\bar{x}_{48} - \bar{x}_I)}{K-1}}$$

$$S_1 = 3.719234$$

Standard deviation of cluster 2, for the variable mean in each country, where the value is $\bar{x}_{II}$ = 28.63.

$$S_2 = \sqrt{\frac{(\bar{x}_1 - \bar{x}_{II})^2 + \ldots + (\bar{x}_{45} - \bar{x}_{II})}{K-1}}$$

$$S_2 = 3.785309$$

Standard deviation of cluster 2, for the variable mean in each country, where the value is $\bar{x}_{III}$ = 57.15.

$$S_3 = \sqrt{\frac{(\bar{x}_1 - \bar{x}_{III})^2 + \ldots + (\bar{x}_{33} - \bar{x}_{III})}{K-1}}$$

$$S_3 = 5.687154$$

So, the standard deviation value in a cluster using the *k*-means method is

$$S_w = \frac{3.719234 + 3.785309 + 5.687154}{3}$$

$$S_w = 4.397232$$

$$\bar{X} = \frac{\bar{x}_I + \bar{x}_{II} + \bar{x}_{III}}{3} = 41.64508$$

$$S_B = \left[(3-1)^{-1} \sum_{k=1}^{K} (\bar{X}_k - \bar{X})^2\right]^{1/2}$$

$$S_B = 14.42526$$

So, the ratio value of standard deviation in cluster and between clusters using *k*-means method is:

$$varians = \frac{S_w}{S_B} = 0.3048$$

3.3.6. Determining the goodness of the clustering method with standard deviation on k-medians
Standard deviation of cluster 1, for the variable mean in each country, where the value is $\bar{x}_I$ = 28.05.

$$S_1 = \sqrt{\frac{(\bar{x}_1 - \bar{x}_I)^2 + \ldots + (\bar{x}_{33} - \bar{x}_I)}{K-1}}$$

$$S_1 = 5.687154$$

Standard deviation of cluster 2, for the variable mean in each country, where the value is $\bar{x}_{II}$ = 38.68.

$$S_2 = \sqrt{\frac{(\bar{x}_1 - \bar{x}_{II})^2 + \ldots + (\bar{x}_{53} - \bar{x}_{II})}{K-1}}$$
$$S_2 = 3.825788$$

Standard deviation of cluster 2, for the variable mean in each country, where the value is $\bar{x}_{III}$ = 57.15.

$$S_3 = \sqrt{\frac{(\bar{x}_1 - \bar{x}_{III})^2 + \ldots + (\bar{x}_{40} - \bar{x}_{III})}{K-1}}$$
$$S_3 = 3.622979$$

So, the standard deviation value in a cluster using the *k*-means method is

$$S_w = \frac{5.687154 + 3.825788 + 3.622979}{3}$$
$$S_w = 4.37864$$

$$\bar{X} = \frac{\bar{x}_I + \bar{x}_{II} + \bar{x}_{III}}{3} = 41.29556$$
$$S_B = \left[(3-1)^{-1} \sum_{k=1}^{K} (\bar{X}_k - \bar{X})^2\right]^{1/2}$$
$$S_B = 14.72897$$

So, the ratio value of standard deviation in cluster and between clusters using *k*-medians method is:

$$varians = \frac{S_w}{S_B} = 0.2973$$

From the results of all clusters using the K-means and *k*-medians methods, cluster validation is sought for both methods using cluster variance values. the cluster variance value will get better when the value gets smaller.

4. Conclusion

The results of the study provided the conclusion based on the analysis that had been carried out, 3 clusters were formed on each method. Cluster 1 in *k*-means method consists of 48 Countries, Cluster 2 consists of 45 Countries and Cluster 3 consists of 33 Countries. Based on the average value, Cluster 1 has the average value of seven variables are quite high. Cluster 2 has the average value of seven variables are low, whereas Cluster 3 has the average value of seven variables are the highest. Furthermore, Cluster 1 in *k*-medians method consists of 33 Countries, Cluster 2 consists of 53 Countries and Cluster 3 consists of 40 Countries, and Cluster 1 has the average value of seven variables are the highest. Cluster 2 has the average value of seven variables are quite high, whereas Cluster 3 has the average value of seven variables are low. From the research that had been done, the result clustering with using *k*-means method and *k*-medians method showed that *k*-medians are better than *k*-means method because the varians value of *k*-medians = 0.297 is smaller than *k*-means = 0.305.

References

[1] Dutta, Soumitra. Lanvin, Bruno, and Wunsch-Vincent, Sacha. 2018. Global Innovation Index 2018. Eleventh edition. WIPO, INSEAD, and Cornell University.

[2] Madhulatha, T.Soni. 2012. An Overview On Clustering Methods. Vol. 2(4) pp: 719-725.

[3] Hair, J. F, Jr. et al. 2006. Multivariate Data Analysis, Sixth Edition. Pearson Prentice Hall: New Jersey

[4] Santoso, S. 2012. Aplikasi SPSS pada Statistik Multivariate, Elex Media Komputindo: Jakarta

[5] Gujarati, D. N., (2004), Basic Econometrics, Fourth Edition. McGraw-Hill: New York.

[6] Chauhan A, Mishra G, and Kumar G. 2011. Survey on Data Mining Techniques in Intrusion Detection vol. 2, no. 7, pp: 2–5.

[7] Riadi I, Istiyanto J E, and Saleh S. 2013. Forensics Framework Based-on Internet Forensics Framework Based-on Clustering.

[8] Eiyanto N S, Mara M N .2013. Characteristics classification by method *k*-means cluster analysis Bul. Llm., vol.02.,no.2, pp.133-136.

[9] Kaufmann L, & Rousseeuw PJ. 1990. Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley.

[10] Handoyo, Rendy., M, R. Rumani., Nasution, Surya Michrandi. 2014. Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan *k*-means Pada Pengelompokkan Dokumen. ISSN. 1412-0100 Vol. 15, No. 2. Pp. 1-10.

[11] Hartigan, J. A. (1975). Clustering Algorithms. New York: John Wiley & Sons.

[12] Rencher, A. C. (2002). Methods of Multivariate Analysis. Canada: John Willey.

[13] Teknomo, K. (2007). *k*-means Clustering Tutorial.

[14] Anderson, Gross, Musicant, Ritz, Smith, dan Steinberg, 2006. Adapting K-Medians to Generate Normalized Cluster Centers.

[15] M. Ackermann, J. Blomer, C. Sohler (2010) Clustering for Metric and Non-Metric Distance Measures ACM Transactions on Algorithms 6:4

[16] Bunkers, et al. (1996). Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique. pp: 130-146.

[17] Fathia, Annisa Nur, Rahmawati, Rita, and Tarno. 2016. Analisis Klaster Kecamatan di Kabupaten Semarang Berdasarkan Potensi Desa Menggunakan Metode Ward dan Single Linkage. Vol.5, No.4, pp:801-810. Semarang

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210108-01 ~ 0210108-07
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8475

"Branket" Design as a Safe Deposit Box Security System using Arduino-Based Tap Sensor

Arsha Raulnadi Trikusuma[1], Mona Rizqa, Dhimas Aria Wardhana, Noora Qotrun Nada*

[1]Faculty of Engineering and Informatics, Universitas PGRI Semarang, Jl. Sidodadi-Timur No.24 Semarang, Central Java 50232, Indonesia

*noora@upgris.ac.id

Abstract. Safe is a safe place to store valuables or documents. Because they are usually made of strong and hard materials, a safe is a place to store valuables and important documents in the event of a natural disaster or fire. In addition, the safe is also equipped with a locking system so that it can also be used to secure valuables or documents from theft. Usually, safes are used by agencies or companies and the general public who have valuable items or documents. Safe security systems that have been used generally use either a manual lock, a rotary lock, or a digital lock. There are several security system developments in the safe, including using a microcontroller-based password and fingerprint code, a fingerprint sensor and an Arduino UNO-based RF remote control, using a microcontroller via SMS and FSK facilities, and other developments in the safe security system. "Branket" (Tap Safe) is a safe with a smart lock system using a knock pattern. The bracket is composed of several electronic components, mainly a microcontroller, a solenoid lock, and a piezoelectric knock sensor. The workflow for using the bracket begins by pressing the power button to turn on the bracket. Then the user sticks his hand into the small space to store or opens the safe by tapping the sensor according to the pattern. Increased security on the bank account includes a locking system with a secret knock pattern, easy to remember by the owner, faster opening of the safe, and the process of opening the safe is difficult for others to know. It is hoped that "branket" will become a new innovation in a unique locking system that still has a high level of security.

Keywords: safes, arduino, knock sensor, smart lock system

(Received 2021-04-30, Accepted 2021-04-30, Available Online by 2021-04-30)

1. Introduction

Safes are one of the most secure storage of valuables or documents [1]. Because it is usually made of strong and hard materials, the safe becomes a place to store valuables and important documents in case of natural disasters or fires [2]. In addition, the safe is also equipped with a locking system so that it can

also be used to secure valuables or documents from theft. Although there are already banks that are also considered safe enough to store money, or the option of digitizing documents, but safes are still used today [3]. This is because safes have advantages such as not having to spend money storage administration fees such as in banks, as well as maintaining the originality of a document to avoid manipulation of documents in digital form. Usually, safes are used by agencies or companies or the general public who have valuable goods or documents [4].

Safes are a convenient storage area but have a high risk of being broken into, so sophisticated security is needed in accordance with technological advances [5]. Safe security systems that have been used generally use locks either with manual locks, rotary locks, or digital locks. The use of keys is already applied as a common security method, but obstacles that occur often find difficulty in opening them. Because the method is arguably still manual and makes it difficult for the user as the need for pull on the lock lever [6]. However, there are several developments in security systems in safes, among others, using password codes, and fingerprints based on microcontrollers, fingerprint sensors and RF control remotes based on Arduino UNO, using microcontrollers through SMS and FSK facilities, and other developments in safe security systems [7].

Along with the development of technology, there is a smart door lock system with a tap pattern [8]. But this system is still reserved for doors or cabinets only. In fact, safes are storage areas that require a secure and unique locking system [9]. So, the tap pattern can also be applied to safes that require security. Knock patterns are a unique innovation but still put forward the security aspect of the safe, where the knock pattern is easy to remember by the owner, the opening of the safe is faster, and has a fairly high confidentiality [10]. So, the process of opening the safe is difficult for others to know. In addition, the tap pattern is also felt faster and more effective than the digital locking system and pins that require longer opening time, as well as fingerprints that need to register each fingerprint [11]. Therefore, "branket" design appeared as a safe with security system using Arduino-based tapping pattern. "Branket" products are expected to be a new innovation in the locking system in a unique safe but still has a high level of security.

2. Introduction

The method used in the writing of this scientific article is by the method of literature studies. According to, research with literature studies is a study that is the same preparation as other research but the source and method of data collection by retrieving data in libraries, reading, recording, and processing research materials [12]. The selection of such methods is based on the purpose of the research to be conducted. This research was conducted to design and develop an application that requires a strong legal basis and theory. Because in the absence of a solid legal and theoretical basis, the application cannot be designed and further developed [13].

The method of literature study uses several steps in its writing. The steps in writing a conceptual article start from extracting information from reliable sources related to what will be discussed. Then all the information that has been obtained is reviewed and analyzed objectively and in accordance with scientific rules [14]. Meanwhile, the necessary literature in the writing of this scientific article, among others, concerning the prevailing laws and regulations in Indonesia, the nature of public policy and democracy, as well as the implementation of the design and design of an application system [15]. The details of the methods or steps used in this scientific article can be described in Table 1.

Table 1. Details of scientific article writing methods

No.	Stages	Description	Method
1.	Literature Study	Collecting various literature related to "Branket" products ranging from the components used includes specifications and uses, and how the system will be implemented.	Discussion

No.	Stages	Description	Method
2.	Analysis product needs	Analyzing the needs of "Branket" components and product features and analyze the suitability of the function of components to be used in "Branket" products.	Discussion
3.	Assembly prototype	Assembling prototype components of "Branket" products in a test board (breadboard) according to the design.	Practice
4.	Testing prototype	Testing the prototype "Branket" product whether it is in accordance with the expected or not.	Practice
5.	Analysis of prototype test results	Analyzing the prototype test results to determine the shortcomings that exist both from the needs of components and the working system of the components.	Discussion

3. Results and Discussion

3.1. Technical Design

Figure 1 shows technical design and features on "Branket". At first glance "branket" does have the same shape as other safes. However, there are some features in "branket" that are different from other safes. At the top there is a power button, a reset button, a reset indicator light, a power indicator, a power indicator light, and a small space to enter a tap pattern. The existence of this small space is intended so that the tap pattern cannot be seen or read by others. The interior of this small space contains piezoelectric beat sensors and silencers so that the sensor does not read taps on other parts of the safe. In addition, "branket" is also equipped with a sealed emergency section. The existence of this emergency section is in anticipation of technical problems such as the owner of the safe who forgot the knock pattern. Later only special technicians can unseal it.

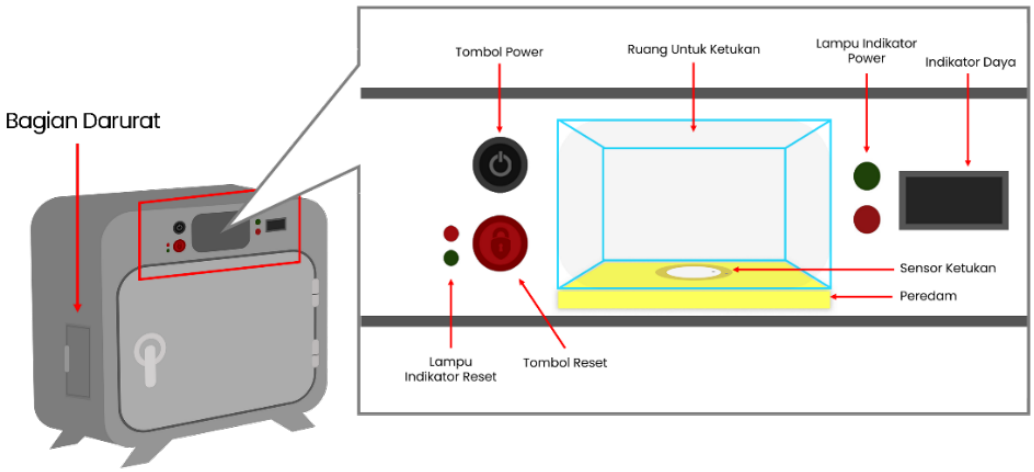


Figure 1. Technical design and features on "Branket"

3.2. Tools and Materials

"Branket" product building materials are divided into two parts: materials for the assembly of prototype series and materials for "branket" products functionally. Materials used for the assembly of the prototype series include Arduino UNO microcontroller, breadboard, piezoelectric beat sensor, 12v solenoid lock, 5v relay, red and yellow/green LED lights, resistors, push-button switches, jumper cables to taste, battery holder case, and two batteries or adapters of 12v each. Then for "branket" products functionally, some raw materials are no longer needed or replaced, among others breadboard, push-button switch, jumper cable, battery holder case, battery, and adapter 12v. Meanwhile, additional raw materials or substitutes include silencers, battery indicator voltmeters, functional buttons, functional cables, as well as a series of power supplies consisting of power supply cables, adapters, PMIC (Power Management Integrated Circuit), and lithium-ion or lithium polymer batteries. Meanwhile, equipment used in the manufacture of "branket" products includes laptops for programming, screws, screwdrivers, solders, etc. Existing equipment is used specifically for the assembly of prototype circuits, assembly circuits on safe bodies, as well as both.

3.3. Creation Process

"Branket" product manufacture is divided into two parts: the assembly of prototype series for testing (Figure 2) and assembly of circuits on the safe body. The assembly process of the prototype circuit for testing is carried out with the help of breadboard. Components assembled on the breadboard include piezoelectric beat sensors, red and yellow/green LED lights, and push-button switches. Then other components are outside the breadboard such as solenoid lock, relay, and Arduino UNO. All of these components are connected to each other by jumper cables and there are also some components that require resistors. The assembly process of the prototype circuit for testing can be divided into 4 parts to make it easier as shown in Figure 3. The first part is the Arduino UNO series with breadboard. The second part is the solenoid, relay, and adapter circuit. The third part is the piezoelectric beat sensor circuit, and the fourth part is a series of knock pattern reset procedures consisting of push-button switches and LED lights.

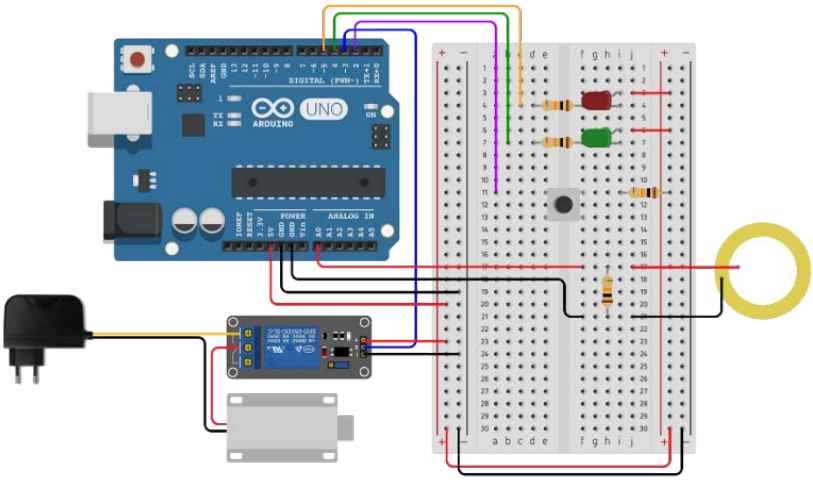


Figure 2. Series of test prototypes

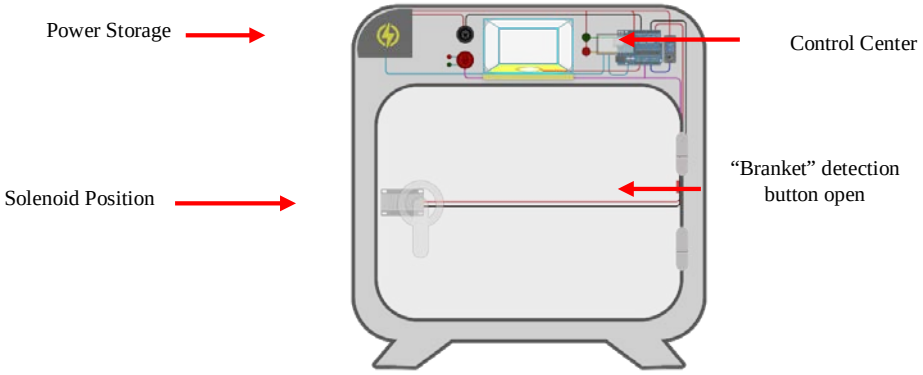


Figure 3. Schematic set of components in “Branket”

Meanwhile, the process of assembling circuits on the safe body is done by selecting components that are still needed, not needed, or replaced with more functional components. Details of component needs in functional products have been explained at the stage of preparation of tools and materials. Most materials or components remain required as the main components of “Branket”. In addition to “Branket” body is made like a typical safe with strong and fire-retardant materials such as titanium which is widely used by safe manufacturers. However, the design of the safe adapts to “Branket”'s features. Component assembly at this stage has also used some additional equipment such as soldering to make the circuit stronger.

3.4. Workflow

“Branket”'s product workflow begins with the testing process on the prototype circuit (Figure 4). The prototype testing process begins by ensuring that the program is uploaded and the components get a power supply, especially for the Arduino and solenoid circuits. Then to save the tap pattern the first time is to press the push-button switch. During pressing, the piezoelectric sensor is tapped with the desired pattern. Push-button switches should not be removed until there is a signal from both LED lights that the pattern has been stored. If the tap pattern is wrong, it will be marked with a red LED light flash. However, if the tap pattern is correct, it will be marked with a green/yellow LED light that lights up and simultaneously the solenoid will open. Overall, “branket” workflows are functionally similar to prototype series workflows in testing. But there are several differentiating procedures. The beat pattern will still be stored in the microcontroller's memory even if “branket” is not supplied with electricity.

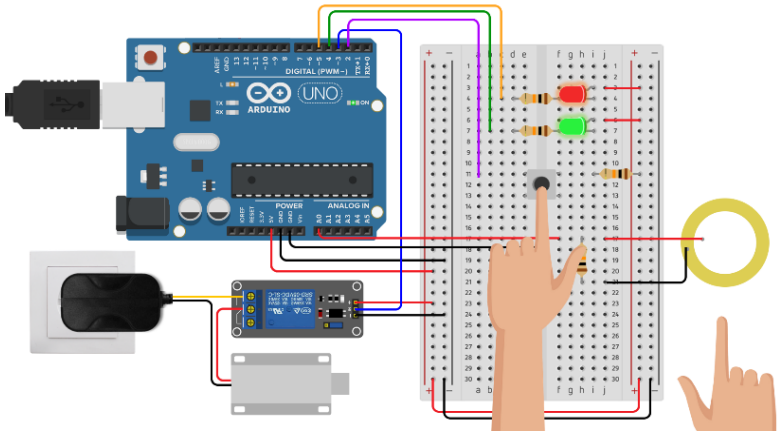


Figure 4. Virtually simulated test pieces

4. Conclusion

Based on the purpose of writing scientific articles and the results of discussions, it can be concluded as follows:

1. "Branket" (Tap Safe) is a safe with a smart lock system using a tap pattern. "Branket" is composed of several electronic components, mainly microcontrollers, solenoid lock, and piezoelectric beat sensors. Before being implemented in the safe, "branket" component sets are first compiled on breadboards for testing. Once tested and able to work as programmed, the circuit can be implemented in the vault.
2. "Branket" usage workflow starts by pressing the power button to turn "branket" on. Then the user puts his hand in a small space to store or open the safe by tapping the sensor according to the pattern. If the pattern is correct, it will be marked with a green LED light that lights up, otherwise if the tap pattern is wrong it will be marked with a red LED light that lights up. The knock pattern reset procedure can only be performed when the vault is open.
3. "Branket" can be a unique and safe innovation in safe locking system. "Branket"'s enhanced security includes a locking system with a secret tap pattern, easy to remember by the owner, faster safe opening, and the process of unlocking the safe is difficult for others to know. "Branket" is expected to be a new innovation in the locking system in a unique safe but still has a high level of security.

References

[1] D. Antonio, "Rancang Bangun Aplikasi Keamanan Brankas Berbasis Sinar Laser Dengan Mikrokontroller Arduino Nano Dan Uno R3," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 2, no. 2, pp. 12–16, 2013.

[2] M. Wijaya and T. Susila, "Sistem keamanan brankas secara otomatis berbasis mikrokontroler dengan menggunakan sms serta pin dan rfid," *Tesla*, vol. 18, no. 2, pp. 139–151, 2016.

[3] I. N. Sukarma, I. G. S. Widarma, and A. S. Wiguna, "Rancang Bangun Sistem Keamanan Brankas Menggunakan Kombinasi Password dan Sidik Jari Berbasis Mikrokontroler ATMEGA328," *J. Matrix*, vol. 6, no. 2, pp. 115–118, 2016.

[4] F. Rabbani, M. K. Resab, and R. Wicaksono, "Sistem Pengamanan Brankas Berbasis GPS Tracking & IoT (Internet of Things)," *J. Autocracy*, vol. 6, no. 1, pp. 36–42, 2019.

[5] S. Sadi and M. Y. M. Pratama, "Sistem Keamanan Buka Tutup Kunci Brankas Menggunakan Bluetooth HC – 05 Berbasis Arduino Mega 2560," *J. Tek.*, vol. 6, no. 2, pp. 99–105, 2017.

[6] F. Yumono, "Sistem Pengamanan Brankas Kantor Perbankan Menggunakan Aktivasi Password Digital Berbasis Mikrokontrol Atmega 8535," *J. Cendekia*, vol. 11, no. 1, pp. 2–6, 2013.

[7] D. Haryanto and B. Nugroho, "Sistem Kunci Pintu Rumah Berbasi Arduino UNO Dengan Irama Ketukan," *Jumantaka*, vol. 3, no. 1, pp. 81–90, 2019.

[8] H. A. Darwito, A. Saleh, and Y. T. Ichwansyah, "Implementasi Finger Vein Recognition pada Sistem Keamanan Brankas," in *Seminar Nasional Teknologi Elektro Terapan 2018*, 2018, vol. 02, no. 01, pp. 147–152.

[9] A. Bazergan and A. Muis, "Rancang Bangun Sistem Pengaman Brankas Berbasis Wireless," in *Prosiding Seminar Hasil Penelitian (SNP2M)*, 2017, vol. 207, pp. 189–195.

[10] A. P. Rantelino, R. R. Isnanto, and E. D. Widianto, "Sistem Pembukaan Kunci Otomatis Menggunakan Identifikasi Pola Ketukan," *J. Teknol. dan Sist. Komput.*, vol. 2, no. 4, pp. 281–287, 2014.

[11] E. Yuliza and T. U. Kalsum, "Alat Keamanan Pintu Brankas Berbasis Sensor Sidik Jari Dan Passoword Digital Dengan Menggunakan," *J. Media Infotama*, vol. 11, no. 1, pp. 1–10, 2015.

[12] A. T. Mahesa, H. Rahmawan, A. Rinharsah, and S. Arifin, "Sistem Keamanan Brankas Berbasis Kartu RFID E-Ktp," *J. Teknol. dan Manaj. Inform.*, vol. 5, no. 1, pp. 1–9, 2019.

[13] M. A. Z. Syafiq, A. A. Rafiq, and H. Susanti, "Pengembangan Metode Haar Cascade Classifier

Pada Pengenalan Mata Untuk Sistem Keamanan Brankas," in *Seminar Nasional Terapan Riset Inovatif (SENTRINOV) Ke-6*, 2020, vol. 6, no. 1, pp. 895–901.
[14] S. Lena and Ridwan, "Sistem Keamanan Brankas Menggunakan SMS (Short Message Service) Berbasis Mikrokontroler," *J. LPKIA*, vol. 1, no. 2, pp. 24–29, 2012.
[15] M. Ichwan, M. G. Husada, and M. Iqbal Ar Rasyid, "Pembangunan Prototipe Sistem Pengendalian Peralatan Listrik Pada Platform Android," *J. Inform.*, vol. 4, no. 1, pp. 13–25, 2013.

Journal ASSET

Advance Sustainable Science, Engineering and Technology

Publisher: Science and Technology Research Centre
Universitas PGRI Semarang INDONESIA

Address : Jl. Lontar No. 1 Semarang
Central Java 50232 INDONESIA

E-Mail : asset@upgris.ac.id

ISSN: 2715-4211