



Utilization of SVM Method and GLCM Feature Extraction in Classifying Fish Images with Formalin

Muhathir^{1*}, Eka Pirdia Wanti², Ayu Pariyandani³, Syed Zulkarnain Syed Idrus⁴, Andre Hasudungan Lubis⁵

^{1,2,3,5}Computer Science Department, Faculty of Engineering, Universitas Medan Area, Indonesia

⁴Faculty of Applied and Human Sciences, Universiti Malaysia Perlis, Perlis Malaysia

Abstract.

Purpose: Fish is a type of animal protein that can be consumed by humans to supplement protein in the body. Due to the fact that there is an abundance of fish in Indonesia, traders often experience losses because of rotting fish. A small proportion of traders tricked the buyers by mixing fish with formaldehyde to preserve fish in order to prevent fish spoilage until it can be consumed. Thus, every fish buyer must be aware of fraud by traders.

Methods: To be able to find out that the fish has been mixed with formalin, the solution offered is computerized by utilizing the GLCM feature extraction as information extraction on the fish image and the SVM method as a classification method.

Result: The results showed an average accuracy of 0.784, precision of 0.799, recall of 0.784, and f-measure of 0.781.

Novelty: The effect of the SVM classification method on the performance measurement of the model is not too big compared to previous studies, but it is better.

Keywords: Fish, Formalin, SVM, GLCM

Received October 2020 / **Revised** April 2021 / **Accepted** May 2021

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



INTRODUCTION

Fish is one of the types of animal protein that can be consumed by humans. Fish is needed by humans because it is to complement the protein in the body. fish are also very easy to find throughout Indonesia. This is because most of Indonesia is the sea, so that the abundant fish income is utilized by the community as a livelihood for fishermen, the protein content in a fish reaches 18% where the content consists of essential amino acids which are very beneficial for the human body. to meet protein needs. Based on the Indonesian List of Food Exchangers (DBMP), one serving of fresh fish (50 g) contains 10 grams of protein (7/3) [1].

With the abundant fish yields from the Indonesian seas, fishermen must sell their catch as quickly as possible. This is because the fish will easily rot if it is not quickly consumed by humans [2]. This should turn the minds of fishermen to make the fish they catch last longer than the usual fresh fish. One of them is fresh fish caught by fishermen mixed with formalin. This is so that if the fish they catch are not sold out, the fish will be preserved using formaldehyde [3]. Formaldehyde substances are chemicals that are harmful if consumed by humans. The nature of these substances, which can damage the vital organs of humans if they are constantly consumed, and at worst can lead to death for those who exceeded within the allowed limit [4].

The fraudulent act of fishermen mixing fresh fish with formaldehyde, requires buyers to be more vigilant in buying the fish. This is because there are many miscreant traders who are interested to gain more profit by mixing fresh fish with formalin excessively. By mixing fish with formaldehyde makes the fish last longer. While fresh fish does not last long and can easily rots, so this technique can reduce losses for fishermen who cannot sell all their catch [5]. The price of formaldehyde used to preserve fish is quite

*Corresponding author.

Email addresses: muhathir@staff.uma.ac.id (Muhathir), syzul@unimap.edu.my (Syed), andrelubis2201@gmail.com (Andre)

DOI: 10.15294/sji.v8i1.26806

affordable, which can provide a big advantage for fraudulent fishermen. However, their actions will not only harm many people, but also may cause a disease for those who had consumed and eventually resulted in death. This is due to the mutagenic and carcinogenic nature of formaldehyde, which if it enters the human body it will trigger the growth of cancer cells and gene defects in the human body [6].

Based on previous research, the authors continued their research to classify fresh fish and fish that had been mixed with formaldehyde by utilizing the support vector machine classification method. The reason for choosing the support vector machine classification method is because this method uses nonlinear mapping to transform data into higher dimensions [9]. With this higher dimension, the data can be separated by a hyperplane [10].

LITERATURE REVIEW

Fish

Fish have various types, shapes, sizes and are different from one fish and another. So that the fish shows that there are certain specifications on the characteristics of each fish [2]. Fish classification is a type or category of fish where each of these species includes one or several groups that are closest, this is because there are many different types of fish that exist on the island of Indonesia. The recognition of this type of fish cannot be separated from the physical characteristics of the fish that are easily seen and can be recognized by distinguishing the types from the fish themselves. The physical form of the fish is closely related to the environment the fish is in. While the body parts of the fish will be used in the identification process, so that it will facilitate the introduction of the fish [1].

Fish are living things that belong to the class of vertebrates and are very diverse. Fish are animals that live in water and can be found starting from river waters, freshwater, to marine waters [11] Fish is known for its delicious taste and besides that fish is a source of vitamins needed for the body. Fish meat stores vitamin A and D which are very high, therefore fish is a source of vitamins A and D [12]. The protein content possessed by fish is very high and good for health. Fish is rich in both essential and non-essential amino acids that can improve blood circulation in the body [13].

Formalin

Formaldehyde has the chemical name Formaldehyde (HCOH), the compound which is the basic ingredient of formaldehyde was discovered by a chemist from Russia, Alexander Mikhailo Vich Butlerov in 1859. Formalin was initially identified by Agust Wilhem Von Hoffam in 1868. Agust Wilhem Von Hoffam It identifies formaldehyde when it passes methanol vapor and water over a hot platinum spiral. The function of formalin as a disinfectant (disinfectant) was only discovered in 1888 (Dapkes RI and Dirjen POM) [4]. Formaldehyde is a liquid that has no color but has a very strong odor. Formalin solution is a dangerous and poisonous compound. Formaldehyde is soluble in water and contains about 37% formaldehyde in water, it is used as a disinfectant and to preserve biological species [14]. Formalin is commonly used to preserve animals, small insects and even humans. However, for some people it turns out that the use of formaldehyde is misused, namely to preserve food although it is clear that the use of formalin for food is prohibited [15]. If the formalin content in the body is high, it will cause death in cells and lead to destruction of the body. In addition, if the formalin in the body is high it will cause other diseases and one of them is that it can trigger cancer (formalin) [6].

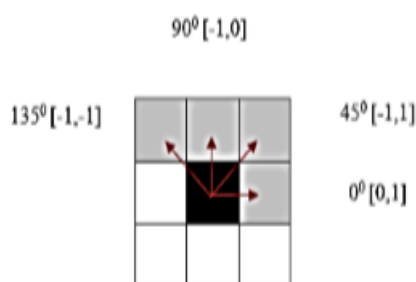
When formaldehyde is exposed to the skin, it can cause discoloration, redness, hardening, numbness, and a burning feeling. When exposed to the eyes, it can cause eye irritation, causing redness, pain, itching, blurred vision, and tears [4][5]. High concentrations of formalin can cause severe tears and damage to the eye lens. If swallowed, the mouth, throat and stomach burn, it hurts to swallow, nausea, vomiting, and diarrhea as well as possible bleeding, severe stomach pain, headache, hypotension (low blood pressure), convulsions, unconsciousness to coma. In addition, damage to the liver, heart, brain, spleen, pancreas, central nervous system, and kidneys can occur [6].

GLCM (Gray Level Co-occurrence Matrix)

Gray Level Co-Occurrence Matrix (GLCM) is the most frequently used feature extraction method, first introduced by Haralick et al. And it has been proposed its use for image classification [16]. The GLCM

method can explore the texture properties of objects and analyze textures that have a relationship between different gray levels [17].

Gray Level Co-Occurrence Matrix (GLCM) is also a method that uses statistical calculations in extracting image texture which considers the spatial relationship of pixels in the image that will be the object of research [18]. The GLCM can be calculated by the following steps, namely as follows. First, the original image texture D is serialized to image G by reducing the number of gray levels, N_g . A typical value for N_g is 16 or 32. Then, the GLCM is calculated from G by scanning the intensity of each pixel and its neighbors, which is defined by displacement d and angle ϕ . A displacement, d can be 1,2,3, ... n while the angle, ϕ is limited to 0°, 45°, 90° and 135° [19]. The GLCM matrix is constructed from ϕ and offset (direction and distance parameters) is shown in Figure 1 [20].



Source: [21]

Figure 1. Extraction of Features in GLCM

The GLCM feature extraction determines the resulting frequency combination of the predetermined pixel brightness values. After extracting features with GLCM, the results of the normalization matrix will be used as input for probabilistic measurements that represent texture features, so that there are several features that can be extracted by GLCM, which are as follows [21]:

1. Angular Second Moment (ASM)
2. Contrast
contrast is a calculation related to the amount of Grayscale intensity diversity
3. Inverse Difference Moment (IDM)
Inverse Difference Moment (IDM) is a calculation in GLCM which serves to measure the homogeneity of the image of the object of research.
4. Entropy
Entropy is calculated to measure the grayscale irregularity in the image. The entropy value will be high if the elements in the GLCM have nearly the same value and the value will be low if the case is close to the value of 1 or 0.
5. Correlation
Correlation is a measure of the linear dependence between grayscale values on the image to be the object of research.
6. IDM (Inverse Different Moment)
IDM (Inverse Different Moment) is a feature extracted by GLCM which focuses on the size of defining the variation of intensity in the image that is the test sample.

The calculation above is a method of extracting the GLCM (Gray Level Co-occurrence Matrix) feature to obtain a feature that is simple and easy to detect images from the object of observation [22].

SVM (Support Vector Machine)

One of the most successful classifications of learning techniques is Support Vector Machine (SVM) which was proposed by Cortes and Vapnik in 1995, where the SVM is enhanced and adapted to different application areas [23]. The accuracy of the SVM method depends on the parameter settings, there are 2 parameters used in SVM, namely the soft margin parameter c and the kernel function independent parameter, where this parameter is a support vector engine that needs to be set to determine a good SVM model [24].

According to Rifaldi [25] Support Vector Machine (SVM) is a supervised machine learning technique and is a popular technique. Training in SVM is carried out with various algorithms, one of which is the Gradient Ascent (GA) or the Sequential Minimal Optimization (SMO). SVM can answer the hyperplane that solves quadratic programming problems [26].

Meanwhile, according to Mabhana [27]. Support Vector Machine (SVM) is a bias classifier defined by a separable hyperplane. In two-dimensional space, a hyperplane is a line that divides the image into two parts. These two parts are located on either side of the hyperplane. In general, SVM and neural networks provide better performance in handling multiclass and continuous features, and logic systems perform better in terms of discrete values. Where a large feature space can be handled with a supporting vector engine and which makes the SVM algorithm the most used in text analysis. SVM is very expensive to train very large data due to its nonlinear kernel [28][31].

METHODS

Dataset

The research data used is in the form of pictures of fresh fish and fish that have been given a mixture of formaldehyde for preservation. The fish used in this study were only pond fish, the sampling process used a smartphone camera with a resolution of 480x640.

Architecture Research

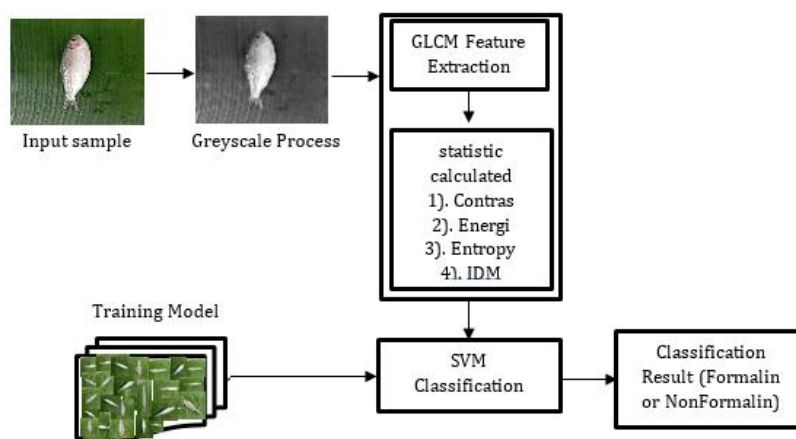


Figure 2. Architecture Research [29][30].

Figure 2 describes the process of classifying images of fresh fish and fish that have been mixed with formaldehyde, the classification process stage begins by entering a sample of fish images then grayscale processing to change from three red, green, blue color spaces to one color space, namely grayscale. The results of the grayscale image are calculated by the co-occurrence matrix and the statistical functions of contrast, energy, entropy, and IDM, the weight gain from the extraction is a reference for classification using the support vector machine method so that it can be seen that the fish image is fresh fish or fish mixed with formalin.

RESULT AND DISCUSSION

The samples in this study used fresh fish images and fish images that had been mixed with formalin. Where the number of each fish image (fresh fish image and fish image mixed with formalin) consists of 250, so the total number of samples used is 500 fish images. From the total sample, 60% is used for training samples and 40% is used for testing samples. Whereas Figures 3 shows part of the sample used in this study.



Figure 3. (a) Fresh fish (b) Formalin fish

After the sample data was collected, which is in the form of fish formalin images, then fish formalin image has feature extraction with GLCM, GLCM explores the feature of fish formalin samples by utilizing energy feature, contrast feature, entropy feature, dan IDM feature. An example of GLCM process using a high-level programming language is shown in Figure 4.

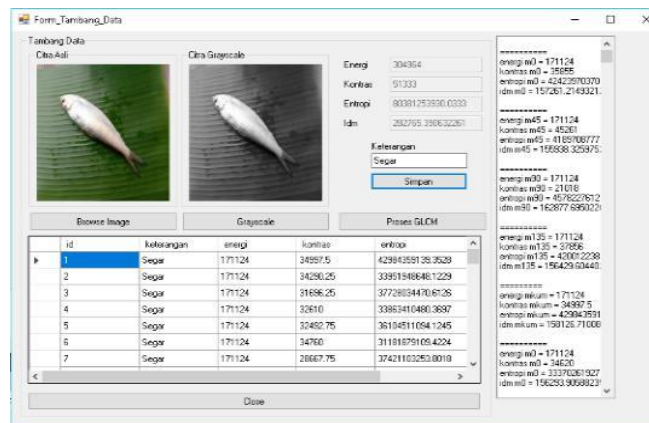


Figure 4. GLCM extraction of fish Imagery

The classification of fish imagery in this study utilizes the SVM classification. Table 1 shows fish formalin classification results with confusion matrix.

Table 1. Test result with confusion matrix

Known	Fresh Fish	Fish Formalin
Fresh Fish	0.896	0.104
Fish Formalin	0.328	0.672

Table 2. Test result using SVM

Known	Accuracy	Precision	Recall	F-Measure
Fresh Fish	0.896	0.732	0.896	0.805
Fish Formalin	0.672	0.866	0.672	0.756
Average	0.784	0.799	0.784	0.781

The results of fish formalin testing using the SVM classification and the extraction of GLCM features is shown in Table 1. Where fresh fish images are identified as fresh fish images with accuracy of 0.896, precision 0.732, recall 0.896 and f-measure 0.805, while images of fish mixed formalin were identified as images of fish mixed formalin with accuracy of 0.672, precision 0.866, recall 0.672 and f-measure 0.756

(refer to Table 2). The effect of the SVM classification method on the performance measurement of the model is not too big compared to previous studies, but it is better.

The comparison of the results of the classification of formaldehyde using SVM with previous research is illustrated in Table 3.

Table 3. Comparison of Classification Results

Method	Accuracy	Precision	Recall	F-Measure
KNN [7]	0.725	0.765	0.725	0.714
Naïve Bayes [8]	0.7	0.795	0.7	0.673
Multilayer Perceptron [32]	0.62	0.654	0.62	0.60
SVM	0.784	0.799	0.784	0.781

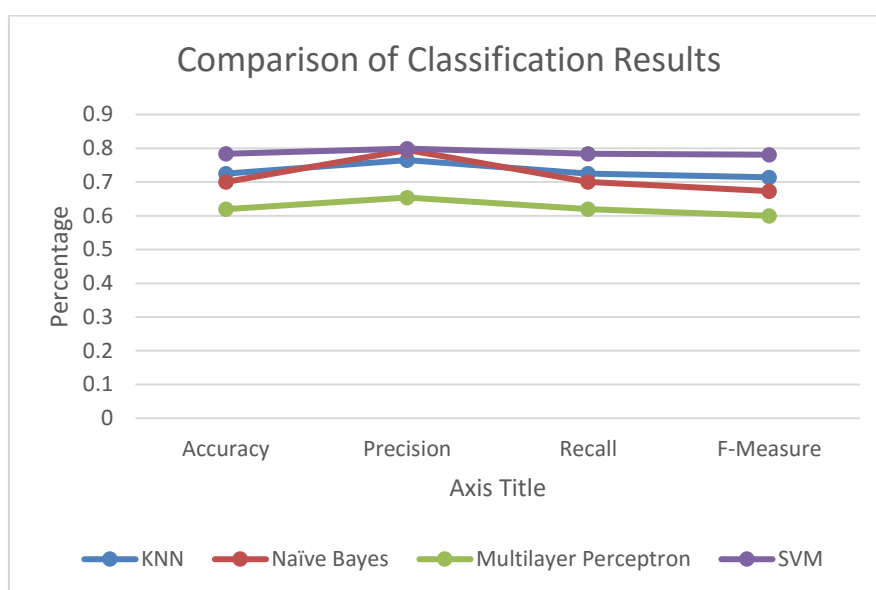


Figure 5. Comparison of Classification Results

Figure 5 illustrates the comparison of the results of the classification test for fresh fish and fish that have been mixed with formalin, the SVM classification method is superior to the KNN, Naïve Bayes, and Multilayer Perceptron classification methods, both in terms of accuracy, recall, precision, and f1-measure.

CONCLUSION

The results from this study show that the SVM classification method is superior with an average accuracy of 0.784, precision of 0.799, recall of 0.784, and f-measure of 0.781. However, this model has not achieved maximum results, but it is better. However, further research development needs be conducted in order to take advantage of deep learning methods by applying them to smartphones.

REFERENCES

- [1] Q. Rachmah, S. Indriani, S. Hidayah, Y. Adhelia and T. Mahmudiono, "Pendidikan Gizi Gemar Makan Ikan Sebagai Upaya Peningkatan Pengetahuan Ibu Tentang Pencegahan Stunting Di Desa Gempolmanis Kecamatan Sambeng Kabupaten Lamongan Provinsi Jawa Timur," *Amarta Nutr. IAGIKMI.*, vol. 1, no. 1, pp. 165-170, 2020.
- [2] M. Ramadhani and D. Heru Murti, "Klasifikasi Ikan Menggunakan Orientasi Fast And Rotation Brief (ORB) dan K-Nearest Neighbor (KNN)," *JUTI : J. Ilm. Teknol. Inf.*, vol. 16 no. 2, pp. 115-120, 2018.
- [3] H. Singgih, "Uji Kandungan Formalin Pada Ikan Asin Menggunakan Sensor Warna Dengan Bantuan FMR (Formalin Main Reagent)," *J. ELTEK.*, vol. 11 no. 1, pp. 55-60, 2013.

- [4] A. Trisnawati and M. S. Arif, "Pelatihan Identifikasi Boraks dan Formalin Pada Makanan Di Desa Bereg, Babadan, Ponorogo," *J. Widya Laksana.*, vol. 8 no. 1, pp. 69-74, 2019.
- [5] E. Puspaningtyas, R. B. Pamungkas and A. Hamad, "Upaya Peningkatan Pengetahuan Bahan Tambahan Pangan Melalui Pelatihan Deteksi Kandungan Formalin dan Boraks," *J. Pengabd. Dan Pemberdaya. Masy.*, vol. 1 no. 1, pp. 46-54, 2017.
- [6] S. L. Poernomo "Analisis Perlindungan Hukum Konsumen Produk Makanan Kemasan Yang Beredar Di Kota Makassar," *Gorontalo Law Rev.*, vol. 3 no. 1, pp. 40-58, 2019.
- [7] A. Pariyandani, D. A. Larasati, E. P. Wanti and M. Muhathir, "Klasifikasi Citra Ikan Berformalin Menggunakan Metode K-NN dan GLCM," *Semin. Nas. Teknol. Inform. (Semant.)*, pp. 42-47, 2019.
- [8] A. Pariyandani, E. P. Wanti, M. Muhathir, "Analysis of the Naïve Bayes Method in Classifying Formalized Fish Images Using GLCM Feature Extraction" *J. Comput. Sci. Inf. Technol. Telecommun. Eng.*, vol. 1, no. 2, 2020.
- [9] M. Muhathir, "Klasifikasi Ekspresi Wajah Menggunakan Bag Of Visual Words," *Jite (J. Inform. Telecommun. Eng.)*, vol. 1, no. 2, pp. 73-82, 2018.
- [10] J. P. Tanjung and M. Muhathir, "Classification of facial expressions using SVM and HOG," *J. Inform. Telecommun. Eng.*, vol. 3, no. 2, pp. 210-215, 2020.
- [11] C. S. U. Dewi, S. and C. J. Harshindhi, Karang dan Ikan Terumbu Karang Pulau Bawean, Malang: UB Press, Maret 2018.
- [12] B. I. Siahaan, "Pendeteksian Ikan Berformalin Melalui Citra Mata Menggunakan Metode Probabilistic Neural Network Berbasis Android," *Respositori Institusi USU.*, 2018.
- [13] N. R. Kinanthi, R. A. Asmara and M. Mentari, "Deteksi Ikan Bandeng Berformalin Berdasarkan Citra Insang Menggunakan Metode Naive Bayes Classifier," *Semin. Nas. Teknol. Inf. dan Apl.*, vol. 10, no. 1, pp. 97-103, 2018.
- [14] H. Singgih, "Uji Kandungan Formalin Pada Ikan Asin Menggunakan Sensor Warna Dengan Bantuan FMR (Formalin Main Reagent)," *J. ELTEK*, vol. 11 no. 1, pp. 55-60, 2013.
- [15] T. Handayani and S. Mutiara, "Pemeriksaan Kandungan Zart Kimia Formalin Pada Baksso Ikan dan Tahu," *J. Katalisator*, Vols. vol.5, no.1, pp. 81-87, 2020
- [16] A. Akhmadiya, N. Nabiyev, K. Moldamurat, K. Dyusekeev and S. Atanov, "Change Detection Based Building Damage Assessment Method Using Radar Imageries With GLCM Textural Parameters," *PREPRINTS*, vol. 3, no. 4, 2020.
- [17] V. A. Gunawan, L. S. Ade Putra and I. I. Fitriani, "Sistem Diagnosis Otomatis Identifikasi Penyakit Jantung Coroner Menggunakan Ekstrasi Ciri GLCM dan Klaasivikasi SVM," *Jurnal Ilmiah Ilmu Komputer*, vol. 15, no. 1, pp. 13-20, 2020.
- [18] J. L. Raheja, S. Kumar and A. Chaudhary, "Fabric Defect Detection Based on GLCM and Gabor Filter : a Comparison," *J. Optik*, pp. 1-6, 2013.
- [19] S. Shalev-Shwartz, Y. Singer, N. Srebro and A. Cotter, "Pegagos : Primal Estimated Sub-Gradient Solver for SVM," *Spinger math. Optim. Soc.*, vol. 3, no. 30, pp. 3-9, 2010.
- [20] J. Wahyudi and I. Maulida, "Pengenaln Pola Citra Kain Tradisional Menggunakan GLCM dan KNN," *JTIULM.*, vo. 4, no. 2, pp. 43-49, 2019.
- [21] D. H. B. Kekre, S. D. Thepade, T. K. Sarode and V. Suryawanshi, "Image Retrieval Using Texture Feature Extracted From GLCM,LBG and KPE," *Int. Journal Comput. Theory Engineering.*, vol. 2 no. 5, pp. 4-9, 2010.
- [22] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image Texture Feature Extration Using GLCM Approach," *Int. J. Sci. Res. Publ.*, vol. 3, no. 5, pp. 290-295, 2013.
- [23] Tuba. E. and Stanimirovic. Z., "Elephant Herding Optimazation Algorithm for Support Vector Machine Parameters Tuning," *Int. Conf.-9th Ed. Electron., Comput. Artif. Intell.*, 2017.
- [24] T. C. Yang, P. S. Yu, K. H. Lin, C. M. Kuo, and H. W. Tseng, "Predictor selection method for the construction of support vector machine (SVM)-based typhoon rainfall forecasting models using a non-dominated sorting genetic algorithm," *Meteorol. Appl.*, vol. 25, no. 4, pp. 510-522, 2018.

- [25] M. H. G. Rifaldi and E. B. Setiawan , "Competence Classification of Twitter Users Using Support Vector Machine (SVM) Method," *Interational Comference Inf. Commun. Technol.*, vol. 1, no. 1, 2019.
- [26] D. F. Indarwati, D. E. Rahmawati and S. Anam, "Klasifikasi Fungsi Senyawa Aktif Berdasarkan Data Simplified Molecular Input Line Entry System (SMILES) Menggunakan Metode Support Vector Machine (SVM)," *Interational Comference Inf. Commun. Technol.*, vol. 3 , no. 8, pp. 7844-7850, 2019.
- [27] M. Akter Mim and Z. K.M Shawkat, "GIS-Based Analysis of Changing Surface Water in Rajshahi City Corporation Area Using Support Vector Machine (SVM), Decision Tree & Random Forest Technique," *Mach. Learn. Res.*, vol. 3, no. 2, pp. 11, 2018.
- [28] E. Tuba and Z. Stanimirovic. , "Elephant Herding Optimazation Algorithm for Support Vector Machine Parameters Tuning," *International Conference-9th Edition Electronics.. Computer and Artificial Intelligence* , 2017.
- [29] M. Muhathir, T. T. S. Sibarani and A.-K. Al-Khowarizmi, "Analysis K-Nearest Neighbors (KNN) in Identifying Tuberculosis Disease (Tb) By Utilizing Hog Feature Extraction," *Al'adzkiya Int. Comput. Sci. Inf. Technol. (AIoCSIT) J.*, vol. 1, no. 1, 2020.
- [30] M. Muhathir. H. Mawengkang and M. Ramli, "Kombinasi Z-Fisher Transform Dan Bray Curtis Distance Untuk Pengenalan Pola Huruf Jar Pada Citra Al-Quran," *J. Bisman Info*, vol. 4, no. 1, 2017.
- [31] S. Fransiska and A. Irham Gufroni, "Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method," *Sci. J. Informatics*, vol. 7, no. 2, pp. 2407–7658, 2020.
- [32] E. P. Wanti and M. Muhathir, "Pengidentifikasian Citra Ikan Berformalin Dengan Menggunakan Metode Multilayer Perceptron," *J. Sains Komput. Inform.*, vol. 5, no. 1, pp. 491–502, 2021.