Caries Level Classification using K-Nearest Neighbor, Support Vector Machine, and Decision Tree using Zernike Moment Invariant Features

Yessi Jusman
Department of Electrical Engineering
Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
yjusman@umy.ac.id or https://orcid.org/0000-0002-1637-087X

Muhammad Ahdan Fawwaz Nurkholid
Department of Electrical Engineering
Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
muhammad.ahdan.2016@ft.umy.ac.id

Muhammad Fajrul Faiz
Department of Electrical Engineering
Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
m.fajrul.ft19@mail.umy.ac.id

Sartika Puspita
Department of Oral Biology, School of Dentistry
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
sartika.puspita@mail.umy.ac.id

Lady Olivia Evellyne
Department of Electrical Engineering
Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
lady.olivia.ft19@mail.umy.ac.id

Kahfi Muhammad
Department of Electrical Engineering
Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
kahfi.muhammad.ft21@mail.umy.ac.id

2022 International Conference on Data Science and Its Applications (ICoDSA) | 978-1-6654-8665-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICoDSA55874.2022.9862879

***Abstract*—Dental caries is the most common disease and is reported as one of the oldest diseases. To avoid the occurrence of dental caries, there are four ways; maintaining oral hygiene, consuming healthy food, adequate fluoride and giving fracture sealers. Regular dental check-ups can also reduce the risk of developing this disease. In detecting this disease, dentists often fail. This failure was due to the inability to detect early enamel lesions that had not yet developed into cavitation. In this regard, new techniques were developed to help detect this disease. This method uses 10-folds cross validation. This cross validation divides 90% (1256 images) for the train data and 10% (132 images) for the test. In this research using the Zernike moment method for feature extraction. The average results of training accuracy are 94.55%, 84.24%, and 88.46% and the average results of training times are 0.74, 1.63, and 0.77 seconds for K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Decision Tree (DT), respectively. This research has obtained perfect performances of classification which are represented with AUC values more than 0.95 for each model.**

Keywords—dental caries, dentistry, machine learning, KNN, SVM, DT

I. Introduction

Dental caries is the most common disease and is reported as one of the oldest diseases [1]. According to the Global Burden of Disease Study 2019 (GBD 2019), 2 billion people are affected by caries in permanent teeth and 520 million children are affected by caries in primary teeth [2]. Caries occurs when plaque builds up on the tooth surface. Plaque is formed from sugars from food and drink that turn into acids that damage teeth over a long period of time [3].

Apart from consuming sugar continuously, lack of fluoride and lack of dental hygiene are also causes [3]. To avoid the occurrence of dental caries, there are four ways; maintaining oral hygiene, consuming healthy food, adequate fluoride and giving fissure sealants [4]. Regular dental check-ups can also reduce the risk of developing this disease [5].

In detecting this disease, dentists often fail. This failure was due to the inability to detect early enamel lesions that had not yet developed into cavitation [6]. Cavitation is caries that has occurred for weeks or even years on the tooth surface [7]. In this regard, new techniques were developed to help detect this disease [6].

Recently, it is known that artificial intelligence has become important in the fields of medicine and dentistry [8]. One of them is the use of artificial neural network which was studied by Geetha et al., to detect the presence of caries or not in 105 radiography images. The result is caries detection accuracy is 97.1% and false positive is 2.8% [9]. This research shows that the neural network is much more accurate than the traditional method [8].

Other studies were also conducted by Elias D Berdouses et al. to detect and classify occlusal caries in 2015. As a result, the methodology correctly detected 337 out of 340 regions in detecting stages with a detection accuracy of 80% and an overall accuracy of 83% [10]. In 2008 there was also research to diagnose proximal dental caries by Devito et al. The accuracy obtained is 88.4% [11]. Machine learning method is also used to diagnose the prediction of root caries by Man Hung et al. The results show that SVM produces an accuracy of 97.1% [12].

Apart from dental caries, AI can also be used to restore teeth in research by Abdalla-Aslan R et al. since 2020. As a result, the algorithm used can detect 93.6% of the restored teeth in 83 panoramic images [13]. Another research was conducted by Javed et al, in 2019 which also used AI. This research predicts pre- and post-Streptococcus mutans using an iOS-based application which produces an accuracy of 99.03% [14]. This shows that AI can solve this problem.

Based on the problems above, it can be concluded that machine learning can be a solution. The solution is expected to help dentists in diagnosing dental caries. Dental caries in this research was divided into four classes whose complete details and methods can be found in the methodology section.

II. Methodology

This research has several stages. This stage is made so that the research can run well. There are two tools used, namely hardware and software. The hardware used in this research is listed in Table I, while the software uses the Matlab 2020a.

TABLE I. SPECIFICATIONS ON HARDWARE

Hardware	Characteristic
RAM	16 GB
Processor	Core i5 9400f
Graphics	Nvidia RTX 2060 6GB

There are four classes in this research, namely 1, 2, 3 and 4. In class 1 there are 125 original image data, class 2 is 94, class 3 is 80 and class 4 is 48, the details are presented in Table II.

TABLE II. IMAGE DATA

	Training	Testing	Total
Class 1	452	48	500
Class 2	340	36	376
Class 3	288	32	320
Class 4	176	16	192
Total Data	**1256**	**132**	**1388**

The data is divided into 2 parts, namely train data and test data. This method uses 10-fold cross validation. This cross-validation technique is commonly used [15].This cross validation divides 90% (1256 images) for the train data and 10% (132 images) for the test [16].

A. Preprocessing

In this research, there is a flowchart shown in Figure 1.

In the pre-processing stage, the image data is prepared to be processed on the system built. This stage consists of five steps, these steps are,

1) Cropping: Image data is cropped to take the important parts that will be used in this system.

2) Grayscale: The cropped image data is then converted into a grayscale image of the original image. The purpose of this stage is to reduce the computational requirements and simplify the algorithm [17].

3) Image Enhancement: Image data is also enhanced; this aims to improve image quality so that it can be read properly on the system used [18].

4) Resizing: The image data to be used is resizing to 455x455 pixels. This goal is done so that all images have the same resolution so that they can be used properly.

5) Augmentation: This stage is used as a process to modify an image. The original image is then changed in position or shape. This process aims so that the system can produce more accurate results. In this research using 3 ways, namely flip vertical, flip horizontal, and flip vertical horizontal [19].

B. Feature Extraction

The pre-processed data is then used by Zernike moment for feature extraction. Zernike moment is a method for projecting an image into an orthogonal function. This function makes this method has advantages, one of which is the ease of image reconstruction. In addition, this method also does not change the value of the rotated image [20]. The results of feature extraction in this research are presented in Table III.

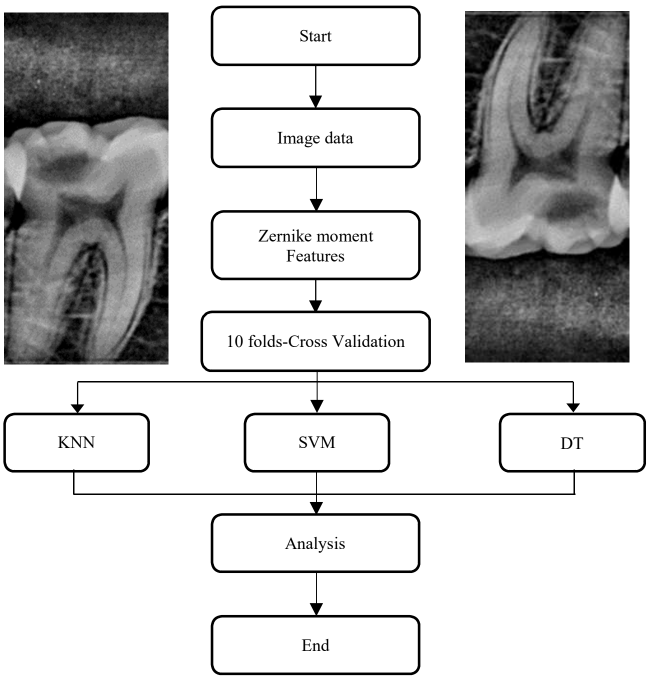

Fig. 1. Flowchart of Research.

TABLE III. AVERAGE VALUE AND STANDARD DEVIATION OF ZERNIKE MOMENT INVARIANT FEATURE EXTRACTION

Feature	Class 1	Class 2	Class 3	Class 4
Feature 1 (Z)	0.003 ± 0.008	-0.001 ± 0.007	-0.002 ± 0.005	-0.003 ± 0.005
Feature 2 (AOH)	0.008 ± 0.006	0.007 ± 0.005	0.006 ± 0.006	0.006 ± 0.006
Feature 3 (PhiOH)	-0.661 ± 90.252	-1.199 ± 119.236	1.266 ± 120.260	-0.089 ± 124.709

C. Classifications

As presented in Figure 1, this study employed three models: Fine K-Nearest Neighbor (KNN), Fine Gaussian Support Vector Machine (SVM) dan Fine Decision Tree (DT). The results are shown the training and the testing results. The training results was analyzed by using the best accuracy training results based on ROC curves. The 10-fold cross validation resulted the 10 values of training accuracies and training times. The testing results are shown by accuracy, precision, recall, specificity, and F-score.

III. RESULTS AND DISCUSSION

Based on the results in Table 3, three features based on Zernike moment invariant algorithm can be significantly differentiated among the four classes of caries levels. Thus, the features are used to be input for machine learning. The results of machine learnings are presented in Figure 2, Table 4, and Table 5. Figure 2 represented the results of training of three models (Fine KNN, Fine Gaussian SVM and Fine DT).

The results are the best and the averages of accuracy training results per models due to the use of 10 folds cross validation method for dataset arrangement. Based on Figure 2, the results of the best and average performance per models are perfect performance of classification which are represented with AUC values more than 0.95 for each model. Several of our previous research can be found in detail in [21-23].

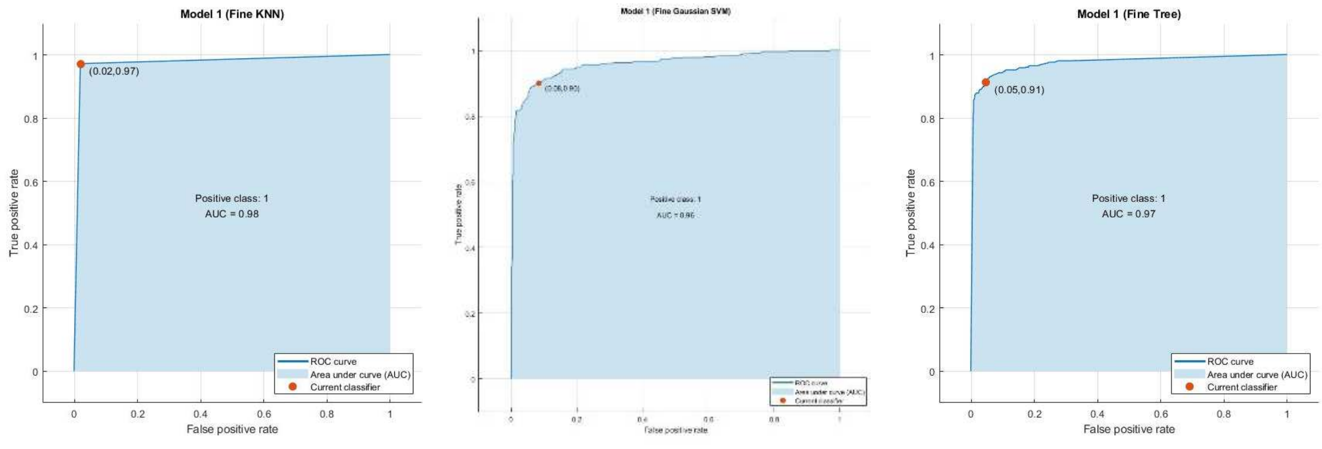

Fig. 2. ROC Training (a) Fine KNN, (b) Fine Gaussian SVM, (c) Fine Tree.

TABLE IV. VALUE OF ACCURACY AND TRAINING TIME WITH FEATURE EXTRACTION METHODS FINE KNN, FINE GAUSSIAN SVM AND FINE DT

	Fine KNN		Fine SVM		Fine DT	
	Acc (%)	*Time (%)*	*Acc (%)*	*Time (%)*	*Acc (%)*	*Time (%)*
Run 1	95.30	0.76	84.20	1.64	**89.70**	0.76
Run 2	94.10	0.74	84.00	1.62	88.70	0.81
Run 3	93.90	0.74	84.40	1.65	87.70	0.75
Run 4	93.70	0.75	84.20	1.61	89.10	0.73
Run 5	94.60	0.74	**85.20**	1.63	88.10	0.75
Run 6	94.90	0.73	83.00	1.61	87.30	0.76
Run 7	94.60	0.73	84.20	1.64	89.00	0.73
Run 8	**95.80**	0.74	84.60	1.62	89.90	0.77
Run 9	94.00	0.73	84.40	1.63	87.50	0.77
Run 10	94.60	0.72	84.60	1.62	87.60	0.77
Average	**94.55**	**0.74**	**84.24**	**1.63**	**88.46**	**0.77**

The results of 10 folds in term of accuracy and training time are tabulated in Table 4. The results are represented good performance of training accuracy and time for three models. The best training results in term accuracy and training time is Fine KNN model. The results achieved the highest accuracy of 95.8% in Run 8 and average accuracy of 94.55%. Whereas the highest and averages of accuracies for Fine Gaussian SVM result are 85.20% in Run 5 and 84.24%. The highest and average results of fine DT results are 89.7% in Run 1 and 88.46%. The average results of training times are 0.74, 1.63, and 0.77 seconds for fine KNN, fine SVM, and fine DT, respectively. Based on the training results, Fine KNN has the best performances.

The results of testing performances are represented in Table 5. The accuracy, precision, recall, specificity, and F-score are compared for three models in the four classes. Based on Table 5, the best result of accuracy is achieved by fine SVM model with value 67% in class 3 and 4 whereas the fine KNN also achieved good value 65% in class 4. The best result of precision is achieved by fine SVM model with value 80% in class 3 and 73% in class 4 whereas the fine KNN model achieved good value 72%.

The best recall performance is achieved by Fine KNN and Fine Gaussian SVM models with value 75% in class 1. Whereas the best specificity and F score values respectively are 73% and 71% for fine Gaussian SVM model in classes 3 and 4. F-score values of fine KNN model are quite same of 67%, 65%, 63%, and 69% for classes 1 to 4.

Based on the results, the performance of Fine Gaussian SVM model is better than Fine KNN model in term of precision, recall, and specificity with value up to 70% respectively.

TABLE V. TESTING PERFORMANCES WITH FEATURE EXTRACTION METHODS FINE KNN, FINE SVM AND FINE DT

Metrics	Models	Class			
		Class 1 (%)	*Class 2 (%)*	*Class 3 (%)*	*Class 4 (%)*
Accuracy	KNN	63	61	59	**65**
	SVM	54	58	**67**	**67**
	DT	**62**	41	55	55
Precision	KNN	61	61	**72**	68
	SVM	48	64	**80**	**73**
	DT	**70**	48	57	57
Recall	KNN	**75**	69	56	69
	SVM	**75**	60	64	69
	DT	**67**	47	**67**	**67**
Specificity	KNN	51	51	**63**	59
	SVM	38	55	**73**	65
	DT	**53**	32	40	40
F-score	KNN	67	65	63	**69**
	SVM	59	62	**71**	**71**
	DT	**68**	48	62	62

IV. CONCLUSION

Performances of Fine KNN and Fine Gaussian SVM models are good for the classification task in caries level X ray images. The features from Zernike moment invariant are fed to be input of three machine learning models. The best training results in term accuracy and average training time is achieved by Fine KNN model with values 95.8% and 0.74 seconds. Whereas the highest Fine Gaussian SVM result is 85.2% of accuracy and training time is 1.63 seconds. The highest fine DT results is 89.7% of accuracy and 0.77 seconds in average training time. The average results of accuracy are 94.55%, 84.24%, and 88.46% for Fine KNN, Fine Gaussian SVM, and Fine DT, respectively. For the testing performances, Fine Gaussian SVM model is better than fine KNN model in term of precision, recall, and specificity with value up to 70% respectively.

The research of caries level classification can be improved in term of the precision, recall, and specificity in testing performances by adding the image processing algorithm and chosen model for classification.

ACKNOWLEDGMENT

This research is supported by Universitas Muhammadiyah Yogyakarta and research project grant from Ministry of Research and Technology of the Republic of Indonesia.

REFERENCES

[1] M. Rathee and A. Sapra, "Dental Caries," *StatPearls*, Oct. 2021, Accessed: Apr. 21, 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK551699/

[2] Global Burden of Disease Collaborative Network, "Global Burden of Disease Study 2019 (GBD 2019)," *Seattle: Institute of Health Metrics and Evaluation (IHME)*, 2020.

[3] World Health Organization (WHO), "Oral health," Mar. 15, 2022. https://www.who.int/news-room/fact-sheets/detail/oral-health (accessed Apr. 21, 2022).

[4] Department of Health and British Association for the Research of Community Dentistry, "Delivering better oral health: an evidence-based toolkit for prevention," London, 2017. Accessed: Apr. 21, 2022. [Online]. Available: https://www.gov.uk/government/publications/delivering-better-oral-health-an-evidence-based-toolkit-for-prevention

[5] J. AlHumaid, Z. Salloot, A. Al-Ansari, M. el Tantawi, Y. AlYousef, and F. Al-Harbi, "Contribution of preventive methods in controlling caries among Saudi primary schoolchildren: a population-based cross-sectional research," *Acta Odontologica Scandinavica*, vol. 76, no. 6, pp. 422–426, Aug. 2018, doi: 10.1080/00016357.2018.1425899.

[6] H. Yılmaz and S. Keleş, "Recent Methods for Diagnosis of Dental Caries in Dentistry," *Meandros Medical and Dental Journal*, vol. 19, no. 1, pp. 1–8, Apr. 2018, doi: 10.4274/meandros.21931.

[7] W. H. Bowen, "Dental caries - not just holes in teeth! A perspective," *Molecular Oral Microbiology*, vol. 31, no. 3, pp. 228–233, Jun. 2016, doi: 10.1111/omi.12132.

[8] A. Ossowska, A. Kusiak, and D. Świetlik, "Artificial Intelligence in Dentistry—Narrative Review," *International Journal of Environmental Research and Public Health*, vol. 19, no. 6. MDPI, Mar. 01, 2022. doi: 10.3390/ijerph19063449.

[9] V. Geetha, K. S. Aprameya, and D. M. Hinduja, "Dental caries diagnosis in digital radiographs using back-propagation neural network," *Health Information Science and Systems*, vol. 8, no. 1, p. 8, Dec. 2020, doi: 10.1007/s13755-019-0096-y.

[10] E. D. Berdouses, G. D. Koutsouri, E. E. Tripoliti, G. K. Matsopoulos, C. J. Oulis, and D. I. Fotiadis, "A computer-aided automated methodology for the detection and classification of occlusal caries from photographic color images," *Computers in Biology and Medicine*, vol. 62, pp. 119–135, Jul. 2015, doi: 10.1016/j.compbiomed.2015.04.016.

[11] K. L. Devito, F. de Souza Barbosa, and W. N. F. Filho, "An artificial multilayer perceptron neural network for diagnosis of proximal dental caries," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, vol. 106, no. 6, pp. 879–884, Dec. 2008, doi: 10.1016/j.tripleo.2008.03.002.

[12] M. Hung *et al.*, "Application of machine learning for diagnostic prediction of root caries," *Gerodontology*, vol. 36, no. 4, pp. 395–404, Dec. 2019, doi: 10.1111/ger.12432.

[13] R. Abdalla-Aslan, T. Yeshua, D. Kabla, I. Leichter, and C. Nadler, "An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 130, no. 5, pp. 593–602, Nov. 2020, doi: 10.1016/j.oooo.2020.05.012.

[14] S. Javed, M. Zakirulla, R. U. Baig, S. M. Asif, and A. B. Meer, "Development of artificial neural network model for prediction of post-streptococcus mutans in dental caries," *Computer Methods and Programs in Biomedicine*, vol. 186, p. 105198, Apr. 2020, doi: 10.1016/j.cmpb.2019.105198.

[15] G. J. McLachlan, K.-A. Do, and C. Ambroise, *Analyzing Microarray Gene Expression Data*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2004. doi: 10.1002/047172842X.

[16] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, Elsevier, 2018, pp. 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X.

[17] C. Kanan and G. W. Cottrell, "Color-to-Grayscale: Does the Method Matter in Image Recognition?," *PLoS ONE*, vol. 7, no. 1, p. e29740, Jan. 2012, doi: 10.1371/journal.pone.0029740.

[18] S. K. Haldar, "Photogeology, Remote Sensing, and Geographic Information System in Mineral Exploration," in *Mineral Exploration*, Elsevier, 2018, pp. 47–68. doi: 10.1016/B978-0-12-814022-2.00003-4.

[19] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *Journal of Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[20] A. Tahmasbi, F. Saki, and S. B. Shokouhi, “An effective breast mass diagnosis system using Zernike moments,” in *2010 17th Iranian Conference of Biomedical Engineering (ICBME)*, Nov. 2010, pp. 1–4. doi: 10.1109/ICBME.2010.5704941.

[21] Y. Jusman, M.K. Anam, S. Puspita , and E. Saleh,” Machine Learnings of Dental Caries Images based on Hu Moment Invariants Features,” in 2021 International Seminar on Applications for Technology of Information and Communication (iSemantic).

[22] Y. Jusman, M.K. Anam, S. Puspita , E. Saleh, S. N. A. M. Kanafiah, and R. I. Tamarena (2021). Comparison of Dental Caries Level Images Classification Performance using KNN and SVM Methods. 2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA).

[23] Y. Jusman, R. I. Tamarena, S. Puspita , E. Saleh, and S. N. A. M. Kanafiah (2020). Analysis of Features Extraction Performance to Differentiate of Dental Caries Types Using Gray Level Co-occurrence Matrix Algorithm. 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE).