MALAYSIAN JOURNAL OF SCIENCE

MJS

ISSN 1394-3065

Malaysian Journal of Science

Vol 45 | No 2 | June 2026

MJS in indexed in Scopus, Google Scholar, Chemical Abstracts Service Database, ASEAN Citation Index (ACI), & MyCite.

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 45(2): 1-12 (June 2026) https://mjs.um.edu.my

Predicting DNA Sequence Similarity Across Species Using Machine Learning: a K-mer Based Approach

Vinodkumar R. Patil[1a]*, Archana S. Vaidya[2a] and Manisha S. Patil[3b]

Abstract: This study examines the effectiveness of machine learning methods for categorizing DNA sequences across human, chimpanzee, and dog samples. We employed k-mer encoding to renovate DNA structures into numerical representations suitable for machine learning models. Four classifiers, such as naive bayes and weighted naive bayes, random forest, K-nearest neighbors, and decision tree, were applied and evaluated using accuracy, precision, recall, and F1-score. The naive bayes classifier consistently outperformed the others across all three datasets, achieving the highest accuracy in classifying human DNA (98.4%), followed by chimpanzee DNA (91.4%), and exhibiting significantly lower accuracy for dog DNA (68.9%). This performance disparity is attributed to the increasing evolutionary distance from human DNA. Additionally, a weighted naive bayes model that was trained on human data showed very high accuracy in predicting chimpanzee (99.3%) and dog (92.6%) DNA sequences. The results presented here show the significance of taking into account evolutionary relations and dataset features whenever developing and training classification models for genetic sequence analysis. The research extends the present research by evaluating the performance of several different algorithms on separate DNA databases, identifying strengths and weaknesses, and suggesting avenues for future research focusing on advanced feature engineering and algorithm selection for improved cross-species classification.

Keywords: *DNA classification, machine learning, K-mer, naive bayes.*

1. Introduction

DNA is a hereditary molecule found in every living species that determines the basic characteristics of bodies and acts as a genomic scheme for a progressing organism. It is comprised of a phosphate category, a sugar category, and nitrogenous compounds (adenine, cytosine, guanine, and thymine) that make up its double-stranded molecule. DNA analysis is decisive in the recognition of diseases, tracing the extent of infections, solving crimes, and paternity tests. Primers, human nucleotide sequences, are used for DNA synthesis, which is an essential tool in molecular biology (Momenzadeh et al., 2020). These primers have been used for the recognition and detection of parasites, bacteria, and viruses. The PCR technology amplifies the DNA segment of a preexisting virus, making it possible to detect it. DNA analysis is now one of the most popular topics in computational biology. DNA sequencing is a way of ordering nucleotides in a DNA fragment that sometimes exists in double strands (Nayak et al., 2022).

Different techniques have been designed for the application of DNA sequences for species identification purposes in different areas, including the score of correspondence, population genetic knowledge, and, particularly for species organizing features, phylogeny. DNA sequencing research is done through shotgun cloning and walking. Classification is one of the means of character recognition of single or groups of characters, like supermolecule sequences. There are several methods applied to classify these sequences into their particular division, secondary category, or family in order to eliminate alternatives, equalize their value, and finally categorize the sequence. Bioinformatics is the analysis of genetic material sequencing information, which refers to the identification of a DNA sequence's structural order. Categorization breaks up a gene into different regions; selected-effect function genes are categorized into functional DNA and rubbish DNA (Benson et al., 2009).

While unusable DNA loses the unselected-effect activity, healthy DNA possesses it. Efficient DNA sequences can be either precise, with the literal DNA choosing the nucleotide order, or indifferent, with the indifferent DNA choosing the existence or inactivity of an arbitrary pattern of DNA. Junk DNA is part of the organism's fitness and is below the variety neutrality, whereas waste DNA decreases it. Functional categories assigned to DNA regions may change over time. K-mers, substrings of length k, are applied in bioinformatics for sequence analysis and computational genomics. They are made up of nucleotides (A, T, C, and G) and are used to assemble DNA sequences (Benson et al., 2020; Solis-Reyes, 2009; Shadab et al., 2020). We apply several classifiers, among them are the Naive Bayes, Random Forest, K-Nearest Neighbour, and Decision Tree classifiers. The research considers three datasets: human, dog, and chimpanzee, based on protein coding sequences. The datasets are processed using k-mer

Authors information:
[a]Department of Computer Engineering, GES's R. H. Sapat College of Engineering, Management Studies and Research, Nashik, INDIA. E-mail: borsevinodkumar@gmail.com[1]; archana.vaidya@ges-coengg.org[2]
[b]Department of Computer Science and Engineering (Data Science), R. C. Patel Institute of Technology, Shirpur, INDIA. E-mail: manishavpatil2007@gmail.com[3]
*Corresponding Author: borsevinodkumar@gmail.com

Received: March, 2025
Accepted: July, 2025
Published: June, 2026

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

encoding and machine learning classifiers, and classified using scikit-learn to determine gene similarity and predict potential relationships. The dataset has two columns: sequence and class, where class is a pre-determined integer value depending on the protein coding sequence (Solis-Reyes et al., 2009). The research paper seeks to investigate the efficacy of machine learning algorithms in DNA classification of protein sequence data and add to the existing literature by comparing results with prior studies, identifying strengths and weaknesses of various algorithms, and proposing areas for improvement (Shadab et al., 2020).

2. Literature Review

Researchers identified new things to implement after evaluating significance, seeking speedy and precise completion. Past scholars contributed a lot to this research area. (Onesime et al., 2021) introduces a new approach to predicting genomic islands (GIs) in bacterial genomes based on seven sequence features extracted with the scikit-learn toolkit. The approach employs a chi-square test to identify significant features, which are subsequently input into a random forest algorithm. The method is appropriate to current approaches in standards of accuracy, precision, recall, and other measures. The article (Alotaibi et al., 2021) discusses the application of six machine learning algorithms (K-nn, RF, SVM, LR, SGD, and GNB) to predict DNA mixture contributors. The provided dataset includes DNA mixtures with a maximum of five contributors. Accuracy, F1-score, recall, and precision have been utilized to measure the metric of each algorithm. Logistic regression (LR) has the best accuracy (95%) for five-contributor mixtures. (Arowolo et al., 2021) Discusses a method for RNA-Seq data classification using genetic algorithm feature selection, decision tree, and K-nearest neighbors (KNN) classifiers. The method is applied to a set of Anopheles gambiae mosquitoes, a malaria vector. The feature selection using GA greatly enhances the performance of the two classifiers, with the decision tree having higher accuracy (98.3%) compared to KNN (88.3%). The paper also discusses literature related to machine learning approaches in the classification of gene expression. (Mathur et al., 2023) Research suggests a DNA sequence classification system for the identification of early diseases, based on NCBI database samples and convolutional neural networks for feature extraction and classification. (Arowolo et al., 2021) Research assesses the application of a genetic algorithm (GA) for feature selection in RNA sequencing data of the Anopheles gambiae malaria vector dataset with 88.3% classification accuracy for KNN and 98.3% for decision tree, showing its viability in bioinformatics. The article (Hamed et al., 2023) explains the application of machine learning models such as random forest, KNN, naïve bayes, decision tree, and SVM in classifying DNA sequences through pattern matching.

The approach (Alshayeji et al., 2023) applies natural language processing and machine learning to automatically detect viruses from human biospecimens at 98.6% classification accuracy, precision, recall, and detection rate. The research employs natural language processing and machine learning to automatically detect viruses in human biospecimens with 98.6% accuracy of classification, precision, recall, and detection rate. The study (Li, F. et al., 2023) introduces an EpiTEAmDNA architecture, a combination of transfer learning and ensemble learning methods, for representing DNA methylation type features in 15 species, outperforming current methods. The study (Yadav, V. et al., 2025) uses genome sequencing to study DNA structure and species similarity between humans, chimpanzees, and dogs, revealing 99.30% and 98.40% similarity, suggesting further research could deepen understanding of life sciences.

3. Methods

The study has two phases of research: preprocessing and postprocessing. The data preprocessing is dealt with in the preprocessing phase, while model learning and framework testing are covered under the article phase. Machine learning and NLP are applied in this research work to process the texting DNA sequence pattern into a series and to evaluate the various ML algorithms.

Data Set

The data file “human_data.txt” was downloaded from the Kaggle database and is used as a public dataset. The data includes two features: DNA sequencing and class. The data size is (4380, 2), consisting of 4380 samples and two features. It contains six gene family classes, including occurrences per class and numeric values. The data is converted from string to numeric classes. In order to demonstrate the occurrences for each class, a count plot was created. Transcription factor (class 6) contained the most data, with 1343 samples, and the Lon channel contained the least class, with 240 samples (Bhushan Bawankar et al., 2024).

Feature Extraction

DNA sequences are written as a set of shorter sub-sequences of length k, known as k-mers. These sub-sequences are similar to words in human language processing (NLP) and help understand DNA sequences and simplify analysis computation. In our model, k-mer encoding generates hexamer “words” with word length adjustable for various situations. In genomics, manipulations like "k-mer counting" are used. Python's natural language processing packages make this process easy. From each pattern string, a function (get_K-mers) is provided to gather all potential overlaid k-mers of a certain length. The data sequences are transformed into repeating k-mers of length six, and repeated operations are performed for every species DNA sequence in the dataset (Hamed et al., 2023).

DNA Classification

This section provides a classification model applied to human, chimpanzee, and dog DNA datasets. We utilize several classifiers; among them are the naive bayes, decision tree classifier, w-naive bayes classifier, random forest, and K-NN. Also, discuss various performance parameters utilized for the evaluation of classifier performance.

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

Naive Bayes Classifier

The bayes principle is applied in the naïve bayes technique, a machine learning statistical methodology, to solve classification issues. It uses an independent variable to establish the range of parameters and requires minimal training data (Peretz et al., 2024). Eq. (1), which employs the Gauss distribution as the decision function, is the most often used naïve bayes classifier (Zuhanda et al., 2025). Eq. (1) utilizes the most probable principle to provide the estimated mean and standard deviation.

$$P(v/y) = \frac{1}{\sqrt{2\pi\sigma^2}}. e^{(v-\mu)^2/2\sigma^2} \quad (1)$$

In order to classify data, naïve bayes considers a Gaussian distribution and computes the mean, variance, and prior probability for each class in training data (Wickramasinghe et al., 2021).

Weighted Naive Bayes Classifier

The NB algorithm assumes condition attributes are independent, which reduces the computational cost. However, practical application varies, and hence, classification accuracy is reduced when the default ownership weight is 1(Xia et al., 2021). In this study, we employ the weighted naïve bayes (WNB) algorithm to assign appropriate attributes based on their classification contribution, preserving computation speed and reducing the impact of attribute conditional independence on classifier performance (Ye, 2021; Wickramasinghe et al., 2021). The weighted NB is calculated using the following Eq. (2).

$$p(C_j/X) = \arg\max p(C_j) \prod_{i=1}^{n} p(A_i / C_j)^{wi} \quad (2)$$

Random Forest Classifier

RF is a technique for classification, an ensemble technique where a number of decision trees are parallel trained with bootstrapping and aggregation of the base learners through bagging. Thus, each decision tree is unequal because of the property of high variance of the decision tree. Aggregation of predictions of the final decision comes from individual base classifiers. The goodness of generalization of the classifiers is also much better and gives higher accuracy in comparison to the random forest, which overcomes overfitting (Onesime et al., 2021; Pazhanikumar et al., 2024). RF classifiers show more robust results in noisy data and are simpler to tune hyperparameters compared to DT classifiers. They give fast computation for the internal measure of variable importance, VIMP, which comes in handy with ranking variables, especially in large multi-dimensional genomic imbalance data sets (Rrmoku et al., 2022).

Decision Tree Classifier

Machine learning models can be classified into parametric and nonparametric algorithms. A decision tree model is a nonparametric model based on supervised learning, where the tree-like structure is used to represent the probability of an event based on training data. The model is arranged in a top-down tree structure, with nodes representing features and branches representing possible values (Benson et al., 2009). This process of building DTs from the training data is called DT induction, which can be represented using a conditional control statement for easy decision-making (Peretz et al., 2024)

K-Nearest Neighbour Classifier

The k-NN algorithm is a memory-based method that stores objects during training, requiring less computational effort than deep learning methods. It constructs an approximation of the objective function, which is different for each new data to be stored, which is advantageous when the objective function is complex (Corso et al., 2021). As a progressive method that stores new data only when it becomes accessible, K-NN may be employed to address challenging issues. Still, it cannot express itself precisely, a concise expression of the objects, and depends on the determination of the distance of any object from all training objects. The redundancy in attributes, along with redundant data, weakens the proposed model. This weakness can be compensated for by checking the reduction in the problem space or testing the variation of distance calculations (Wang et al., 2023).

4. Results and Discussion

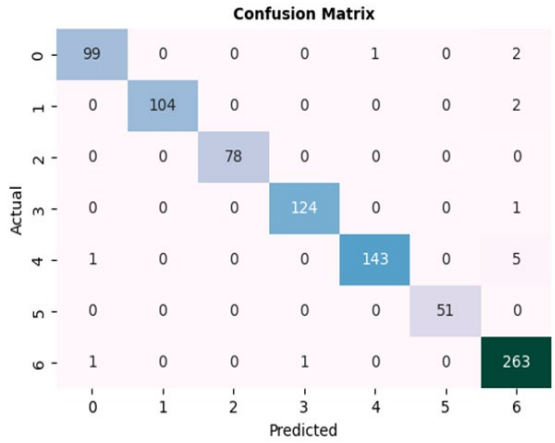
In this stage, we provide an analysis of the classification model performance by our results with human, chimpanzee, and dog DNA datasets.

Table 1. Performance analysis of various classifiers based on human DNA data.

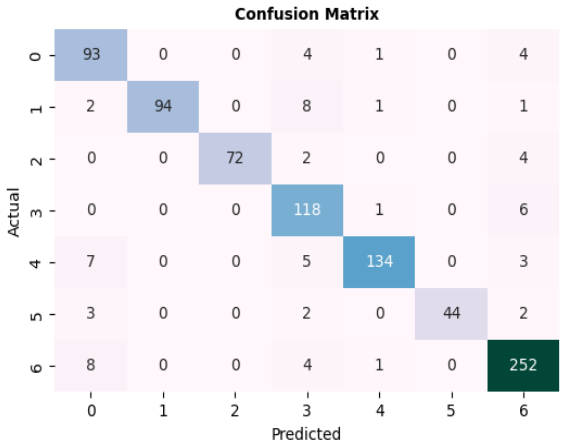
Classifier	Accuracy	Precision	Recall	F1 Score
Naive Bayes	0.984	0.984	0.984	0.984
Random Forest	0.918	0.926	0.918	0.919
K-Nearest Neighbour	0.824	0.874	0.824	0.827
Decision Tree	0.807	0.827	0.807	0.813

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

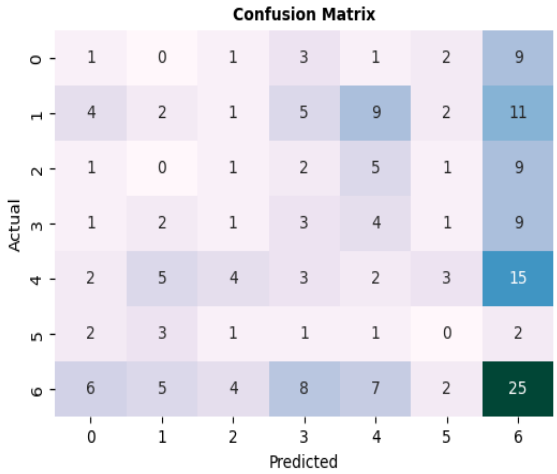
Table 2. Confusion matrix of various classifier based on Human DNA data.



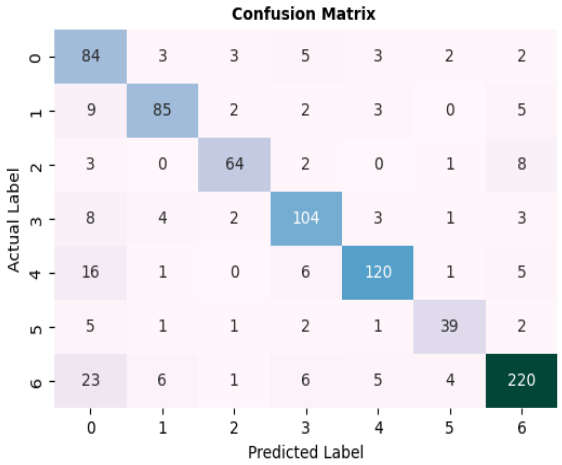
1. Naive bayes classifier



2. Random forest classifier



3. K-nearest neighbour classifier



4. Decision tree classifier

We utilize several classifiers, among them are the naive bayes, decision tree classifiers, k-nearest neighbour, and random forest. As a part of the standard parameters, the test performance for classifiers was also executed in terms of accuracy, precision, recall, and F1 score. We compute the receiver operating characteristic and the Area under the curve values in each classifier. Table 1 shows the actual performance of a classifier, and Table 2 shows the confusion matrix of all classifiers based on human DNA data.

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

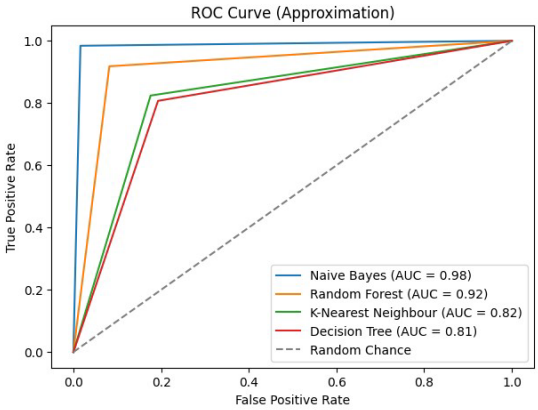


Figure 1. ROC and AUC of different classifier based on human DNA data.

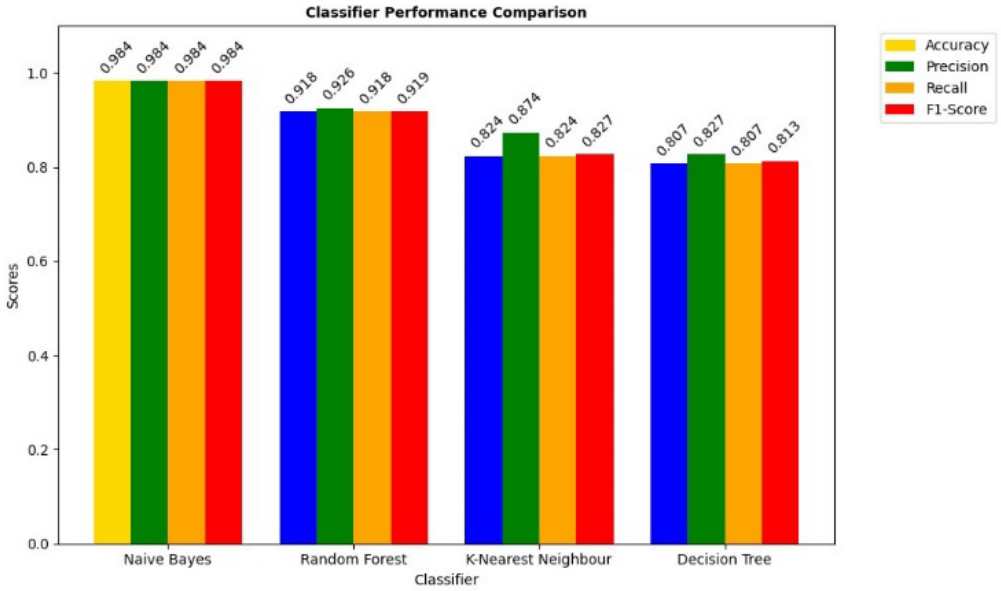


Figure 2. Comparison of various classifier metrics of human DNA data.

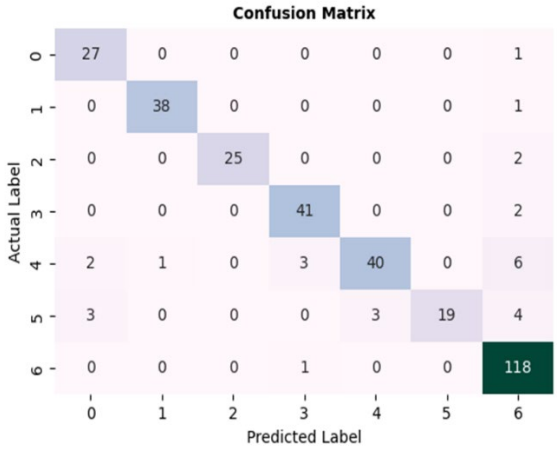
The naive bayes classifier performed better with an accuracy, precision, recall, and F1-score of 98.4 %. Random forest classifier performed fairly well with an accuracy of 91.8%. K-nearest neighbour and decision tree classifiers performed poorly, with an accuracy of 82.4% and 80.7 % respectively, as shown in Table 1.

Table 3. Performance analysis of various classifiers based on chimpanzee DNA data.

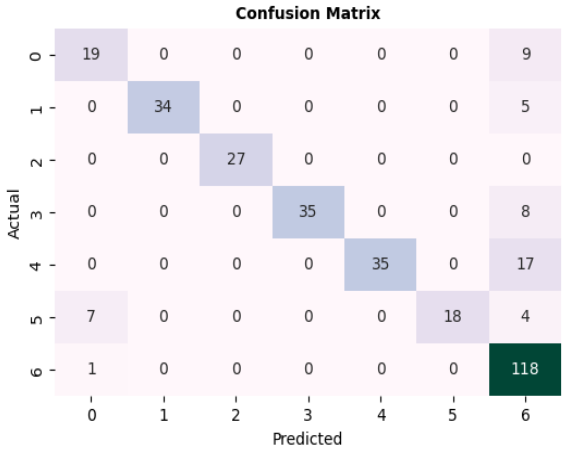
Classifier	Accuracy	Precision	Recall	F1 Score
Naive Bayes	0.914	0.920	0.914	0.911
Random Forest	0.843	0.874	0.843	0.841
Decision Tree	0.769	0.767	0.769	0.766
K-Nearest Neighbour	0.682	0.852	0.682	0.705

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

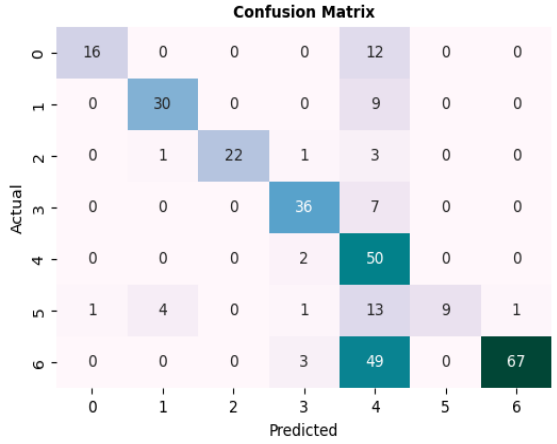
Table 4. Confusion matrix of various classifiers based on chimpanzee DNA data.



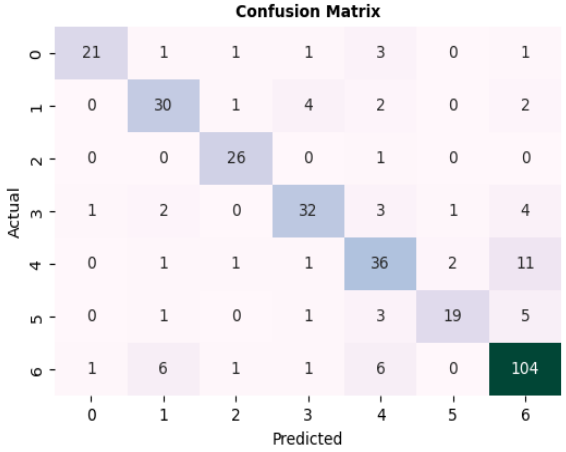
1. Naive bayes classifier



2. Random forest classifier



3. K-nearest neighbour classifier



4. Decision tree classifier

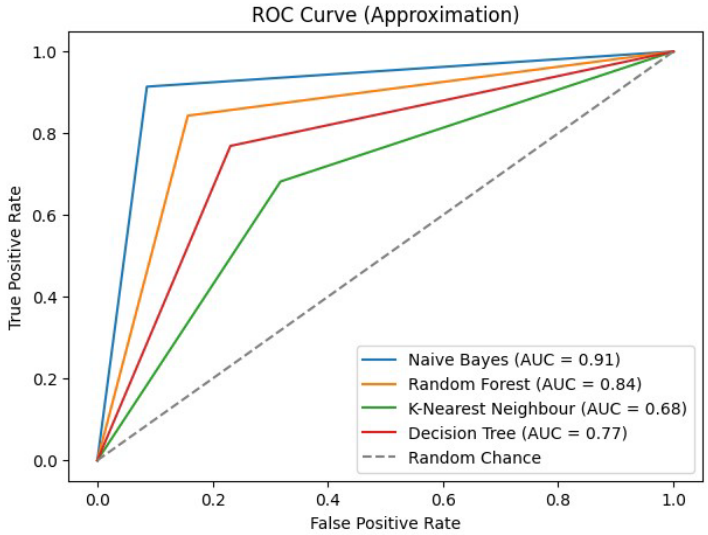


Figure 3. ROC and AUC of different classifiers based on chimpanzee DNA data.

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

These results indicate that, because of its ability to classify high-dimensional data and its underlying assumptions that align with the data's structure, the naive bayes classifier is the most appropriate for classifying human DNA data. Figures 1 and 2 are the graphical representations of ROC curves and comparison of different classifier metrics, respectively shown in Table 1.

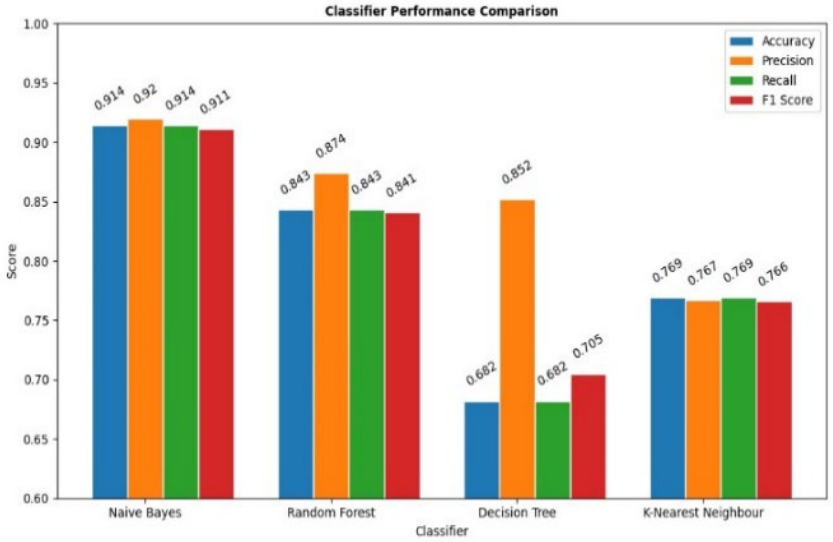


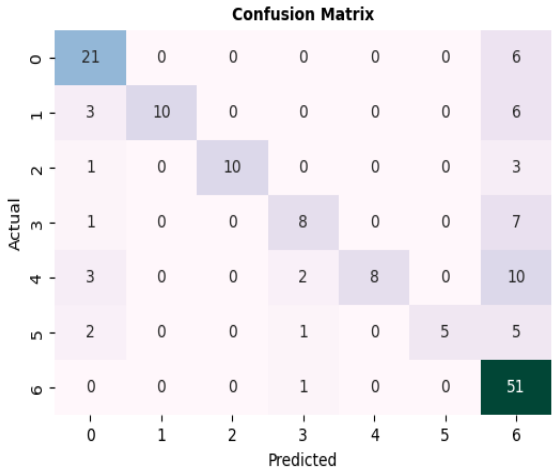
Figure 4. Comparison of various classifier metrics of chimpanzee DNA data.

Table 5. Performance analysis of various classifiers based on dog DNA Data.

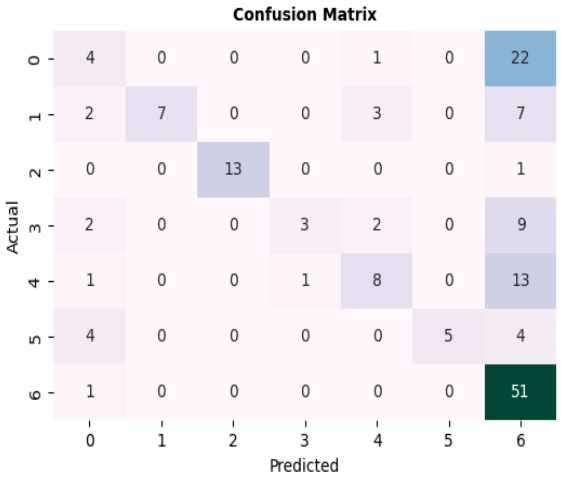
Classifier	Accuracy	Precision	Recall	F1 Score
Naive Bayes	0.689	0.781	0.689	0.673
Decision Tree	0.506	0.500	0.506	0.496
Random Forest	0.573	0.670	0.571	0.541
K-Nearest Neighbour	0.323	0.406	0.323	0.22

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

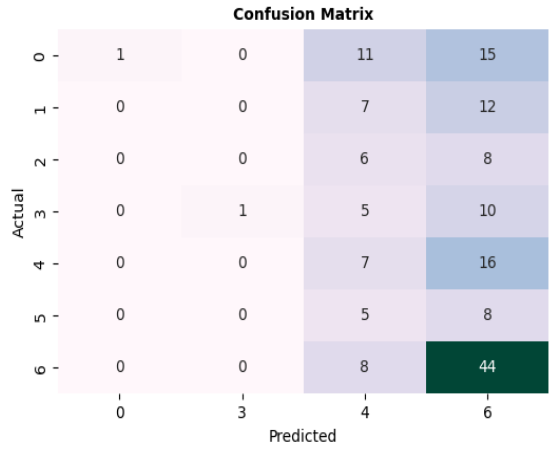
Table 6. Confusion matrix of various classifiers based on dog DNA Data.



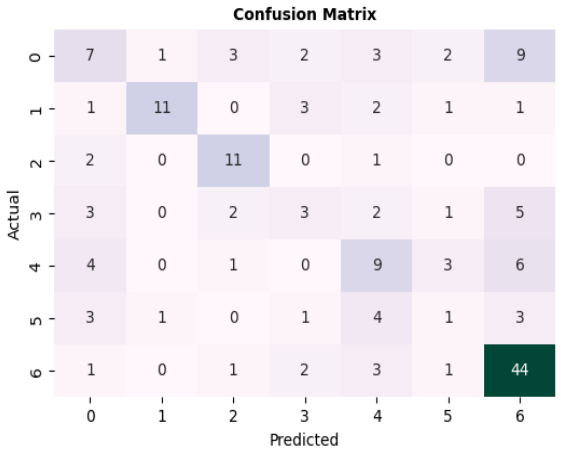
1. Naive bayes classifier



2. Random forest classifier



3. K-nearest neighbour classifier



4. Decision tree classifier

The performance of the classifiers on chimpanzee DNA data was generally lower compared to human DNA data. The Naive Bayes classifier was once more the best, at 91.4% as illustrated in Table 3. Random forest at 84.3%, the decision tree at 76.9%, and K-nearest neighbour at 68.2% had lower accuracies as indicated in Table 3. It could be the differences in the genetic structure and complexity between human and chimpanzee DNA that result in reduced performance in comparison with human data. Figure 3 and Figure 4 illustrate the ROC curves and the comparison of the classifier metrics for the chimpanzee DNA data shown in Table 3. Table 4 shows the confusion matrix of all classifiers.

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

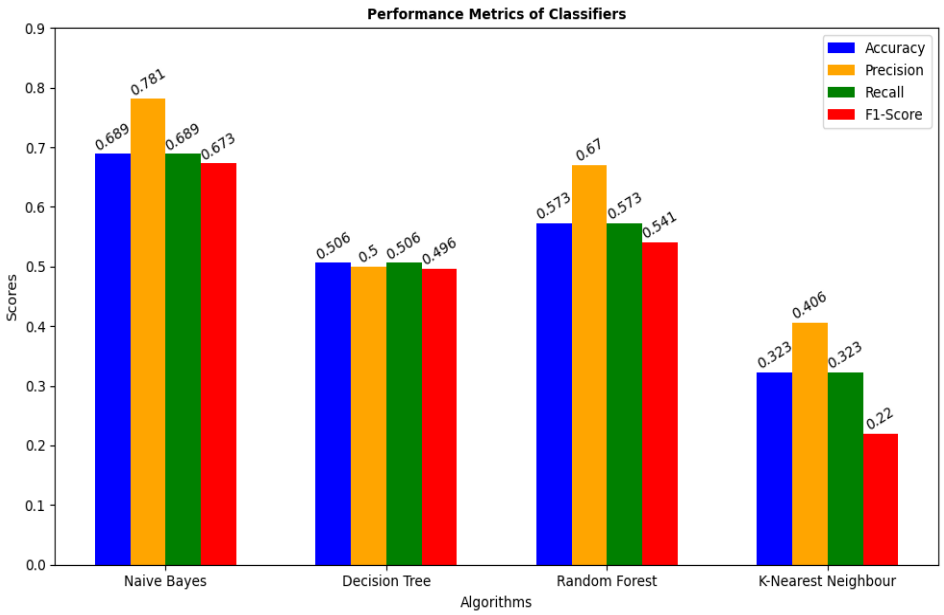


Figure 5. Comparison of various classifier metrics of dog DNA data.

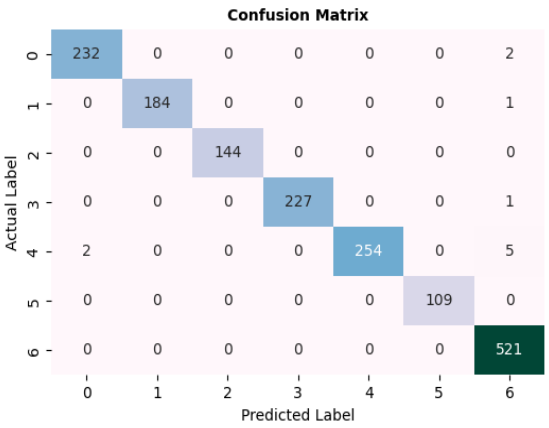
The classifiers performed substantially lower in dog DNA data compared to both human and chimpanzee data. The highest accuracy was observed for the naive bayes classifier at 68.9%, followed by the random forest classifier at 57.3% as shown in Table 5. The Decision Tree and K-Nearest Neighbour classifiers had poor accuracies of 50.6% and 32.3%, respectively, as shown in Table 5. This can be due to major differences in genetic structure and evolutionary distance between dogs and humans. Figures 5 and 6 compare the classifier metrics and ROC curves for dog DNA data shown in Table 5, and Table 6 shows the performance comparison and confusion matrix of all classifiers.

We used the trained weighted naive bayes human model from the human DNA data to predict chimpanzee and dog DNA types in order to further investigate the generalizability of our models. The model obtained high accuracy in predicting chimpanzee DNA at 99.3% and dog DNA at 92.6% Table 7. High accuracy in predicting chimpanzee DNA may indicate some similarity in genetic features between humans and chimpanzees that the model can effectively capture. Lower accuracy in predicting dog DNA is consistent with previous findings indicating greater genetic divergence between dogs and humans. Performance comparison of the Weighted Naive Bayes model for the three species shown in Figure 7. Table 7 and Table 8 show the Prediction metrics and confusion matrix of all classifiers

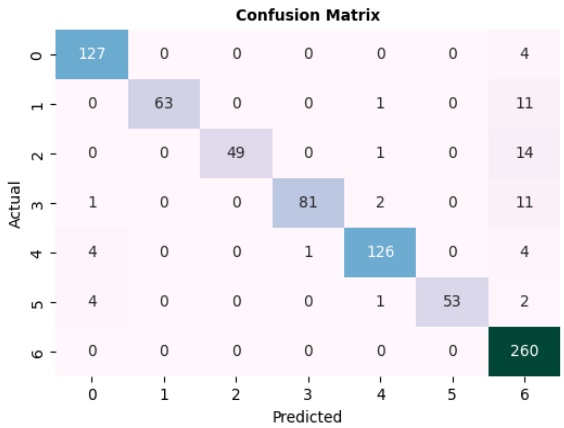
Table 7. Prediction metrics using a weighted naive bayes classifier (human model).

DNA Type	Accuracy	Precision	Recall	F1 Score
Chimpanzee	0.993	0.994	0.993	0.993
Dog	0.926	0.934	0.926	0.925
Human	0.996	0.996	0.996	0.996

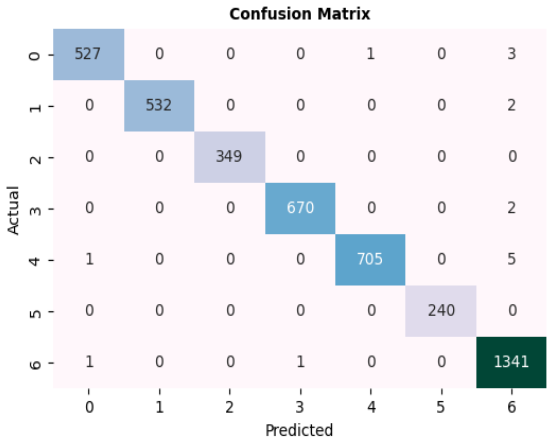
DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)



1. Chimpanzee DNA



2. Dog DNA



3. Human DNA

Table 8. Confusion matrix of weighted naive bayes classifier (human model).

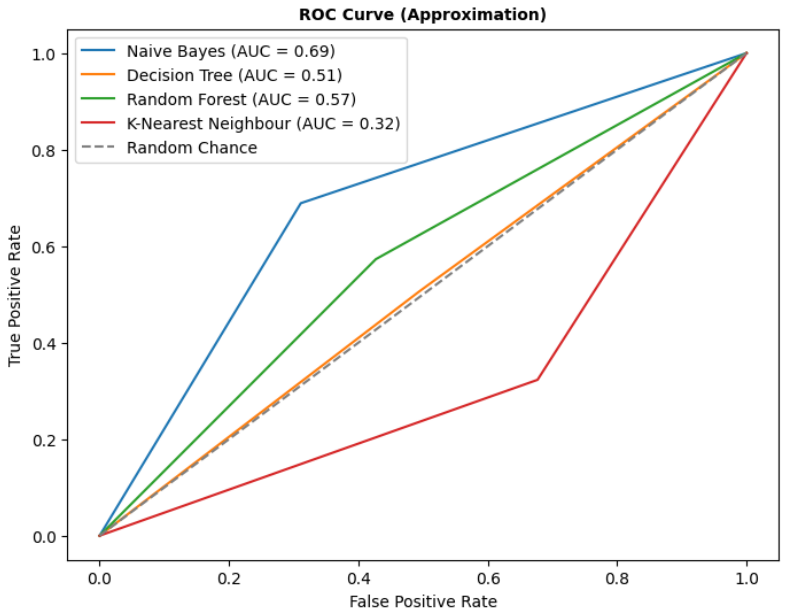


Figure 6. ROC and AUC of different classifiers based on dog DNA data.

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

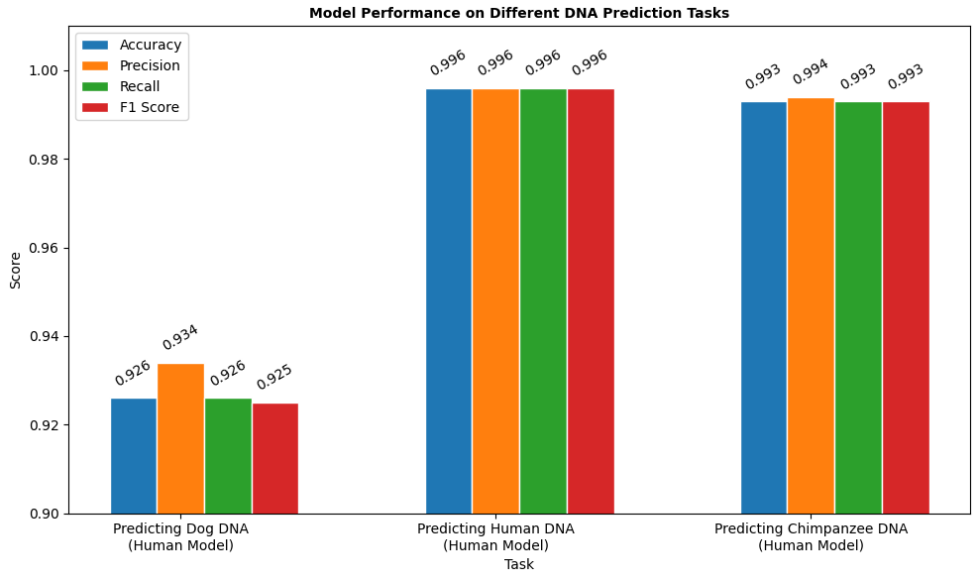


Figure 7. Performance comparison of human, chimpanzee, and dog DNA data using weighted naive byes (human model).

4. Conclusion

In conclusion, our results show that the weighted naive bayes classifier outperforms other classifiers for all three datasets. However, the performance is significantly reduced with an increase in evolutionary distance from human DNA. This shows the need to pay special attention to the characteristics of the DNA data while choosing and training classification models. Future research may include more advanced feature engineering techniques and other classification algorithms that may improve performance on more distantly related species.

7. References

Alotaibi, H., Alsolami, F., & Mehmood, R. (2021). DNA profiling: An investigation of six machine learning algorithms for estimating the number of contributors in DNA mixtures. *International Journal of Advanced Computer Science and Applications, 12*(11).

Alshayeji, M. H., Sindhu, S. C., & Abed, S. (2023). Viral genome prediction from raw human DNA sequence samples by combining natural language processing and machine learning techniques. *Expert Systems with Applications*, 218, 119641. https://doi.org/10.1016/j.eswa.2023.119641.

Arowolo, M. O., Adebiyi, M., Adebiyi, A. A., & OKesola, J. O. (2021). Predicting RNA-Seq data using genetic algorithm and ensemble classification algorithms. Indonesian Journal of *Electrical Engineering and Computer Science*, 21(2), 1073-1081.

Arowolo, M. O., Adebiyi, M. O., & Adebiyi, A. A. (2021). A genetic algorithm approach for predicting ribonucleic acid sequencing data classification using KNN and decision tree. *TELKOMNIKA, Telecommunication Computing Electronics and Control*, 19(1), 310-316.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2009). GenBank. *Nucleic Acids Research, 37*(suppl_1), D26-D31.

Beskopylny, A. N., Stel'makh, S. A., Shcherban', E. M., Mailyan, L. R., Meskhi, B., Razveeva, I., & Beskopylny, N. (2022). Concrete strength prediction using machine learning methods CatBoost, k-nearest neighbors, support vector regression. *Applied Sciences, 12*(21), 10864.

Bhushan Bawankar. (2024). Analysis of machine learning approaches for dna sequencing and classification: An optimized approach. *Communications on Applied Nonlinear Analysis, 31*(2s), 436–453. https://doi.org/10.52783/cana.v31.659.

Coscia, A., Dentamaro, V., Galantucci, S., Maci, A., & Pirlo, G. (2024). Automatic decision tree-based NIDPS ruleset generation for DoS/DDoS attacks. *Journal of Information Security and Applications, 82*, 103736.

Corso, M. P., Perez, F. L., Stefenon, S. F., Yow, K. C., García Ovejero, R., & Leithardt, V. R. Q. (2021). Classification of contaminated insulators using k-nearest neighbors based on computer vision. *Computers*, 10(9), 112.

Gao, L., Li, D., Liu, X., & Liu, G. (2022). Enhanced chiller faults detection and isolation method based on independent component analysis and k-nearest neighbors classifier. *Building and Environment*, 216, 109010.

Hamed, B. A., Ibrahim, O. A. S., & Abd El-Hafeez, T. (2023). Optimizing classification efficiency with machine learning techniques for pattern matching. *Journal of Big Data*, 10(1), 124.

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, 77(5), 5198-5219.

Li, F., Liu, S., Li, K., Zhang, Y., Duan, M., Yao, Z., Zhu, G., Guo, Y., Wang, Y., Huang, L., & Zhou, F. (2023). EpiTEAmDNA: Sequence feature representation via transfer learning and ensemble learning for identifying multiple DNA epigenetic modification types across species. *Computers in Biology and Medicine*, 160, 107030. https://doi.org/10.1016/j.compbiomed.2023.107030

Mathur, G., Pandey, A., & Goyal, S. (2023). A comprehensive tool for rapid and accurate prediction of disease using DNA sequence classifier. *Journal of Ambient Intelligence and Humanized Computing*, 14(10), 13869-13885.

Momenzadeh, M., Sehhati, M., & Rabbani, H. (2020). Using hidden Markov model to predict recurrence of breast cancer based on sequential patterns in gene expression profiles. *Journal of Biomedical Informatics*, 111, 103570.

Nayak, J., Mishra, M., Naik, B., Swapnarekha, H., Cengiz, K., & Shanmuganathan, V. (2022). An impact study of COVID-19 on six different industries: Automobile, energy and power, agriculture, education, travel and tourism and consumer electronics. *Expert Systems*, 39(3), e12677.

Onesime, M., Yang, Z., & Dai, Q. (2021). Genomic Island Prediction via Chi-Square Test and Random Forest Algorithm. *Computational and Mathematical Methods in Medicine*, 2021(1), 9969751.

Pazhanikumar, K., & KuzhalVoiMozhi, S. N. (2024). Remote sensing image classification using modified random forest with empirical loss function through crowd-sourced data. *Multimedia Tools and Applications*, 83(18), 53899-53921.

Peretz, O., Koren, M., & Koren, O. (2024). Naive Bayes classifier–An ensemble procedure for recall and precision enrichment. *Engineering Applications of Artificial Intelligence*, 136, 108972.

Rrmoku, K., Selimi, B., & Ahmedi, L. (2022). Application of trust in recommender systems—utilizing naive Bayes classifier. *Computation*, 10(1), 6.

Shadab, S., Khan, M. T. A., Neezi, N. A., Adilina, S., & Shatabda, S. (2020). DeepDBP: deep neural networks for identification of DNA-binding proteins. *Informatics in Medicine Unlocked*, 19, 100318.

Solis-Reyes, S., Avino, M., Poon, A., & Kari, L. (2018). An open-source k-mer based machine learning tool for fast and accurate subtyping of HIV-1 genomes. *PloS One*, 13(11), e0206409.

Wang, A. X., Chukova, S. S., & Nguyen, B. P. (2023). Ensemble k-nearest neighbors based on centroid displacement. *Information Sciences*, 629, 313-323.

Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing*, 25(3), 2277-2293.

Xia, X., & Yan, J. (2021). Construction of music teaching evaluation model based on weighted naïve bayes. *Scientific Programming*, 2021(1), 7196197.

Yadav, V., Kumar, V., Gupta, S., & Kaushik, V. D. (Eds.). (2025). Computational intelligence and its applications. *BENTHAM SCIENCE PUBLISHERS*. https://doi.org/10.2174/9789815313321125010101.

Ye, Y. (2024, May). Design and Implementation of an English Mobile Learning System Based on Weighted Naive Bayes. In 2024 *5th International Conference on Big Data and Informatization Education (ICBDIE 2024)* (pp. 187-196). Atlantis Press.

Zuhanda, M. K., Permata, L., & Ongko, E. (2025). Impact of Adaptive Synthetic on Naïve Bayes Accuracy in Imbalanced Anemia Detection Datasets. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi),* 9(1), 85-93.

DOI:https//doi.org/10.22452/mjs.vol45no2.1
Malaysian Journal of Science 45(2): 1-12 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 45(2): 13-22 (June 2026) https://mjs.um.edu.my

Physicochemical Properties of Potassium Carbonate-Ethylene Glycol Mixtures at Different Molar Ratios and Temperatures

Syarifah Nursyimi Azlina Syed Ismail[1a]*, Mohd Azlan Mohd Ishak[2a], Nur Nasulhah Kasim[3a], Azil Bahari Alias[4b] and Razi Ahmad[5c], Khudzir Ismail[6a]

Abstract: Deep Eutectic Solvents (DESs) are highly effective and environmentally benign for various industrial applications, including extraction, catalysis, and synthesis. The success of DESs relies on the types of hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD), as well as the molar ratios and temperatures used during the process. The study investigated the solubility of PC-EG mixture at various molar ratios to determine suitable molar ratios for achieving a eutectic mixture. Then, the physicochemical properties of the PC-EG mixtures, such as phase behavior, H-bonding, density, viscosity, conductivity and thermal stability across different molar ratios and temperatures, were analyzed. Differential Scanning Calorimetry (DSC) analysis confirmed that all mixtures were DESs, as their freezing points were lower than those of their individual components. The solid-liquid phase diagram showed that DES-8, DES-12, DES-16, and DES-19 remained in the liquid phase at temperatures of 470°C, 350°C, 220°C and 180°C, respectively. Meanwhile, FTIR analyses verified the presence of H-bonding within the DESs, but DES-19 showed minimal interaction between PC and EG. The results indicated that adding more EG content raised the molar ratio but weakened H-bonding. Similarly, increasing the molar ratio and temperature decreased the pH, density, and viscosity of the DESs, while ionic conductivity increased. Thermal analysis indicated the thermal instability of the mixtures at very high temperatures, suggesting their suitability for use at moderate temperatures. The result found that DES-12 and DES-16 had low viscosity, with strong H-bonding and excellent ionic conductivity within 30°C–40°C and 50°C–70°C. The findings of this study provide significant guidance on the effective use of PC-EG at appropriate molar ratios and temperatures to enhance desulfurization, carbon capture, and various industrial applications to achieve optimal results.

Keywords: *Deep Eutectic Solvent, Intermolecular H-bonding, Ionic conductivity, Physicochemical properties, Potassium carbonate-ethylene glycol.*

1. Introduction

Industrial organic solvents present significant hazards due to their high flammability, volatility, and persistence. The emission of volatile organic compounds (VOCs) contributes to atmospheric pollution, global warming, and ozone depletion. DESs are safe and environmentally benign alternatives to traditional solvents. They are formed by mixing an HBA and an HBD that result in a eutectic mixture stabilized by strong hydrogen bonding interactions. These interactions significantly lower the melting point of the mixture compared to its individual components, enabling DESs to remain liquid under ambient conditions. Compared with conventional organic solvents, they exhibit advantageous properties such as low flammability, low volatility, thermal stability, biodegradability, low toxicity, cost-effectiveness, and ease of preparation, making them a preferable choice (Ali et al., 2026).

Authors information:
[a]Faculty of Applied Sciences, Universiti Teknologi Mara, Arau Campus, 02600 Perlis, MALAYSIA. Email: syarifah_nursyimi@uitm.edu.my[1]; azlanishak@uitm.edu.my[2]; nurnasulhah@uitm.edu.my[3]; khudzir@uitm.edu.my[6]
[b]Faculty of Chemical Engineering, Universiti Teknologi Mara, Shah Alam Campus, 40450 Selangor, MALAYSIA. Email: azilbahari@uitm.edu.my[4]
[c]Faculty of Civil Engineering Technology, Universiti Malaysia Perlis (UniMAP), 02600 Arau, Perlis, MALAYSIA. Email: razi@unimap.edu.my[5]
*Corresponding Author: syarifah_nursyimi@uitm.edu.my

The concept of DESs has been extended to low-transition-temperature mixtures (LTTMs), which are eutectic solvents with low transition temperatures (Wang & Baker, 2018). Their preparation is similar to DESs, but the choice of components and molar ratios determines whether the result is a DES or an LTTM (Francisco et al., 2013). Both possess low transition temperatures, wide liquid ranges, low vapor pressures, and excellent thermal stability. However, DESs show a freezing temperature peak in DSC, while LTTMs only display a glass transition temperature without a crystallization peak. Glass transitions behavior can occur even at low cooling rates, making it significant for characterizing DESs (Ghaedi et al., 2019).

DESs are effective for removing carbon dioxide (CO_2) and sulfur dioxide (SO_2). Their gas absorption depends on the types of HBAs and HBDs, their molar ratios, and temperature (Manafpour et al., 2024). Therefore, selecting suitable HBAs and HBDs with proper molar ratios is important for optimal performance at specific temperatures.

Matjie et al. (2018) showed that potassium carbonate (PC) effectively reduces SO_2 emissions during coal pyrolysis due to its high thermal stability, maintaining performance at temperatures of up to 950°C. The carbonate ion (CO_3^{2-}), with its two negatively charged oxygen atoms, exhibits strong basicity and high electron density (Zhou et al., 2024), allowing it to accept H-bonds from

Received: April, 2024
Accepted: Jan, 2025
Published: June, 2026

DOI:https//doi.org/10.22452/mjs.vol45no2.2
Malaysian Journal of Science 45(2): 13-22 (June 2026)

HBDs like ethylene glycol (EG) and promote stable DES formation. EG has two OH groups at both ends, which promote strong hydrogen bonding, making it an effective HBD and an innovative eutectic solvent. Rogošić and Kučan (2019) found that increasing the percentage of EG improved thermal conductivity and temperature diffusivity while lowering heat capacity.

The choice of PC–EG DES for this study is supported by previous reports on its effectiveness in CO_2 absorption at molar ratios of 1:6 and 1:10 (Ghaedi et al., 2022;Saeed et al., 2022). The liquid solvent aids in selective component extraction, while a solid salt increases separation through enhanced volatility differences (Naser J. et al., 2013). The PC–EG eutectic mixture has a basic pH, indicating strong alkalinity, which makes alkaline DESs effective in absorbing acidic gases (Mjalli et al., 2014). For example, the choline chloride (ChCl)–urea mixture's moderate alkalinity enables it to capture small amounts of CO_2 (Zhang et al., 2021). In comparison, the higher alkalinity of the PC–EG is expected to further enhance its interaction with CO_2 and improve its removal efficiency (Ghaedi et al., 2022). Available data on PC-EG behavior near the eutectic composition remain limited, highlighting the need for a more comprehensive analysis across different molar ratios. The HBA-to-HBD molar ratio significantly affects the physicochemical properties and extraction performance of DESs. Increasing the HBD molar ratio can enhance SO_2 solubility by providing more active hydrogen sites, thereby facilitating stronger interactions with sulfur-containing species (Zou et al., 2026).

The molar ratio of DES components significantly affects its viscosity, thereby influencing practical applications through mass transfer rates. High-viscosity DESs decrease extraction efficiency by hindering molecular diffusion. Low-viscosity DESs, while exhibiting poor solubility, can enhance extraction efficiency due to crucial hydrogen bonding interactions with target compounds (Şahin et al., 2026). Thus, a balanced optimization of solvent properties is essential. Experimental studies show that ChCl-EG with higher HBD ratios (1:2, 1:3, and 1:3.5) enhances desulfurization efficiency. The ChCl-EG systems achieved a 70.9% extraction efficiency at a 1:3.5 molar ratio in one cycle and reached 99% efficiency after four cycles (Rogošić & Kučan, 2019). These findings highlight the importance of optimizing molar ratios in designing effective DES systems, as they significantly influence physicochemical properties like phase behavior, hydrogen-bonding interactions, and thermal stability (Mamtani et al., 2023).

Thermal stability is another critical parameter that ensures DES and the target compound maintain their structural integrity and perform reliably during the extraction process. Studies show that optimal extraction occurs between 30°C and 60°C. If the temperature exceeds this range, the extraction efficiency may decrease due to changes in the solvent's structure (Lakhmir et al., 2022).

Our previous study (Ismail et al., 2025) showed that a PC-EG mixture at a 1:16 molar ratio and 30°C optimally extracted sulfur from coal. This research builds on those findings by examining the physicochemical properties of the DESs for broader applications. The solubility of the DESs was evaluated at various molar ratios (1:2, 1:4, 1:6, 1:8, 1:10, 1:12, 1:14, and 1:16) to identify the eutectic composition. Additionally, Response Surface Methodology-Centered Composite Design (RSM-CCD) produced ratios of 1:8, 1:12, 1:16, and 1:19 to analyze their respective properties, including phase behavior, H-bonding, density, viscosity, conductivity, and thermal stability, across different molar ratios and temperatures. Understanding the eutectic mixture and the physicochemical properties of DESs clarifies the interactions between PC and EG, allowing for performance prediction in desulfurization and extraction at optimal molar ratios and temperatures. This research also supports safety and environmental goals of the 2030 Agenda for Sustainable Development by addressing existing gaps in the limited studies on PC-EG mixtures.

2. Materials and Methods

Preparation of PC-EG Mixtures

Potassium carbonate (PC) (R&M Chemicals) and 98% ethylene glycol (EG) (Merck) were mixed at 30°C in the following molar ratios (mol/mol): 1:2, 1:4, 1:6, 1:8, 1:10, 1:12, 1:14, and 1:16 until a clear solution was obtained. Then, the solutions were stored in sealed Schott bottles and kept in a desiccator for 24 hours to reach the eutectic phase (Albayati et al., 2021). The molar ratios of DESs without precipitation were analyzed using Response Surface Methodology-Central Composite Design (RSM-CCD), generating ratios of 1:8, 1:12, 1:16, and 1:19. A significant variance of at least 2 g was observed between the 1:8 and 1:12 ratios. This difference is significant, whereas the variance between the 1:8 and 1:10 ratios was smaller and likely less impactful on physicochemical properties. The 1:19 ratio represents the upper limit set by RSM-CCD.

Differential Scanning Calorimetry (DSC) for Freezing Points Determination

The melting or freezing point of DES was measured using a Netzsch DSC 200 F3 DSC. About 10 mg of the sample was first cooled to −196°C at a rate of 20°C/min, then cooled from 25°C to −60°C and reheated to 25°C at 10°C/min (Majid et al., 2020).

Construction of Solid-Liquid Phase Diagram

A phase diagram was constructed to identify the eutectic point (Majid et al., 2020). The diagram is based on the mole fraction of PC, as defined in Eq. 1. For instance, a molar ratio of 1:8 means 1 mol of PC for every 8 mol of EG.

$$\text{Mole fraction of PC} = \frac{\text{Mole number of PC}}{\text{Total mole number of PC and EG}} \quad (1)$$

DOI:https//doi.org/10.22452/mjs.vol45no2.2
Malaysian Journal of Science 45(2): 13-22 (June 2026)

Determination of pH, Density, Viscosity, and Ionic Conductivity

The pH, density, viscosity, and ionic conductivity of DESs were measured using a Hanna pH meter, an Attension digital density meter, a Brookfield rotational viscometer, and a Mettler Toledo Seven Compact S230 conductivity meter with a 2-platinum-pole conductivity cell, respectively. Measurements were taken at temperatures of 30°C, 40°C, 50°C, 60°C, and 70°C. Due to the instability of the mixtures, two measurements were recorded to calculate an average.

Fourier Transform Infrared (FTIR) Spectroscopy for Functional Group Characterization

Attenuated Total Reflectance-Fourier Transform Infrared (ATR-FTIR) spectroscopy was used to analyze liquid samples of DES. A small amount of DES was placed on the pad, and spectral data were recorded from 4000 to 400 cm^{-1} at a resolution of 2 cm^{-1}, with a total of 64 scans to enhance accuracy (Biernacki et al., 2020).

Thermogravimetric Analysis (TGA) for Assessing Thermal Stability of DESs

A TG analyzer from NETZSCH was used to study the mass loss of DESs. Approximately 10–12 mg of sample was placed in an alumina crucible and heated from room temperature to 400°C at a rate of 20°C/min under nitrogen (Khan & Srivastava, 2022).

3. Results and Discussion

Solubility of PC-EG Mixtures as a DES

After 4 hours of mixing at 30°C–35°C, it was observed that DES-2 and DES-4 remained insoluble, while DES-6, DES-8, DES-10, DES-12, DES-14, and DES-16 were soluble. However, after 24 hours, DES-6 showed white precipitation, indicating poor mixing and instability. This phase transition led to the formation of crystals and precipitates in some DES formulations. It is essential to keep the solvent as a clear liquid at room temperature without solid recrystallization after 24 hours. Although the solubility of the 1:19 ratio was not tested, it is likely soluble, as reported by Zeng et al. (2024). Figure 1 illustrates the various molar ratios of DESs, and Table 1 displays the composition of the studied DESs.

Figure 1. PC-EG mixture at various molar ratios.

Table 1. Composition of DESs

Molar ratio	Abbreviations	Appearance at room temperature
1:2	DES-2	Turbid white liquid
1:4	DES-4	Turbid white liquid
1:6	DES-6	A colorless liquid, but a residue appeared after 24 hours
1:8	DES-8	Colorless liquid
1:10	DES-10	Colorless liquid
1:12	DES-12	Colorless liquid
1:14	DES-14	Colorless liquid
1:16	DES-16	Colorless liquid
1:19	DES-19	Colorless liquid

Phase Behavior

Figure 2 shows the phase transitions of DESs at different molar ratios, highlighting three thermal transitions identified by DSC. The freezing point (T_f) is the highest exothermic peak during cooling (Pinho et al., 2024), observed between −36°C and −38°C for all DESs. The second exothermic peak near −58°C corresponds to the cold crystallization temperature (T_{C-C}), occurring during reheating after rapid cooling. While both T_f and T_{C-C} involve crystallization, T_f occurs during cooling and T_{C-C} during heating (Pinho et al., 2024). The endothermic peak around 17°C indicates the glass transition temperature (T_g). Heating causes the DESs to transition from a rigid state to a fluid state, leading to reduced viscosity and increased molecular mobility (Zeng et al., 2024). These results confirm that the PC–EG mixtures function as DESs due to their significantly lower freezing points compared to pure PC (891°C) and EG (−12.9°C), highlighting strong H-bonding and eutectic behavior (Zeng et al., 2024). Table 2 compares the T_{C-C}, T_f, and T_g temperatures for all DESs.

DOI:https//doi.org/10.22452/mjs.vol45no2.2
Malaysian Journal of Science 45(2): 13-22 (June 2026)

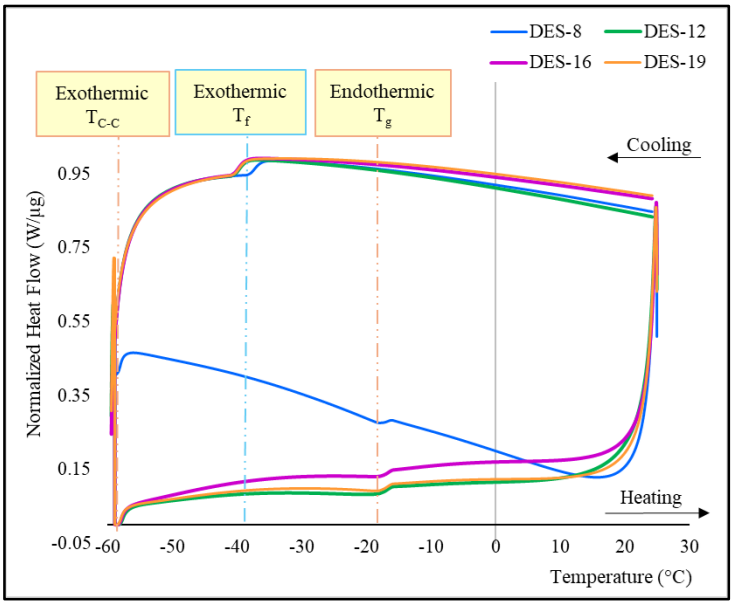


Figure 2. Phase transitions of DESs at molar ratios 1:8, 1:12, 1:16 and 1:19.

Table 2. Cold-Crystallization, Glass-Transition and Freezing Temperatures of PC-EG eutectic mixtures at different molar ratios.

Phase behavior	Thermic type	DES-8	DES-12	DES-16	DES-19
T_{C-C}	Exo	−58.84 ±0.47	−58.26 ±0.11	−58.26 ±0.06	−58.47 ±0.10
T_f	Exo	−35.88 ±1.95	−38.51 ±0.18	−38.26 ±0.30	−38.18 ±0.19
T_g	Endo	−7.52 ±0.37	−8.42 ±0.95	−8.25 ±0.94	−8.14 ±0.67

Construction of Solid-Liquid Phase Diagram

The construction of the solid-liquid phase diagram (Majid et al., 2020) highlights the importance of understanding DES behavior to avoid solid-phase formation during applications. The diagram was created using Eq. (1), and Table 3 displays the molar fractions and mole numbers of PC-EG DESs at different molar ratios.

Table 3. Molar fraction and mole number of PC-EG DESs at various molar ratios.

DES	Mole number of PC	Mole number of EG	The molar fraction of PC
DES-8	1	8	1/ (1 + 8) = 0.11
DES-12	1	12	1/ (1 + 12) = 0.08
DES-16	1	16	1/ (1 + 16) = 0.06
DES-19	1	19	1/ (1 + 19) = 0.05

The red asterisk (*) in the phase diagram indicates the mole fraction of PC at 1:8, 1:12, 1:16, and 1:19 molar ratios. The pure freezing points of PC (891°C) and EG (−12.9°C) are marked with blue asterisks (*). Initial eutectic point is detected at a PC molar fraction of 0.11 (molar ratio 1:8) with a freezing point of around −35°C, similar to Figure 2. DES-12, DES-16, and DES-19 show slight freezing point reductions, likely due to decreased intermolecular H-bonding (Şahin et al., 2026). However, no visible precipitation was observed, and the DESs remained liquid (Figure 1). Freezing point depression (ΔT_f) occurs when the PC-EG mixture's freezing point is lower than its components (Majid et al., 2020).

The EG + liquid mixture region exists below EG's freezing point (−12.9°C), where solid EG coexists with liquid EG. The PC + liquid region appears below PC's freezing point (891°C), with solid PC coexisting with liquid. In this area, solid PC forms first due to its higher mole fraction, but some of the mixture remains liquid because of EG's presence. The green line in the phase diagram indicates the temperature range for the desulfurization study, confirming the mixture stayed liquid between 30°C and 80°C. The solid EG and solid PC regions below the dashed blue line represent a binary solid phase (Majid et al., 2020). The solid–liquid phase diagram verifies that the DES mixtures remained liquid at 30°C–78°C.

DOI:https//doi.org/10.22452/mjs.vol45no2.2
Malaysian Journal of Science 45(2): 13-22 (June 2026)

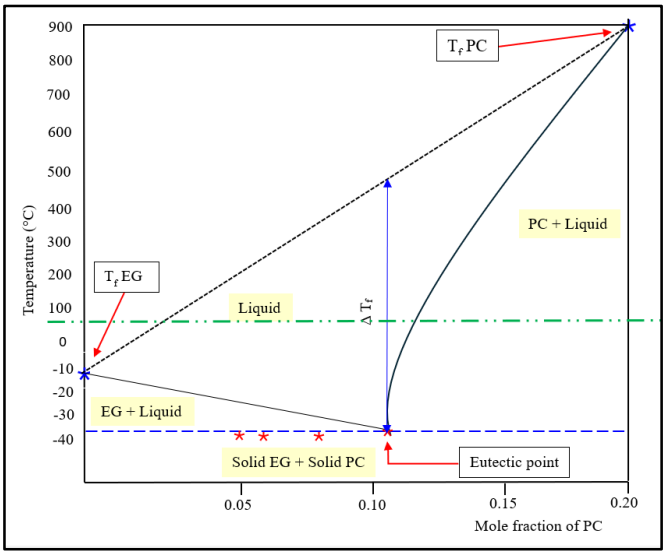
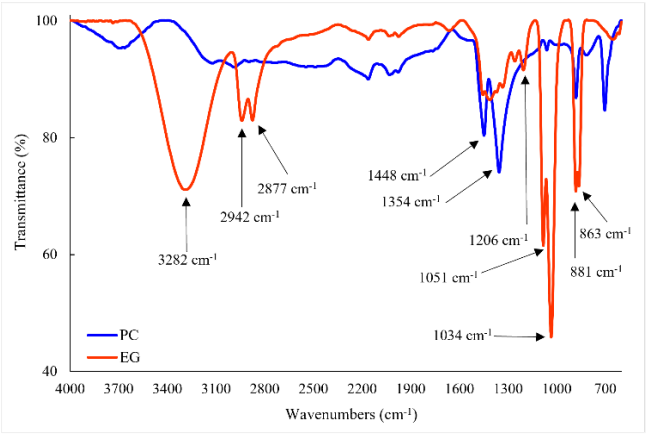


Figure 3. Phase diagram of PC-EG eutectic mixtures at various molar ratios and temperatures.

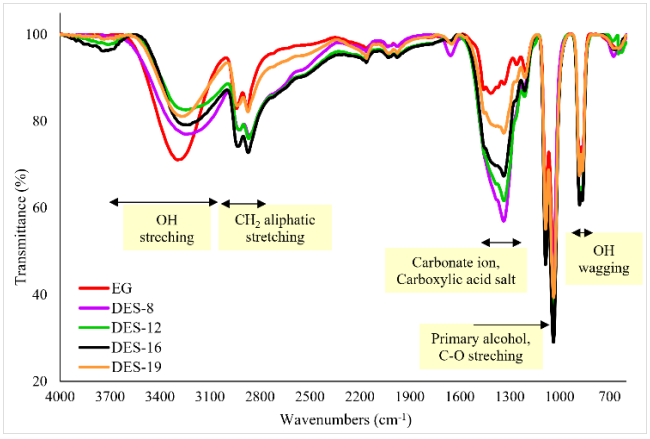
Intermolecular Interaction (H-Bonding) of PC-EG Eutectic Mixtures

Figures 4(a) and 4(b) show the FTIR spectroscopy of PC, EG, and PC-EG mixtures at various molar ratios. The O-H stretching vibrations are observed around 3800–3100 cm^{-1} in all DES samples, including EG, which contains a diol group, consistent with prior studies (Ghaedi et al., 2022). As the EG concentration increased, significant changes in the OH peaks of the DESs were observed, with shifts to lower wavenumbers for DES-8 (3229 cm^{-1}), DES-12 (3234 cm^{-1}), DES-16 (3238 cm^{-1}), and DES-19 (3249 cm^{-1}). The peaks broadened, indicating stronger H-bonding interactions between PC and EG. However, DES-19 showed a slight shift due to the high EG presence, limiting interactions and reducing H-bonding at 19 mol EG. This observation aligns with previous research (Zeng et al., 2024; Saeed et al., 2022). The peaks at 2942–2877 cm^{-1}, 1051–1034 cm^{-1}, and 881–863 cm^{-1} represent C-H stretching of aliphatic hydrocarbons, C-O stretching of primary alcohols, and OH wagging vibrations, respectively. These peaks, found in EG, are present in all DES components. When mixed with PC, the wavenumbers shifted to higher values, indicating increased bond strength between the compounds (Zeng et al., 2024).

In the PC sample, the peak at 1448–1354 cm^{-1} indicates C=O stretching vibrations linked to carboxyl groups (carbonate ions). In the DES mixtures, peaks for carbonate ions and carboxylic acid salts appear at 1434–1336 cm^{-1} (Sultana et al., 2022). The peak shift to higher wavenumbers suggests stronger bonding between the carbonate ion and the OH group of EG (Zeng et al., 2024). Overall, FTIR analysis confirmed the interaction between PC and EG, revealing eutectic mixtures formed through hydrogen bonding, which strengthened with higher PC-EG molar ratios but decreased at an EG ratio of 19 mol.



(a)



(b)

Figure 4. FTIR Spectroscopy of (a) PC and EG components; (b) PC-EG eutectic mixtures at molar ratios 1:8, 1:12, 1:16 and 1:19.

DOI:https//doi.org/10.22452/mjs.vol45no2.2
Malaysian Journal of Science 45(2): 13-22 (June 2026)

pH of PC-EG Eutectic Mixtures

Figure 5 shows the pH values of EG and various DESs at different molar ratios and temperatures. All DESs had a very alkaline pH between 12.5 and 13, while the pH of EG was between 7.5 and 8.0. At room temperature, DES-8 had the highest pH, and DES-19 had the lowest. As the molar ratio of EG increased from 8 to 19 mol, pH values decreased, consistent with Mjalli et al. (2014), who noted that increasing HBDs such as glycerol and EG reduces basicity. EG can donate protons (H^+) via its OH groups, forming bicarbonate ions (HCO_3^-), which are less basic than carbonate ions. This interaction further decreases basicity, particularly in higher molar-ratio DESs like DES-16 and DES-19. Zhang et al. (2021) found that alkaline DESs are highly effective at separating and purifying acidic gases, with moderate alkalinity in the ChCl-urea mixture enabling the absorption of CO_2.

The pH of the DESs is also affected by temperature. As the temperature increases from 50°C to 70°C, the pH of all DESs and EG decreases slightly. Overall, pH levels across all DESs remain stable at different temperatures, consistent with a study showing that alcohol-based DESs also exhibit a gradual pH decrease with increasing temperature (Hansen et al., 2021).

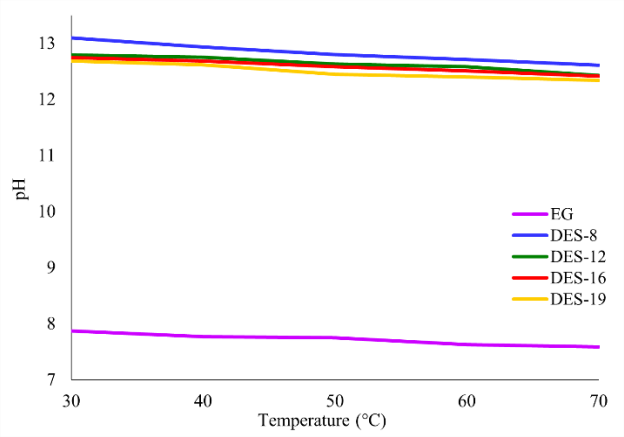


Figure 5. pH of PC-EG eutectic mixtures at various molar ratios and temperatures.

Density of PC-EG Eutectic Mixtures

Figure 6 shows the density of EG and DESs at different molar ratios and temperatures. At ambient temperature, DES-19 has the lowest density, while DES-8 has the highest. The strong interaction between EG molecules and PC in DES-8 leads to the strongest H-bonding among the analyzed solvents, consistent with Ghaedi et al. (2019), who noted that higher-density DESs exhibit stronger H-bonding. FTIR analysis also confirmed strong H-bonding in DES-8. While Hansen et al. (2021) suggested that DES density increases with increasing hydroxyl (OH) content, our findings suggest otherwise. At a molar ratio of 1:8, OH groups in EG form new H-bonds with PC, restricting molecular movement. As the EG molar ratio increases, density decreases due to excess unbound OH groups, allowing more free movement within the DES.

The density of DESs is also affected by temperature. The density of DESs decreases from DES-8 to DES-19 as temperature rises from 30°C to 70°C. Increased thermal energy allows molecules to move more freely, leading to larger intermolecular distances and lower density (Şahin et al., 2026). Lower temperatures can reduce pore size and change molecular packing, resulting in increased density (Zhang et al., 2021).

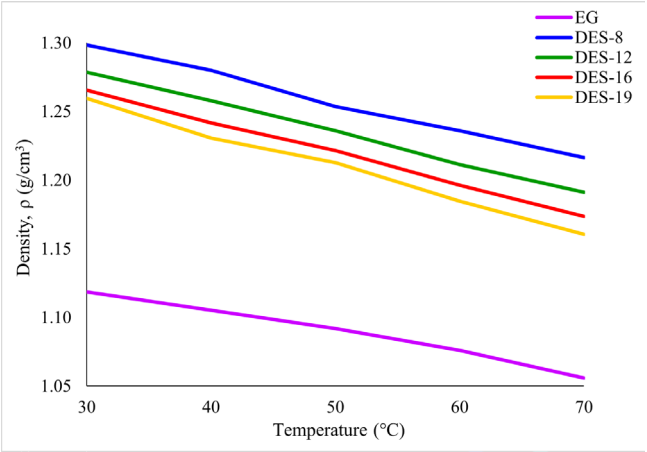


Figure 6. Density of PC-EG eutectic mixtures at various molar ratios and temperatures.

Viscosity of PC-EG Eutectic Mixtures

Figure 7 illustrates the viscosities of EG and DESs at different molar ratios and temperatures. Viscosity measures a fluid's resistance to flow (Şahin et al., 2026). DES-8 has the highest viscosity, while DES-12, DES-16, and DES-19 demonstrate a decrease in viscosity as the molar ratio of EG increases. The increase in viscosity indicates the formation of a strong H-bond network between HBA and HBD, resulting in reduced DES mobility (Şahin et al., 2026). Thus, the decrease in DES-19 viscosity indicates reduced PC-EG interactions. This occurs because unbound molecules near the PC-EG groups actively enhance the movement of unbound molecules in the solvent. The finding is also supported by several studies (Şahin et al., 2026; Ijardar et al., 2022).

The viscosity of DESs also decreases as temperature increases. At 70°C, the viscosities of all DESs become comparable to EG. The findings suggest that the thermal energy is sufficient to weaken or break the H-bonds between the PC-EG components, leading to a significant reduction in viscosity and flow resistance (Ghaedi et al., 2022;Yan et al., 2022), as observed in DES-16 and DES-19. DESs with low viscosities have have greater potential for green technology because their properties can be tailored by varying the sizes of the HBD and cation components (Şahin et al., 2026). Furthermore, the low viscosity DES increases the rate of SO_2 absorption (Yan et al., 2022). The low viscosity of DES-16 and DES-19 provides an advantage in desulfurization. The high viscosity of the liquid impedes chemical reactions due to poor mass transfer. For example, extracting substances from solids becomes challenging because reactions occur at the interface between the solid and liquid phases. Diffusion affects these reactions, adding another layer of difficulty to the process. Once stirring stops, separating solid particles from highly viscous liquids becomes challenging. The particles settle slowly, and the liquids adhere to

DOI:https//doi.org/10.22452/mjs.vol45no2.2
Malaysian Journal of Science 45(2): 13-22 (June 2026)

their surfaces, increasing the risk of fluid loss (Binnemans & Jones, 2023). Additionally, there is a positive correlation between conductivity and viscosity. The higher viscosity of DESs can lead to poor conductivity. Thus, the molar ratio of DES components significantly affects both viscosity and conductivity (Xiaoyu et al., 2023).

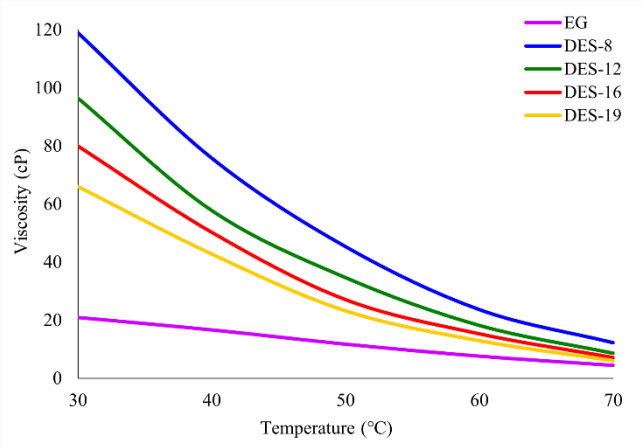


Figure 7. Viscosity of PC-EG Eutectic Mixtures at Various Molar Ratios and Temperatures.

Ionic Conductivity of PC-EG Eutectic Mixtures

Figure 8 illustrates the ionic conductivities of EG and PC-EG at various molar ratios and temperatures. At lower temperatures, the ionic conductivity is stable and consistent across all DESs, despite varying viscosities. High viscosity restricts ionic mobility, directly affecting ionic conductivity (Xiaoyu et al., 2023). At 30°C, DES-16 has the highest ionic conductivity, followed by DES-19, DES-12, and DES-8. The results indicate that increasing the EG molar ratio enhances conductivity, aligning with Xiaoyu et al. (2023), who observed that reducing the ChCl (HBA) ratio from 1:2 to 1:10 improves ionic conductivity.

The graph shows a positive correlation between higher temperatures and increased conductivity. As the temperature increased from 30°C to 40°C, the conductivity of the DESs rose, with DES-12 showing the biggest rise, followed by DES-16 and DES-19. The result suggests that a higher molar ratio of EG may enhance ionic conductivity. DES-8 exhibited the lowest conductivity, likely due to its viscosity hindering ion diffusion (Talip et al., 2021). While increasing the temperature from 40°C to 50°C did not enhance conductivity, significant increases were noted between 60°C and 70°C. The results indicate that higher temperatures disrupt the H-bond network in DES-8, thereby enhancing ionic mobility and conductivity. At 50°C, the ionic conductivity of DES-12 slightly decreases, while DES-16 and DES-19 maintain stable conductivity between 40°C–50°C. At 70°C, DES-12 outperformed DES-16, making it the best-performing DES. The findings also suggest that increased EG concentration results in less consistent thermal conductivity, and DES-19, despite having the highest EG concentration, does not show the highest conductivity. Additional EG in DES-19 increases the number of conductive ions but limits conductivity at high temperatures.

The thermal conductivity of DESs can be enhanced by OH groups and potassium ions (K^+). However, higher temperatures can increase electron motion and lower ionic conductivity (Albayati et al., 2021). EG shows almost zero conductivity between 30°C and 70°C, indicating that it is a poor conductor. Zeng et al. (2024) found that the ionic conductivity of PC-EG correlates with its viscosity at a 1:20 molar ratio, which yields the highest conductivity before it declines. In summary, DES conductivity depends on PC-EG interaction, viscosity, and temperature, with a higher EG ratio improving ionic conductivity.

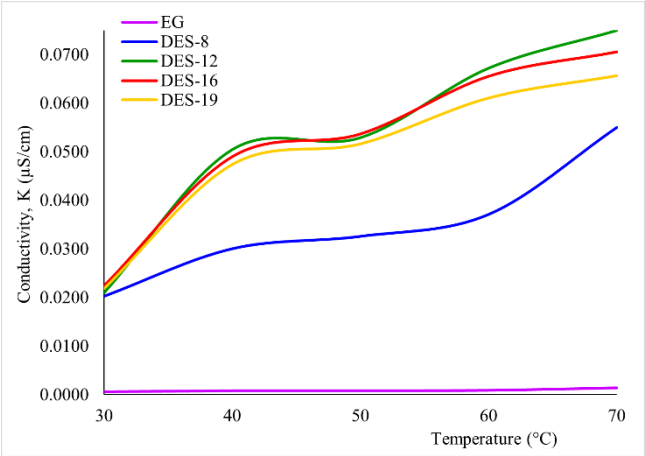


Figure 8. Ionic conductivity of PC-EG eutectic mixtures at various molar ratios and temperatures.

Thermal Analysis of PC-EG Eutectic Mixtures

The thermal stability of DESs at various molar ratios was assessed using thermogravimetric analysis. The goal was to determine the temperature range in which structural integrity is preserved and thermal degradation is prevented. Figure 9 illustrates the TGA results, showing that PC remained stable at up to 300°C, while EG began to degrade at around 98°C. The result indicates a notable difference in thermal stability, with PC being more stable than EG. According to Figure 9, DES-12 and DES-16 were thermally stable up to 130°C, while DES-8 remained stable until 160°C before degrading due to the EG volatilization. DES-19 started to lose weight at 110°C, which highlights the influence of EG. Zeng et al. (2024) noted that PC decomposes at higher temperatures than EG, and increasing HBD content leads to greater weight loss during initial decomposition. All DESs were fully degraded at around 350°C, with final weights of 21.52% for DES-8, 12.64% for DES-12, 16.90% for DES-16, and 2.56% for DES-19. The findings suggest interactions between PC and EG, with H-bond strength decreasing at higher temperatures (Saeed et al., 2022). Overall, PC-EG mixtures are effective at moderate temperatures but lose structural integrity at 130°C. Although the 1:8 molar ratio is stable at temperatures of up to 180°C, its high viscosity limits ionic conductivity and efficiency.

DOI:https//doi.org/10.22452/mjs.vol45no2.2
Malaysian Journal of Science 45(2): 13-22 (June 2026)

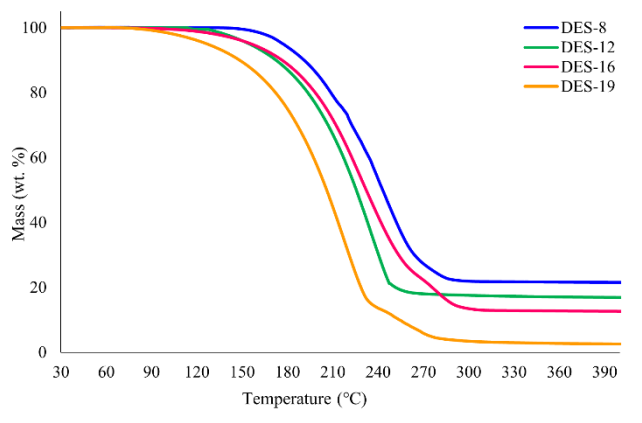


Figure 9. TGA analysis of PC-EG eutectic mixtures at different molar ratios.

4. Conclusions

The PC-EG mixture formed a eutectic at a 1:8 molar ratio, remaining a clear liquid at room temperature even after 24 hours. DSC analysis revealed lower freezing points, confirming they are DESs. The solid-liquid phase diagram indicated that DES-8, DES-12, DES-16, and DES-19 remained liquid at 470°C, 350°C, 220°C, and 180°C, respectively, but degraded at around 350°C. This analysis suggests that DESs are suitable for moderate temperatures but unstable at high temperatures. FTIR analysis revealed H-bonding in the mixtures, with bond strength increasing at higher PC-EG ratios before decreasing at an EG molar ratio of 19. The molar ratio and temperature also influenced the pH, density, viscosity, and conductivity of the DESs. DES-8 had the highest pH, density, and viscosity, while DES-19 had the lowest. Higher molar ratios and temperatures decreased viscosity but enhanced ionic conductivity. DES-12 and DES-16 exhibited low viscosity, strong H-bonding, and excellent conductivity at 30°C–40°C and 50°C–70°C, suggesting potential for extraction and industrial use.

5. Acknowledgements

The author expresses gratitude to Universiti Teknologi MARA for its valuable support and for providing the necessary facilities.

6. References

Albayati, N., Kadhom, M., Abdullah, G., & Salih, S. (2021). Thermal Conductivity of Room Temperature Deep Eutectic Solvents. *Thermal Science*, 30(6), 1960-1972.

Ali, A., Li, R., Zhu, R., Mahmood, S., Chen, Q., & Yao, S. (2026). Deep eutectic solvents for green extraction and separation of bioactive compounds from traditional Chinese medicines. *Chinese Medicine*, 21(1), 38.

Biernacki, K., Souza, H. K. S., Almeida, C. M. R., Magalhães, A. L., & Goncąlves, M. P. (2020). Physicochemical properties of choline chloride-based deep eutectic solvents with polyols: An experimental and theoretical investigation. *ACS Sustainable Chemistry & Engineering*, 8(50), 18712-18728.

Binnemans, K., & Jones, P. T. (2023). Ionic liquids and deep-eutectic solvents in extractive metallurgy: mismatch between academic research and industrial applicability. *Sustainable Metallurgy*, 9(2), 423-438.

Francisco, M., Van Den Bruinhorst, A., & Kroon, M. C. (2013). Low-transition-temperature mixtures (LTTMs): A new generation of designer solvents. *Angewandte Chemie International edition*, 52(11), 3074-3085.

Ghaedi, H., Kalhor, P., Zhao, M., Clough, P. T., Anthony, E. J., & Fennell, P. S. (2022). Potassium carbonate-based ternary transition temperature mixture (deep eutectic analogues) for CO2 absorption: Characterizations and DFT analysis. *Frontiers of Environmental Science & Engineering*, 16(7), 92.

Ghaedi, H., Zhao, M., Ayoub, M., Zahraa, D., Shariff, A. M., & Inayat, A. (2019). Preparation and characterization of amine (N-methyl diethanolamine)-based transition temperature mixtures (deep eutectic analogues solvents). *Chemical Thermodynamics*, 137, 108-118.

Hansen, B. B., Spittle, S., Chen, B., Poe, D., Zhang, Y., Klein, J. M., Horton, A., Adhikari, L. I., Zelovich, T., Doherty, B. W., Gurkan, B., Maginn, E. ., Ragauskas, A., Dadmun, M., Zawodzinski, T. A., Baker, G. A., Tuckerman, M. E., Savinell, R. F., & Sangoro, J. R. (2021). Deep eutectic solvents: A review of fundamentals and applications. *Chemical reviews*, 121(3), 1232-1285.

DOI:https//doi.org/10.22452/mjs.vol45no2.2
Malaysian Journal of Science 45(2): 13-22 (June 2026)

Ijardar, S. P., Singh, V., & Gardas, R. L. (2022). Revisiting the physicochemical properties and applications of deep eutectic solvents. *Molecules*, 27(4), 1368.

Ismail, S. N. A. S., Ishak, M. A. M., Ahmad, R., & Ismail, K. (2025). Impact of optimized molar ratio, temperature and time on organic sulfur transformation and thermophysical properties of high-sulfur coal. *Technology*, 16(2), 395-410.

Khan, N., & Srivastava, V. C. (2022). Extractive desulfurization using ethylene glycol and glycerol-based deep eutectic solvents: engineering aspects and intensification using ultrasound. *Chemical Engineering and Processing-Process Intensification*, 180, 108973.

Lakhmira, M. A., Soomroa, S. A., Unara, I. N., & Akhterb, F. (2022). A review of chemical demineralization and desulphurization of high ash & high sulphur lignite coal. *Jurnal Kejuruteraan*, 34(3), 353-364.

Majid, M. F., Mohd Zaid, H. F., Fai Kait, C., Jumbri, K., Lim, J. W., Masri, A. N., & Yuliarto, B. (2020). Liquid polymer eutectic mixture for integrated extractive-oxidative desulfurization of fuel oil: An optimization study via response surface methodology. *Processes*, 8(7), 848.

Mamtani, K., Stevenson, S. J., Galvosas, P., Weber, C. C., Farid, M. M., & Shahbaz, K. (2023). Synthesis and characterization of acidic deep eutectic solvents based on p-Toluenesulfonic acid. *Molecular Structure*, 1294, 136378.

Manafpour, A. A., Feyzi, F., & Rezaee, M. (2024). An environmentally friendly deep eutectic solvent for CO_2 capture. *Scientific Reports*, 14(1), 19744.

Matjie, R. H., Lesufi, J. M., Bunt, J. R., Strydom, C. A., Schobert, H. H., & Uwaoma, R. (2018). In situ capturing and absorption of sulfur gases formed during thermal treatment of South African coals. *ACS Omega*, 3(10), 14201-14212.

Mjalli, F. S., Naser, J., Jibril, B., Al-Hatmi, S. S., & Gano, Z. S. (2014). Ionic liquids analogues based on potassium carbonate. *Thermochimica Acta,* 575, 135-143.

Naser, J., Mjalli, F., Jibril, B., Al-Hatmi, S., & Gano, Z. (2013). Potassium carbonate as a salt for deep eutectic solvents. *Chemical Engineering and Applications*, 4(3), 114.

Pinho, M. R., Lima, A. S., de Almeida Ribeiro Oliveira, G., Liao, L. M., Franceschi, E., Silva, R. D., & Cardozo-Filho, L. (2024). Choline chloride-and organic acids-based deep eutectic solvents: Exploring chemical and thermophysical properties. *Chemical & Engineering Data*, 69(10), 3403-3414.

Rogošić, M., & Kučan, K. Z. (2019). Deep eutectic solvents based on choline chloride and ethylene glycol as media for extractive denitrification/desulfurization/dearomatization of motor fuels. *Industrial and engineering chemistry*, 72, 87-99.

Saeed, U., Khan, A. U., Khan, A. L., Gilani, M. A., & Bilad, M. R. (2022). Separation of Carbon Dioxide by Potassium Carbonate Based Supported Deep Eutectic Liquid Membranes: Influence of Hydrogen Bond Donor. *Membrane Science and Research*, 8(1), e244300.

Şahin, S., Kurtulbaş, E., Toprakçı, İ., Anwar, F., Khan, R., Ciğeroğlu, Z., & Falsafi, S. R. (2026). Deep Eutectic Solvents for Sustainable Extraction of Bioactive Compounds from Biomass: Mechanistic Insights and Scale-Up Challenges. *Molecules*, 31(5), 880.

Sultana, K., Rahman, M. T., Habib, K., & Das, L. (2022). Recent advances in deep eutectic solvents as shale swelling inhibitors: A comprehensive review. *ACS Omega*, 7(33), 28723-28755.

Talip, R. A. A., Yahya, W. Z. N., & Bustam, M. A. (2021). Viscosity and ionic conductivity of imidazolium based ionic liquids bearing triiodide anion. *In E3S Web of Conferences* (Vol. 287, p. 02015). EDP Sciences.

DOI:https//doi.org/10.22452/mjs.vol45no2.2
Malaysian Journal of Science 45(2): 13-22 (June 2026)

Wang, X., Wang, Y., Dong, M., Fang, Z., Hu, Y., Xue, K., & Peng, X. (2023). A two-dimensional nanochannel facilitates ionic conductivity of a deep eutectic solvent for an efficient supercapacitor. *Materials Today Energy*, 34, 101285.

Wang, J., & Baker, S. N. (2018). Pyrrolidinium salt based binary and ternary deep eutectic solvents: green preparations and physiochemical property characterizations. *Green Processing and Synthesis*, 7(4), 353-359.

Xiaoyu, W. Yuqi, D., M., F., Z., H., Y., X., K., Y., Z., & P., X. (2023). A two-dimensional nanochannel facilitates ionic conductivity of a deep eutectic solvent for an efficient supercapacitor. Materials Today Energy, 34, 101285.

Yan, Z., Lai, S. Y., Ngan, C. L., Li, H., & Mohamed, A. R. (2022). Recent advances in energy-efficient and regenerative SO_2 absorption over deep eutectic solvents. *Environmental Chemical Engineering*, 10(6), 108967.

Zeng, P., Wu, D., Wang, T., Liu, P., & Jia, D. (2024). Redefine the existence form and function of water in potassium carbonate-based deep eutectic electrolyte. *Fuel*, 357, 129738.

Zhang, M., Zhang, X., Liu, Y., Wu, K., Zhu, Y., Lu, H., & Liang, B. (2021). Insights into the relationships between physicochemical properties, solvent performance, and applications of deep eutectic solvents. *Environmental Science and Pollution Research*, 28(27), 35537-35563.

Zhou, Y., Chen, M., Dong, X., & Yang, D. (2024). The reaction between K_2CO_3 and ethylene glycol in deep eutectic solvents. *Molecules*, 29(17), 4113.

Zou, Y., Xing, X., Wang, C., Tao, D., Ji, H., Wu, P., & Zhu, W. (2026). Investigating the SO_2 absorption behavior of pyrimidine-based deep eutectic solvents via a dual-site thermodynamic model. *Industrial Chemistry & Materials*.

DOI:https//doi.org/10.22452/mjs.vol45no2.2
Malaysian Journal of Science 45(2): 13-22 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 45(2): 23-32 (June 2026) https://mjs.um.edu.my

A Novel of Green Synthesis Silver Nanoparticles Using Kecemcem Leaves (*Spondias pinnata* (L.F) Kurz.) Extract for Tetracycline Biosensor

Gusti Ayu Dewi Lestari[1a]*, Ni Luh Ayu Arsita Dewi[2a], Ni Ketut Esati[3a], Iryanti Eka Suprihatin[4b], Prastika Krisma Jiwanti[5c] and Laurencia Gabrielle Sutanto[6c]

Abstract: Tetracycline is one of antibiotics that widely used inappropriately in livestock. The long-term use of tetracycline may cause bacterial resistance. Biosensor can be applied as one of the detection methods for antibiotics in livestock. Development of biosensor materials can be obtained by nanoparticles formation through a synthesis process. Kecemcem leaves are materials that can be used to synthesize silver nanoparticles. This study aimed to synthesize silver nanoparticles using kecemcem leaves (*Spondias pinnata* (L.f) Kurz.) and its application as tetracycline biosensor. Aqueous extract of kecemcem leaves and 1 mM $AgNO_3$ solution were combined in a 1:10 (v/v) ratio to create silver nanoparticles. The biosensor capability of silver nanoparticles was evaluated against tetracyclines after they were characterized using a UV-Vis spectrophotometer, particle size analyzer, scanning electron microscope-EDS, transmision electron microscope, and FTIR spectrophotometer. The maximum wavelength of the AgNPs scanning results on the UV-Vis spectrophotometer was in the 405 nm region. The results of characterization on TEM, SEM-EDS and PSA morphology of the formed AgNPs were spherical, the smallest particle size was 26.886 nm on average. In the characterization using FTIR there were found hydroxyl groups in AgNPs. The resulting nanoparticles could act as tetracycline biosensors, as evidenced by the shift in wavelength when AgNPs interacted with tetracycline. Detection of tetracycline was established based on the SPR technique. The brownish red color of the AgNPs solution turned to clear slightly reddish in the presence of tetracycline, accompanying the decrease in intensity of SPR band. Based on the method validation carried out, the linearity value was obtained in the form of r value of 0.983, LOD value of 0.879 ppm, and LOQ of 2.93 ppm, while a precision value of 3.356%, and an accuracy of 82.58%.

Keywords: *Biosensors, green synthesis, kecemcem leaves, silver nanoparticles, tetracyclines.*

1. Introduction

Tetracyclines are antibiotics used for respiratory and digestive tract infections in human. Tetracyclines are also used for the treatment of livestock, thereby reducing the risk of death and returning livestock to a healthy condition. Tetracycline is irrationally frequently used as a feed supplement since it maximizes broiler development and boosts feed intake efficiency *(*Widiasih *et al*., 2019). Extensive use without following indications can lead to accumulation of tetracycline residues in livestock organs. Over 75% of the tetracycline utilized will be released into the environment in an active form through feces and urine. Tetracycline which is used globally has led to findings of tetracycline contamination in water sources, and the toxicity of tetracycline contamination in water sources has a more dangerous level than parasitic fish and *Daphnia sp*. (Xu *et al*., 2021) can cause serious harm to human health who consume livestock and water contaminated with tetracyclines.

Authors information:
[a]Department of Pharmacy, Faculty of Pharmacy and Health Sciences, Universitas Pendidikan Nasional, 80224, Bali, INDONESIA. Email: dewilestari@undiknas.ac.id[1]; arsitadewiayu14@gmail.com[2]; esati@undiknas.ac.id[3]
[b]Department of Chemistry, Faculty of Mathematics, and Natural Sciences, Udayana University, Bandung, INDONESIA. Email: eka_suprihatin@unud.ac.id[4]
[c]Nanotechnology Engineering, Faculty of Advanced Technology and Multidiscipline, Airlangga University, Surabaya 60115, INDONESIA. Email: prastika.krisma@ftmm.unair.ac.id[5]; laurencia.gabrielle.sutanto-2020@ftmm.unair.ac.id[6]
*Corresponding Author: dewilestari@undiknas.ac.id

In Indonesia, the residue of tetracycline antibiotics was found in pig feces as much as almost 95%, pig farm waste as much as 66.7% and pig slaughterhouse wastewater as much as 50% (Pazra *et al.,* 2023). In free-range and laying hens there were findings of meat containing antibiotic residues, especially tetracyclines of 12.5% (Widiasih *et al.,* 2019). These findings prove that tetracycline is still widely used and contained in parts of livestock in several provinces in Indonesia. In Indonesia, to prevent this case from occurring, the maximum residue limit for tetracycline in livestock products must comply with the safe limit in meat and eggs of 0.05 mg/kg, so many efforts have been made to develop a system for detecting tetracycline residues in livestock parts (Widiasih *et al*., 2019).

A method that can detect an antibiotic in livestock parts is a biosensor (Ghodake *et al*., 2020). Tetracyclines and other antibiotics can be detected using biosensors (Jalalian *et al.,* 2018). According to (Kim *et al.,* 2010) who stated that tetracycline can be detected by biosensors derived from gold nanoparticles, but this method is difficult to develop given the price of gold which is quite expensive and the stability of the nanoparticles is not good. Antibiotics can be easily, quickly, and affordably detected by biosensors in tainted food products as well as pharmaceutical preparation. Silver nanoparticles are one of the materials that can be developed as a biosensor (nanoprobe). Antibiotics analysis with AgNPs can be carried out by visual observation (colorimetry)

Received: December, 2024
Accepted: August, 2025
Published: June, 2026

DOI:https//doi.org/10.22452/mjs.vol45no2.3
Malaysian Journal of Science 45(2): 23-32 (June 2026)

and observation of the shift in the AgNPs wavelength on a UV-Vis spectrophotometer (Ghodake *et al.*, 2020). Nano-based biosensors have better biosensor characteristics (Malhotra & Ali, 2018). Silver nanoparticles have unique properties, it exhibits as an efficient sensor probe with enhanced sensitivity and selectivity (Gurunathan *et al.,* 2016; Haes *et al.*, 2002). During the sensing process, the silver nanoparticles exhibit a visible color change in the presence of particular external agents including ions or antibiotics. The visual color change suggests that the targeted ions or molecules are present in the analysis. Therefore, herein, efforts are being made in developing silver nanoparticles based sensors which are cost-effective, easy and rapid monitoring system to detect hazardous antibiotics in environmental samples (Singh *et al*., 2018)

In applying the benefits of silver nanoparticles as a biosensor, synthesis is necessary. The chemical synthesis method using biological compounds is a method that can be applied in the nanoparticles preparation (Deng *et al*., 2016). This method is broadly utilized, because it has the advantages of being cheap, easy to carry out, sustainable, and renewable. Plant extracts are the most commonly used ingredients. Apart from being cheap, abundant and environmentally friendly resources are also advantage of plant extracts (Sharma *et al*., 2019).

One of the plants that can be utilized in the production of silver nanoparticles is kecemcem leaves (*Spondias pinnata (L.f)* Kurz.). It has several other names, including cemcem and forest kedondong. The distribution of kecemcem plants in Bali is dominantly found in Kintamani, Bangli and some can be found in other areas of Bali such as Gianyar and Klungkung. The local community grows kecemcem plants along the fields or on the edges of rice fields which can play a role in reforestation and as a traditional health drink (loloh). However, the use of kecemcem leaves is not optimal, so the selling value of these leaves is very low (Sujarwo *et al*., 2017).

It has been demonstrated that silver nanoparticles can be biosynthesised utilizing kecemcem leaf bioreductors. Based on research by (Purnamasari *et al*., 2021) optimal AgNPs results are formed at 60^{o}C with extract concentration of 0.5%. This description suggests that the bioreductor found in kecemcem leaves was used as reducing/stabilizing agent for the production of stable, small size AgNPs. Nanoparticles were characterized by various analytical methods. The possible

application of AgNPs as rapid, selective, easy-to-use and cost effective colorimetric sensors for tetracycline was investigated.

2. Materials and Methods

Kecemcem Leaf Sample Preparation

After being cleaned under running water, the kecemcem leaves were chopped into small pieces and allowed to dry for seven days away from the sun. Kecemcem leaves were dried, ground into a powder, and then sieved through a 50 mesh screen. In a 250 mL beaker, 20 grams of powdered kecemcem leaf and 100 mL of demineralized aqua were boiled for 15 minutes at 60°C. The mixture was then allowed to cool before being filtered through filter paper. The filtrate is concentrated to 0.5 percent and utilized to create silver nanoparticles.

Phytochemical Test

The purpose of the phytochemical screening test is to identify secondary metabolites present in kecemcem leaves. Alkaloids, tannins, terpenoids/steroids, flavonoids as well as saponins were all examined in the phytochemical screening test of the kecemcem leaf aqueous extract.

Silver Nanoparticle Synthesis

Weighing 170 mg of $AgNO_3$ powder, dissolving it in aquadem in a beaker glass, and then placing it in a 1000 mL volumetric flask, demineralized aqua was added, and the mixture was shaken until it was homogenous. This produced a 1 mM $AgNO_3$ solution. A 0.5% kecemcem leaf water extract was combined with a 1 mM $AgNO_3$ solution. Kecemcem leaf water extract and $AgNO_3$ solution had a 1:10 (v/v) ratio. The mixture was heated to 60°C until it turned a brownish red color. Subsequently, the AgNPs that has been obtained was measured for its wavelength and absorbance on UV-Vis spectrophotometry and can then be characterized using PSA, SEM-EDS, FTIR, and TEM instruments.

Silver Nanoparticles Characterization

The purpose of the UV-Vis spectrophotometer's characterisation test is to ascertain whether silver nanoparticles have developed. The AgNPs sample was put into the cuvette and the absorbance was examined. The maximum wavelength between the range of 400 – 450 nm indicates that the sample contains AgNPs. The functional groups of the secondary metabolites in the pure kecemcem leaf water extract and the colloidal silver nanoparticles, which were produced from the aqueous extract of kecemcem, were compared using FTIR analysis.

The morphology as well as size of the resultant pictures were ascertained using TEM characterization. A few drops of colloidal silver nanoparticles were dropped on the TEM grid and then analyzed. Particle sizes range from 1-100 nm was determined wheter it has spherical, triangular, crystalline or anisotropic particle shapes. The goal of SEM-EDS characterization is to observe the composition and shape of colloidal silver nanoparticles. A precipitate was created by centrifuging the sample solution for ten minutes at 10,000 rpm, washing the resultant precipitate with ten milliliters of aqua demineralized, and then centrifuging it once more. After that, the precipitate was dried at 50°C in an oven. SEM-EDS was then used to evaluate the resulting powder. The PSA characterisation test seeks to determine the zeta potential of silver nanoparticles as well as their size or diameter. A PSA device was used to examine the AgNPs solution after it had been placed in a cuvette. The resulting size ranges from 0.6 nm – 7 µm.

DOI:https//doi.org/10.22452/mjs.vol45no2.3
Malaysian Journal of Science 45(2): 23-32 (June 2026)

AgNPs Analysis as a Tetracycline Biosensor

A concentration series solution of tetracycline 1; 2; 3; 4; 5; 6; 7; 8; 9; and 10 ppm were prepared with 0.1 N HCl as solvent. The 1.8 mL of demineralized water and 0.2 mL of AgNPs were combined with each tetracycline concentration. Then the mixture was analyzed for its wavelength and absorbance on a UV-Vis spectrophotometer. The shift between the peak wavelengths of AgNPs and the wavelengths of the mixture of nanoparticles and tetracycline (MNT) is an indication of a sensory response by AgNPs to tetracyclines.

Method Validation

Method validation was carried out to ensure and confirm the method used is good and can be used continuously. Method validation includes: linearity, determination of LOD and LOQ, determination of precision, and determination of accuracy in MNT (Metal Nanoparticles Tetracycline).

3. Results and Discussion

Silver Nanoparticles Synthesis

The extraction process made an initial concentration of 20%, then diluted to a concentration of 0.5% which was then used for synthesis. Heating was carried out on the extract for 15 minutes at 60^{o}C, which aims to speed up the extraction process. This method was selected since it is simple, affordable, and doesn't call for specialized equipment. A chemical found in plant extracts may become more soluble if the water temperature is raised during the extraction process (Sulaiman *et al*., 2017). While the steroid test yielded negative results, the qualitative test using phytochemical screening on an aqueous extract of kecemcem leaves revealed that the sample included terpenoids, flavonoids, tannins, and alkaloids and saponins (Figure 1). This positive result occurred due to the suitability of the solvent during the extraction process. The appropriate solvent is one of the factors that influence whether the desired compound is dissolved well or not (Sulaiman *et al*., 2017). The solvent used is aqua demineralization which is polar, so the extract obtained is polar. The results of flavonoids and tannins are more concentrated because these compounds are polar, so they dissolve in polar solvents. Alkaloids, terpenoids, and saponins also showed positive results, but the results were not very concentrated or significant. This is probably because there are not many of these substances in the kecemcem leaf aqueous extract. Because the steroid component is non-polar, it is insoluble in solvents and does not show up in the steroid test on the aqueous extract of kecemcem leaves, which is why the test yielded negative results. The amount of steroid content that is not much makes this steroid compound must be tested more specifically to see the quantity (Swathi & Lakshman, 2022).

Based on studies on the synthesis process, it was evident that the $AgNO_3$ solution containing kecemcem leaf water extract changed color; initially yellow, the solution turned brownish red when heated, as seen in Figure 2. One sign that silver nanoparticles are forming is this change in appearance. This is one of the features of the reduction process in silver ions that results in a color shift and the production of silver nanoparticles (Badi'ah *et al.,* 2019).

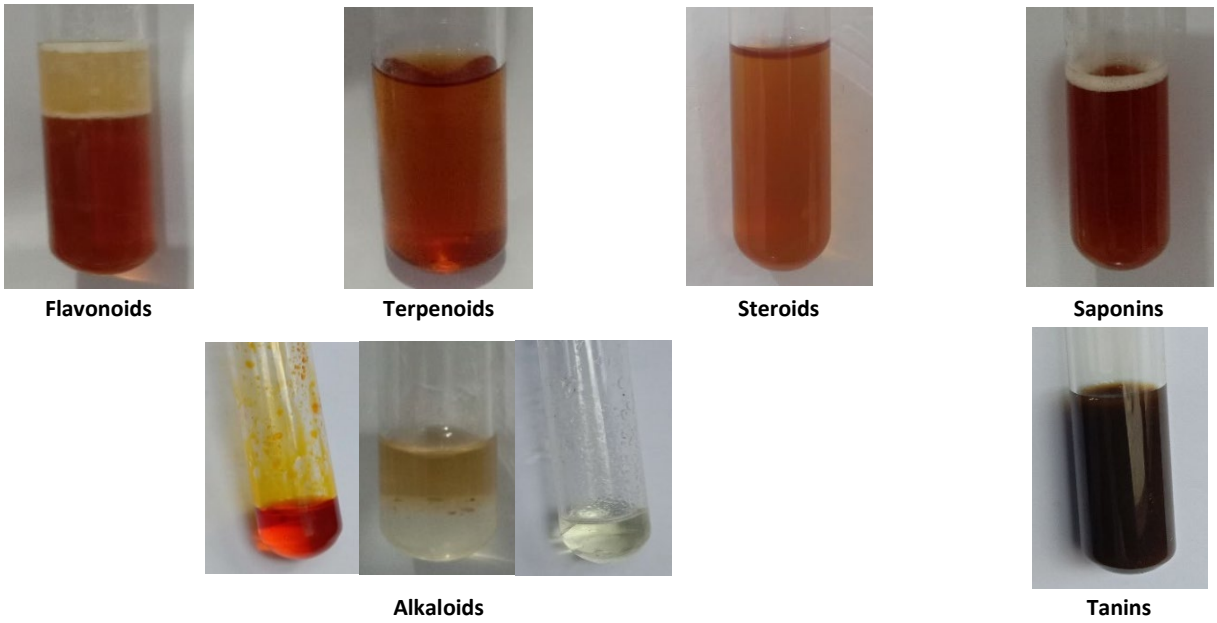


Figure 1. Phytochemical screening on aqueous extract of kecemcem leaves.

DOI:https//doi.org/10.22452/mjs.vol45no2.3
Malaysian Journal of Science 45(2): 23-32 (June 2026)

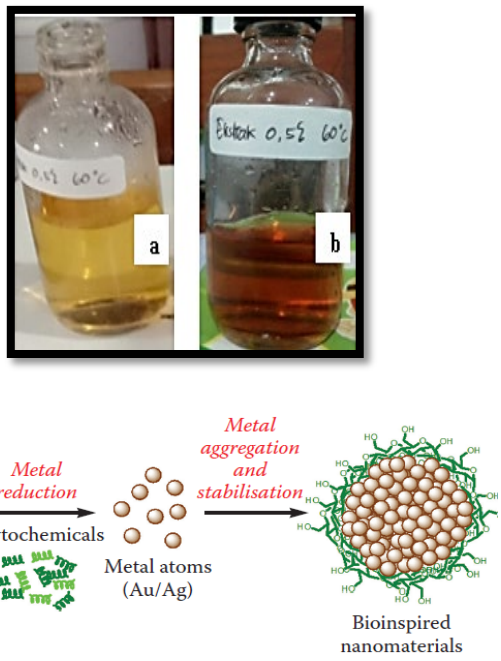


Figure 2. Formation of AgNPs colloid (a) initial mixing between $AgNO_3$ solution with plant extract, (b) formation of AgNPs after synthesis for 15 minutes, and (c) AgNPs formation scheme (Bhaumik *et al.,* 2015).

Silver Nanoparticles' Characterization

The study's findings demonstrated that silver nanoparticles created by reducing kecemcem leaf aqueous extract exhibited SPR intensity at a wavelength of 400-450 nm, with a maximum wavelength of 405 nm, as measured by a UV-Vis spectrophotometer (Figure 3). The fact that the produced silver nanoparticles' SPR intensity satisfied the necessary conditions for nanoparticle production indicates that the synthesis was successful.

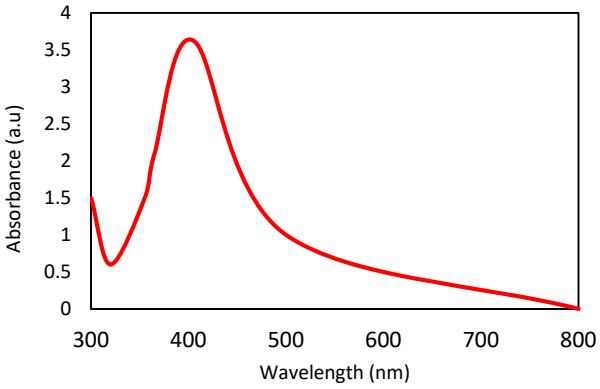


Figure 3. AgNPs absorbance curve

Characterization of aqueous extract of kecemcem leaves with FTIR showed absorptions at certain wavenumbers which are characteristic of flavonoid compounds that are similar to the flavonol group (Figure 4a). Absorption occurs in several areas, namely at 3626.50 cm^{-1} and 3460.06 cm^{-1} indicating -OH vibrations of the hydroxyl groups and C-H aromatics group. A chromophore group is present in the tested sample when the wave number is 2075.12 cm^{-1} (Suchithra *et al*., 2021). Absorption at wavenumbers 562.14 cm^{-1} and 501.04 cm^{-1} shows the existence of phenolic compounds (Niraimathi *et al*., 2013), while the presence of absorption in the 1633.96 cm^{-1} area suggests C=C stretching of the alkenes group and C=O stretching of the ketone group (Periasamy *et al*., 2022) .

Multiple absorptions in specific regions, which suggested the existence of flavonoid compound characteristics, were the findings of the characterization analysis on AgNPs (Figure 4b). The presence of absorption in the 3445.32 cm^{-1} range is indicative of the -OH group, which acts as a reducing agent. In addition to -OH, aromatic C-H also absorbs in this region. An absorption in the 1634.01 cm^{-1} region indicates C=C stretching of the alkenes group as well as C=O stretching of the ketone group (Periasamy *et al.,* 2022), a wave number of 2066.73 cm^{-1} demonstrates the presence of a chromophore group (Suchithra *et al*., 2021), and at a wave number 562.64 cm^{-1} showing the presence of silver ions' presence (Niraimathi *et al*., 2013).

According to the preceding FTIR study results on the characterization of the aqueous extract of kecemcem leaves, flavonoid compounds with -OH groups are present and can donate electrons to create AgNPs. According to earlier studies (Bhaumik *et al*., 2015; Ghodake *et al*., 2020; Mittal *et al*., 2014) the presence of hydroxyl groups (-OH), which were demonstrated in the characterization of AgNPs, could function as capping agents by generating hydrogen bonds.

DOI:https//doi.org/10.22452/mjs.vol45no2.3
Malaysian Journal of Science 45(2): 23-32 (June 2026)

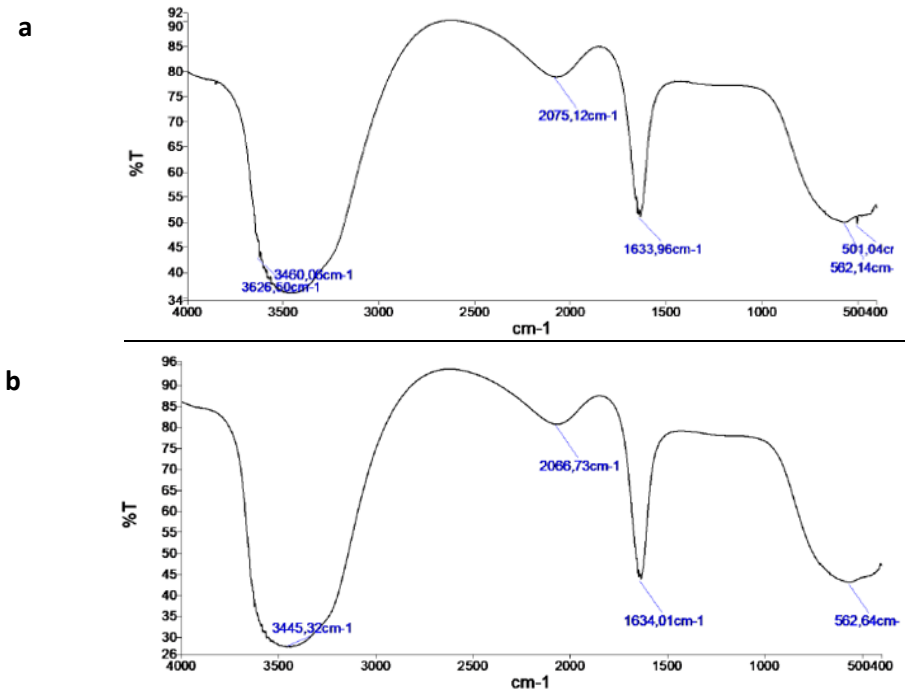


Figure 4. FTIR Spectrum Results from (a) aqueous extract of kecemcem leaves and (b) AgNPs.

A strong method for high-resolution imaging of a thin layer on a solid sample for structural as well as compositional investigation of nanomaterials is TEM (Mittal *et al*., 2014). AgNPs made from kecemcem leaf extract were shown in a TEM micrograph to be spherical, around 100 nm in size, and capped with plant components that inhibited particle aggregation (Figure 5). The size of nanoparticles made from plant bioreductors ranges from 1 to 100 nm, and they have a spherical morphology, smaller the size of AgNPs, the more active they are (Restrepo & Villa, 2021).

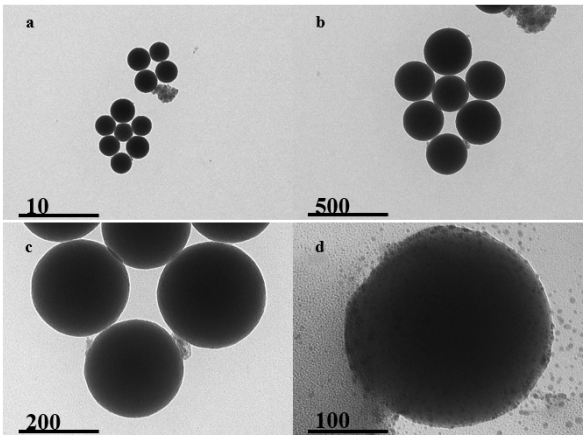


Figure 5. AgNPs morphological analysis using TEM at magnification of (a) 10,000x, (b) 20,000x, (c) 50,000x and (d) 100,000x

The results of SEM measurements (Figure 6.) provide an overview of the formation in the form of micro to Ag^0 nanoparticles in the synthesized colloid. The SEM image at 30x magnification shows random-shaped crystal particles with a fairly large size (Figure 6a). In contrast to the SEM at 10,000x magnification, Figure 6d shows randomly shaped particles with a size range of 1µm that are close to each other (agglomeration). Based on the research of (Restrepo & Villa, 2021) the results of the morphology of silver nanoparticles are random and diverse, this can be caused by the aggregation effect of nanoparticles which can cause agglomeration (Restrepo & Villa, 2021). The agglomeration that occurs causes large nanoparticles, and this is influenced by the presence of Van der Waals forces.

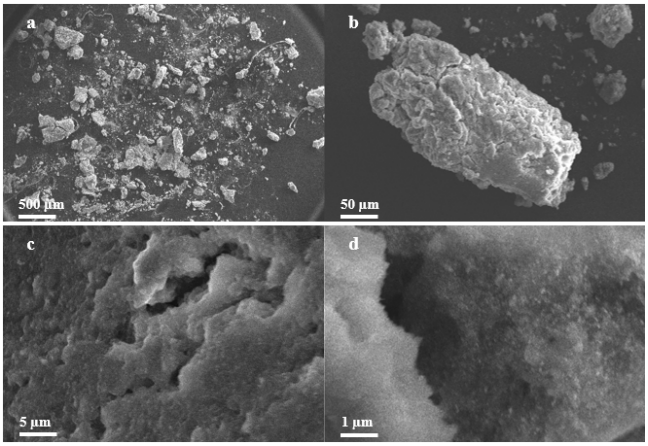


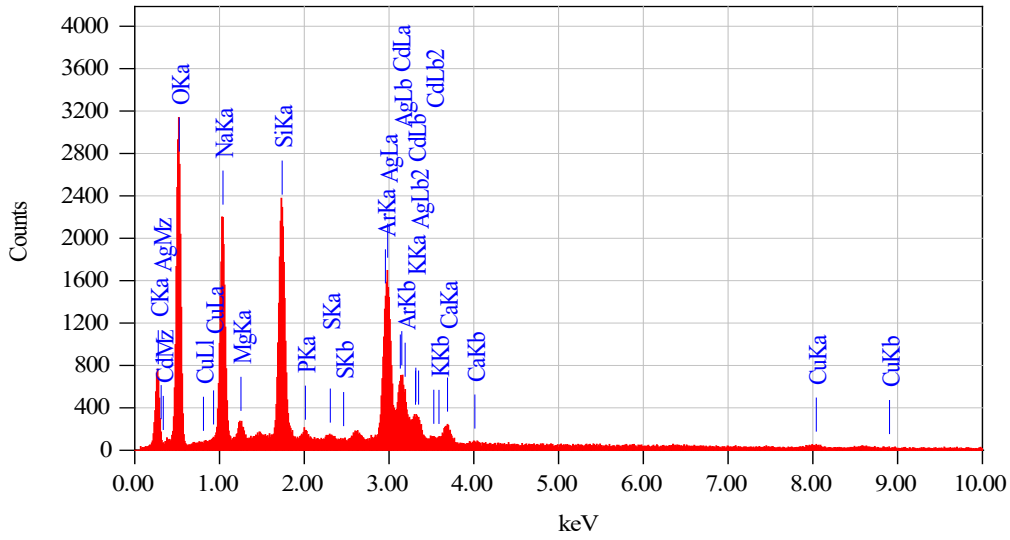
Figure 6. AgNPs morphological analysis using SEM (a) 30x magnification, (b) 500x magnification, (c) 5,000x magnification and (d) 10,000x magnification

The components detected in EDS in the synthesized AgNPs (Figure 7.) are carbon (C), oxygen (O), sodium (Na), magnesium (Mg), silicon (Si), phosphorus (P), sulfur (S), argon (Ar), potassium (K), calcium (Ca), copper (Cu), silver (Ag), cadmium (Cd). Based on

DOI:https//doi.org/10.22452/mjs.vol45no2.3
Malaysian Journal of Science 45(2): 23-32 (June 2026)

the elemental composition data of silver nanoparticles in Figure 7, it shows that the ratio of each element (K) of the Ag compound has a high ratio of 13.6014, so the biosynthetic results do indeed contain silver. In terms of mass percentage, it can be observed that the elements carbon (C) and oxygen (O) have a mass percentage of, namely; 15.92% and 44.53% and the percentage of silver mass (Ag) is 9.91%. These results refer to the biosynthetic process during the formation of nanoparticles, where the use of a water extract bioreductor of kecemcem leaves will affect the results of the elemental composition of AgNPs. Several other elements were also detected, which probably came from other secondary metabolites in kecemcem leaves. Residual elements arising from the tool used which was not cleaned completely or used several times with different samples, is detected by the EDS instrument

a



b

ZAF Method Standardless Quantitative Analysis

Fitting Coefficient : 0.1725

Element	(keV)	Mass%	Sigma	Atom%	Compound	Mass%	Cation	K
C K	0.277	15.92	0.14	25.57				4.9744
O K	0.525	44.53	0.34	53.69				46.5120
Na K	1.041	11.13	0.12	9.33				10.8168
Mg K	1.253	0.79	0.04	0.63				0.8626
Si K	1.739	7.53	0.08	5.17				8.8772
P K	2.013	0.31	0.02	0.20				0.4631
S K	2.307	0.17	0.01	0.10				0.2657
Ar K	2.957	4.33	0.07	2.09				5.9388
K K	3.312	0.58	0.04	0.28				0.9433
Ca K	3.690	0.85	0.03	0.41				1.3664
Cu K	8.040	0.65	0.06	0.20				0.9093
Ag L	2.983	9.91	0.15	1.77				13.6014
Cd L	3.132	3.30	0.10	0.57				4.4690
Total		100.00		100.00				

Figure 7. (a) Spectrum EDS and (b) elemental composition of silver nanoparticles.

Based on the research conducted, the particle size formed has an average size of 26.88 nm, a polydispersity index (PdI) of 0.597 (Table 1.) and a zeta potential value of AgNPs is -25.3 mV. A good PdI value indicates good long-term stability and a PdI value close to zero indicates a homogeneous or monodisperse particle size, while a PdI value that exceeds 0.5 to 1 indicates that the particles have a high degree of heterogeneity or polydisperse (Jalab *et al.*, 2021). These results indicate that the formed AgNPs colloidal particles have poor stability, while the negative value (-) in the results indicates that the formed AgNPs layer is negatively charged. Another study by Singh et al., (2018), zeta potential value of AgNPs was found to be -24.7mV, suggesting the presence of negative charges on the nanoparticle surface. Zeta potential value also indicates strong electrostatic repulsion between the synthesized AgNPs, confirming stable nanoparticles formation. According to (Helmlinger *et al*., 2016), AgNPs has a small size, below 100nm.

DOI:https//doi.org/10.22452/mjs.vol45no2.3
Malaysian Journal of Science 45(2): 23-32 (June 2026)

Table 1. AgNPs PSA Results

Concentration	Replication	Z-average (nm)	(PdI)
Concentration	I	27.09	0.599
0,5%	II	26.78	0.591
	III	26.79	0.603
Average		26.88	0.597

AgNPs Analysis as a Tetracycline Biosensor

The detection ability of AgNPs was performed by the addition of antibiotic tetracycline at a fixed concentration of 10 ppm solutions into AgNPs solution. The change of absorption intensity was evaluated using UV-vis spectroscopy. The noteworthy change in the intensity of the SPR band was clearly observed upon the addition of tetracycline and the color of the solution instantly (within one minute) changed from brownish red color to clear slightly reddish (Figure 8.), indicating the high selectivity of AgNPs toward the detection of tetracycline.

Based on the research, the AgNPs formed and analyzed on a UV-Vis spectrophotometer has a peak at a wavelength of 405 nm, whereas when AgNPs is mixed with tetracycline (MNT), the wavelength shifts to 366 nm (Figure 9.). This is due to a shift in wavelength to a shorter wavelength region (hypsochromic shift) due to substitution or solvent effects (Buncel & Rajagopal, 1990). This is in line with the research of (Kaur *et al.*, 2022), there was a hypsochromic shift in the AgNPs peak from 415 nm to 390 nm when mixed with cephalexin. This was due to a hypsochomic jump, namely the blue shift, which is thought to be the peak for the formation of AgNPs bonds with cephalexin.

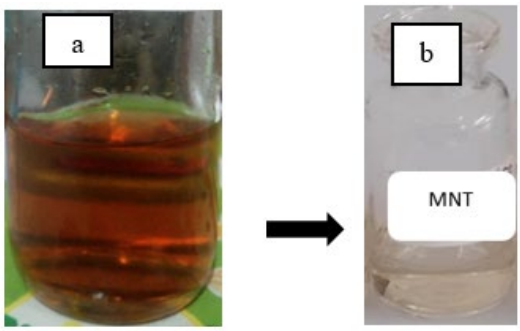


Figure 8. AgNPs (a) and MNT solution (b)

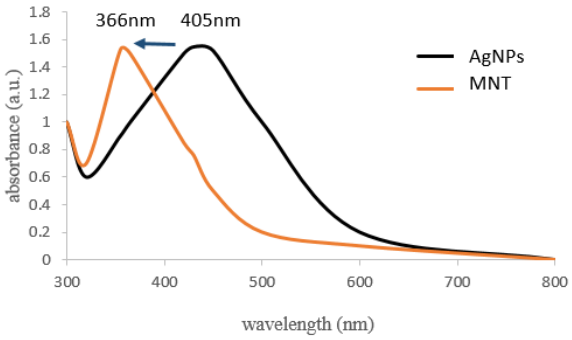


Figure 9. Comparison of AgNPs and MNT peaks.

Hipsochromic shift can occur because AgNPs binds to tetracycline causing AgNPs to be substituted in the tetracycline structure. This causes the conjugation structure of tetracycline to change into an organic compound that has a single bond (σ), thereby reducing the conjugation process in the tetracycline structure. This will absorb UV wavelengths in the direction of wavelengths between 200-400nm. The more AgNPs that binds to tetracycline or the number of tetracycline conjugate structures decreases, the wavelength will continue to shift below 200nm (ultraviolet vacuum). This is what causes a blue shift or hypsochromic in the interaction of AgNPs with tetracycline ((Kaur *et al.*, 2022; Singh *et al.*, 2023). Prediction of possible interactions between nanoparticles and tetracyclines as shown in Figure 10. In the recent research, the presence of the antibiotic kanamycin in milk samples can be detected using AgNPs. That suggesting negligible interference by the other components present in the milk, which makes the proposed method applicable for the direct determination of kanamycin in real samples (Singh *et al.*, 2018). It is hoped that the AgNPs obtained from this research can be used to detect tetracycline in real samples.

DOI:https//doi.org/10.22452/mjs.vol45no2.3
Malaysian Journal of Science 45(2): 23-32 (June 2026)

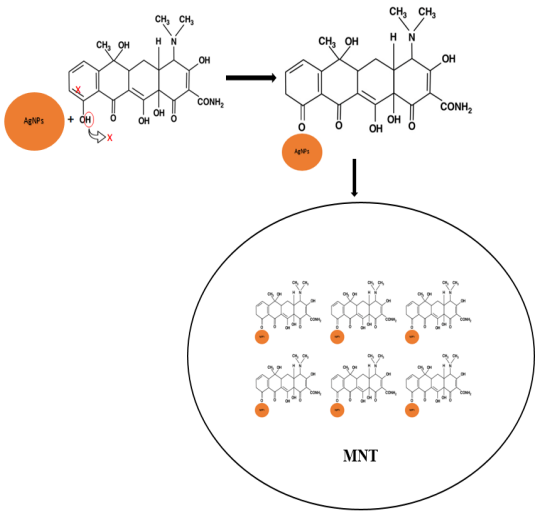


Figure 10. Predicted reaction mechanism of AgNPs with tetracycline.

Validation of Analysis Methods

Validation of the analytical method begins with linearity and concentration series which is carried out by measuring 10 MNT concentrations with a UV-Vis spectrophotometer at a wavelength range of 300-800 nm. The results of the linear regression equation obtained are y = 0.0142x + 0.1399 with a correlation coefficient (r) of 0.983. According to (Kadian *et al*., 2016) linear regression analysis has a minimum value of 0.98 to meet the requirements of good linearity, whereas according to (Araujo, 2009) the value indicates a correlation coefficient (r) which is close to 1, thus indicating a linear relationship between the measured absorbance value with the analyte concentration value (Kadian *et al.,* 2016; Araujo, 2009). Based on the research results, and compared with the literature, the data obtained has a linear correlation. The higher the sample concentration, the greater the absorbance.

The validation analysis of the LOD and LOQ methods was carried out by observing the absorbance obtained from the MNT regression curve, then calculating the MNT LOD and LOQ. Based on the calculations carried out, the LOD value is 0.8792 ppm and LOQ value was obtained at 2.930 ppm, this means that the sample concentration measurement obtained must be more than 2.930 ppm so that the measurement results can be said to be accurate. From the calculations made based on the precision absorbance data, the RSD result is 3.36%. Based on the results of the calculation of the precision value obtained, it is quite precise. This is in accordance with the range in the literature, where the precision of the RSD value is below 16%. Based on the (Bruce *et al*., 1998), the acceptable %RSD value for an analyte concentration of 1 ppm is 11% (Bruce *et al*., 1998). This result means that the method has closeness or suitability of the test results with the supposed value, so that it can be said that the method used is appropriate and precise. Based on the calculation results, the %recovery value obtained is 82.58%. Based on the analyte requirements, should have an acceptable %recovery value of 80-120% so it can be concluded that the method has good accuracy (FDA, 2001).

4. Conclusion

Water extract of kecemcem leaves can be used as a bioreductor in the synthesize of silver nanoparticles. AgNPs formed at a maximum wavelength of 405 nm has an absorbance of 3.627. Based on TEM, PSA, and SEM-EDS analysis, the morphology formed was spherical, with an average particle size of 26.886 nm with a zeta potential value of -25mV. In the characterization of AgNPs using FTIR, there are hydroxyl groups. Nanoparticles formed from the biosynthesis of kecemcem leaves can act as a tetracycline biosensor, in which a shift in wavelength occurs when AgNPs interacts with tetracycline. Synthesized AgNPs have been demonstrated as potential colorimetric sensors for the selective detection of tetracycline, respectively based on SPR. Based on the method validation carried out, it was obtained that linearity was in the form of a value (r) of 0.983, an LOD value of 0.879 ppm, and an LOQ of 2.93 ppm. The precision value obtained is 3.356%, and an accuracy of 82.58%.

5. References

Araujo, Pedro. (2009). Key aspects of analytical method validation and linearity evaluation. *Journal of Chromatography B*, 877(23), 2224-2234

Badi'ah H., Seedeh F., Supriyanto G., Zaidan A. (2019). Synthesis of Silver Nanoparticles and the Development in Analysis Method. *Iop Conference Series: Materials Science and Engineering*, 217(1), 012005)

Bhaumik J., Thakur N.S., Aili P.K., Ghanghoriya A., Mittal A.K., Banerjee U.C. (2015). Bioinspired Nanotheranostic Agents: Synthesis, Surface Functionalization, And Antioxidant Potential. *Acs Biomaterials Science & Engineering*, 1, 382–392

Bruce P., Minkkinen P., Riekkola M.L. (1998). Practical method validation: Validation sufficient for an analysis method. *Mikrochim Acta*, 128, 93–106. https://doi.org/10.1007/BF01242196

DOI:https//doi.org/10.22452/mjs.vol45no2.3
Malaysian Journal of Science 45(2): 23-32 (June 2026)

Buncel E., Rajagopal S. (1990). Solvatochromism and Solvent Polarity Scales. *Acc Chem Res*, 23, 226-231

Deng H., Mcshan D., Zhang Y., Sinha S.S, Arslan Z., Ray P.C., Yu H. (2016). Mechanistic Study of the Synergistic Antibacterial Activity of Combined Silver Nanoparticles and Common Antibiotics. *Environmental Science & Technology*, 50(16), 8840-8848.

Food And Drug Administration. (2001). Guidance For Industry: Bioanalytical Method Validation. Rockville: Center For Veterinary Medicine

Ghodake G., Shinde S., Saratale R.G., Kadam A., Saratale G.D., Syed A., Kim D.Y. (2020). Silver nanoparticle probe for colorimetric detection of aminoglycoside antibiotics: picomolar-level sensitivity toward streptomycin in water, serum, and milk samples. *Journal Of The Science Of Food And Agriculture*, 100(2), 874-884

Gurunathan S., Kim J.H. (2016). Synthesis, toxicity, biocompatibility, and biomedical applications of graphene and graphene-related materials. *Int. J. Nanomed*, 11, 1927

Haes A.J., Van Duyne R.P. (2002). A nanoscale optical biosensor: sensitivity and selectivity of an approach based on the localized surface plasmon resonance spectroscopy of triangular silver nanoparticles. *J. Am.Chem. Soc*, 124, 10596-10604

Helmlinger J., Sengstock C., Groß-Heitfeld C., Mayer C., Schildhauer T.A., Köller M., Epple M. (2016). Silver nanoparticles with different size and shape: equal cytotoxicity, but different antibacterial effects. *RSC Adv*, 6, 18490-18501, DOI: 10.1039/C5RA27836H

Jalab J., Abdelwahed W., Kitaz A., Al-Kayali, R. (2021). Green synthesis of silver nanoparticles using aqueous extract of Acacia cyanophylla and its antibacterial activity. *Heliyon*, 7(9).

Jalalian S.H., Karimabadi N., Ramezani M., Abnous K., Taghdisi S.M. (2018). Electrochemical and optical aptamer-based sensors for detection of tetracyclines. *Trends In Food Science & Technology*, 73, 45-57

Kadian N., Raju K.S.R., Rashid M., Malik M.Y., Taneja I., Wahajuddin M. (2016). Comparative assessment of bioanalytical method validation guidelines for pharmaceutical industry. *Journal of Pharmaceutical and Biomedical Analysis*, 126, 83-97. https://doi.org/10.1016/j.jpba.2016.03.052

Kaur M., Gautam A., Guleria P., Singh K., Kumar V. (2022). Green synthesis of metal nanoparticles and their environmental applications. *Current Opinion in Environmental Science & Health*, 29, 100390

Kim Y.S., Kim J.H., Kim I.A., Lee S.J., Jurng J., Gu M.B. (2010). A novel colorimetric aptasensor using gold nanoparticle for a highly sensitive and specific detection of oxytetracycline. *Biosens Bioelectron*, 26(4),1644-9. http://doi10.1016/j.bios.2010.08.046

Malhotra B.D., Ali M.A. (2018). Nanomaterials In Biosensors: Fundamentals And Applications. http://doi10.1016/B978-0-323-44923-6.00001-7

Mittal A.K., J Bhaumik, S Kumar, UC Banerjee. (2014). Biosynthesis Of Silver Nanoparticles: Elucidation Of Prospective Mechanism And Therapeutic Potential. *Journal Of Colloid And Interface Science*, 415, 39–47.

Niraimathi K.L., Sudha V., Lavanya R., Brindha P. (2013). Biosynthesis of silver nanoparticles using *Alternanthera sessilis* (Linn.) extract and their antimicrobial, antioxidant activities. *Colloids and Surfaces B: Biointerfaces*,102, 288-291

Pazra D.F., Latif H., Basri C., Wibawan I.W.T. (2023). Tetrasiklin Resistance In Escherichia Coli Isolated From Pig Farm, Pig Slaugterhouse, And The Environment In Banten Province. *Jurnal Kedokteran Hewan*, 17(4), 121-126

Periasamy S., Jegadeesan U., Sundaramoorthi K., Rajeswari T., Tokala V.N.B., Bhattacharya S., Muthusamy S., Sankoh M., Nellore M.K. (2022). Comparative Analysis of Synthesis and Characterization of Silver Nanoparticles Extracted Using Leaf, Flower, and Bark of Hibiscus rosasinensis and Examine Its Antimicrobicidal Activity. *Journal of Nanomaterials*, 8123854

Purnamasari G.A.P.P., Lestari G.A.D., Cahyadi K.D., Esati N.K., Suprihatin I.E. (2021) Biosintesis Nanopartikel Perak Menggunakan Ekstrak Air Daun Cemmem (Spondias Pinnata (Lf) Kurz.) Dan Aktivitasnya Sebagai Antibakteri. *Indonesian E-Journal Of Applied Chemistry*, 9(2), 75-79

Restrepo C.V., Villa C.C. (2021). Synthesis of silver nanoparticles, influence of capping agents, and dependence on size and shape: A review. *Environmental Nanotechnology, Monitoring & Management*, 15, 100428

Sharma D., Kanchi S., Bisetty K. (2019). Biogenic synthesis of nanoparticles: A review. Arabian *Journal of Chemistry*, 12(8), 3576-3600

Singh R.K., Panigrahi B., Mishra S., Das B., Jayabalan R., Parhi P.K., Mandal D. (2018). pH triggered green synthesized silver nanoparticles toward selective colorimetric detection of kanamycin and hazardous sulfide ions. *Molliq*, http://doi:10.1016/j.molliq.2018.08.056

DOI:https//doi.org/10.22452/mjs.vol45no2.3
Malaysian Journal of Science 45(2): 23-32 (June 2026)

Singh B., Bhat A., Dutta L., Pati K.R., Korpan Y., Dahiya I. (2023). Electrochemical Biosensors for the Detection of Antibiotics in Milk: Recent Trends and Future Perspectives. *Biosensors*, 13, 867. https://doi.org/10.3390/ bios13090867

Suchithta M.R., Bhuvaneswari S., Sampathkumar P., Dineshkumar R., Chithradevi K., Beevi, M., Madhumitha R., Kavisri M. (2021). In vitro study of antioxidant, antidiabetic and antiurolithiatic activity of synthesized silver nanoparticles using stem bark extracts of Hybanthus enneaspermus. *Biocatalysis and Agricultural Biotechnology*, 38, 102219

Sujarwo W., Saraswaty V., Keim A.P., Caneva G., Tofani D. (2017). Ethnobotanical uses of 'cemcem' (Spondias pinnata (L.F.) kurz; anacardiaceae) leaves in bali (Indonesia) and its antioxidant activity. *Pharmacologyonline*, 1, 113-123.

Sulaiman I.S., Basri M., Fard H.R.M, Chee W.J., Ashari S.E., Ismail M. (2017) Effects of temperature, time, and solvent ratio on the extraction of phenolic compounds and the anti-radical activity of Clinacanthus nutans Lindau leaves by response surface methodology. *Chemistry Central Journal*, 11(1), 1-11.

Swathi S., Lakshman K. (2022) Phytopharmacological and Biological Exertion of Spondias pinnata: (A Review). *Oriental Journal Of Chemistry*, 38(2), 271

Widiasih D.A., Drastini Y., Yudhabuntara D., Maya F.L.R.D., Sivalingham P.L., Susetya H., Sumiarto. (2019) Detection of antibiotic residues in chicken meat and eggs from traditional markets at Yogyakarta City using bioassay method. *B. Acta Veterinaria Indonesiana*, 1-6.

Xu L., Zhang H., Xiong P., Zhu Q., Liao C., Jiang G (2021). Occurrence, fate, and risk assessment of typical tetracycline antibiotics in the aquatic environment: A review. Science Of The Total Environment, 753.

DOI:https//doi.org/10.22452/mjs.vol45no2.3
Malaysian Journal of Science 45(2): 23-32 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 45(2): 33-41 (June 2026) https://mjs.um.edu.my

Cocoa Agro-Industry Development Based on Pyrolysis Technology with Cocoa Liquid Smoke, Biochar and Bio-Oil Product

Mohammad Wijaya[1a]*, Muhammad Wiharto[2b], Muhammad Anwar[3a] and Vika Puji Cahyani[4a]

Abstract: This study aims to analyze the utilization of cocoa husk through fermentation and pyrolysis processes into three main derivative products, namely liquid smoke, charcoal (biochar), and bio-oil. Cocoa husk was collected from cocoa fruit in Central Mamuju Regency, West Sulawesi Province. Before the pyrolysis process, the cocoa husk was fermented to reduce water content, increase homogeneity, and reduce interfering compounds such as complex lignin. Fermentation also increases pyrolysis efficiency by opening the cellulose and hemicellulose structures. After fermentation, the cocoa husk was dried and fed into a pyrolysis reactor with combustion temperatures of 112, 212, 312, 412, and 512°C for 5 hours. During the pyrolysis process, the cocoa husk decomposed into three main phases: the solid phase produced biochar, the liquid phase produced bio-oil, and the gas formed during pyrolysis was condensed into liquid smoke. The experimental results showed that temperature plays an important role in determining the amount of liquid smoke product produced. A temperature of 212°C is considered the optimum temperature to obtain the highest liquid smoke yield. This process not only increases the efficiency of cocoa shell waste utilization, but also produces products that can be applied in various industries, such as filtration, air treatment, alternative energy, and preservation. SEM and FTIR analysis show the potential and value of the resulting products in accordance with the principles of sustainable agro-industrial development and circular bioeconomy.

Keywords: *Agro-industry, cocoa husk, cocoa pod, liquid smoke, pyrolysis.*

1. Introduction

Indonesia, as the third-largest cocoa producer in the world after Ghana and Côte d'Ivoire, plays a vital role in the global cocoa market and significantly contributes to the national economy by providing employment, income, and foreign exchange (Fahmid et al., 2022; Muslimin et al., 2017). Nevertheless, challenges such as low-quality cocoa due to pest and disease infestations and limited processing technology hinder the industry's progress (Cilas & Bastide, 2020). A more innovative approach in estate development that emphasizes productivity, efficiency, and competitiveness is thus essential (Aneani et al., 2017). Despite its potential, Indonesia's cocoa export performance still lags behind other producing countries, indicating lower comparative competitiveness (Fahmid et al., 2022). To overcome this, strategies such as replanting, improving fermentation techniques, enhancing seed quality, using pest-resistant varieties, and expanding downstream industries must be adopted to boost product quality, increase added value, and improve farmers' incomes (Feby et al., 2021; Zulfiandri, 2023).

A thermochemical process called pyrolysis uses high temperatures and no oxygen to break down organic materials. Cocoa bean and pod shells can be converted into more valuable products in the cocoa industry by the process of pyrolysis. Bioenergy in the form of bio-oil and bio-gas can be produced by pyrolyzing cocoa shells. Because of its high calorific value, the resulting bio-oil has the potential to replace fossil fuels (Akinola et al., 2018; Mumbach et al., 2022). Moreover, pyrolysis can also produce biochar which can be used as an alternative fuel or fertiliser (Tsai et al., 2018). Moreover, pyrolysis can also produce liquid smoke that can be used in the food and cosmetic industries (Handojo et al., 2020; Wijaya et al., 2019). Although pyrolysis offers many benefits, there are some challenges that need to be overcome. One of them is the optimisation of pyrolysis conditions to maximise the yield of the desired product. Research shows that the right temperature and heating rate are crucial to improve the quality of pyrolysis products. The study by Handojo et al. (2020) has proven that cocoa bean shells can be utilized through a pyrolysis process into liquid smoke containing bioactive compounds with various benefits. This finding opens a great opportunity to develop a cocoa agro-industry model based on pyrolysis technology as a strategy for product diversification and increasing added value. However, until now there has not been much research to develop an integrated agro-industrial approach with pyrolysis technology on an applicable scale, especially with the orientation of sustainable cocoa derivative product development (Handojo et al., 2020).

Authors information:
[a]Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Indonesia. E-mail: wijasumi@unm.ac.id[1]; m.anwar@unm.ac.id[3]; vika.puji.cahyani@unm.ac.id[4]
[b]Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Indonesia. E-mail: wiharto@unm.ac.id[2]
*Corresponding Author: wijasumi@unm.ac.id

Received: January, 2025
Accepted: May, 2025
Published: June, 2026

DOI:https//doi.org/10.22452/mjs.vol45no2.4
Malaysian Journal of Science 45(2): 33-41 (June 2026)

This study aims to analyse the utilisation of cocoa shells through fermentation and pyrolysis processes into three main derivative products, namely liquid smoke, charcoal (biochar), and bio-oil. This study presents an innovative approach to cocoa shell waste-based agro-industrial development through the integration of fermentation and pyrolysis processes to produce three economically valuable products: liquid smoke, biochar and bio-oil. This approach differs from previous studies that only focused on one processing method or one type of product. For example, the study by Villasanaet al. (2023) only explored the value enhancement of cocoa shells through thermochemical pyrolysis and catalytic upgrading to produce volatile compounds, without the involvement of fermentation or integrated multi-product production (Villasana et al., 2023). Meanwhile, the study by Wijaya et al. (2019) did produce liquid smoke and charcoal from cocoa waste, but the approach used is general and does not prioritize agro-industry development systematically and has not touched on detailed product analysis such as charcoal porosity (Wijaya et al., 2019).

In addition, this study also examined the pore structure of cocoa shell pyrolyzed charcoal as part of the product quality evaluation. This focus has not been found in many similar studies, where generally biochar from other biomass such as apricot or heavy oil is studied in terms of quality improvement through co-pyrolysis without specific exploration of functional materials (Zhuang et al., 2022). This integrated approach provides significant added value not only in the aspect of biomass waste utilization technology, but also in the context of local economic sustainability based on agricultural products such as cocoa. Thus, this article makes a novel contribution to the development of a sustainable cocoa agro-industry through integrated technology engineering.

2. Methods

Cocoa shells were collected from cocoa pods in Central Mamuju Regency, West Sulawesi Province. The cocoa beans were then fermented by adding media and the skins were dried until the moisture content reached 10-20% (w/b) according to Indonesian national standards. Prior to the pyrolysis process, cocoa pods are fermented to reduce moisture content, improve homogeneity, and reduce the content of interfering compounds such as complex lignin (Chen et al., 2019; S. Wang et al., 2017). Fermentation can also improve pyrolysis efficiency by opening up the structure of cellulose and hemicellulose. After fermentation, the cocoa pods are dried and put into the pyrolysis reactor with the combustion temperatures used are 112, 212, 312, 412, and 512°C with a combustion time of 5 hours (Handojo et al., 2020; Kouadio et al., 2019). During this process, the cocoa shells will decompose into three main phases, namely 1) the solid phase produces biochar (charcoal) in the form of solid carbon residue which is rich in pores and has potential as an adsorbent, fertiliser, or active ingredient. 2) The liquid phase produces bio-oil, a complex mixture of organic compounds that can be used as fuel or a source of bioactive chemical compounds. 3) The gas phase, where some of the pyrolysis gas is condensed into liquid smoke, contains phenolic, carbonyl, and acidic compounds that are antimicrobial and used as natural preservatives or disinfectants (Anca-Couce, 2016; Moldoveanu, 2019).

Important aspects in this process are temperature regulation, residence time, and gas condensation system. The gas formed during pyrolysis will be directed to the cooling system to produce liquid smoke through the condensation process, while the solid residue in the reactor becomes biochar, and the other non-condensate liquid fraction is collected as bio-oil (Song et al., 2019; Q. Wang et al., 2020; Zong et al., 2020). The cocoa shell charcoal from the pyrolysis process was then analysed by SEM to see the morphological structure, and then the liquid smoke product was analysed using Fourier Transform Infrared (FTIR) analysis to characterise the chemical compounds. The analytical method was chosen because each molecule has a unique spectral 'fingerprint', allowing identification of organic and inorganic compounds with high accuracy. In addition, it is very useful to determine the functional groups in the molecule, such as hydroxyl, carbonyl, or amine. This information is important in understanding the chemical structure of compounds and their reactivity (Du et al., 2014; Mumbach et al., 2022). Through this approach, previously unutilised cocoa shell waste can be transformed into three products that not only have high economic value, but are also environmentally friendly, thus supporting the principles of circular bioeconomy and sustainable agro-industrial development.

3. Results and Discussion

Pyrolysis Process

Cocoa shells collected from Central Mamuju Regency, West Sulawesi Province, went through a series of stages involving fermentation and drying. Fermentation is carried out to reduce the moisture content of the cocoa shells, which are then dried to a moisture content of 10-20% (w/b), in accordance with Indonesian national standards. The purpose of fermentation is to improve the homogeneity of the cocoa shells and reduce the content of interfering compounds such as complex lignin. The fermentation process is also effective in opening up the structure of cellulose and hemicellulose, thus increasing the efficiency of subsequent pyrolysis. After the fermentation and drying process, the cocoa pods were put into the pyrolysis reactor with various combustion temperatures, namely 112°C, 212°C, 312°C, 412°C, and 512°C, with a combustion time of 5 hours. The results of pyrolysis of cocoa pod shells with a moisture content of 9.83%, while cocoa pods of Central Mamuju Regency with lignin content for cocoa pod shells 50.34%, alpha cellulose 28.82 and hemicellulose 1.25%. Lignin content depends on different types of raw materials and burner temperature conditions. Lignin does not have repeat units like hemicellulose and cellulose, but consists of complex phenolate units. Lignin is an amorphous copolymer complex consisting of guaiacyl (G), p-hydroxyphenyl (H), and syringyl (S) units that interconnect as éter bonds and carbon-carbon linkages (D. Shen et al., 2015). During the pyrolysis process, the cocoa shells decompose into three main phases:

DOI:https//doi.org/10.22452/mjs.vol45no2.4
Malaysian Journal of Science 45(2): 33-41 (June 2026)

1. Solid Phase: This phase produces biochar or charcoal that is rich in pores. This biochar has potential as an adsorbent, fertiliser, or active ingredient for various industrial applications. The biochar results show success in increasing sorption capacity, which can be utilised in water or air treatment.
2. Liquid Phase: The liquid phase produces bio-oil, which is a complex mixture of organic compounds. These bio-oils have potential as alternative fuels or sources of bioactive chemical compounds that can be used in various chemical and energy industries (Mohanty et al., 2024).
3. Gas Phase: The gas formed during pyrolysis is condensed into liquid smoke, which contains phenolic, carbonyl, and acidic compounds. These compounds have antimicrobial properties and can be used as natural preservatives or disinfectants, further adding to the economic value of the resulting product. Due to the potential advantages of these substances, cocoa liquid smoke has significant commercial value. Cocoa liquid smoke is also known to have strong antioxidant properties. According to research, liquid smoke made from cocoa shells having a moisture content between 10% and 25% showed strong antioxidant activity. Cocoa liquid smoke has the potential to be used as a natural food preservative due to its antioxidant properties (Budaraga et al., 2019).

Setting the combustion temperature is a key aspect in the pyrolysis process. The higher the temperature used, the greater the amount of biochar produced, but with a decrease in bio-oil and gas content. In addition, the combustion time set for 5 hours has provided optimal results in producing the desired product fraction. The gas condensation process is also very important, where the pyrolysis gas formed is directed to the cooling system to produce liquid smoke rich in chemical compounds with antimicrobial activity. This stage was conducted to determine the effect of temperature variation on the weight and yield (%) of liquid smoke produced from the dry distillation process of cocoa pod skin. The experimental results show that temperature plays an important role in determining the amount of liquid smoke product produced.

Table 1. Yield of Cocoa Pod Liquid Smoke

Temperature (°C)	Cocoa Liquid Smoke	
	Weight (g)	Yield (%)
112	28	3.17
212	177	20.05
312	98.5	11.15
412	96	10.87
512	36	4.07

To calculate the yield (%) of the experiment as in the table above, the formula commonly used in biomass distillation or pyrolysis processes is:

$$Yield\ (\%) = \left(\frac{Weight\ of\ yield\ (gram)}{Weight\ of\ starting\ material\ (gram)}\right) \times 100\%$$

Steps to Calculate Yield (%) is by weighing the initial weight of raw materials in the form of dried cocoa peels as much as 883 grams. Then pyrolysis/destillation is carried out at a certain temperature. Then collect the condensation results in the form of liquid smoke and weigh the weight using the formula above to calculate the yield. The results of dry distillation of cocoa pod samples showed that increasing the temperature had a significant effect on the weight and yield of liquid smoke produced. The results showed that dry distillation temperature has a significant effect on the weight and percentage yield of liquid smoke produced from cocoa pods. At 112°C, the yield produced was 3.17% (28 g), while at 212°C there was a drastic increase to reach 20.05% (177 g). This shows that 212°C is the optimum point in the dry distillation process to produce liquid smoke from these raw materials. After the temperature increased past the optimum point, the yield decreased gradually. At 312°C and 412°C, the yield dropped to 11.15% and 10.87%, respectively. Another significant decrease occurred at 512°C, where the yield was only 4.07% (36 g). This decrease in yield is most likely due to the thermal degradation of volatile compounds and the formation of products that do not re-condense into liquid smoke. In general, the trend of the data forms a downward open parabolic curve, which can be represented by a quadratic regression model. Quadratic regression provides a good predictive approach to determine the ideal temperature in cocoa biomass-based liquid smoke production. This model shows that there is a non-linear relationship between temperature and yield, where there is a maximum point at a certain temperature.

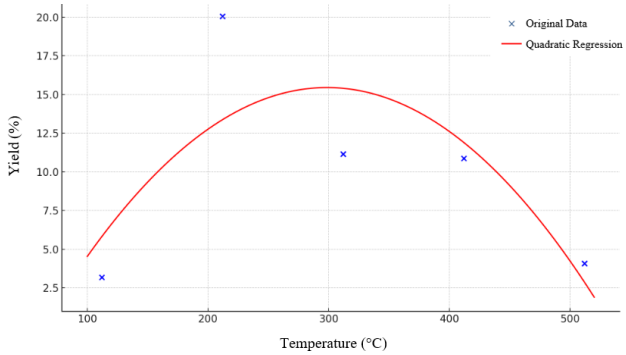


Figure 2. Quadratic regression model

The quadratic regression model obtained is:
$Y = -0.00027671T2+0.16529T-9.2376$

where:
Y = yield (%)
T = temperature (°C)

The analysis showed that the relationship between heating temperature and yield (%) of liquid smoke followed a quadratic non-linear pattern. The yield of liquid smoke increased

DOI:https//doi.org/10.22452/mjs.vol45no2.4
Malaysian Journal of Science 45(2): 33-41 (June 2026)

significantly until it peaked at 212°C (20.05%), then decreased at higher temperatures. This can be explained from a thermal point of view. At low temperatures (112°C) very little product is produced, the decomposition of organic matter is not optimal, resulting in only 3.17% yield of liquid smoke. The temperature is not enough to break down complex compounds such as lignin, cellulose, and hemicellulose efficiently (Kawamoto, 2007). At 212°C, there is maximum decomposition of volatile and semi-volatile compounds, especially lignin and cellulose, which are the main components of cocoa biomass. At this temperature, the thermal decomposition of lignocellulose reaches its peak, producing the maximum amount of liquid smoke conductivity (Bridgwater, 2012). This explains the highest yield values. After 212°C, further degradation or even carbonisation occurs, so the volatile compounds turn into gas or solids (charcoal) instead of liquid smoke. The yield drops dramatically, to only 4.07% at 512°C. This can be explained scientifically that at high temperatures, the volatile compounds previously formed undergo further degradation into gases such as carbon monoxide (CO), carbon dioxide (CO_2), and methane (CH_4), which cannot be condensed into liquid form. In addition, at high temperatures, phenolic compounds tend to polymerise to form tar, or even decompose completely so that they no longer contribute to the formation of liquid smoke.

The dominant carbonisation process at high temperatures also causes most of the organic components to turn into solid residues in the form of char. Therefore, the higher the distillation temperature, the lower the liquid fraction that can be obtained, as most of the product turns into the gas or solid phase. This is in line with previous research which states that excessive heating in the pyrolysis or dry distillation process can cause loss of volatile compounds due to thermal decomposition (Liu et al., 2022; Oginni & Singh, 2020; Rathika et al., 2024). Thus, a temperature of 212°C can be considered as the optimum temperature in the dry distillation process of cocoa pods to obtain the highest yield of liquid smoke. Temperatures above this value tend to be inefficient in liquid smoke production and produce more gas and solid residue. Increasing pyrolysis temperature tends to reduce biochar yield but increase carbon content (Zhang et al., 2020). This finding is in line with previous literature on pyrolysis of lignocellulosic biomass, where the middle temperature (200-250°C) is the optimal range for liquid smoke production with good quality and quantity (Bridgwater, 2012; Mohan et al., 2006).

FTIR Analysis

The next stage is analysis using Fourier Transform Infrared (FTIR). This analysis technique is used to identify functional groups in a material, which is very useful in the pyrolysis process. FTIR analysis can provide deep insight into the chemical and structural changes of the material that occur during the process. With FTIR, researchers can ensure that pyrolysis products have the desired properties for specific applications, such as in the manufacture of activated carbon for supercapacitors or catalysts for liquid fuel production. FTIR analysis results for cocoa pods of Mamuju Tengah Regency can be seen in Figure 1.

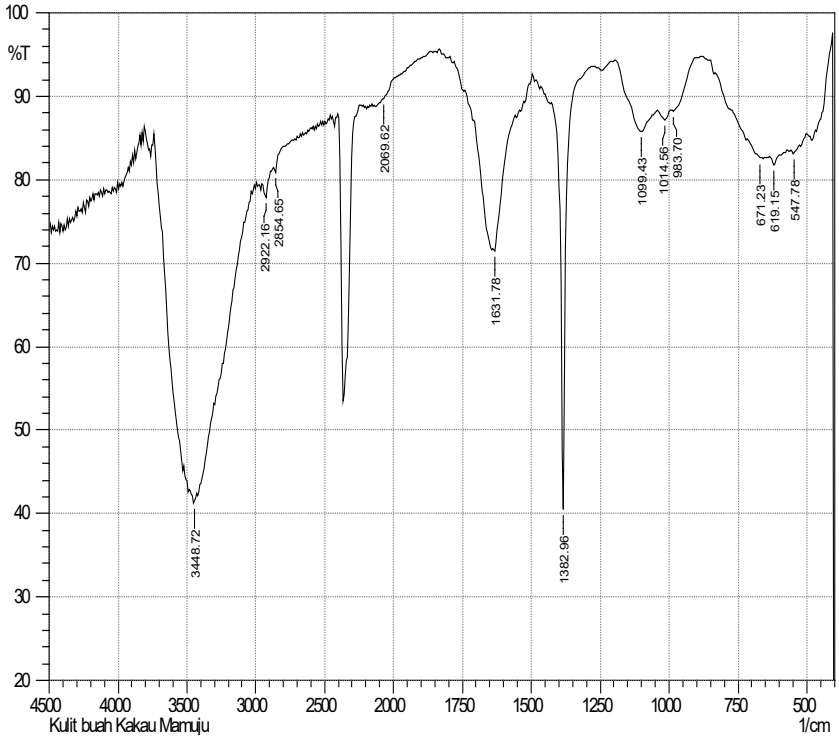


Figure 2. FTIR analysis for cocoa pod skin of Mamuju Tengah district

Based on the FTIR spectrum in Figure 1, significant peaks can be interpreted in the spectrum of cocoa pod skin (with absorbance numbers in cm^{-1}):

DOI:https//doi.org/10.22452/mjs.vol45no2.4
Malaysian Journal of Science 45(2): 33-41 (June 2026)

Table 2. Interpretation of FTIR Analysis of Cocoa Pod Peels

Wavelength (cm^{-1})	Interpretation
3448.72	Indicates the presence of hydroxyl (-OH) groups, usually associated with hydrogen bonded alcohol, phenol, or water compounds.
2924.65	This peak is generally attributed to the C-H stretch of aliphatic groups (CH_2 and CH_3), which are often found in lipid or fat compounds.
2089.62	These peaks can come from triple bond groups such as C≡C (alkyne) or C≡N (nitrile), although they tend to be smaller in intensity than other groups.
1631.78	Indicates C=O (carbonyl) stretch vibrations, which are often found in compounds such as carboxylic acids, esters, or amides.
1382.98	This peak can be attributed to C-H symmetrical stretching of the methyl group (CH_3) or deformation vibrations.
1009.43–1041.49	This range generally indicates C-O vibrations of alcohols, esters, or sugars. This is consistent with the presence of polysaccharides or lignocellulose-derived compounds.
671.23–647.78	Indicates aromatic C-H out-of-plane bending vibrations, which often occur in aromatic compounds such as lignin.

The spectrum in Figure 1 shows the presence of major organic compounds such as water (hydroxyl groups), lipids or fats (C-H aliphatic), polysaccharides (C-O), aromatic compounds or lignin (C-H aromatic). These FTIR spectra indicate that the cocoa pods have lignocellulose components, lipids, and possibly phenolic compounds that are relevant to their applications, such as for bioactive materials or biomaterials. These compounds could be used in various industrial applications, including cosmetics and food (Londoño-Larrea et al., 2022).

The next step is to conduct FTIR analysis for cocoa pod shell charcoal. Fourier Transform Infrared (FTIR) spectroscopy is an analytical technique used to identify and characterise chemical components in various materials, including charcoal. This technique utilises the infrared spectrum to detect various functional groups in the sample, which can provide important information about the chemical structure and physical properties of the charcoal. The results of FTIR analysis for cocoa pod shell charcoal can be seen in Figure 2.

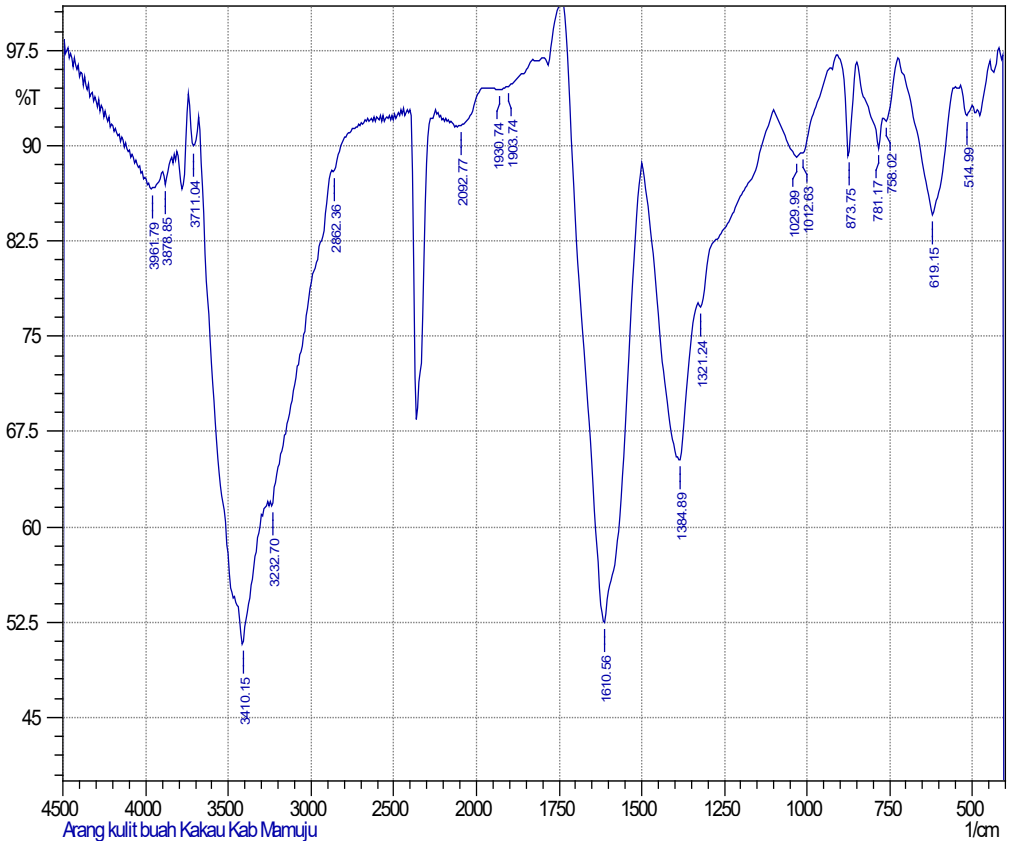


Figure 3. FT IR analysis for cocoa pod shell charcoal of Mamuju Tengah Regency

Based on the FTIR spectrum in Figure 2, the main peaks can be analysed. The analysis results can be seen in Table 3:

DOI:https//doi.org/10.22452/mjs.vol45no2.4
Malaysian Journal of Science 45(2): 33-41 (June 2026)

Table 3. Interpretation of FTIR Analysis of Cocoa Pod Peels

Wavelength (cm^{-1})	Interpretation
3870.78–3711.04	This broad peak indicates the presence of hydroxyl (-OH) groups, which are often indicated in alcohol or water compounds, although the intensity may decrease due to the carbonisation process.
3401.05	More pronounced -OH stretching vibrations, indicating adsorbed water or remaining phenolic groups. The hydroxyl group at 3401.05 cm^{-1} indicates that cocoa shell charcoal has the ability to adsorb water or polar compounds. This can be utilised for adsorption applications in water filtration or removal of organic pollutants such as dyes or heavy metals.
2922.70	Indicates aliphatic C-H stretch vibrations (CH_2 or CH_3), which tend to decrease in carbonised materials.
2022.77	This peak may indicate the presence of triple bonds (C≡C or C≡N) in small amounts.
1631.74	C=O (carbonyl) stretch vibrations or C=C asymmetric vibrations in aromatic compounds. These are common in carbon materials containing conjugated aromatic structures.
1384.90	This peak is related to the C-H deformation vibrations of the methyl group (CH_3) or aromatic structure.
1212.41	Indicates C-O vibrations of esters, phenols, or other compounds with oxygen groups.
1029.59	A more dominant C-O stretch vibration, indicating the possible presence of ether or alcohol groups.
871.75; 781.17; and 619.16	An aromatic C-H out-of-plane bending vibration, which confirms the presence of an aromatic structure or conjugated carbon compound.

The FTIR spectrum of the cocoa pod shell charcoal shows the presence of hydroxyl (-OH) and carbonyl (C=O) groups, although their intensity decreases due to carbonization (Omole et al., 2024). Aromatic structures are dominant, which is typical for charcoal with a high carbon content. Other oxygen compounds such as phenols or ethers are less likely (León et al., 2022; Wijaya et al., 2024). The carbonisation process appears to have removed most of the aliphatic groups, but left aromatic structures and conjugated carbon compounds, which are typical for activated carbon materials. The hydroxyl group at 3401.05 cm^{-1} indicates that cocoa shell charcoal has the ability to adsorb water or polar compounds. This can be utilised for adsorption applications in water filtration or removal of organic pollutants such as dyes or heavy metals. Vibrations at 1631.74 cm^{-1} (aromatic C=C) and peaks in the range 871.75-619.16 cm^{-1} indicate the presence of conjugated aromatic carbon structures. This structure is commonly found in activated carbon and provides good chemical properties, such as high adsorption capacity useful for adsorbing organic molecules, toxic gases, or vapours as well as thermal and chemical stability to make this charcoal suitable for applications at high temperatures or extreme environments. Vibrations at 1212.41 cm^{-1} (C-O) and 1631.74 cm^{-1} (C=O) indicate oxygenated functional groups, which are important for enhancing the hydrophilic properties and polarity of the charcoal (Jiang et al., 2023; C. Wang et al., 2021). This supports its application in the adsorption of certain chemicals, such as acid gases (H_2S, CO_2), and organic and inorganic pollutants in aqueous solutions (J. Shen et al., 2018; Tofighy & Mohammadi, 2022). The reduced intensity of the peak at 2922.70 cm^{-1} (C-H aliphatic) indicates the degradation of non-aromatic organic matter during carbonisation. This indicates the formation of a graphitic carbon structure, which improves electrical conductivity.

SEM Analysis

To analyse the morphology of the biochar produced from the solid phase, Scanning Electron Microscopy (SEM) analysis was conducted. By heating organic materials without oxygen, a thermochemical process known as pyrolysis converts them into charcoal. In this study, pyrolysis was used to make charcoal using cocoa pod shells (CPH) as raw material. Scanning electron microscope (SEM) analysis was required to understand the morphological structure of the resulting charcoal. Figure 4, 5 and 6 displays the findings from the SEM investigation.

DOI:https//doi.org/10.22452/mjs.vol45no2.4
Malaysian Journal of Science 45(2): 33-41 (June 2026)

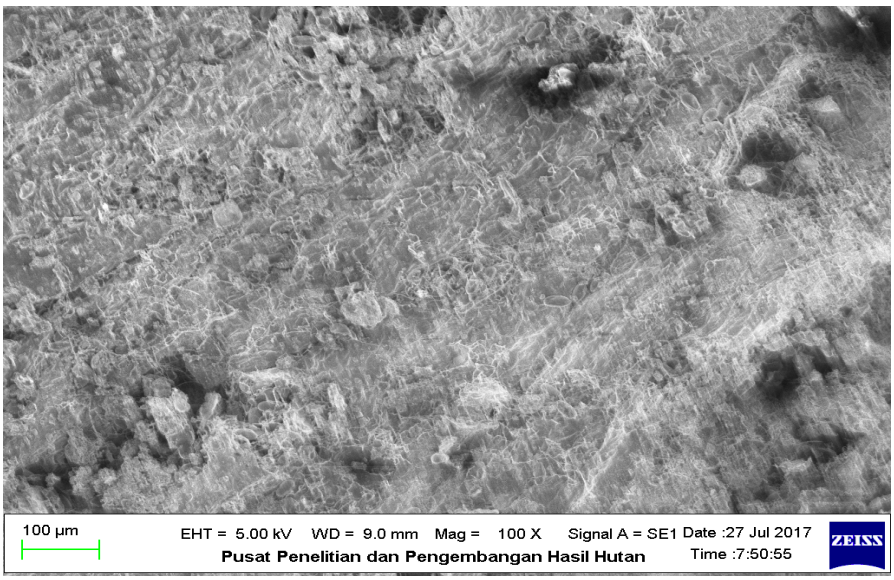


Figure 4. SEM analysis results for cocoa pod shell with 100x Magnification, 100 µm scale

Figure 4 shows a rough and irregular surface with visible structures resembling fibers and clusters of material. The surface texture indicates a relatively porous nature with some clear channel-like structures, which is typical for adsorbent materials. This structure is beneficial for applications in adsorption, such as water filtration or pollutant removal, due to the availability of voids for trapping particles. This morphology shows that the charcoal has an ample surface area, which is desirable for its adsorptive capacity, as more surface area allows for more contaminants to be adsorbed.

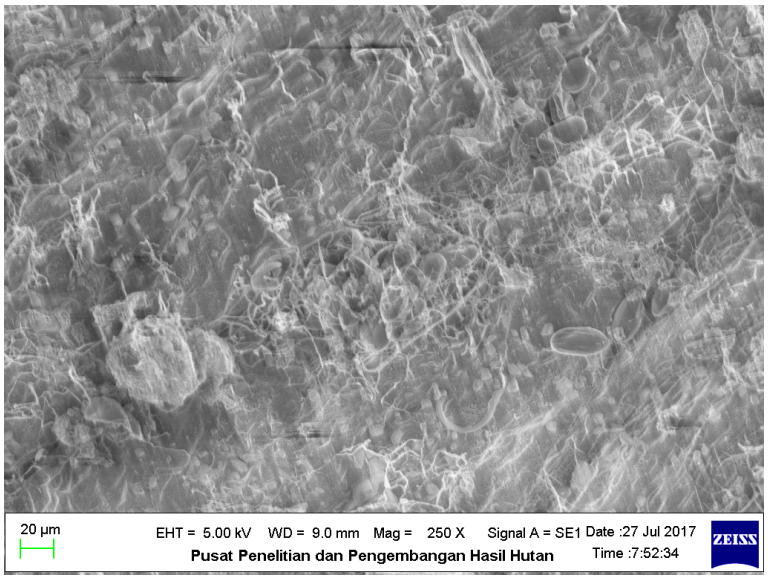


Figure 5. SEM analysis results for cocoa pod shell with 250x Magnification, 100 µm scale

Figure 5 shows a combination of smooth and rough areas, with structures resembling fibers and rounded aggregates. This suggests that the material has a hybrid structure, which could be beneficial for both adsorption and catalytic reactions. The larger structures in the image could enhance bulk adsorption properties, while the finer ones might be better at absorbing specific smaller molecules. The varied surface features suggest that the charcoal could be versatile in terms of the types of contaminants it can adsorb, ranging from larger organic compounds to smaller, more chemically active species.

DOI:https//doi.org/10.22452/mjs.vol45no2.4
Malaysian Journal of Science 45(2): 33-41 (June 2026)

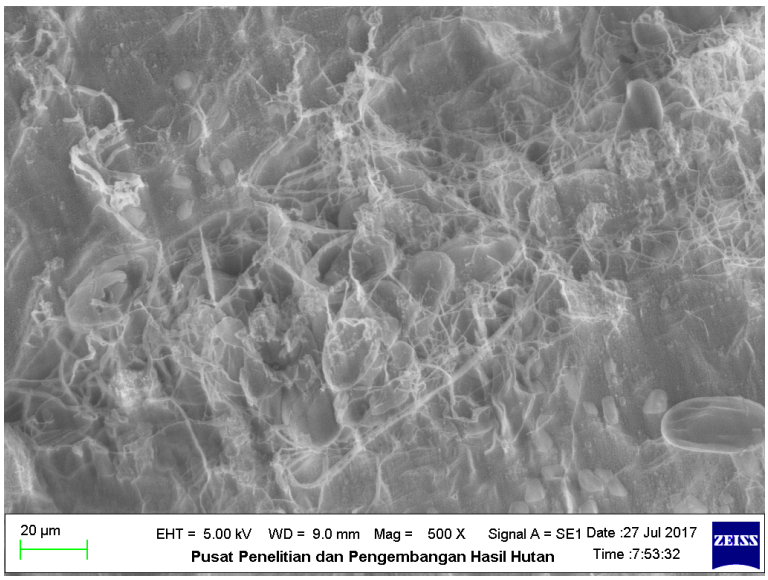


Figure 6. SEM analysis results for cocoa pod shell with 500x Magnification, 100 µm scale

At a higher magnification, the fine structure of the charcoal becomes more apparent. You can see more intricate network patterns and fibrous-like formations. The porosity is more pronounced here, with visible interconnecting pores that provide additional surface area for adsorption. These fine pores are especially suitable for micro and mesopore adsorption, such as absorbing organic molecules or small inorganic pollutants. The detailed structure observed at this level indicates that the charcoal can interact effectively with a variety of molecules, including organic contaminants and heavy metals, due to its high porosity.

SEM analyses of the cocoa pod charcoal shown in Figures 4, 5 and 6 provide valuable insights into its morphological structure and potential applications. SEM analysis reveals the small pore structure of the cocoa pod charcoal. This indicates that the charcoal has a highly porous surface, which is important for its sorption ability. Such small pores are particularly suitable for trapping and adsorbing inorganic and organic contaminants, which is beneficial for filtration applications. Additionally, high porosity is a key feature of charcoal, indicating that it has a large surface area available for absorption. This characteristic is particularly valuable in industries that require efficient filtration, such as water treatment and pollution control. The micro and mesoporous structure of charcoal, makes it effective in adsorbing organic and inorganic contaminants. This includes the removal of heavy metals such as Pb and Cd, as well as pollutants such as nitrates and phosphates. Additionally, charcoal can serve as an activated carbon medium to remove odour, colour, and organic contaminants from wastewater (Dolatkhah et al., 2022; Labied et al., 2022; Mopoung et al., 2014; Tuaprakone et al., 2011).

4. Conclusion

The cocoa shell pyrolysis process with prior fermentation and drying can produce three main products: biochar, bio-oil, and liquid smoke. This process not only increases the efficiency of using cocoa shell waste, but also produces products that can be applied in various industries, such as filtration, water treatment, alternative energy, and preservation.

212°C can be considered as the optimum temperature in the dry distillation process of cocoa pods to obtain the highest yield of liquid smoke. Temperatures above this value tend to be inefficient in liquid smoke production and produce more gas and solid residue. Increasing the pyrolysis temperature tends to decrease the biochar yield but increase the carbon content. SEM and FTIR analyses provide a clear picture of the potential and value of the resulting products, which is in line with the principles of sustainable agro-industrial development and circular bioeconomy.

5. Acknowledgements

The author would like to thank the Director General of Ditlitabmas Kemristek Dikti RI for the PSNI National Strategy Research Grant and the author expresses his gratitude and highest appreciation for the facilities and infrastructure in this research activity to Prof. (R). Dr. Gustan Pari, MS and Mr Dadang Setiawan, and the Department of Chemistry FMIPA Makassar State University.

7. References

Achlioptas, D., McSherry, F. (2001). Fast computation of low rank matrix approximations. *In Proceedings of the thirty-third annual ACM symposium on Theory of computing (STOC '01). Association for Computing Machinery*, New York, NY, USA, 611-618. https://doi.org/10.1145/380752.380858

Advani, R., O'Hagan, S. (2022). Efficient algorithms for constructing an interpolative decomposition. arXiv.

Aruchunan, E., Siri, Z., Noor Aziz, M. H. B., Ab Wahab, M. H. B., Muthuvalu, M. S., Sulaiman, J. (2022). Solution of peak junction temperature with Crank-Nicolson and SOR approach. *In Intelligent Systems Modeling and Simulation II: Machine*

DOI:https//doi.org/10.22452/mjs.vol45no2.4
Malaysian Journal of Science 45(2): 33-41 (June 2026)

Learning, Neural Networks, Efficient Numerical Algorithm and Statistical Models (pp. 225-234). Springer. https://doi.org/10.1007/978-3-031-04028-3_15

Bentbib, A. H., Kreit, K., Labaali, I. (2022). Randomized tensor singular value decomposition for multidimensional data compression. *In 2022 11th International Symposium on Signal, Image, Video and Communications (ISIVC)* (pp. 1-6). El Jadida, Morocco: IEEE.

Bhaskara, A., Lattanzi, S., Vassilvitskii, S., Zadimoghaddam, M. (2020). Residual based sampling for online low rank approximation. In *2020 Information Theory and Applications Workshop (ITA)*. San Diego, CA, USA: IEEE. 1-19. https://doi.org/10.1109/ITA50056.2020.9244974.

Damle, A., Lin, L., Ying, L. (2017). SCDM-k: Localized orbitals for solids via selected columns of the density matrix. *Journal of Computational Physics*, 334, 1-15. https://doi.org/10.1016/j.jcp.2016.12.053.

Golub, G. (1965). Numerical methods for solving linear least squares problems. *Numerische Mathematik*, 7(3), 206-216.

Halko, N., Martinsson, P. G., Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review,* 53(2), 217-288. https://doi.org/10.1137/090771806.

Kawamura, H., Suda, R. (2021). Least upper bound of truncation error of low-rank matrix approximation algorithm using QR decomposition with pivoting. *Japan Journal of Industrial and Applied Mathematics*, 38(3), 757-779. https://doi.org/10.1007/s13160-021-00459-x.

Khoei, T. T., Singh, A. (2024). Data reduction in big data: a survey of methods, challenges and future directions, *International Journal of Data Science and Analytics*, 2364-4168.

Libal, U., Baras, J. S., Johansson, K. H. (2020). Dimensionality reduction of volterra kernels by tensor decomposition using higher-order SVD. In *2020 59th IEEE Conference on Decision and Control (CDC)*. Jeju, Korea (South): IEEE. 5935-5941. https://doi.org/10.1109/CDC42340.2020.9303951.

Li, S., Lu, J., Hu, Y., Mattos, L. S., Li, Z. (2025). Towards scalable medical image compression using hybrid model analysis. *Journal of Big Data*, 12(45). https://doi.org/10.1186/s40537-025-01073-1

Liberty, E., Woolfe, F., Martinsson, P. G., Rokhlin, V., Tygert, M. (2007). Randomized algorithms for the low-rank approximation of matrices. *Proceedings of the National Academy of Sciences* 104(51), 20167-20172. https://doi.org/10.1073/pnas.0709640104(2007).

Liu, W., He, M. (2019). Accelerating solution of volume-surface integral equations with multiple right-hand sides by improved skeletonization techniques. *IEEE Antennas and Wireless Propagation Letters* 18(10), 2006-2010.

Lu, J., Ying, L. (2015). Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost. *Journal of Computational Physics* 302, 329-335. https://doi.org/10.1016/j.jcp.2015.09.014.

Mersereau, R. M. (1979). The processing of hexagonally sampled two-dimensional signals. *Proceedings of the IEEE* 67(6), 930-949. https://doi.org/10.1109/PROC.1979.11356.

Muravev, A., Tran, D. T., Iosifidis, A., Kiranyaz, S., Gabbouj, M. (2018). Acceleration approaches for big data analysis. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. Athens: IEEE, 311-315. https://doi.org/10.1109/ICIP.2018.8451082.

Parameshachari, B. D., Kumar, D. S., Prafulla, P. S., Yashwanth, J. (2023). Singular Value Decomposition (SVD) based optimal image compression technique. *In 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*. Bengaluru, India: IEEE, 1-6.

Strang, G., 2006. Linear algebra and its applications. 4th ed. Boston: Cengage Learning.

Su, Q., Wang, G., Zhang, X., Lv, G., Chen, B. (2018). A new algorithm of blind color image watermarking based on LU decomposition. *Multidimensional Systems and Signal Processing*, 29(3), 1055-1074. https://doi.org/10.1007/s11045-017-0487-7.

Tang, W. K. A., Ng, W. S., Liew, H. H. (2023). Separation of two musical instruments using matrix factorisation techniques. *International Journal of Applied Mathematics* 36(3), 425. http://dx.doi.org/10.12732/ijam.v36i3.8.

Varghese, P., Saroja, G. A. S. (2021). Hexagonal image compression using singular value decomposition in Python. In *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)* IEEE. 211-215. https://doi.org/10.1109/ACCESS51619.2021.9563312.

Wang, Z., Simoncelli, E. P., Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. *In The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers 2003*, 2, 1398-1402. IEEE. https://doi.org/10.1109/ACSSC.2003.1292216.

Wüthrich, C. A., Stucki, P. (1991). An algorithmic comparison between square-and hexagonal-based grids. *CVGIP: Graphical Models and Image Processing*, 53(4), 324-339. https://doi.org/10.1016/1049-9652(91)90036-J.

DOI:https//doi.org/10.22452/mjs.vol45no2.4
Malaysian Journal of Science 45(2): 33-41 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 45(2): 42-54 (June 2026) https://mjs.um.edu.my

Utilization of Landsat-8 Image Classification Results with the Random Forest Algorithm to Determine Groundwater Recharge Areas (Case Study: Singgahan and Montong Districts, Tuban Regency)

Moh. Singgih Purwanto[1ab], Adi Susilo[2a*], Agus Naba[3a], Ayi Syaeful Bahri[4b], Muhammad Ari Sanggam[5c], and Noorlaila Hayati[6c]

Abstract: Groundwater is highly beneficial for society, and fulfills essential daily living needs and activities. With increasing population growth, groundwater use has increased, necessitating the creation of a map depicting the distribution of groundwater recharge areas. This map is expected to aid the government and communities to identify all the locations of groundwater distribution and ensure the sustainability of groundwater discharge in the area. The method employed involves using Landsat-8 image classification results to produce land cover maps of the study area. Land cover data processing was conducted using the open platform Google Earth Engine with a Random Forest algorithm. The classification of Landsat-8 images resulted in six land cover classes with an overall accuracy of 99.60% and a Kappa value of 0.994. Overlay and weighting were performed using parameters that determine the distribution of groundwater recharge areas. The parameters utilized included supporting data such as rainfall, soil type, and slope, which were obtained from DEMNAS data. The overlay and weighting results indicate that the potential recharge area covered 13,010.33 hectares (59.21%), the transition zone covered 8,935.27 hectares (40.66%), and the discharge area covered 28.64 hectares (0.13%).

Keywords: *Landsat-8, image classification, random forest, groundwater recharge area.*

1. Introduction

Groundwater, which is also known as subsurface water, is a natural resource that is crucial for humanity. According to Law Number 7 of 2004 on Water Resources, Chapter 1, Article 1, Paragraph 4 emphasizes that groundwater is defined as water contained in layers of soil or rocks below the surface of the land (Undang-Undang no 7, 2004). The benefits of groundwater are significant for society, as it meets essential daily living needs. With an increasing population growth, groundwater use has risen accordingly. Similar to other natural resources, groundwater is being exploited at an increasing rate worldwide. It is generally preferred as a source of water for household purposes because its availability is relatively constant and it is more naturally protected from pollution. Therefore, information regarding the distribution of water recharge areas in a region, such as Tuban Regency, is essential.

Tuban Regency is located in East Java, where land dips into bowls, then rises again into rocky limestone stretches that run sideways across the map. Based on data from the Meteorology, Climatology, and Geophysics Agency in 2021, rain falls here most months, around 1100 to 1500 millimeters each year. That same report counted roughly 90 to 120 wet days annually. Water bubbles out of the ground in places like Singgahan and Montong, spots home to natural flows such as Krawak spring. Bowl-shaped terrain mixed with porous rock makes seepage likely, feeding those outlets. Additionally, there are springs in several sub-districts within this district, such as the Krawak spring located between the Singgahan and Montong sub-districts. There has been no previous research in this area, and the presence of springs, supported by the basin topography and karst areas, justifies the need for research on the recharge areas in these two sub-districts. To determine the potential recharge areas, spatial analysis is required. One of the accurate methods for conducting spatial analysis is Remote sensing, which is a long-established and effective tool for land cover monitoring, due to its ability to quickly, widely, precisely, and easily provide information about spatial variability on the Earth's surface (Shin et al., 2022). Landsat-8 pictures step into the role perfectly for sorting ground types as they carry red, green, blue eyes, plus near and far infrared sight, each tuned to catch how surfaces bounce back light. Because certain bands react strongly to what covers the land, spotting differences becomes possible (Zulfajri et al., 2021).

There are several previous studies related to the research being conducted, including the work of Purwanto et al. (2022). Their approach pulled in Landsat-8 OLI/TIRS satellite images along with factors like rainfall, soil type, land cover, and slope. They used a points-based system to analyze land cover in the study area of Tanah Datar Regency, North Sumatra. The study concluded that the use of rainfall data, slope, land cover, and soil type can effectively indicate recharge areas. On investigating surface

Authors information:
[a]Brawijaya University, St. Veteran Malang, Lowokwaru, Malang, Indonesia. E-mail: singgih16@student.ub.ac.id[1]; adisusilo@ub.ac.id[2]; anaba@ub.ac.id[3]
[b]Geophysics Engineering Department, Institut Teknologi Sepuluh Nopember, St. Teknik Kimia, Sukolilo, Surabaya, Indonesia. E-mail: singgih@geofisika.its.ac.id[1]; syaeful_b@geofisika.its.ac.id[4]
[c]Geomatics Engineering Department, Institut Teknologi Sepuluh Nopember, St. Teknik Kimia, Sukolilo, Surabaya, Indonesia. E-mail: sanggamari@geofisika.its.ac.id[5]; noorlaila@geodesy.its.ac.id[6]
*Corresponding Author: adisusilo@ub.ac.id

Received: June, 2024
Accepted: November, 2024
Published: June, 2026

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

plants, some kinds stood out that allowed water to seep down into the dirt layers without much resistance. Soil type also has an impact due to its permeability, and slope data influences the formation of land where groundwater accumulates. However, rainfall data had a minor influence.

Furthermore, the study by Zulfajri et al. (2021) classified land cover using Landsat-8 OLI data in Pidie Regency, Aceh Province, with a Random Forest algorithm. The algorithm used 70% of the data for training and 30% for testing. The accuracy of the results was tested using the Confusion Matrix method. The study found that Landsat-8 OLI imagery is limited when cloud cover is present, affecting the quality and classification results. Therefore, careful attention is needed when selecting and determining the training data area. Instead of falling into typical traps with decision trees, the Random Forest method handled overfitting well. The results showed an accuracy of 89.53% and a kappa value of 0.91, indicating that the classification was reliable and accurate, making it useful as a reference for future research. Based on previous studies, this work differs from earlier findings as it dives into groundwater recharge zones by pairing Landsat-8 images with a Machine Learning approach, the Random Forest model, which is new in Singgahan and Montong, Tuban Regency. Additionally, the parameter data used in this study are the most recent, covering the year 2022 in the study area.

Based on the background and several previous studies, research will be conducted on the spatial analysis of groundwater recharge areas in Singgahan and Montong Districts, Tuban Regency. Satellite pictures from Landsat-8 form the base for sorting how land is covered there. Sorting these images uses Machine Learning, specifically Random Forest, to label surface types more accurately followed by a validation of the classification results against the actual physical conditions (ground truth) at the research location. Subsequently, overlay and weighting will be conducted using parameters that determine the distribution of groundwater recharge areas, including rainfall, soil type, and slope (Vasileva, 2019). It is hoped that this map of groundwater recharge area distribution will be useful for the government and local communities in identifying groundwater distribution locations, especially in Singgahan and Montong Districts, Tuban Regency.

2. Research Description

Research Location

The research location for the assignment is in Singgahan District and Montong District, Tuban Regency, East Java Province, Indonesia. Astronomically, Singgahan District is located at coordinates 6° 55' - 7° 1' South Latitude and 111° 44' - 111° 48' East Longitude. Singgahan District covers an area of 79.05 km² and consists of 12 villages. Meanwhile, Montong District is located at coordinates 6° 55' - 7° 1' South Latitude and 111° 48' - 111° 55' East Longitude. Montong District covers an area of 147.98 km² and consists of 13 villages. Both sub-districts in the research area contain numerous water sources, such as Krawak, Nglirip Waterfall, and Nganget Hot Springs. However, no research or exploration regarding groundwater recharge has been conducted in these areas, and that gap sparked the motivation for this study.

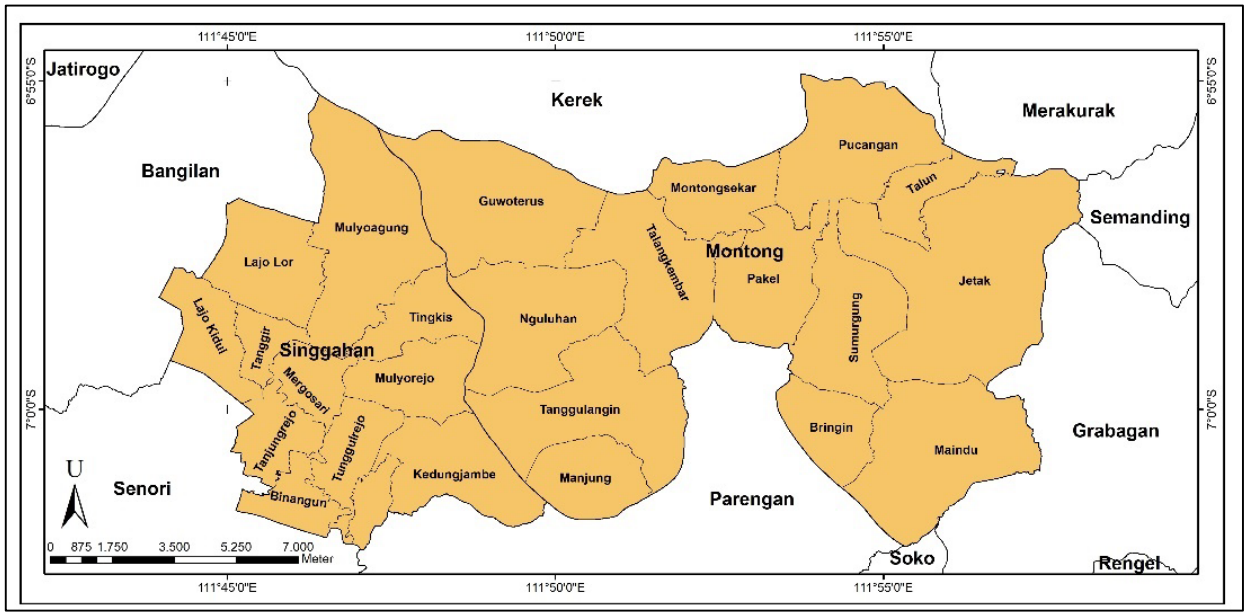


Figure 1. Research Location Map.

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

Data and Equipment

In conducting this research, the data used were as follows:

1. Landsat-8 level 2 tier 1 satellite image data at 30-meter spatial resolution, obtained from the Google Earth Engine platform, covers the whole research area with the date filters specific and cloud cover less than 40%, sourced from USGS and accessed through earthexplorer.usgs.gov.
2. Vector map of the administrative boundaries of Singgahan District and Montong District at a scale of 1:25,000, obtained from Inageoportal accessed through Ina-Geoportal (indonesia.go.id).
3. Data on rainfall levels for the Tuban Regency area, 2022, sourced from BMKG.
4. Soil type data using soil type shapefile data obtained from the Ministry of Environment and Forestry.
5. Slope data using Digital Elevation Model National data with a spatial resolution of 8 meters obtained from Inageoportal.

The equipment used in this research includes:

1. A laptop and mouse.
2. Google Earth Engine for satellite image processing; geotaging software on mobile phones for validation sampling points; ArcGIS for studying patterns, showing results, and arranging layouts; and Microsoft Office 365 for number and word processing.

Stages of Processing Land Cover Maps with Landsat-8 Image Data

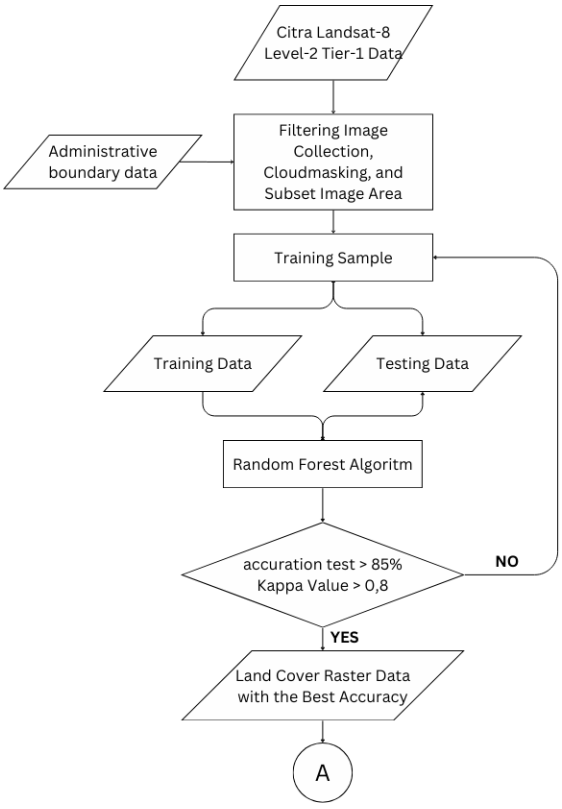


Figure 2. Flowchart of Land Cover Map Processing

The following is an explanation of the flow diagram above:

1. Create a script for accessing Landsat-8 Level 2 Tier 1 data by applying several filters, like time filters, cloud filters, and area filters. The selected image is cropped using the boundary of the study area extracted from InaGeoportal and given in shapefile format (subset data). Thus, an image without any cloud cover and adjusted to the study location area is obtained.
2. Create training samples by taking point samples in each land cover class, consisting of six land cover classes: water bodies, rice fields, settlements, moors, plantations, and dense forests.
3. Write a script to divide the training sample data into training data and testing data, with a ratio of 0.7 for training data to 0.3 for testing data.
4. Perform classification by creating a Random Forest algorithm script. In the classification process, band stacking is carried out by adding the Modified Normalized Difference Vegetation Index (MNDVI), Modified Normalized Difference Water Index (MNDWI), and Simple Ratio (SR) indices. These three indices are used to differentiate several land covers that will be classified. The MNDVI index is used to differentiate forest areas from grassland and agriculture (Wang & Fu, 2023). The MNDWI index is used to increase the contrast between water and built-up land, where water features will have increased values and built-up land will have decreased values (Ramji, 2022). The Simple Ratio (SR) is an index that has a good correlation with foliage density and is sensitive to three main external factors: sun and display geometry, soil background, and atmospheric effects (Melillos & Hadjimitsis, 2020).
5. Conduct an accuracy test after the classification results are available, displaying the accuracy value with Overall Accuracy (OA) and Kappa values. The classification accuracy is considered adequate if the OA value is >85% and the Kappa value is >0.80. If the classification accuracy is less than 85%, reclassification will be performed by improving the training area (Wiggers et al., 2020).
6. The result from processing the Landsat-8 satellite images is a raster map showing the land cover classification for the Singgahan District and Montong District in Tuban Regency.

Stages of Random Forest Processing

The land cover data processing in this research is done by applying the Random Forest algorithm. The outcome of the classifier, Random Forest, is chosen because it is a highly efficient and powerful classification method. Furthermore, it is capable of handling large training datasets as well as estimating missing data. The steps of utilizing Random Forest in this work are as follows:

1) The initial steps involve pre-processing as outlined below:
   a. Cloud masking: This accomplishes removing pixels covered by clouds or their shadows, through functions like cloudmask and maskL8sr. At this point, bitwise masking is done on the Cloud Shadow bit (bit 3) and Clouds (bit 5), both set to 0, meaning cloud-free pixels.

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

Only pixels where bits 3 and 5 are both 0 (clear condition) are kept.

b. Advanced masking: This step utilizes the QA_PIXEL band to define and identify the pixel quality condition information. The first five bits (0-4) are inspected, where bit 0 represents fill, bit 1 is dilated cloud, bit 2 is cirrus, bit 3 is cloud, and bit 4 is cloud shadow. The results are compared with 0, and clear pixels are retained. Saturation masking is also used to detect saturated pixels.

c. Radiometric correction: Applied to optical and thermal bands to convert digital values into representative physical values. For the optical bands of Landsat-8, a scale factor correction is applied by multiplying each pixel by 0.0000275 and subtracting 0.2. This converts Digital Number (DN) values into surface reflectance. For the thermal bands, pixels are multiplied by 0.00341802, and 149.0 is added to convert DN into surface temperature in Kelvin.

2) Adding vegetation and water indices.

a. NDVI (Normalized Difference Vegetation Index): Used to measure vegetation density in a region by utilizing the ratio between the Near-infrared (NIR) band 5 and red bands using band 4. The NDVI value ranges from -1 to 1, where values closer to 1 indicate dense vegetation, while values closer to -1 indicate areas without vegetation, such as water or bare land. The equation is:

$$NVDI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

b. MNDWI (Modified Normalized Difference Water Index): Used to detect surface water presence, such as lakes, rivers, or water bodies. This index compares the GREEN band and shortwave infrared (SWIR). In Landsat-8 images, GREEN corresponds to band 3, and SWIR corresponds to band 6. Positive MNDWI values indicate water presence, while negative values indicate dry land or vegetation:

$$MNDWI\ 0 = \frac{GREEN - SWIR}{GREEN + SWIR} \quad (2)$$

c. MNDVI (Modified Normalized Difference Vegetation Index): Measures vegetation with more smoothness, reducing the impact of reflected light variation. It works by averaging the visible bands, RED, GREEN, and BLUE, compared to NIR:

$$MNVDI = \frac{NIR - (\frac{RED + GREEN + BLUE}{3})}{NIR + (\frac{RED + GREEN + BLUE}{3})} \quad (3)$$

d. SR (Simple Ratio): A simpler vegetation index is calculated by the ratio of reflected NIR and RED light. This index signals the presence of vegetation where higher values characterize more green or denser vegetation.

$$SR = \frac{NIR}{RED} \quad (4)$$

3) Satellite image acquisition: Landsat-8 Level 2 satellite data from April 1, 2022, to October 31, 2023, have been acquired. Only scenes with a cloud cover of less than 40% that mainly cover the chosen Region of Interest (ROI) are kept after filtering the data.

4) Model development: A model is constructed based on the Bagging, or Bootstrap Aggregating, concept. The first step is to train a model with a satellite image sample data to predict the land cover classes using spectral indices and bands. The model is designed to produce 100 decision trees to stabilize and improve the accuracy of the prediction. 70% of the data is utilized for training purposes, and the remaining 30% is kept for testing. During this step, the decision trees are formed by the use of random data subsets. Each tree identifies different features to divide the data, for instance, classifying the areas as forest if NDVI > x. The classification output of each tree is aggregated through a voting method to lower bias.

5) Training and feature association: The model was trained with the given training data, in which the input features consist of spectral bands and vegetation indices. Then the model links these features to different land cover classes, such as forest, water bodies, etc.

6) Testing and accuracy evaluation: Model accuracy is evaluated using the testing data, and various accuracy matrices are computed, such as:

a. Overall Accuracy: The ratio of correct predictions to the total data.

b. User Accuracy: The percentage of correct predictions for each class.

c. Producer Accuracy: The percentage of true data correctly classified by the model.

d. Kappa Coefficient: Measures model accuracy compared to random classification.

7) Application of the model: The trained model is applied to the entire study area, predicting land cover classes for each pixel based on the spectral features of the image. The results are displayed as land cover classes with different colours.

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

Stages of Processing Parameters for Determining Groundwater Recharge Area

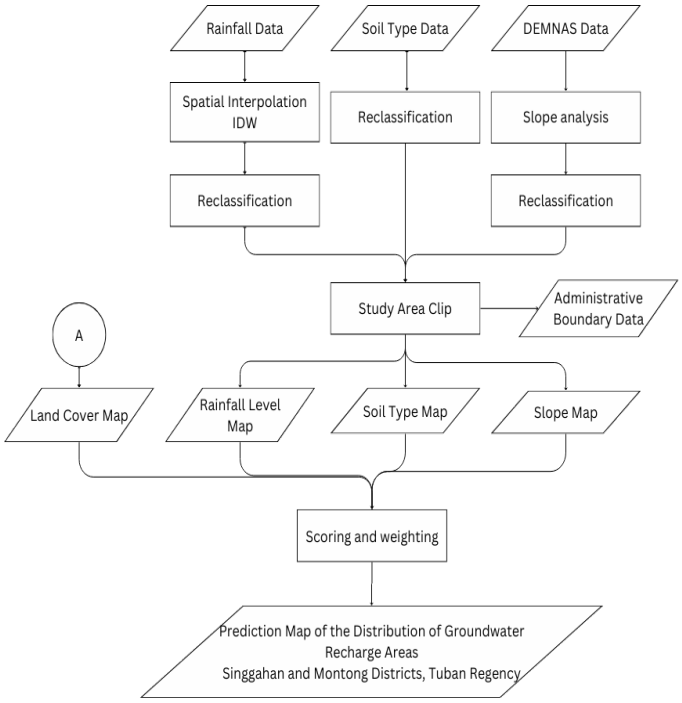


Figure 3. Parameter Data Flow Diagram

The following is an explanation of the following flow diagram:

1) Rainfall Data
   Rainfall data from several rain gauge stations in Tuban Regency are obtained from the official BMKG website. The station points undergo an Inverse Distance Weighted (IDW) interpolation process to estimate values for the surrounding areas. Reclassification is then conducted into three classes: <2000 mm/year, 2000-3000 mm/year, and >3000 mm/year.
2) Soil Type Data
   Soil type data was taken from the Ministry of Environment and Forestry, which provided, in the form of a shapefile, data depicting Indonesian Soil types. This data was then further processed by cropping using the area of the research location to generate a map of soil types of Singgahan and Montong Districts, Tuban Regency.
3) DEMNAS data
   Slope data is acquired from DEMNAS data with a spatial resolution of 8 meters. A slope analysis process is conducted to determine the slope values, followed by a reclassification process into five classes: <8% (flat), 8-15% (ramps), 15-25% (wavy), 25-40% (slightly steep), and >40% (steep).
4) Scoring and Weighting: Each zone is marked out based on the administrative boundaries of the study locale and subsequently changed from raster to vector data for the overlay operation. The 'Weight' column represents the extent of each parameter's contribution to the determination of the Groundwater Recharge Area, whereas the 'Value' column in the table represents the score given to each classification according to the parameters set. The score indicates the degree of probability that the area is a recharge zone. This research employs a scoring procedure to ascertain recharge areas, wherein every parameter is given a score and weighted differently according to its significance. The different layers are then combined, and the area with the highest value is considered the recharge area. There are four maps as parameters, and they are from the Indonesian Ministry of Forestry's 2009 regulations, with some changes. The classification of parameters of groundwater recharge areas, together with the scores and weights employed in this research, is consistent with these regulations as stated in Table 1 (Menkeu RI, 2009). However, the parameters for rainfall and soil type have been changed to be in line with the research area.

Table 1. Scoring and Weighting of Parameters Determining Groundwater Recharge Areas.

Parameter	Weight	Classification	Value
Land Cover	30%	Residential areas, rice fields	1
		Field, moor	2
		Shrubs, meadows	3
		Production forests, plantations	4
		Dense forest	5
Rainfall	15%	<2000	1
		2000 - 3000	2
		>3000	3
Soil Type	35%	Grummosol	1
		Litosol, Mediterran	2
		Latosol, Kambisol, Mollisol, Gleisol	3
		Aluvial	4
		Regosol, Rezina	5
Slope	20%	>40%	1
		25 – 40%	2

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

Parameter	Weight	Classification	Value
		15 – 25%	3
		8 – 15%	4
		<8%	5

The final weighting is the accumulated result of the weight percentage multiplied by the value of each parameter, which then produces a final value based on equation 5 below.

$$S_T = B_1\,(S_1) + B_2\,(S_2) + \cdots . + B_n\,(S_n) \quad (5)$$

S_T = Total score
$B_1, B_2, \ldots, B_n$ = Weight of each parameter
$S_1, S_2, \ldots, S_n$ = Score for each parameter
n = Number of parameters

If all data processing has been completed, the result will be a map showing the distribution of groundwater recharge areas in Singgahan and Montong Districts, Tuban Regency.

3. Result and Analysis

Land Cover

Land cover data processing based on Landsat-8 imagery is conducted using the Google Earth Engine (GEE) platform employing the Random Forest algorithm (Arikan, 2023). Subsequently, an accuracy test is performed using an error matrix (confusion matrix) to determine the accuracy level resulting from the Landsat-8 image classification process using the Random Forest algorithm. Figure 4 below shows the land cover map of Singgahan and Montong sub-districts resulting from the classification of Landsat-8 images by the Random Forest (RF) algorithm. Meanwhile, the area and percentage distribution of land cover in the two sub-districts are presented in Table 2.

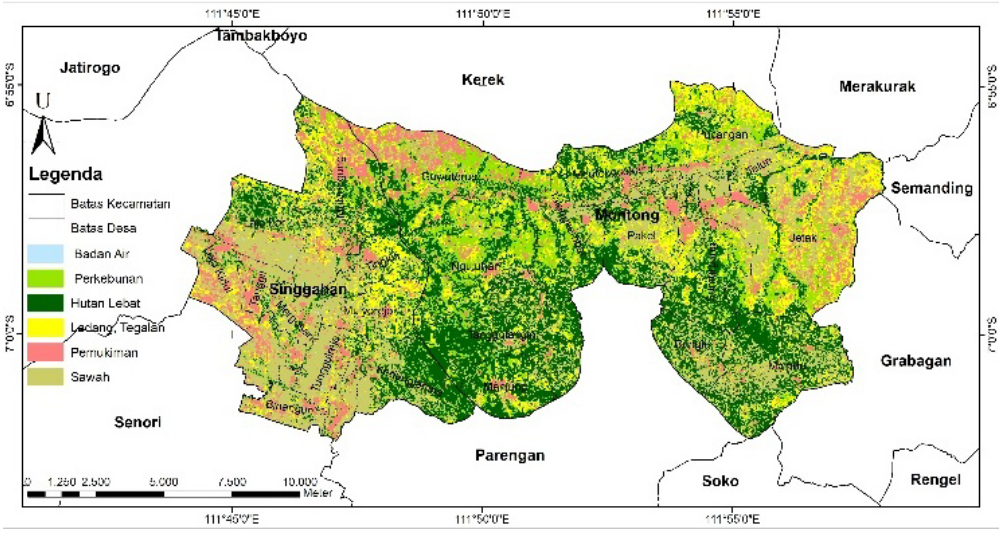


Figure 4. Land Cover Map Using the RF Algorithm for the Singgahan and Montong District Areas

Table 2 shows the area, percentage of land cover types, and the classification of values for each land cover type. A scoring range of 0 to 5 is used, depending on how well each land cover type can function as a recharge area or an area with high permeability. In this study, a score of 0 is assigned to the land cover parameter representing water bodies. This parameter is given a score of 0 because water bodies are areas where water cannot infiltrate the soil. Hence, a score of 0 in the context of this parameter is indicative of non-permeable soil in that area. Dense forest land cover is given a score of 5 because a dense forest is an excellent area for soil absorption. In this case, a plant root factor and plant root together with the soil in the forest area are loose so that the surface water can be absorbed well and retained underground. So, it is concluded that the dense forest area has great potential to become a recharge area. This entire classification is based on the Regulation of the Minister of Forestry of the Republic of Indonesia Number: P.32/MENHUT-II/2009. It was found that the land cover in Montong and Singgahan sub-districts is predominantly comprised of rice fields and dense forests, with percentages of 28.05% and 28.22%, respectively. The area covered by paddy fields is 6,159.37 hectares, while the area of dense forests is 6,196.65 hectares.

Table 2. Area and Land Cover Score Using the RF Algorithm in the Montong and Singgahan Districts

Land Cover	Area (Ha)	Percentage	Value
Water body	25,42	0,12%	0
Rice field	6159,37	28,05%	1
Settlement	2297,91	10,46%	2
Farm, Moor	3550,72	16,17%	3

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

Plantation	3729,84	16,98%	4
Dense Forest	6196,65	28,22%	5

Accuracy testing of the land cover results was conducted by splitting the data into two sets: 70% of the data was used for training, and 30% was used for testing. Calculations were then performed using a confusion matrix, which was constructed to estimate the values for producer's accuracy, user's accuracy, and overall accuracy (Wiggers et al., 2020). By using the training data and Google Earth Engine software, the following confusion matrix results were obtained:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 188 & 0 & 0 & 0 & 0 \\ 0 & 0 & 111 & 0 & 0 & 0 \\ 0 & 1 & 1 & 159 & 1 & 2 \\ 0 & 0 & 0 & 0 & 257 & 4 \\ 0 & 0 & 0 & 0 & 0 & 531 \end{bmatrix}$$

Figure 5. Confusion Matrix for Land Cover Classification Using Random Forest Algorithm

Based on the confusion matrix results, calculations for overall accuracy and the Kappa value were performed to assess the accuracy of the land cover classification. The classification accuracy standards used are as follows: (1) Adequate if the OA value is 85-89%; (2) Moderate if the OA value is 90-94%; (3) High if the OA value is >95% (Zulfajri et al., 2021). If the algorithm achieves an overall accuracy (OA) value >85% and a Kappa value >0.80, it is considered to have high accuracy, and the data processing can proceed. The calculations yielded an Overall Accuracy (OA) value of 99.60% and a Kappa value of 0.994. The results of the classification accuracy test are presented in Table 3.

Table 3. Image Classification Accuracy Test Results with RF

Land Cover	User's Accuracy	Producer's Accuracy
Water body	1,000	1,000
Rice field	0,994	1,000
Settlement	1,000	1,000
Farm, Moor	0,994	1,000
Plantation	1,000	0,989
Dense Forest	0,994	0,996
Overall Accuracy	0,996	
Kappa	0,994	

Therefore, the land cover map resulting from image classification using the RF algorithm was selected as a parameter in determining groundwater recharge areas. The land cover parameter holds a weight of 30% with five classification classes as outlined in Table 2.

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

Rainfall

Rainfall parameters were compiled based on the average rainfall data for 2021-2022 collected from 15 rainfed stations around the research area, sourced from BMKG. The rainfall parameters were classified into three classes: <2000 mm/year, 2000-3000 mm/year, and >3000 mm/year, with a weight of 15% to determine groundwater recharge areas.

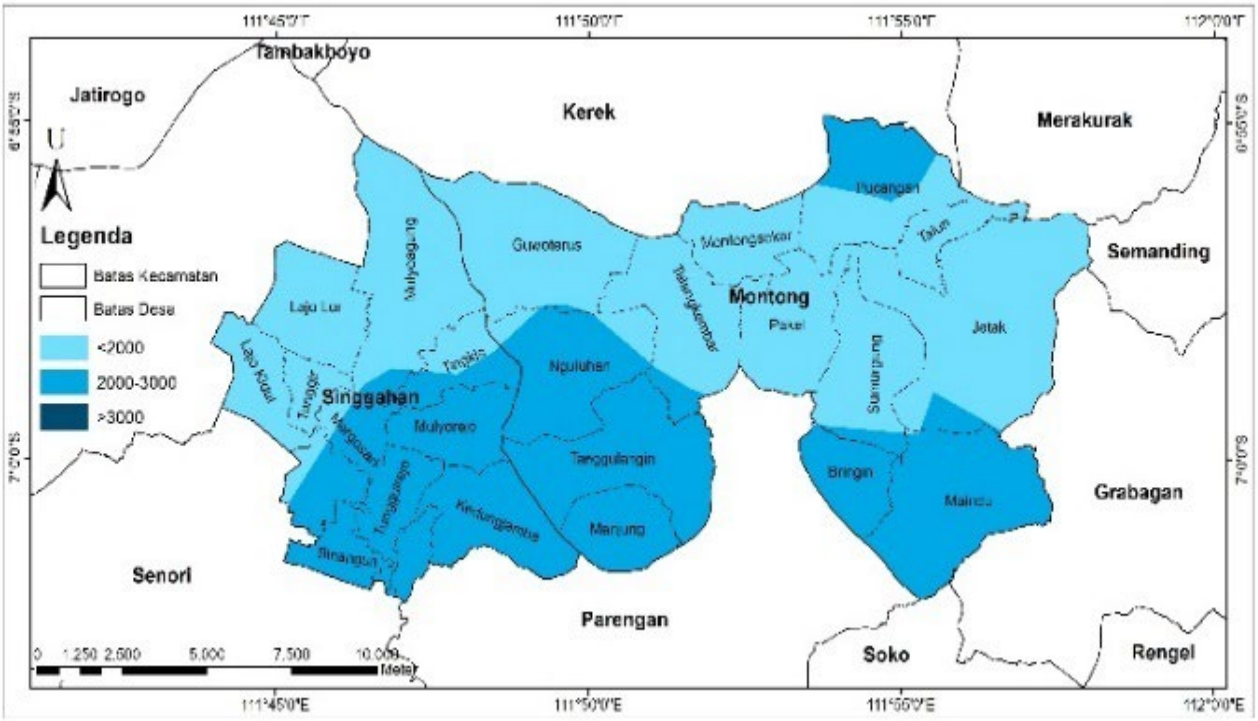


Figure 6. Land Cover Map Using the RF Algorithm for the Singgahan and Montong District Area

Based on Figure 6, it is evident that Singgahan and Montong sub-districts only have two bulk level classes spread across the two sub-districts. This is because the rainfall spread over the research area does not exceed 3000 mm/year. The area and percentage of rainfall data are presented in Table 4.

Table 4. Area and Rainfall Score for Singgahan and Montong Districts

Rainfall (mm/year)	Area (Ha)	Percentage	Value
<2000	12374,64	56,34%	1
2000-3000	9588,66	43,66%	2
>3000	0	0%	3

Soil Type

Parameters of soil type were sourced from a vector map of soil types, which was then overlain with the administrative boundaries of the research area, namely Singgahan and Montong Districts. Soil type is a decisive factor in locating recharge areas because it depends on the grain size of the soil. Bigger grain sizes widen the spaces between particles, thereby increasing porosity. Additionally, higher organic content in the soil can affect vegetation density in the area. Soil Research Institute in Bogor has classified various soil types in Indonesia, with one notable classification introduced by Dudal and Soepraptohardjo in 1957 (Fiantis et al., 2017), including the following soil groups: (1) Mediterranean soils, which are compact in texture and have low organic content, leading to reduced water absorption capacity; (2) Litosols, derived from weathered rock, with thin, rocky soil layers that are less capable of retaining water; (3) Gleysols, which possess a loose structure and high organic content, resulting in good porosity and making them effective absorbers; (4) Cambisols, which have a relatively fine texture and are moderately capable of functioning as areas for water absorption and storage.

The soil type map of Singgahan and Montong Districts is displayed in Figure 6 below.

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

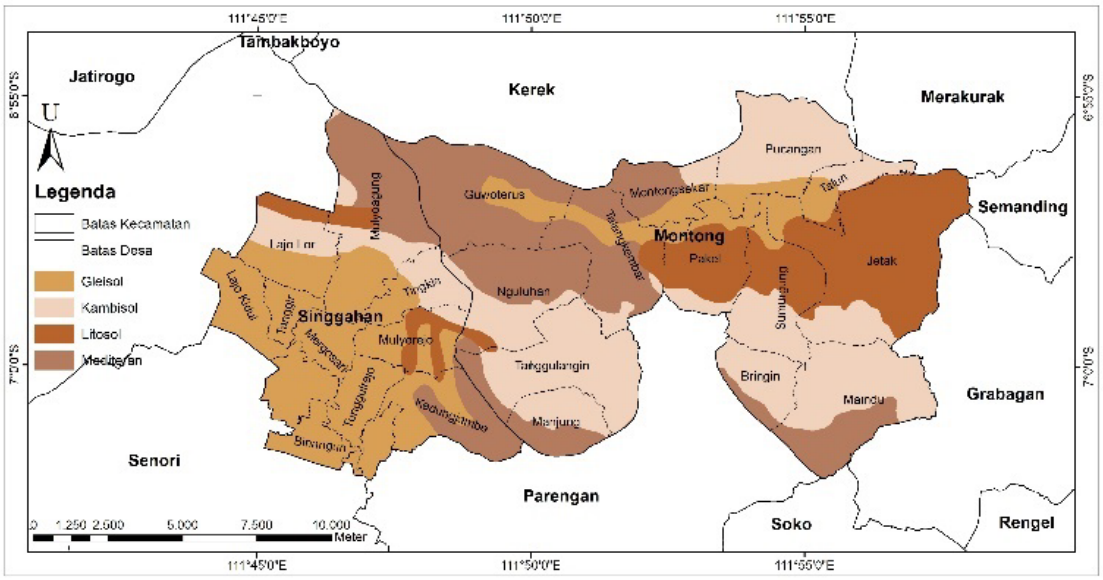


Figure 7. District Soil Type Map for Singgahan and Montong District Area

Figure 7 shows that Singgahan and Montong Districts consist of four types of soil spread across several villages, with the area and value of each type of soil outlined in Table 5 below.

Table 5. Area and Rainfall Score for Singgahan and Montong Districts

Soil Type	Area (Ha)	Percentage	Value
Litosol	3780,66	17,21%	2
Mediterranean	5561,97	25,32%	2
Cambisole	6952,93	31,66%	3
Gleisol	5667,74	25,81%	3

Soil type carries the maximum influence as a factor in identifying groundwater recharge zones. It accounts for 35% of the total impact. This is mainly because various soil types differ in their inherent ability to allow water to infiltrate.

Slope

The slope parameters are derived from slope analysis of DEMNAS data. Subsequently, the slope levels were reclassified into five classes: <8% (flat), 8-15% (ramps), 15-25% (wavy), 25-40% (slightly steep), and >40% (steep). The land slope map in the Singgahan and Montong Districts can be observed in Figure 7.

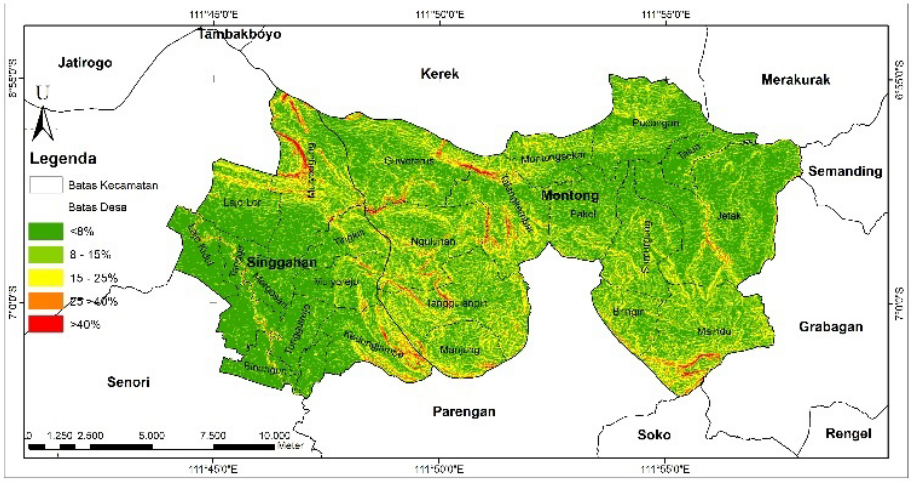


Figure 8. Slope Map of the Singgahan and Montong District Areas

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

From Figure 8, it is evident that most of the slopes in the Singgahan and Montong Districts are characterized by flat (<8%) and gentle slopes (8-15%). The slope parameter holds a weight of 20%. The area and percentage of slope in the Singgahan and Montong Districts are presented in Table 6 below.

Table 6. Area and Slope Score for Singgahan and Montong Districts

Slope	Area (Ha)	Percentage	Value
>40%	102,98	0,47%	1
25-40%	610,33	2,78%	2
15-25%	3039,51	13,84%	3
8-15%	7697,16	35,05%	4
<8%	10511,46	47,86%	5

Map of the Distribution of Groundwater Recharge Areas

The map depicting the distribution of groundwater recharge areas was generated through the overlay of the four determining parameters: land cover, rainfall, soil type, and slope. The overlay was made according to the scoring and weighting values as shown in Table 1. Subsequently, class determination was carried out according to research by Purwanto et al. (2022), where groundwater recharge areas were divided into three classes: Recharge Area, Transition Zone, and Discharge Area, as outlined in Table 7.

Table 7. Groundwater Recharge Area Classification Interval Value for Singgahan and Montong Districts

Classification of Groundwater Recharge Areas	Interval
Recharge Area	2,63 – 3,85
Transition Zone	1,42 – 2,63
Discharge Area	0,20 – 1,42

Following the process of determining the value intervals for each class, a map illustrating the groundwater recharge area was generated, as depicted in Figure 8. In the map, the recharge area is represented in green, the transition zone in yellow, and the discharge area in red.

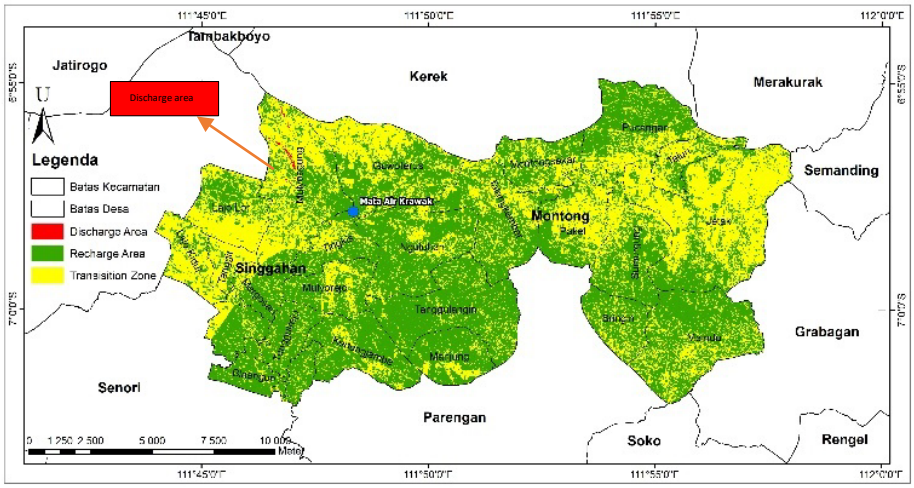


Figure 9. Regional Groundwater Recharge Area Distribution Map, Singgahan and Montong Districts

The area and percentage distribution of groundwater recharge areas are presented in Table 8. The largest percentage being 59.21%, covering an area of 13010.33 ha for the recharge area, followed by 40.66% covering 8935.27 ha for the transition zone, and 0.13% covering 28.64 ha for the discharge area (Northwest area).

Table 8. Groundwater Recharge Area Classification, Area, and Percentage for Singgahan and Montong Districts

Classification	Area(Ha)	Percentage
Recharge Area	13010,33	59,21%
Transition Zone	8935,27	40,66%
Discharge Area	28,64	0,13%

Based on the results of the map processing of groundwater recharge area distribution, it is clear that the research location, especially Singgahan and Montong Districts, has great potential to be used as groundwater recharge areas. This is due to the fact that the groundwater recharge parameters have a high weighting. Land cover parameters correlation figures show that high potential zones are generally located in areas of dense vegetation such as forests, plantations, and fields. This is because plants help the infiltration process so that raindrops can flow properly. Regarding the soil type, the infiltration rate is basically conditioned by the permeability of the soil. Soils containing a high percentage of clay have low permeability, which is the general rule. Hence, water will flow better in soils with less clay and more coarse textures like Cambisols and Gleysols. The slope parameter is the primary determinant of groundwater recharge areas since the slope can influence infiltration of water, where gentle slopes will allow more water to be absorbed and retained. Most of the slopes in the Singgahan District are flat, thus having the highest groundwater recharge potential. Besides the parameters mentioned above, the physical form of the research area is also

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

considered, which mainly comprises karst formations with high rock porosity or porous regions, leading to high permeability and more effective infiltration compared to other land types. Thus, the relatively low average rainfall over the past two decades does not significantly affect the recharge potential. Validation beyond overlaying the four parameters is necessary. Currently, there are no published maps of groundwater recharge areas from relevant ministries or local government authorities, making the final validation of the created recharge area map challenging.

The research results, when compared to the Groundwater Basin map of Indonesia (MyPatriot, 2017) in Figure 10, reveal that the Singgahan and Montong sub-districts are located within the groundwater basin area (marked by red dots in the orange region). Although the groundwater basin data are presented on a regional scale, these data can serve as supporting evidence, which indicates that the research findings are sufficiently accurate. The accuracy was supported by the fact that this study identifies Montong and Singgahan as areas with potential to become recharge zones with more specific results. This delineates the level of potential, including recharge areas (green), transition zones (yellow), and discharge areas (red) (Figure 9).

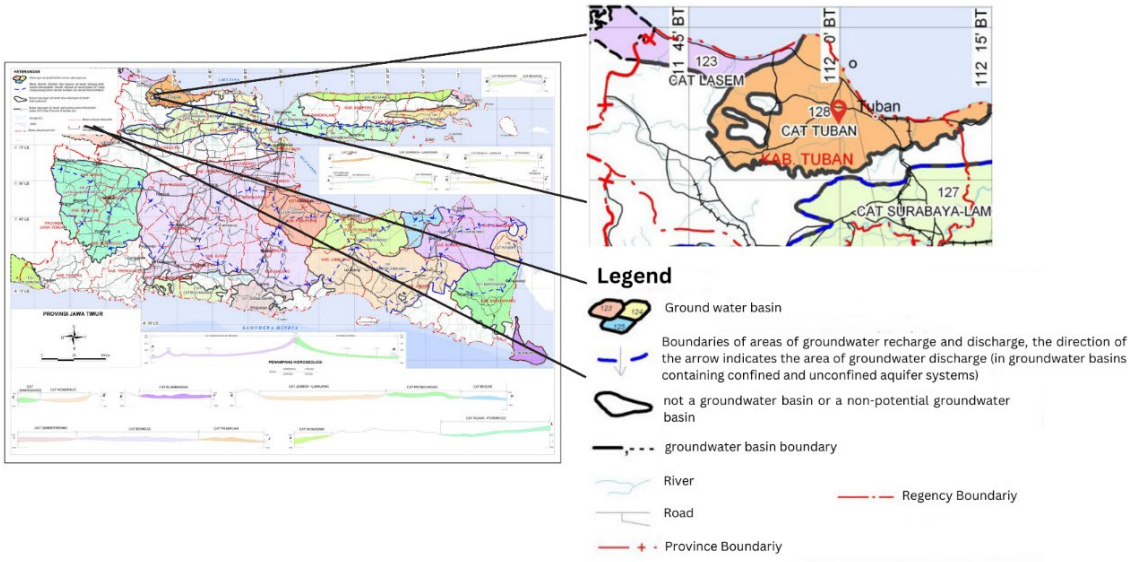
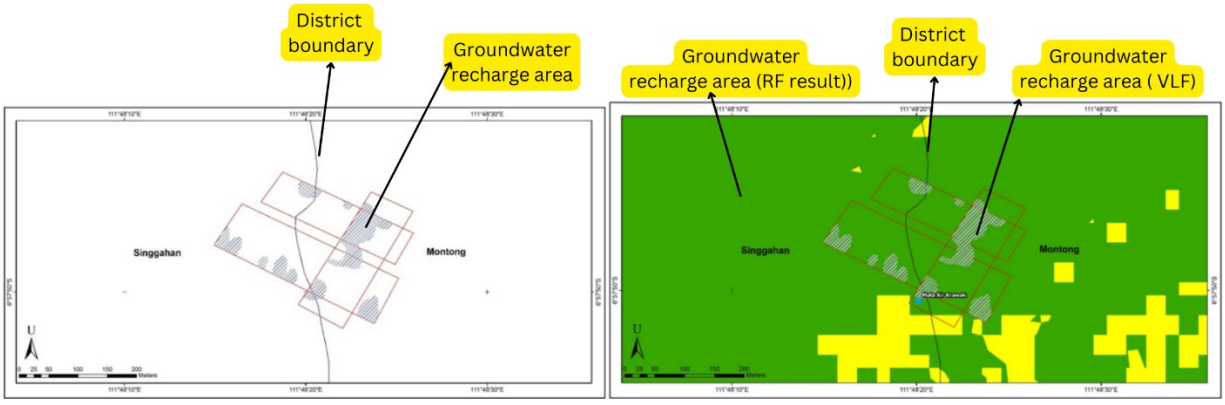


Figure 10. Map of Groundwater Basin in Indonesia

The validation was conducted by comparing the satellite image processing results with data from research using the VLF geophysical method (Purwanto et al., 2024). Figure 11a shows the distribution map of groundwater recharge areas in the Krawak spring region, Singgahan sub-district, Tuban Regency. The satellite image and VLF results were combined (Figure 11b). The results indicate that the green areas represent groundwater recharge zones, while the hatched areas (gray) denote recharge zones based on the VLF geophysical method. The results of this study are validated; however, the limited scope of the previous research resulted in the comparison data being less optimal, which is a limitation of this study. Under these circumstances, it is expected that future research will cover a wider area, allowing for a more accurate and comprehensive mapping of recharge areas in the Singgahan and Montong sub-districts. This will provide a valuable reference for well placement and drought mitigation efforts in the future.



(a) (b)

Figure 11. (a) Map of the Distribution of Groundwater Recharge Areas Using VLF; (b) Overlay map of VLF results and satellite imagery (Modified).

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

4. Conclusion

Based on the results obtained, it was observed that Landsat-8 image classification for land cover maps can be determined using several algorithms. The Random Forest algorithm was employed and it is found that it exhibits the highest Overall Accuracy (OA) value of 99.75% and a Kappa value of 0.99. Subsequently, to delineate the groundwater recharge area, four parameters are considered, including land cover, rainfall, soil type, and slope. Scoring and weighting are conducted, resulting in a recharge area of 13010.33 ha, a transition zone of 8935.27 ha, and a discharge area of 28.64 ha. From the processing results, it can be concluded that the research location, namely Singgahan and Montong Districts, harbors a significant potential to serve as a source of groundwater recharge area. The findings of this research can be beneficial for the government and local communities in preserving and upholding the primary function of groundwater recharge areas. Based on the overall results, the author has several recommendations. First, the study should be followed up by incorporating additional parameters, such as the type of rock in the research area, since rock type is closely related to the water infiltration process. Moreover, the threshold values for the classification of potential groundwater recharge areas need to be derived from an official standard or reference.

5. Acknowledgements

Thankful to God Almighty, whose mercy and grace were abundantly granted through the duration of the research to be completed without any hindrance. A heartfelt thanks to parents whose unending prayers and support have been the source of strength in this journey. Besides that, the author highly appreciates Prof. Dr. Ir. Adi Susilo, M.Si., Ph.D., Prof. Dr. Eng. Agus Naba, S.Sc., M.T., and Dr. Ayi Syaeful Bahri, S.Si., M.T., as Promoter and Co-Promoters, for their consistent guidance, support, and pearls of wisdom throughout the research work. Moreover, the author would like to extend gratitude to Mr. Ari Sanggam and Mrs. Dr. Ing. Noorlaila Hayati, S.T., M.T., for their help and contributions towards the research to a great extent.

6. References

Adugna T., Xu W., & Fan, J. (2022). Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. *Remote Sensing*, *14*(3), 1–22. https://doi.org/10.3390/rs14030574

Arikan D. (2023). Land cover generation using random forest algorithm on Google Earth Engine platform. Proceedings of the International Turkic World Congress On Science and Engineering, Bishkek Kyrgyztan.

Badan Meteorologi, Klimatologi, dan Geofisika. (2021). Badan Meteorologi, Klimatologi, dan Geofisika. Retrieved from https://www.bmkg.go.id/

Forestier H. (2007). Ribuan Gunung, Ribuan Alat Batu: Prasejarah Song Keplek, Gunung Sewu, Jawa Timur (No. 7). Kepustakaan Populer Gramedia.

Lachaud A., Adam M., & Mišković I. (2023). Comparative Study of Random Forest and Support Vector Machine Algorithms in Mineral Prospectivity Mapping with Limited Training Data. Minerals, 13(8). https://doi.org/10.3390/min13081073

Melillos G., & Hadjimitsis D. (2020). Using simple ratio (SR) vegetation index to detect deep man-made infrastructures in Cyprus. https://doi.org/10.1117/12.2557893

Mypatriot.id. (2017). Peta cekungan air tanah Indonesia tahun 2017. Retrieved from https://mypatriot.id/patgtl/datapubliclist

Peraturan Menteri Kehutanan RI. (2009). P.32/MENHUT-II/2009 tentang Tata Cara Penyusunan Teknik Rehabilitasi Hutan dan Lahan Daerah Aliran Sungai (RTkRHLDAS). Jakarta: Menteri Kehutanan RI.

Purwanto M. S., Haq, Z. M., Fachyesi, A. S., Dewi, S., & Amalina, S. (2022). Penentuan recharge area pada Kabupaten Tanah Datar menggunakan citra Landsat-8 dan sistem informasi geografis (SIG). Journal Geosaintek, 8(3), 242–248.

Purwanto M. S., Susilo, A., Bahri, A. S., Naba, A., Sari, U. I., & Almais, A. T. W. (2024). Mapping underground river flows in karst areas with the VLF-EM method (Case study of the Krawak Region, Singgahan Tuban). IOP Conference Series: Earth and Environmental Science, 1307(1), Article 012005. https://doi.org/10.1088/1755-1315/1307/1/012005

Ramji Satyaji Rao Y. (2022). Mapping and Change Detection of Water Bodies in the Godavari Delta using Geospatial Technology. *ADBU-Journal of Engineering Technology*, *11*(1), 1–6. www.earthexplorer.usgs.gov

Shin N., Saitoh T. M., Takeuchi Y., Miura T., Aiba M., Kurokawa H., Onoda Y., Ichii K., Nasahara K. N., Suzuki R., Nakashizuka T., & Muraoka H. (2022). Review: Monitoring of land cover changes and plant phenology by remote-sensing in East Asia. *Ecological Research*, *38*(1), 111–133. https://doi.org/10.1111/1440-1703.12371

Fiantis D. (2017). Morfologi dan Klasifikasi Tanah. Padang. Lembaga Pengembangan Teknologi Informasi dan Komunikasi (LPTIK) Universitas Andalas.

Undang-undang Nomor 7 Tahun 2004 tentang Sumber Daya Air. Jakarta, Indonesia: Pemerintah Pusat.

Vasileva T. (2019). An assessment of potential groundwater recharge zones in Bulgaria. *Geologica Balcanica*, *48*(1), 43–61. https://doi.org/10.52321/geolbalc.48.1.43

Wang S., & Fu G. (2023). Modelling soil moisture using climate data and normalized difference vegetation index based on nine algorithms in alpine grasslands. *Frontiers in Environmental Science*, *11*(February), 1–12. https://doi.org/10.3389/fenvs.2023.1130448

Wiggers M. J., Nuarsa I. W., & Putra I. N. (2020). Monitoring perubahan penggunaan lahan pesisir di Kecamatan Batu Layar, Kabupaten Lombok Barat pada tahun 2002 dan 2019. Journal of Marine Research and Technology, 3(2), 68-74.

Zulfajri, Danoedoro P., & Murti S. H. (2021). Klasifikasi tutupan lahan data Landsat-8 Oli menggunakan metode Random Forest. Journal Penginderaan Jauh Indonesia, 03(01), 1–7. http://journal.mapin.or.id/index.php/jpji/issue/archive

DOI:https//doi.org/10.22452/mjs.vol45no2.5
Malaysian Journal of Science 45(2): 42-54 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 45(2): 55-61 (June 2026) https://mjs.um.edu.my

Examine Horizontal Resolution of Dipole-Dipole Array Using Synthetic Models Of 2D and 3D Tomography Techniques

Ahmed S. Al-Zubedi[1a] and Zaidoon T. Abdulrazzaq[2b]*

Abstract: The horizontal resolution of the array is the most important factor in accurately mapping subsurface features and ensuring reliable geological interpretations in electrical resistivity surveys. Therefore, six two-dimensional (2D) and three-dimensional (3D) synthetic models were used to assess the horizontal resolution of the Dipole-Dipole array. In each synthetic model, two identical structures with a resistivity of 100Ωm are embedded within a medium of 30Ωm. The horizontal distance between the structures is varied between 2m, 3m, and 4m. The results revealed that the Dipole-Dipole array shows difficulties in delineating the exact location and size of the structures. It provides a distorted image of the structures' locations. It shows better resolution with 2D inverse models compared to 3D inverse models when the horizontal distance becomes small. The sensitivity of this array to the horizontal change in resistivity is increased when the distance between the structures is increased. In the 3D inverse models, Dipole-Dipole array generated a distorted image of the structures when the horizontal distance between the structures is small, while it provides reasonable images when the distance between the structures is increased. So, for investigating a subsurface horizontal structure that is separated with small distances, 2D tomography technique is a better approach than 3D tomography technique when using Dipole-Dipole array.

Keywords: *Horizontal Resolution; Dipole-Dipole Array; 2D and 3D Tomography Techniques.*

1. Introduction

At the present time, the 2D and 3D resistivity techniques are considered the most important techniques of the electrical resistivity method that are mainly used for exploration of the shallow underground structures (Kemna *et al*., 2002; Zhou *et al*., 2004; Loke *et al*., 2013; Al-Zubedi and Thabit, 2016, Al-Awsi & Abdulrazzaq, 2022). These techniques are usually carried out using different electrode arrays, for example, Wenner, Wenner-Schlumberger, Dipole-dipole, Pole-dipole, and Gradient arrays. Various factors, including the target depth, the array's sensitivity function, the array's resolution, the sensitivity of the resistivity meter, and the background noise level, influence the selection of the optimal array for 2D and 3D imaging surveys (Roy and Apparao, 1971; Loke, 2012; Al-Zubedi and Abdulrazzaq, 2025). Accordingly, there have been many studies carried out to determine the best array's response in delineating targets within different situations (Zhou *et al.*, 2002; Dahlin and Zhou, 2004; Tamssar, 2013; Al-Hameedawi, 2013; Thabit and Al-Zubedi, 2015; Al-Zubedi, 2016; Al-Khersan, *et al., 2016*). These studies showed that Dipole-Dipole array has the ability to offer the best resolution and highest sensitivity to target details for a shallow investigation compared with other arrays. On contrary, Wenner-Schlumberger is superior than other arrays for deep investigation. In studies conducted in Malaysia, the Enhancing Horizontal Resolution (EHR) technique significantly improves horizontal resolution in 2D resistivity studies by optimizing data acquisition and inversion processes. This technique allows for deeper penetration and more detailed imaging of subsurface structures. The technique has been validated against borehole data, confirming its effectiveness in accurately mapping subsurface structures (Nordiana et al., 2013; Nordiana et al., 2014).

The dipole-dipole is the array that is most commonly utilized for 2D and 3D tomography techniques, as it is easy to apply in fieldwork. The distance between the current and potential electrodes in this array is known as the n-factor, whereas the spacing between the current electrode pair, AB, is provided as (a), which is the same as the distance between the potential electrode pair, MN. The n-factor controls the depth of investigation and noise level, as with a large n-factor, Dipole-Dipole array will gain more noisy data because of low signal strength Alpine (1950; Kunetz (1966); Keller and Frischknecht (1966) ; Look (2020).

Dahlin and Zhou, (2004); Chitea and Georgescu, (2009); and Loke, (2020), indicated that the Dipole-Dipole array is sensitive when lateral resistivity variation is encountered. Therefore, it is the best choice when mapping vertical structures. However, in mapping horizontal structures, such as sedimentary layers, it is relatively poor. This means that the Dipole-Dipole array has a good horizontal resolution compared to vertical resolution, as it is resolution decrease with a depth.

The resolution of electrical arrays is the ability to distinguish variations in electrical resistivity. This is based on electrode

Authors information:
[a]Remote Sensing and Geophysics College, AlKarkh University of Science, Baghdad, IRAQ. E-mail: AhmedSrdah@yahoo.com
[b]Space Research and Technology Center, Scientific Research Commission, Baghdad, IRAQ. E-mail: zaidoon.taha@live.com
*Corresponding Author: zaidoon.taha@live.com

Received: April, 2025
Accepted: June, 2025
Published: June, 2026

DOI:https//doi.org/10.22452/mjs.vol45no2.6
Malaysian Journal of Science 45(2):55-61 (June 2026)

spacing and spatial locations, so that the two targets can be separately recognized instead of being one target. Further, the resolution can be classified into two kinds, vertical and horizontal resolution. Vertical resolution is the ability of the array to distinguish two targets located at different depth levels. Conversely, horizontal resolution describes how two neighboring targets can be positioned horizontally and still be distinguished as two distinct targets as opposed to one (Kallweit and Wood, 1982).

This study will evaluate the Dipole-Dipole array's horizontal resolution through synthetic models to examine its ability in defining shallow buried structures that are separated horizontally using 2D and 3D tomography techniques.

2. Materials and Methods

Horizontal resolution of arrays is a critical parameter in geophysical surveys. The purpose of the present research is to estimate the resolution of dipole-dipole arrays related to the horizontal discontinuity based on the synthetic models using two-dimensional and three-dimensional tomography techniques. The models used were built with contrast rectangles representing an anomalous body in a uniformly layered Earth model. The middle depth position of the rectangle is taken as the level of the array setup called the observation position. A somewhat different approach related to the array position was chosen for the presentation of a different electrical nature. Synthetic models are an integral part of advances in different fields. These models help to understand and interpret the data and methodologies. Tomography is an ill-posed inverse problem that retrieves subsurface geophysical parameters using the observed data. Different tomography techniques use different mathematical approaches. Hence, one can use different synthetic models to test the behavior and capability of these techniques. Various models such as 1D, 2D, 2.5D, and 3D setups are used in the tomography technique based on the availability of data. The effect of the complexity of the model on the desired resolution can be calculated easily by testing the technique on these synthetic models.

In order to determine the horizontal resolution of Dipole-Dipole array in 2D and 3D tomography techniques. Six synthetic models are prepared to represent two buried structures. The synthetic models were created using the 2D and 3D forward modeling software RES2DMOD and RES3DMOD (Al-Zubedi and Thabit, 2015).

Three 2D resistivity synthetic models were designed with a survey line of 35 m long and minimum electrode spacing of 1 m. In these synthetic models, two identical structures were built. The width and the height for each one are 2m, while their resistivity value is set to be 100 Ωm. These structures are embedded within a host medium characterized by a resistivity of 30 Ωm (Fig. 1). In this case, this host medium may represent sand materials. For the 2D synthetic model, the horizontal distance between the two structures is set to be 2 m, 3 m, and 4m. Figure (1), shows the 2D synthetic model of the two structures when the horizontal distance is 2 m.

As shown in the 2D synthetic models, the same resistivity values for the structures and the host medium are used to create the 3D synthetic models. Further, the same horizontal distances of 2m, 3m, and 4m between the structures are used for the 3D synthetic models. To build the 3D synthetic models, the distances between the structures are changed in the x-direction to be 2m, 3m, and 4m, while the distance in the y-direction remains fixed. Figure (2) shows a 3D synthetic model of the horizontal structures when the distance between them equals 2m, and they are 1.6 m thick.

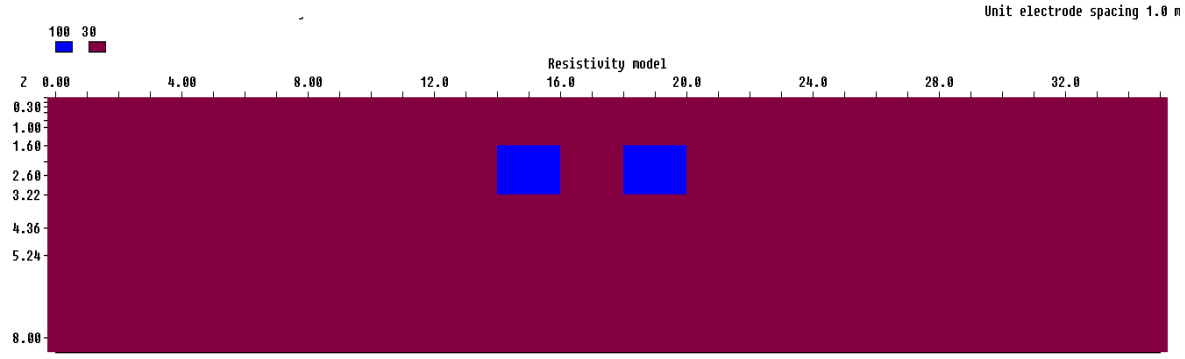


Figure 1. The 2D synthetic model of the two structures when the horizontal distance is 2 m.

DOI:https//doi.org/10.22452/mjs.vol45no2.6
Malaysian Journal of Science 45(2):55-61 (June 2026)

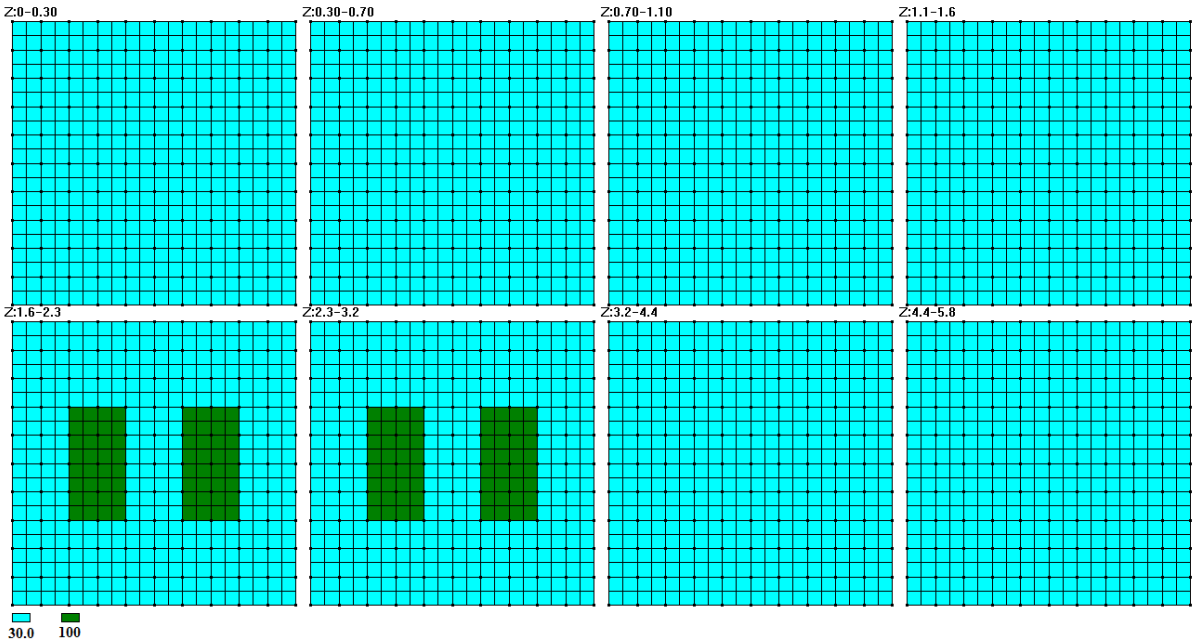


Figure 2. 3D synthetic model of horizontal structures with (2m) distance.

3. Results and Discussion

After synthetic models are created, the apparent resistivity measurements are calculated for each synthetic model. These measurements are carried out on the basis that the n-factor is 8a while the a-spacing is varied between 1a to a maximum spacing of 12a. This procedure allows for performing a higher resolution and a maximum investigation depth. Because of Dipole-Dipole array has a very weak signal strength when the distance between current and potential electrodes becomes very large (Dahlin and Zhou, 2004; Chitea and Georgescu, 2009; Loke, 2020), random noise of 5% is added to the collected data.

The 2D inverse models are created using the RES2DINV ver. 3.59 Software, by the L1 norm (robust) inversion method. This method is useful when there are sharp variations between resistivity values due to sharp boundaries between the buried structures. The inversion reduces the absolute values of the data misfit (Claerbout and Muir, 1973; Olayinka and Yaramanci, 2000 and Loke et al., 2013). Al-Zubedi, (2016) confirmed that the robust inversion method can give the best boundary resolution results for buried walls that have sharp geologic boundaries, while investigated by the Dipole-Dipole array. Therefore, this method will improve the horizontal resolution of the Dipole-Dipole array and help in identifying the buried structures.

At a horizontal distance of 2m between the buried structures, the inverse models of the Dipole-Dipole array successfully identify the left-hand buried structure. The right-hand structure is determined also, but its location is less accurate as it appeared under electrodes 19-21 (or the distance 18m-20m). The inverse model shows a misfit error of 3.8%. The Dipole-Dipole array is able to identify the two structures separately instead of being to appear one structure concerning the minimum distance (Fig. 3).

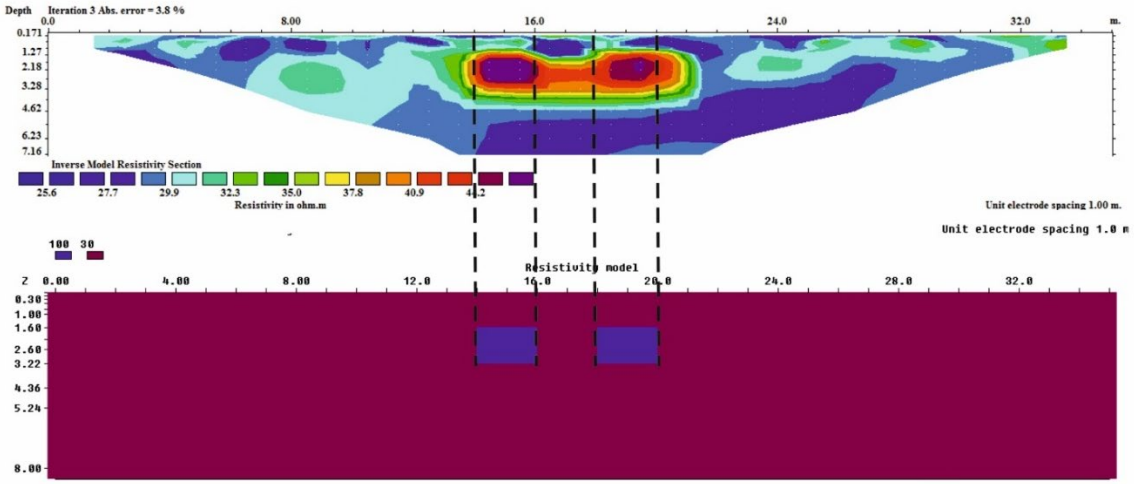


Figure 3. The inverse and the 2D synthetic model of the structures when the horizontal distance is (2m).

DOI:https//doi.org/10.22452/mjs.vol45no2.6
Malaysian Journal of Science 45(2):55-61 (June 2026)

The second and third inverse models, which derived from the second and third synthetic models of 3m and 4m horizontal distance between the buried structures, are not successfully delineating the right-hand structure that occurs under electrodes 19-21 (Fig. 4, and 5 respectively). They provide a distorted image of the location of the right-hand structure. The misfit error of both inverse models is 3.8% after three iterations. This indicated that the Dipole-Dipole array has difficulties in delineating the exact size and location of the buried structures. Accordingly, the Dipole-Dipole array is indeed sensitive to changes in resistivity that occur horizontally. This array still has the ability to identify both structures when the horizontal distance between the buried structures is changed from 2m to 4m.

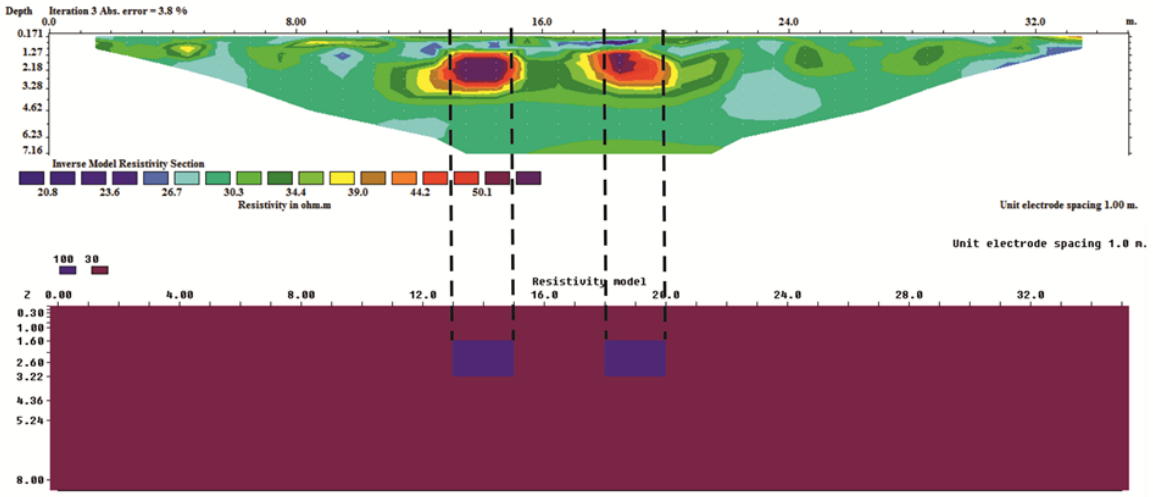


Figure 4. The inverse and the 2D synthetic model of the structures when the horizontal distance is (3m).

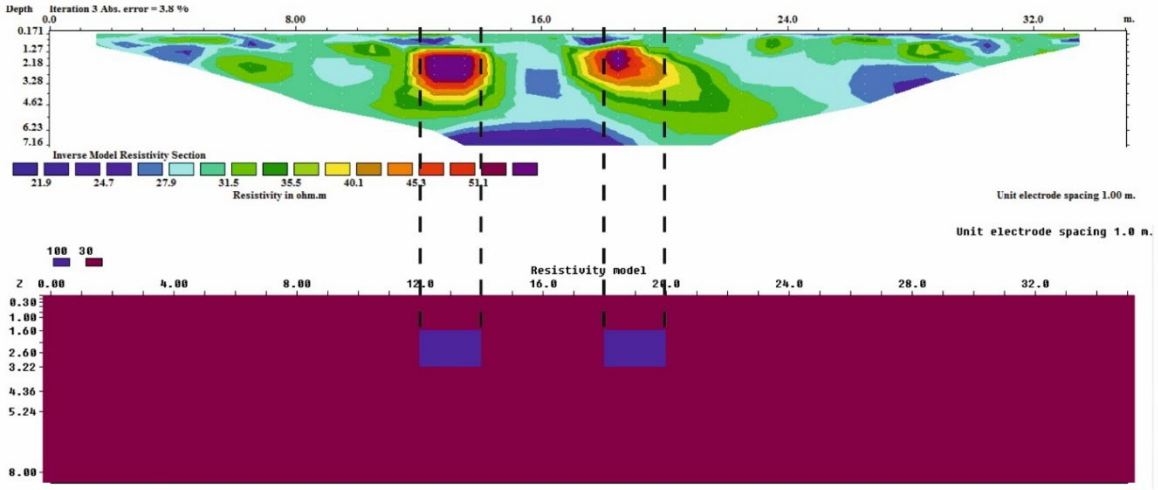


Figure 5. The inverse and the 2D synthetic model of the structures when the horizontal distance is (4m).

For the 3D inverse models, the apparent resistivity measurements are derived from the 3D synthetic models after adding 5% random noise. The 3D inverse models are created using RES3DINV ver. 3.18.2 software. When the horizontal distance between the structures is 2m, the 3D inverse model of the Dipole-Dipole seems incapable of delineating the buried structures. It shows a distorted image of the size and location of these structures. Both are merged and appeared as one structure (Fig. 6). This indicates that the Dipole-Dipole has a lower resolution and incapable of imaging the underground structures when the horizontal distance between the buried structures is small as compared to the electrode spacing used for the survey. Finally, this 3D inverse model shows a misfit error of 3.49% after three iterations.

The 3D inverse model with a 3m horizontal distance between the buried structures does not provide a very obvious image of the two structures. It partially successes in delineating the structures, but it was not able to recognize the shape and the size of the right-hand structure under electrodes (8-10) (Fig. 7). It is relatively delineated by the left-hand structure. Furthermore, it successfully recognizes and separates the buried structures. This inverse model shows a misfit error equals to 3.58%. The 3D inverse model of the Dipole-Dipole array when the horizontal distance between the structures is 4m, successfully determines the two structures, but their sizes and shapes look smeared. The

DOI:https//doi.org/10.22452/mjs.vol45no2.6
Malaysian Journal of Science 45(2):55-61 (June 2026)

right-hand structure is delineated better compared with the previous models (Fig. 8). It is indicated that the resolution of this array is increasing when the horizontal distance between these structures is increasing. This array can provide reasonable images for subsurface structures.

The theoretical results showed that the inverse models yield resistivity values lower than the actual ones. All models depicted the resistivity of the structures as less than 100 Ω·m, whereas the true resistivity of each structure is 100 Ω·m. This discrepancy arises from the calculation method used by the software to compute resistivity values within the inverted model. Therefore, the resistivity values in the inverted model do not accurately represent the true resistivity of the underground structures.

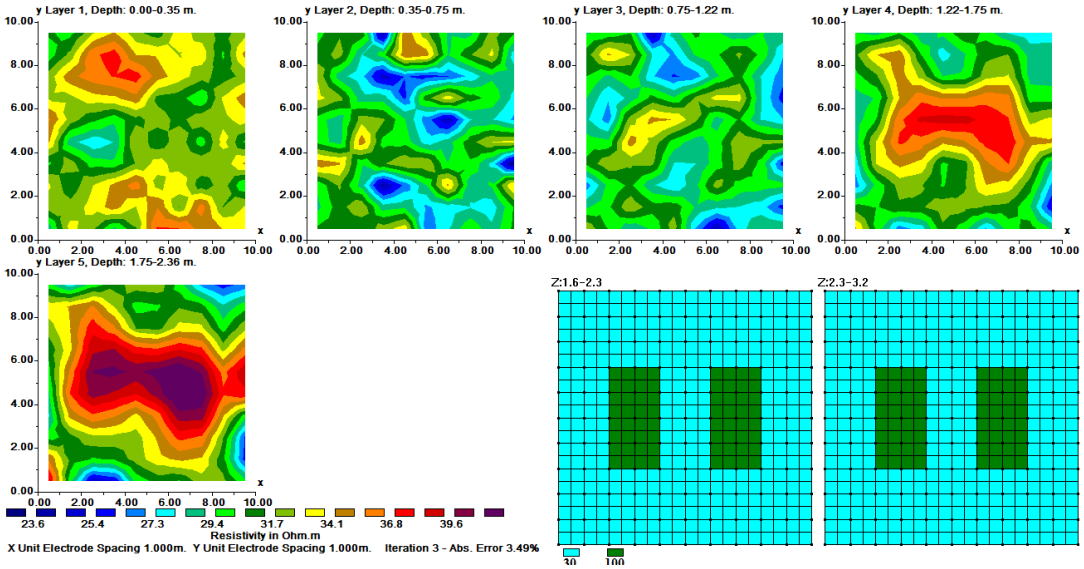


Figure 6. The inverse and synthetic 3D model for structures when separated by (2m).

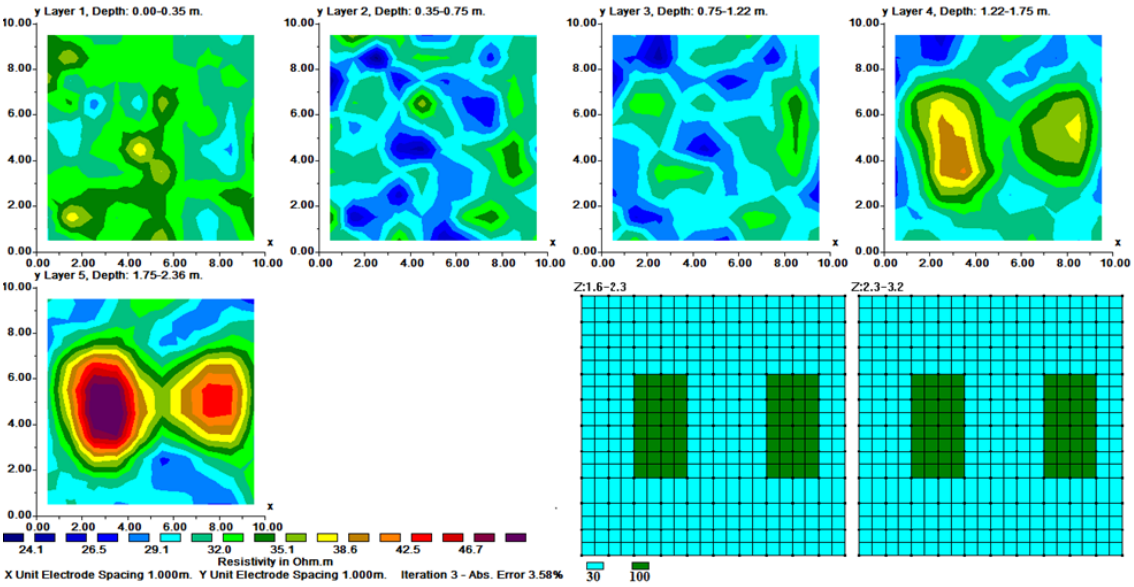


Figure 7. The inverse and synthetic 3D model for structures when separated by (3m).

DOI:https//doi.org/10.22452/mjs.vol45no2.6
Malaysian Journal of Science 45(2):55-61 (June 2026)

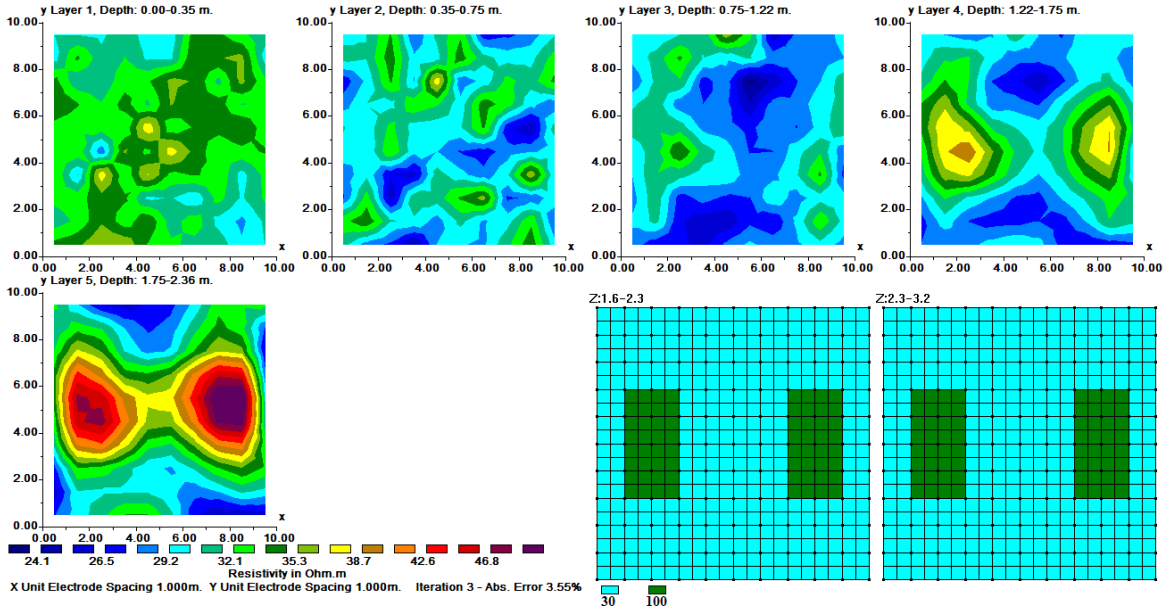


Figure 8. The inverse and synthetic 3D model for structures when separated by (4m).

4. Conclusion

The comparison of horizontal resolution for the Dipole-Dipole array using 2D and 3D tomography techniques on synthetic models show some difficulty in delineating the exact size and location of the structures. Its horizontal resolution within 2D inverse models is slightly better than it in the 3D inverse models, especially when the distance between the structures becomes too small. The Dipole-Dipole's sensitivity to the horizontal change in resistivity is increased when the distance between the structures is increased, with optimal resolution typically achieved when the inter-structure spacing is approximately 1 to 1.5 times their dimensions. In the 3D inverse models, while the Dipole-Dipole array generated a distorted image of the structures when the horizontal distance between the structures is small. It provides reasonable images when the distance between the structures is increased. So, for investigating subsurface horizontal structures that are separated with small distances, 2D tomography technique is a better approach than 3D tomography technique when using the Dipole-Dipole array.

5. References

Al-Awsi, M. D., & Abdulrazzaq, Z. T. (2022). Implementing electrical resistivity tomography to delineate soil contamination zone, Southern Baqubah City, Iraq. *Kuwait Journal of Science*, *49*(2), pp. 1-15. https://doi.org/10.48129/kjs.10674

Al-Hameedawie, M. A. (2013). *Comparison between different electrode arrays in delineating aquifer boundary by using 1D and 2D techniques in north Badra area eastern Iraq* (Doctoral dissertation, M. Sc. Thesis Department of Geology, college of Science, University of Baghdad, Iraq, 142p).

Al-Khersan, E. H., Al-Ani, J. M., & Abrahem, S. N. (2016). Integrated GPR and ERT as enhanced detection for subsurface historical structures inside Babylonian houses site, Uruk City, southern Iraq. *Pure and Applied Geophysics*, *173*, 963-982. https://doi.org/10.1007/s00024-015-1162-2

Alpin L.M. (1966). The theory of dipole sounding. In: *Dipole Methods for Measuring Earth Conductivity*, pp. 1–60. Consult Bureau, New York.

Al-Zubedi, A. S. (2016). Evaluation of five electrode arrays in imaging subsurface shallow targets: A case study. *Iraqi Bulletin of Geology and Mining*, *12*(2), 39-46.

Al-Zubedi, A. S., & Thabit, J. M. (2016). A comparison between 2D azimuthal and 3D resistivity imaging techniques in determining the subsurface fracture zones within Abu-Jir Fault Zone, Southwest Karbala, Central Iraq. *Near Surface Geophysics*, *14*(5), 413-421. https://doi.org/10.3997/1873-0604.2016020

Al-Zubedi, A. S., & Thabit, J. M. (2016). Use of 2D azimuthal resistivity imaging in delineation of the fracture characteristics in Dammam aquifer within and out of Abu-Jir fault zone, central Iraq. *Arabian Journal of Geosciences*, *9*(1), 22. https://doi.org/10.1007/s12517-015-2070-6

Al-Zubedi, A., & Abdulrazzaq, Z. (2025). The Resolution and Sensitivity Function of Electrode Arrays in 2D Resistivity Imaging Technique. *Malaysian Journal of Science,* 44(1), 55–62. https//doi.org/10.22452/mjs.vol44no1.7

Chitea, F. and Georgescu, P., (2009). Sensitivity function for various geoelectric arrays. In *Geophysical Research Abstracts*. University of Bucharest, Faculty of Geology and Geophysics.

DOI:https//doi.org/10.22452/mjs.vol45no2.6
Malaysian Journal of Science 45(2):55-61 (June 2026)

Claerbout, J. F., & Muir, F. (1973). Robust modeling with erratic data. *Geophysics*, *38*(5), 826-844.

Dahlin, T., & Zhou, B. (2004). A numerical comparison of 2D resistivity imaging with 10 electrode arrays. *Geophysical prospecting*, *52*(5), 379-398. https://doi.org/10.1111/j.1365-2478.2004.00423.x

Kallweit, R. S., & Wood, L. C. (1982). The limits of resolution of zero-phase wavelets. *Geophysics*, *47*(7), 1035-1046. https://doi.org/10.1190/1.1441367

Keller, G.V., Frischknecht, F.C., 1966. Electrical methods in geophysical prospecting. Pergamon Press, London, England, Oxford, 519p.

Kemna, A., Vanderborght, J., Kulessa, B. and Vereecken, H., 2002. Imaging and characterisation of subsurface solute transport using electrical resistivity tomography (ERT) and equivalent transport models. *Journal of hydrology*, *267*(3-4), pp.125-146. https://doi.org/10.1016/S0022-1694(02)00145-2

Kunetz, G., 1966. Principles of direct current resistivity prospecting. Borntraeger, Stuttgart, Germany, 106 p.

Loke, M. H., (2012). Tutorial: 2-D and 3D Electrical Imaging Surveys, 172p.

Loke, M. H., (2020). Tutorial: 2-D and 3D Electrical Imaging Surveys, 221p.

Loke, M. H., Chambers, J.E., Rucker, D.F., Kuras, O. and Wilkinson, P.B., (2013). Recent developments in the direct-current geoelectrical imaging method. *Journal of applied geophysics*, *95*, pp.135-156. https://doi.org/10.1016/j.jappgeo.2013.02.017

Nordiana, M. M., Rosli, S., & Nawawi, M. N. M. (2014). A numerical comparison of enhancing horizontal resolution (EHR) technique utilizing 2D resistivity imaging. *Arabian Journal of Geosciences*, *7*, 299-309. https://doi.org/10.1007/S12517-012-0792-2

Nordiana, M. M., Saad, R., Nawawi, M. N. M., Azwin, I. N., & Saufia, A. T. (2013). 2-D Resistivity Study: The Horizontal Resolution Improvement by Introducing the Enhancing Horizontal Resolution (EHR) Technique. *Open Journal of Geology*, *3*(02), 1-6. https://doi.org/10.4236/OJG.2013.32B001

Olayinka, A. I., & Yaramanci, U. (2000). Assessment of the reliability of 2D inversion of apparent resistivity data [Link]. *Geophysical Prospecting*, *48*(2), 293-316. https://doi.org/10.1046/j.1365-2478.2000.00173.x

Roy, A., & Apparao, A. (1971). Depth of investigation in direct current methods. *Geophysics*, *36*(5), 943-959. https://doi.org/10.1190/1.1440226

Tamssar, A. H. (2013). *An evaluation of the suitability of different electrode arrays for geohydrological studies in karoo rocks using electrical resistivity tomography* (Doctoral dissertation, University of the Free State).

Thabit, J. M., & Al-Zubedi, A. S. (2015). Evaluation of three important electrode arrays in defining the vertical and horizontal structures in 2D imaging surveys. *Iraqi Journal of Science*, 1465-1470.

Zhou, Q. Y., Matsui, H., & Shimada, J. (2004). Characterization of the unsaturated zone around a cavity in fractured rocks using electrical resistivity tomography. *Journal of Hydraulic Research*, *42*(S1), 25-31. https://doi.org/10.1080/00221680409500044

Zhou, W., Beck, B. F., & Adams, A. L. (2002). Effective electrode array in mapping karst hazards in electrical resistivity tomography. *Environmental geology*, *42*, 922-928. https://doi.org/10.1007/s00254-002-0594-z

DOI:https//doi.org/10.22452/mjs.vol45no2.6
Malaysian Journal of Science 45(2):55-61 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 45(2): 62-74 (June 2026) https://mjs.um.edu.my

Direct Numerical Algorithm as First, Second and Third-Order IVPS Solver

Emmanuel Oluseye Adeyefa[1a], Victoria Oluwatoyin Ojo[2b], Rotimi O. Folaranmi[3c] and Adebisi A. Ibrahim[4d]*

Abstract: This paper presents a novel numerical model for accurately integrating initial value problems of multi-order ordinary differential equations (ODEs) of first, second, and third orders. Chebyshev polynomials are employed as the basis functions, and the collocation technique is used to develop continuous schemes that are evaluated at selected points to formulate the proposed multi-order ODE solver, which is applied in a block-by-block manner. The convergence analysis is carried out to establish the zero-stability and consistency of the method. Comparisons with existing methods show the superior performance of the proposed method. The results indicate its ability to solve multi-order ODEs more effectively while reducing the computational cost. This work represents a significant advance in numerical integration for ODEs, providing improved accuracy and efficiency in solving a wide range of multi-order ODE problems.

Keywords: *Block method, collocation, consistency, convergence, zero-stability.*

Authors information:
[a]Department of Mathematics, Federal University Oye-Ekiti, Ekiti State, NIGERIA. E-mail: adeoluman@yahoo.com[1]
[b]Department of Statistics, Oyo State College of Agriculture and Technology, Igbo-Ora, Oyo State, NIGERIA. E-mail: ojovictoria2@yahoo.com[2]
[c]Department of Mathematical and Computing Sciences, Thomas Adewumi University, Oko, Kwara State, NIGERIA. E-mail: rotimifolaranmi@gmail.com[3]
[d]Department of Mathematical Sciences, Oduduwa University, Ipetumodu, Osun State, NIGERIA. E-mail: adebisiibrahim97@gmail.com[4]
*Corresponding Author: ayu.arista.dewi@undiksha.ac.id

Received: July, 2024
Accepted: March, 2025
Published: June, 2026

1. Introduction

This paper aims to develop a linear multistep method of the form

$$\left\{ y(x) = h\left(\sum_{r=0}^{s} \beta_r(t) m_{n+r}\right) + h^2\left(\sum_{r=0}^{s} \mu_r(t) l_{n+r}\right) + h^3\left(\sum_{r=0}^{s} \delta_r(t) k_{n+r}\right) \right. \quad (1)$$

for the direct integration of ordinary differential equations (ODEs), where in (1) either of $\sigma_0(t)$ and $\beta_0(t)$ do not vanish, $\sigma_s(t) = 1$, $\beta_s(t) \neq 0$ and $s = 1$.

Our focus is on three distinct types of ordinary differential equations (ODEs), denoted Equation (2) in this paper.

$$\omega'(x) = f(x, \omega(x)), \omega(x_0) = \omega_0 \quad (2a)$$

$$\omega''(x) = f(x, \omega(x), \omega'(x)), \omega(x_0) = \omega_0, \omega'(x_0) = \omega_0' \quad (2b)$$

$$\omega'''(x) = f(x, \omega(x), \omega'(x), \omega''(x)), \omega(x_0) = \omega_0, \omega'(x_0) = \omega_0', \omega''(x_0) = \omega_0'' \quad (2c)$$

Numerical methods have become essential tools for approximating solutions to differential equations, particularly for real-world problems whose analytical solutions do not exist (Adeyefa & Kuboye, 2020; Jator, 2010). As a result, there is a growing need for more efficient and accurate computational methods to solve such initial value ordinary differential equation (ODE) problems.

Many authors have proposed different numerical methods for solving Equation (2a), as documented in Ajileye et al. (2018). Conventionally, the solutions of Equations (2b) and (2c) are obtained by reducing and into equivalent systems of first-order ODEs. However, this approach often increases computational time and effort (Henrichi, 1962; Lambert, 1973). In contrast, direct integration methods for Equation (2b) have

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

been proposed (Awoyemi, 1999; Fatunla, 1991; Koboye, 2015), employing self-starting numerical algorithms that remove the need for predictors. Equation (2c) has received considerable attention from researchers, leading to the development of various schemes (Allogmany, 2020; Hussain, 2017; Olabode, 2009; Rufai, 2023, Duromola, 2022, Rufai et al., 2023, Rufai et al., 2024) for solving third-order ODEs. In addition, Ramos & Rufai (2021) introduced a one-step Lobatto-type hybrid block method incorporating a second derivative to obtain approximate solutions in an embedded-like form executed in adaptive mode, thereby improving the performance of higher-order block methods.

Notably, a recent trend among scholars is the development of a unified scheme for solving equations of the form given in Equation (2). This emerging area of research has attracted the interest of many investigators. Matthew et al. (2024) formulated a single numerical method for the integration of first- and second-order ordinary differential equations. Adeyefa et al. (2020) proposed a method for the direct integration of second- and third-order ordinary differential equations. Abolarin et al. developed implicit hybrid block methods for solving second-, third, and fourth-order ordinary differential equations directly using power series. Given the elegant properties of Chebyshev polynomials, the proposed method employs them as basis functions.

The main novelty of this work is the formulation of a two-step hybrid block method (THBM), a single numerical model that can efficiently and accurately handle first-, second-, and third-order ODEs. Section 2 sets out the formulation of the continuous algorithm that gives rise to the discrete method for solving Equation (2). The convergence analysis of the proposed method is presented in Section 3, followed by the results in Section 4. Finally, Section 5 provides concluding remarks.

2. Development of Two-Step Hybrid Block Method (THBM)

In this section, THBM is derived using the well-known Chebyshev polynomial of the first degree, given by

$$y(x)=\sum_{r=0}^{s+9}\sigma_r X_r \tag{3}$$

where X^r are the parameters of Chebyshev polynomials and σ_r are unknown constants to be determined. The Chebyshev polynomials are chosen because of their elegant properties, such as an efficient even distribution of error that leads to rapid convergence, minimization of the maximum error, and economization of the power series.

An interpolation of Equation (3) at point $x=x_n$ and collocation of its 1st and 2nd derivatives at point $x=x_{n+v}$ yields a system of equations. Thus, we have

$$\sum_{r=3}^{s+9}(r^3-3r^2+2r)\sigma_r X_{n+w}^{r-3}=k_{n+w}.$$

where k_{n+w} denotes the third derivative of Equation (3).

The interpolation and collocation of (3) produce a system of equations represented as

$$\begin{cases}\sum_{r=0}^{s+9}\sigma_r X_n^r=y_n\\ \sum_{r=1}^{s+9}r\sigma_r X_{n+v}^{r-1}=m_{n+v}\\ \sum_{r=2}^{s+9}(r^2-r)\sigma_r X_{n+v}^{r-2}=l_{n+v}\\ \sum_{r=3}^{s+9}(r^3-3r^2+2r)\sigma_r X_{n+w}^{r-3}=k_{n+w}\end{cases} \tag{4}$$

where m_{n+v} is the first derivative of (3), l_{n+v} is its second derivative, and s is the step number (s = 2).

The proposed method is formulated by specifying parameters v and w. Thus, $v=0,\frac{2}{8},\frac{5}{8},1$ and $w=2$.

By solving the system of ten equations for $\sigma's$ using MAPLE and substituting the resulting values into Equation (3), the continuous implicit form of THBM is obtained as

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

$$\sigma_0(t)y_n = h\left(\sum_{r=0}^{2}\beta_1(t)m_{n+1} + \beta_{\frac{2}{8}}(t)m_{n+\frac{2}{8}} + \beta_{\frac{5}{8}}(t)m_{n+\frac{5}{8}}\right) + h^2\left(\sum_{r=0}^{2}\mu_1(t)l_{n+1} + \mu_{\frac{2}{8}}(t)l_{n+\frac{2}{8}} + \mu_{\frac{5}{8}}(t)l_{n+\frac{5}{8}}\right) + h^3\left(\delta_2(t)k_{n+2}\right) \quad (5)$$

where, $t = \dfrac{2x - 2x_n - h}{h}$, $\sigma_0(t) = 1$.

$$\begin{cases}
\alpha_0 - \beta_0 + \beta_{\frac{2}{8}} - \beta_{\frac{5}{8}} + \beta_1 - \lambda_0 + \lambda_{\frac{2}{8}} - \lambda_{\frac{5}{8}} + \lambda_1 - \delta_2 = y_n \\
2\beta_0 - 8\beta_{\frac{2}{8}} + 18\beta_{\frac{5}{8}} - 32\beta_1 + 50\lambda_0 - 72\lambda_{\frac{2}{8}} + 98\lambda_{\frac{5}{8}} - 128\lambda_1 + 162\delta_2 = hm_n \\
2\beta_0 - 4\beta_{\frac{2}{8}} + 0\beta_{\frac{5}{8}} + 8\beta_1 - 10\lambda_0 + 0\lambda_{\frac{2}{8}} + 14\lambda_{\frac{5}{8}} - `16\lambda_1 + 0\delta_2 = hm_{n+\frac{2}{8}} \\
2\beta_0 + 8\beta_{\frac{2}{8}} + 18\beta_{\frac{5}{8}} + 32\beta_1 + 50\lambda_0 + 72\lambda_{\frac{2}{8}} + 98\lambda_{\frac{5}{8}} + 128\lambda_1 + 162\delta_2 = hm_{n+\frac{5}{8}} \\
2\beta_0 + 2\beta_{\frac{2}{8}} - \dfrac{9}{2}\beta_{\frac{5}{8}} - 7\beta_1 + \dfrac{25}{8}\lambda_0 + \dfrac{99}{8}\lambda_{\frac{2}{8}} + \dfrac{91}{32}\lambda_{\frac{5}{8}} - \dfrac{119}{8}\lambda_1 - \dfrac{1539}{128}\delta_2 = hm_{n+1} \\
16\beta_{\frac{2}{8}} - 96\beta_{\frac{5}{8}} + 320\beta_1 - 800\lambda_0 + 1680\lambda_{\frac{2}{8}} - 3136\lambda_{\frac{5}{8}} + 5376\lambda_1 - 8640\delta_2 = h^2 l_n \\
16\beta_{\frac{2}{8}} - 48\beta_{\frac{5}{8}} + 32\beta_1 - 80\lambda_0 - 192\lambda_{\frac{2}{8}} + 122\lambda_{\frac{5}{8}} + 192\lambda_1 - 432\delta_2 = h^2 l_{n+\frac{2}{8}} \\
16\beta_{\frac{2}{8}} + 24\beta_{\frac{5}{8}} - 40\beta_1 - 100\lambda_0 + 15\lambda_{\frac{2}{8}} + \dfrac{413}{2}\lambda_{\frac{5}{8}} + 111\lambda_1 - 270\delta_2 = h^2 l_{n+\frac{5}{8}} \\
16\beta_{\frac{2}{8}} + 96\beta_{\frac{5}{8}} + 320\beta_1 + 800\lambda_0 + 1680\lambda_{\frac{2}{8}} + 3136\lambda_{\frac{5}{8}} + 5376\lambda_1 + 8640\delta_2 = h^2 l_{n+1} \\
192\beta_{\frac{5}{8}} + 4608\beta_1 + 68160\lambda_0 + 801792\lambda_{\frac{2}{8}} + 8227968\lambda_{\frac{5}{8}} + 77064192\lambda_1 + 675946368\delta_2 = h^3 k_{n+2}
\end{cases}$$

Equation (5) is evaluated at $x = x_{n+1}(t=1)$, $x = x_{n+\frac{1}{4}}(t=\dfrac{-1}{2})$, $x = x_{n+\frac{5}{8}}(t=\dfrac{1}{4})$ and $x = x_{n+2}(t=3)$. This yields the following formulae:

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

$$
\begin{aligned}
y_{n+\frac{1}{4}} &= y_n - \frac{88137791}{7502028192} l_{n+\frac{1}{4}} h^2 - \frac{2444852359}{16804543150080} l_{n+1} h^2 - \frac{854061367}{410267166750} l_{n+\frac{5}{8}} h^2 \\
&\quad + \frac{864884159429}{8644312320000} h m_n + \frac{25652175317}{16804543150080} m_{n+1} h + \frac{9108020977}{65642746680} m_{n+\frac{1}{4}} h \\
&\quad + \frac{1101992704}{113963101875} m_{n+\frac{5}{8}} h + \frac{607}{59275284480} k_{n+2} h^3 + \frac{855155449}{288143744000} l_n h^2 \\
y_{n+\frac{5}{8}} &= y_n - \frac{513919379375}{860392609284096} l_{n+1} h^2 + \frac{5112396875}{960259608576} l_{n+\frac{1}{4}} h^2 - \frac{13252571975}{840227157504} l_{n+\frac{5}{8}} h^2 \\
&\quad + \frac{418129035395}{3540710326272} m_n h + \frac{5667019294375}{860392609284096} m_{n+} h + \frac{1048820088125}{3360908630016} m_{n+\frac{1}{4}} h \\
&\quad + \frac{274615925}{1458727704} m_{n+\frac{5}{8}} h + \frac{100625}{3034894565376} k_{n+2} h^3 + \frac{604142225}{147529596928} l_n h^2 \\
y_{n+1} &= y_n - \frac{224396423}{32821373340} l_{n+1} h^2 + \frac{4552256}{234438381} l_{n+\frac{1}{4}} h^2 + \frac{3855855616}{205133583375} l_{n+\frac{5}{8}} h^2 \\
&\quad + \frac{610085017}{4220855625} m_n h + \frac{1241893456}{8205343335} m_{n+1} h + \frac{2815556096}{8205343335} m_{n+\frac{1}{4}} h \\
&\quad + \frac{41137340416}{113963101875} m_{n+\frac{5}{8}} h + \frac{7}{57886020} k_{n+2} h^3 + \frac{6617941}{1125561500} l_n h^2 \\
y_{n+2} &= y_n - \frac{1605440031568}{8205343335} m_{n+1} h + \frac{1199003656192}{8205343335} m_{n+\frac{1}{4}} h + \frac{30183451648}{234438381} l_{n+\frac{1}{4}} h^2 \\
&\quad + \frac{202566136196}{8205343335} l_{n+1} h^2 + \frac{83816}{14471505} k_{n+2} h^3 - \frac{8833629945856}{37987700625} m_{n+\frac{5}{8}} h \\
&\quad + \frac{1199028359234}{4220855625} m_{n+\frac{1}{4}} h + \frac{5631909658}{281390375} l_n h^2 + \frac{27285660434432}{205133583375} l_{n+\frac{5}{8}} h^2
\end{aligned}
\tag{6}
$$

which constitutes the proposed THBM.

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

3. Convergence Analysis of THBM

The convergence analysis of THBM is presented in this section. The order, error constant, zero-stability, and consistency of THBM are all investigated.

Error Constant of THBM

The derived scheme belongs to the class of linear multistep methods of the form:

$$\sum_{r=0}^{s}\sigma_r y_{n+r} = h\left(\sum_{r=0}^{s}\beta_r(t)m_{n+r}\right)+h^2\left(\sum_{r=0}^{s}\mu_r(t)l_{n+r}\right)+h^3\left(\sum_{r=0}^{s}\delta_r(t)k_{n+r}\right) \quad (7)$$

Theorem 3.1

The local truncation error (LTE) of THBM is $C^{10}h^{10}y^{10}(x_n)+O(h^{11}).$

Proof

The LTE associated with Equation (7) is defined by the operator

$$LTE = \sum_{r=0}^{k}\left[\sigma_r y(x_n+rh)-h^2\beta_r m(x_n+rh)-h^3\gamma_r l(x_n+rh)\right] \quad (8)$$

The $y(x)$ in (8) is an arbitrary function. Equation (8) is expanded using the Taylor series about the point x, giving the expression below.

$$LTE = \lambda_0 y(x)+\lambda_1 hy'(x)+\lambda_2 h^2 y''(x)+\ldots+\lambda_{p+3}h^{p+3}y^{p+3}(x)$$

where the λ_0, $\lambda_1,\lambda_2\ldots\lambda_p\ldots\lambda_{p+2}$ are obtained as $\lambda_0=\sum_{r=0}^{s}\sigma_r$, $\lambda_1=\sum_{r=1}^{s}r\sigma_r$, $\lambda_2=\frac{1}{2!}\sum_{r=1}^{s}r^2\sigma_r$, $\lambda_q=\frac{1}{q!}\left[\sum_{r=1}^{s}r^q\alpha_r-q(q-1)\sum_{r=1}^{s}\beta_r r^{q-2}-q(q-1)(q-2)\sum_{r=1}^{s}\gamma_r r^{q-3}\right]$.

According to [16], Equation (8) is of order p if $\lambda_0=\lambda_1=\lambda_2=\ldots\lambda_p=\lambda_{p+1}=0$ and $\lambda_{p+j}\neq 0$. The $\lambda_{p+j}\neq 0$ is called the error constant, and $\lambda_{p+j}h^{p+j}y^{p+j}(x_n)$ is the principal local truncation error at the point x_n.

Thus, with further simplification of the above equation, the order of THBM is obtained as $p=7$ with an error constant

$$\lambda_{10}=\left[\frac{16772791859}{411152948223226675200},\frac{500772879875}{4210206189805841154048},\frac{17950369}{50189568874905600},\frac{17378108219}{3136848054681600}\right]^T$$

.

Zero-Stability of the THBM

This property concerns the ability of the proposed numerical method to remain bounded and avoid the exponential growth of errors when applied to differential equations with zero initial conditions.

Equation (7) is presented below in the form of column vectors in order to analyze the zero-stability of the scheme:

$\varepsilon=(\varepsilon_1\ldots\varepsilon_r)^T$, $d=(d_1\ldots d_j)^T$, $y_m=(y_{n+1}\ldots y_{n+j})^T$, $M(y_m)=(m_{n+1}\ldots m_{n+j})^T$, $L(y_m)=(l_{n+1}\ldots l_{n+j})^T$, $K(y_m)=(k_{n+1}\ldots k_{n+j})^T$ and matrices $A=(a_{ij})$, $B=(b_{ij})$.

Thus, Equation (6) is presented as

$$A^0 y_m = hRV(y_m)+A^1 y_n+hrv_n+h^2UL(y_m)+h^2Ul_n+h^3MW(y_m)+h^3uK_n \quad (9)$$

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

where h is the fixed mesh size within a block.

In line with Equation (9),

$$A^0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, A^1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$b = \begin{pmatrix} \frac{864884159429}{8644312320000} \\ \frac{9108020977}{65642746680} \\ \frac{1101992704}{113963101875} \\ \frac{25652175317}{16804543150080} \\ 0 \end{pmatrix}, d = \begin{pmatrix} \frac{855155449}{288143744000} \\ \frac{604142225}{147529596928} \\ \frac{6617941}{1125561500} \\ \frac{5631909658}{281390375} \end{pmatrix}, u = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{607}{59275284480} \end{pmatrix},$$

$$B = \begin{pmatrix} \frac{418129035395}{3540710326272} & \frac{610085017}{4220855625} & \frac{1199028359234}{4220855625} \\ \frac{1048820088125}{3360908630016} & \frac{2815556096}{8205343335} & \frac{1199003656192}{8205343335} \\ \frac{274615925}{10714587277041} & \frac{41137340416}{113963101875} & \frac{-8833629945856}{37987700625} \\ \frac{5667019294375}{860392609284096} & \frac{1241893456}{8205343335} & \frac{-1605440031568}{8205343335} \\ 0 & 0 & 0 \end{pmatrix}$$

$$D = \begin{pmatrix} \frac{604142225}{147529596928} & \frac{6617941}{1125561500} & \frac{5631909658}{281390375} \\ \frac{5112396875}{960259608576} & \frac{4552256}{234438381} & \frac{30183451648}{234438381} \\ \frac{-13252571975}{840227157504} & \frac{3855855616}{205133583375} & \frac{27285660434432}{205133583375} \\ \frac{-513919379375}{860392609284096} & \frac{-224396423}{32821373340} & \frac{202566136196}{8205343335} \\ 0 & 0 & 0 \end{pmatrix} \text{ and}$$

$$E = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{100625}{3034894565376} & \frac{7}{57886020} & \frac{6875}{33312384} & \frac{83816}{14471505} \end{pmatrix}$$

Equation (10) is the first characteristic polynomial of THBM, given by

$$\rho(R) = \det(RA^0 - A^1) \tag{10}$$

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

where

$$A^0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } A^1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Substituting A^0 and A^1 in Equation (10) and solving for R, the values of R are obtained as 0, 0, 0, and 1.

According to Lambert (1991), the zero-stability of THBM is guaranteed since $|R_j| \leq 1$, $j = 1$ is satisfied by $\rho(R) = 0$ and the multiplicity of the roots with $|R_j| = 1$ does not exceed three.

Consistency of THBM

By introducing the exact solution into the proposed THBM and requiring that the residual vanishes as the step size (h) approaches zero, the consistency of the model is examined. The order of consistency, established above to be seven, is determined by the rate at which the residual tends to zero.

An LMM must be zero-stable and consistent in order to be convergent (Dahlquist, 1979). These properties confirm the convergence of THBM.

4. Numerical Experiments

In this section, the accuracy of THBM is examined using five test problems. Example 3, a second-order initial value problem (IVP), and Example 4, a third-order IVP, presented in Adeyefa (2020) and Kuboye (2015), respectively, were solved using the Second- and Third-Order Model (STOM) and the Single-Order Model (SOM). Similarly, Examples 2 and 5, both third-order problems from Kuboye (2015), were solved using SOM. In addition, the solutions to these examples were obtained using the First-, Second-, and Third-Order Model (FSTOM) and the Third-Order Integrator (TOI), respectively. Test Problem 1, a first-order epidemiological model, was solved in Adeyefa (2020) and Ajileye et al. (2018) using FSTOM and the First-Order Model (FOM), respectively.

THBM is compared with all of these existing methods, and efficiency curves are presented for each problem. These curves plot the logarithm of the maximum error against the number of function evaluations. The comparative analysis and efficiency curves provide useful insights into the performance and accuracy of THBM relative to other numerical methods for the test problems considered.

Test Problems (TP)

TP 1: The *epidemiological model* **known as the** *SIR model is considered here, where* $S(t)$ *is the number of susceptible people,* $I(t)$ *is the number of people infected and* $R(t)$ *is the number of people who have recovered. It is given by the following three coupled equations:*

$$\frac{dS}{dt} = \mu(1 - S) - \beta IS \quad \text{(i)}$$

$$\frac{dI}{dt} = -\mu I - \gamma I + \beta IS \quad \text{(ii)}$$

$$\frac{dR}{dt} = -\mu R + \gamma I \quad \text{(iii)}$$

where μ, γ and β are positive parameters and y is defined as

$$y = S + I + R \quad \text{(iv)}$$

Adding Equations (i), (ii), and (iii), the evolution equation for y is obtained as

$$y' = \mu(1 - y) \quad \text{(v)}$$

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

Taking μ and y as 0.5 each, the IVP

$$y' = \frac{1-y}{2}, \quad y(0) = 0.5, h = 0.1 \tag{vi}$$

whose exact solution is

$$y(t) = 1 - \frac{1}{2}e^{-0.5t}, \tag{vii}$$

is obtained.

Table 1. TP 1 Errors Comparison

X	THBM	[2]	[3]
0.1	1.497820E-08	1.0E-10	1.218026E-13
0.2	2.849542E-08	1.0E-10	1.399991E-13
0.3	4.065853E-08	1.0E-10	1.184941E-12
0.4	5.156745E-08	2.0E-10	1.538991E-12
0.5	6.131560E-08	3.0E-10	1.110001E-12
0.6	6.999025E-08	3.0E-10	5.270229E-12
0.7	7.767291E-08	2.0E-10	2.10898E-12
0.8	8.443973E-08	3.0E-10	1.297895E-11
0.9	9.036176E-08	3.0E-10	3.08229E-11
1.0	9.550529E-08	2.0E-10	4.121925E-11

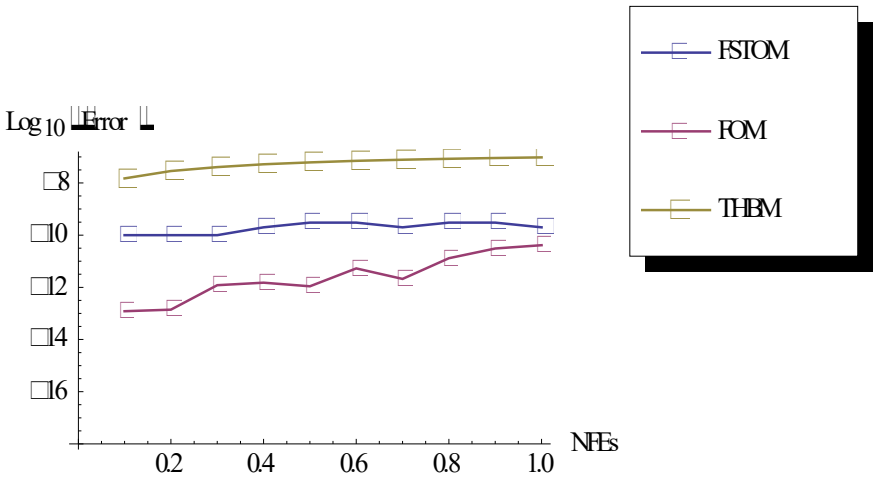


Figure 1. Efficiency Curves for TP 1

This problem is a *first-order epidemiological model solved using* FSTOM and FOM. The proposed THBM was also applied to TP1. Figure 1 shows that the two existing methods are more accurate, but they cannot handle multi-order ODEs.

TP 2: $\frac{d^3y}{dx^3} = 3\sin x, \quad y(0) = 1, y'(0) = 0, y''(0) = -2, h = 0.1$

Analytical Solution: $y(x) = 3\cos x + \frac{x^2}{2} - 2$

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

Table 2. TP 2 Errors Comparison

x- values	THBM	[2]	[14]
0.1	1.070284E-19	2.000000E-010	6.370460E-13
0.2	2.127975E-19	4.000000E-010	4.052980E-12
0.3	3.162495E-19	2.000000E-010	1.009326E-11
0.4	4.163510E-19	2.000000E-010	1.890366E-11
0.5	5.121032E-19	9.000000E-010	3.033807E-11
0.6	6.025481E-19	1.100000E-009	4.455258E-11
0.7	6.867816E-19	1.500000E-009	5.987466E-11
0.8	7.639632E-19	1.300000E-009	7.711903E-11
0.9	8.333217E-19	1.500000E-009	9.618412E-11
1.0	8.941628E-19	2.000000E-009	1.171654E-10

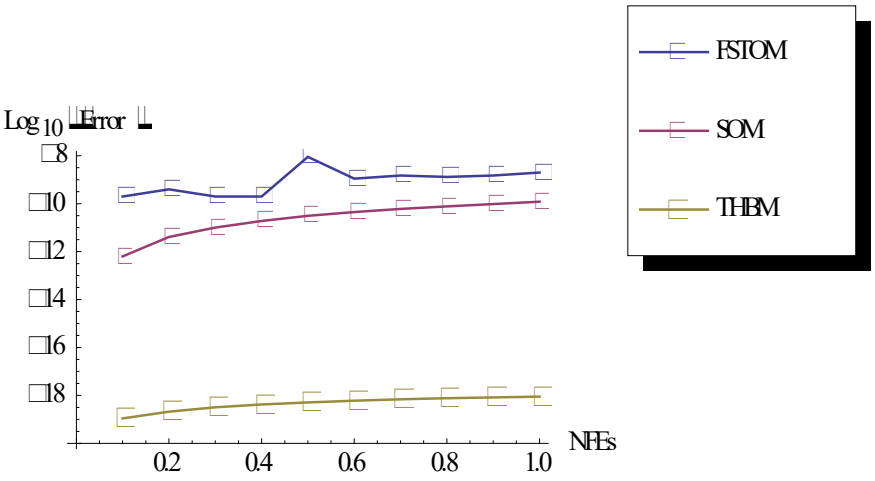


Figure 2. Efficiency Curves for TP 2

TP2 is a third-order IVP. Table 2 compares the errors of THBM with those of the existing methods (FSTOM and SOM), and the efficiency curve is displayed in Figure 2. THBM is clearly more accurate.

TP 3 : $y'' = y',\ \ y(0) = 0, y'(0) = -1,\ h = 0.1$

True Solution: $y(x) = 1 - e^{x}$

Table 3. TP 3 Errors Comparison

x- values	THBM	[1]	[14]
0.1	2.649757E-10	2.095826E-10	2.508826E-13
0.2	5.352775E-10	2.092718E-09	6.493175E-11
0.3	8.109857E-10	7.842546E-09	1.683146E-09
0.4	1.092181E-09	2.009500E-08	1.700635E-08
0.5	1.378947E-09	4.199771E-08	1.025454E-07
0.6	1.671367E-09	7.728842E-08	2.558711E-06
0.7	1.969526E-09	1.303844E-07	5.273300E-06
0.8	2.273509E-09	2.064839E-07	8.275935E-06
0.9	2.583402E-09	3.116817E-07	1.161667E-05
1.0	2.899296E-09	4.531001E-07	1.542187E-05

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

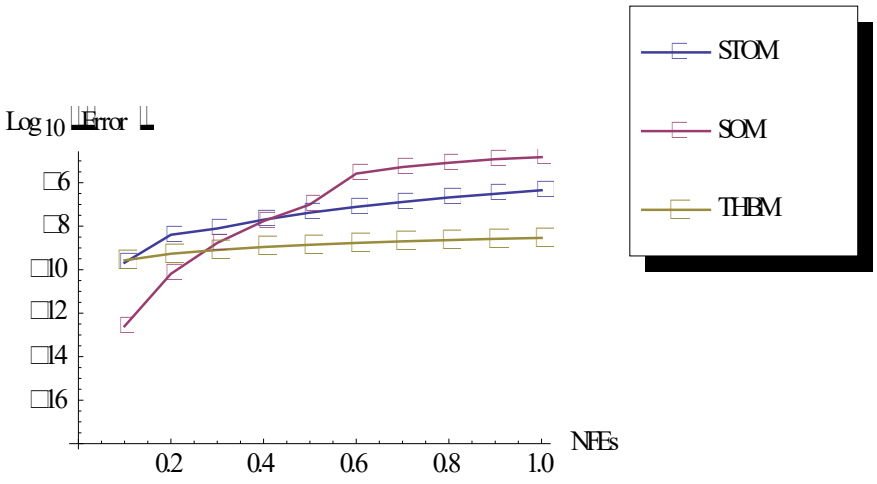


Figure 3. Efficiency Curves for TP 3

TP3 is a second-order IVP. Table 3 compares the errors of THBM with those of the existing methods (STOM and SOM), and the efficiency curve is displayed in Figure 3. The superiority of THBM in terms of accuracy is clear.

TP 4: $\frac{d^3z}{dx^3} = e^x \quad z(0) = 3, z'(0) = 1, z''(0) = 5, h = 0.1$

True Solution: $z(x) = 2(1+x^2) + e^x$

Table 4. TP 4 Errors Comparison

x- value	THBM	[1]	[14]
0.1	3.827E-20	8.881E-015	3.369E-12
0.2	8.058E-20	3.552E-014	2.160E-11
0.3	1.273E-19	8.304E-014	5.333E-11
0.4	1.790E-19	1.527E-013	9.988E-11
0.5	2.361E-19	2.460E-013	1.598E-10
0.6	2.992E-19	3.668E-013	2.511E-10
0.7	3.689E-19	5.178E-013	3.961E-10
0.8	4.460E-19	7.025E-013	5.926E-10
0.9	5.312E-19	9.254E-013	8.429E-10
1.0	6.254E-19	1.187E-012	1.144E-09

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

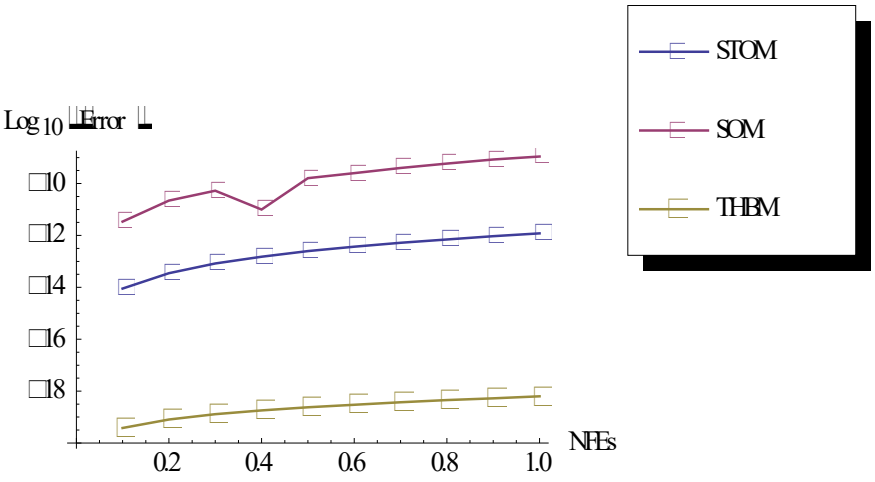


Figure 4. Efficiency Curves for TP 4

TP4 is a third-order IVP. Table 4 compares the errors of THBM with those of the existing methods (STOM and SOM), and Figure 4 shows the efficiency curve. THBM is again shown to be superior in terms of accuracy.

TP 5 : $y'''= y'(2xy''+y'),\quad y(0)=1, y'(0)=0.5, y''(0)=0,\ h=0.01.$

True Solution: $y(x)=1+\frac{1}{2}In\left(\frac{2+x}{2-x}\right)$.

Table 5: TP 5 Errors Comparison

x- values	THBM	[14]	[8]
0.01	0.000000E+000	2.508826E-13	1.194048E-013
0.02	0.000000E+000	6.493175E-11	4.086842E-013
0.03	1.000000E-024	1.683146E-09	1.016689E-012
0.04	0.000000E+000	1.700635E-08	2.139484E-012
0.05	0.000000E+000	1.025454E-07	4.083580E-012
0.06	1.000000E-024	2.558711E-06	7.350069E-012
0.07	0.000000E+000	5.273300E-06	1.279204E-011

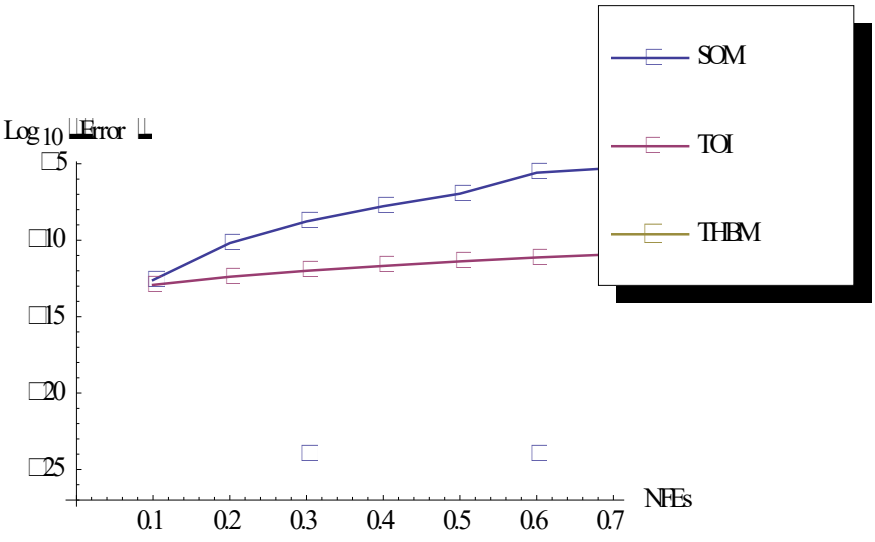


Figure 5. Efficiency Curves for TP 5

TP5 is a third-order IVP. Table 5 and Figure 5 show that THBM reproduces the exact results.

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

Discussion of Results

Overall, Tables I to V display the numerical solutions obtained from the application of THBM.

TP1 is a first-order ODE, TP3 is a second-order ODE, and TP2, TP4, and TP5 are third-order ODEs. THBM solved all of these test problems. Comparison with existing methods shows that THBM consistently delivers favorable performance, as it handles three different orders compared with the single order of the existing methods. Also, there is a reduction in error in the majority of the cases considered, with the exception of TP1, where the existing methods compared are more accurate but are limited to a single order. These tabular results provide strong evidence of the capabilities of THBM in solving first-, second-, and third-order ODEs.

5. Conclusion

This study focuses on the direct solution of first-, second-, and third-order ordinary differential equations (ODEs) using THBM. With its eighth-order accuracy, THBM is shown to be consistent. Its capability to solve differential equations of various orders provides a significant advantage over existing numerical methods. Several ODEs of orders 1, 2, and 3 are used to illustrate the efficacy of the proposed methodology. The results, presented in Tables I to V, surpass those obtained by existing methods in terms of accuracy. This demonstrates the superior performance of THBM in tackling ODEs of multiple orders and supports its potential as an advanced numerical solver in various scientific and engineering applications. Although THBM cannot handle ODEs of orders higher than three, the approach can be extended to solve stiff IVPs and partial differential equations.

6. References

Adeyefa, E.O., Olajide, O.A., Akinola, L.S., Abolarin, O.E., Ibrahim, A.A. & Haruna, Y. (2020), On direct integration of second and third order ordinary differential equations, *J. Eng. Appl. Sci*. 15 1972-1976.

Ajileye, G., Amoo, S.A. & Ogwumu, O.D. (2018), Hybrid block method algorithms for solution of first order initial value problems in ordinary differential equation, *J. Appl. Comput. Math*. 7 1-4. https://doi.org/10.4172/2168-9679.1000390.

Allogmany, R. & Ismail F. (2020), Implicit three-point block numerical algorithm for solving third order initial value problem directly with applications, *Math. 8* 1771. doi:10.3390/math8101771.

Awoyemi, D.O. (1999), A class of continuous linear multistep methods for general second-order initial value problems in ordinary differential equations, *Intern. J. Comput. Math.* 72 29-37.

Dahlquist, G. (1979), Some properties of linear multistep and one leg method for ordinary differential equations. *Department of comput. sci. Royal institute of tech*., Stockholm

Duromola, M.K., Momoh, A.L., Rufai, M.A. & Animasaun, L.L. (2022), Insight into 2-step continuous block method for solving mixture model and SIR model, *International Journal of Computing Science and Mathematics*, 14(4) 347-356.

Fatunla, S.O. (1991), Block method for second-order initial value problem, *Intern. J. Compt. Math*. 41 55-63.

Folaranmi, R.O. Adeniyi, R.B. & Adeyefa, E.O. (2016), An orthogonal based self-starting numerical integrator for third order IVPs in ordinary differential equations, *The Pacific J. Sci. Tech*. 17 73 – 86.

Henrichi, P. (1962), Discrete variable methods in ordinary differential equations, {*it John Wiley and Sons*}, New York

Hussain, K.A. Ismail, F. Senu, N. & Rabiei, F. (2017), Fourth-order improved runge–kutta method for directly solving special third order ordinary differential equations, *Iranian J. Sci. Tech.* 41 429-437. DOI 10.1007/s40995-017-0258-1

Ismail, F. Ken, Y.L. & Othman, M. (2009), Explicit and implicit 3-point block methods for solving special second order ordinary differential equations directly, *Intern. J. Math. Analy*. 3 239-254.

Jator, S.N. (2010), Solving second order initial value problems by a hybrid multistep method without predictors, *Appl. Math. Comput*. 217 4036–4046

Kayode, S.J. & Adegboro, J.O. (2018), Predictor-corrector linear multistep method for direct solution of initial value problems of second order ordinary differential equations, *Asian J. Phys. Chem. Sci*. 6 1-9.

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

Kuboye, J.O. (2015), Block methods for direct solution of higher-order ordinary differential equations using interpolation and collocation approach, *Ph.D. thesis, Universiti Utara Malaysia*.. http://etd.uum.edu.my/id/eprint/5789.

Lambert, J.D. (1991), Numerical methods for ordinary differential systems, *John Willey and Sons New York*.

Lambert, J.D. (1973), Computational methods for ordinary differential equations, *John Willey and Sons New York*.

Mohammed, U. & Adeniyi, R.B. (2014), Derivation of five-step block hybrid backward differential formulas through the continuous multi-step collocation for solving second order differential equation, *Pacific J. Sci. Tech*. 15 89 – 95.

Olabode, B.T. (2009), An accurate scheme by block method for the third-order ordinary differential equation, *Pacific J. Sci. Tech*. 10 136 – 142.

Ramos, H., Mehta, S. Vigo-Aguiar, J. A unified approach for the development of k-Step block Falkner-type methods for solving general second-order initial-value problems in ordinary differential equations, *J. Comput. Appl. Math*. Article in Press.

Ramos, H. & Rufai, M.A. (2021) An adaptive one-point second derivative lobatto-type hybrid method for solving efficiently differential systems, *Intern. J. Comput. Math*. 1-19. DOI: 10.1080/00207160.2021.1999429.

Rufai, M. A. Bruno, C. & Ramos, H. (2024), A new pair of block techniques for direct integration of third-order singular IVPs, *Applied Numerical Mathematics,* 204, 222-231.

Rufai, M. A. Bruno, C. & Ramos, H. (2023), A new hybrid block method for solving first-order differential system models in applied sciences and engineering, *Fractal and Fractional*, 7(10), 703.

Rufai, M. A. (2023), A variable step-size implementation of the hybrid Nyström method for integrating Hamiltonian and stiff differential systems, *Computational and Applied Mathematics*, 42, 156.

DOI:https//doi.org/10.22452/mjs.vol45no2.7
Malaysian Journal of Science 45(2): 62-74 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 45(2): 75-86 (June 2026) https://mjs.um.edu.my

Comparative Performance Analysis of Interpolative Decomposition and Singular Value Decomposition for Image Reduction

Wei Shean Ng[1a]*, How Hui Liew[2a], Huey Voon Chen[3a] and Hao Tian Ng[4a]

Abstract: Matrix decomposition techniques, such as Singular Value Decomposition (SVD), have traditionally been used for image compression, achieving data reduction while preserving image fidelity. Recently, randomised linear algebra algorithms, including Interpolative Decomposition (ID), have gained attention due to their efficiency in computational expense and memory consumption. This study presents a comparative analysis of ID and randomised SVD for image compression using a large collection of images from the USC-SIPI Image Database. Experimental results demonstrate that both methods effectively reduce image sizes, with SVD yielding greater compression ratios, while ID better preserves image quality at lower computational costs. The analysis also demonstrated that both techniques faced limitations when compressing images with dark or low-contrast areas, with ID performing best on images exhibiting repetitive or structured patterns. These findings indicate that ID is more suitable for memory-limited applications, such as reducing large tabular data, which is important in marketing data analysis, whereas SVD is preferable for image compression.

Keywords: *Interpolative decomposition, singular value decomposition, image reduction.*

1. Introduction

In the digital era, data reduction plays a crucial role in managing big data (Khoei & Singh, 2024). Much tabular data can contain repeated columns and rows. A collection of survey data is one such example. Duplicate data is important since it represents the information collected from the general public. Randomised matrix decomposition has emerged as an essential approach for data decomposition. By deconstructing complex matrices into simpler components, these techniques provide a framework for analysing and approximating data, improving computational efficiency and insight (Muravev et al., 2018; Kawamura & Suda, 2021).

Among randomised matrix decomposition methods (Achlioptas & McSherry, 2001), interpolative decomposition (ID) and randomised Singular Value Decomposition (SVD) have been found to be versatile in data analysis. SVD works by decomposing a matrix into singular values and orthogonal vectors, making it highly effective for tasks such as dimensionality reduction, collaborative filtering, and image compression (Su et al., 2018). Recent research by Parameshachari et al. (2023) explored SVD-based image compression using a modified truncated SVD approach, where truncated components were set to zero. Their findings showed that this method effectively reduces storage requirements while maintaining image fidelity. Hybrid models combining SVD with convolutional neural networks have shown potential in balancing fidelity and compression rate (Li et al., 2025).

Authors information:
[a]Department of Mathematical and Actuarial Sciences, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Jalan Sungai Long, Bandar Sungai Long, Cheras 43000, Kajang, Selangor, MALAYSIA. E-mail: ngws@utar.edu.my[1]; liewhh@utar.edu.my[2]; chenhv@utar.edu.my[3]; haotianng@utar.my[4]
*Corresponding Author: ngws@utar.edu.my

Further advancements in SVD-based image compression include its application to hexagonal grid images, as demonstrated by Varghese and Saroja (2021). Hexagonal grids require fewer sampling points than conventional square grids, reducing storage and computational demands (Mersereau, 1979). Inspired by the hexagonal patterns in human retinal photoreceptors, this approach converts a square-lattice image into a pseudo-hexagonal structure using methods described by Wüthrich and Stucki (1991). By applying SVD and rank-k approximations, the study demonstrated improved image compression performance on hexagonal grids compared to conventional square grids. A recent advancement by Li et al. (2024) introduced an efficient online randomised ID algorithm with a single-pass error estimator, enabling real-time image compression in streaming and large-scale applications. Moreover, matrix factorisation and low-rank approximations have also demonstrated effectiveness in addressing complex numerical models in thermal analysis and image domains (Aruchunan et al., 2022). These foundational contributions support the motivation to extend interpolative and singular value decompositions to high-dimensional data applications, including image compression.

The Interpolative Decomposition (ID), a lesser-known randomised matrix decomposition method, is found to be a basic randomised algorithm for the low-rank approximation of data with repeated entries. ID selects representative columns (or rows) from the matrix A, known as the skeleton matrix C, and interpolates across these columns to generate low-rank approximations $A \approx CZ$, where Z is called the interpolation matrix. Since the skeleton matrix C is a random selection of the columns from A, there is no requirement to allocate computer memory to store C. This approach greatly reduces computational

Received: December, 2024
Accepted: August, 2025
Published: June, 2026

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

complexity and memory usage while retaining essential data properties. Its potential for tabular data is obvious, but we are interested in experimenting with it for image reduction.

In the domain of data reduction, ID is the foundation for the CUR matrix factorisation, $A \approx CUR$, where C is the subset of the columns of A, R is the subset of the rows of A, and U refines the approximation. Beyond data reduction, ID has found applications in fields such as quantum mechanics. For example, Damle et al. (2017) employed ID in the Kohn-Sham Density Functional Theory (DFT) for solving nonlinear eigenvalue problems. They used the Selected Columns of the Density Matrix (SCDM) technique to efficiently approximate solutions by focusing on sparse and localised columns of the density matrix. Additionally, Lu and Ying (2015) demonstrated ID's effectiveness in compressing low-rank matrices representing electron systems through its use in the electron repulsion integral tensor, combined with Fast Fourier Transforms (FFTs). These studies highlight ID's ability to decrease computational demands in quantum simulations.

In the area of signal processing, Tang et al. (2023) examined ID's role in musical instrument source separation, comparing its performance with Non-negative Matrix Factorisation (NMF) and Convolutive Non-negative Matrix Factorisation (CNMF). Their results showed that ID underperformed relative to NMF and CNMF due to its random column selection, which may introduce redundant or irrelevant data into the factorised matrix. This finding highlights the importance of understanding ID's inherent limitations and adapting its application to specific problems.

Each matrix decomposition method offers distinct strengths, weaknesses, and compromises in terms of computational efficiency, accuracy, and suitability for particular fields. This paper aims to provide a comparative analysis of Interpolative Decomposition (ID) and Singular Value Decomposition (SVD) for image reduction and compression, investigating their performance, benefits, and limitations in image processing.

2. Matrix Decompositions

Randomised numerical linear algebra methods are gaining popularity in modern data analysis because they typically use less memory and have lower computational complexity (Muravev et al., 2018). They involve sampling column vectors and row vectors from the original matrix and subsequently performing random embedding. Interpolative Decomposition (ID) and Singular Value Decomposition (SVD) are examined in detail in the following sections.

Interpolative Decomposition (ID)

ID is calculated by modifying the pivoted QR, as proposed by Golub (1965). The computation cost of ID is $O(mnk)$, which is cheaper than SVD ($O(mn \cdot \min(m,n))$), where m, n, k are the number of rows, the number of columns, and the rank of the original matrix, respectively. As $k \leq \min(m,n)$, ID is beneficial when solving rank-deficient least squares problems. Many algorithms have been proposed to solve the interpolative decomposition problem (Liu & He, 2019; Bhaskara et al., 2020; Advani & O'Hagan, 2022). Python includes a library called scipy.linalg.interpolative, which calculates the interpolative decomposition of a given matrix.

Algorithm 1. An outline of the randomised interpolative decomposition (ID) for an $m \times n$ matrix A (Liberty et al., 2007):

Step 1. Randomly sample $p = 1.2k$ nonnegative integers from the range 0 to $n-1$ as indices, idx.

Step 2. Perform pivoted QR decomposition of the submatrix $A[:, \text{idx}]$ to obtain Q, R, P.

Step 3. $C = A\left[:, \text{idx}\left[P[:k]\right]\right]$ where $P[:k]$ is the first k values in P.

Step 4. Let $R_k = R[:k, :k]$ be the $k \times k$ submatrix of R, obtain the matrix Z by solving $\left(R_k^T R_k\right)Z = C^T A$ and return the ID approximation CZ as well as the indices $\text{idx}\left[P[:k]\right]$ and the $k \times n$ interpolation matrix Z.

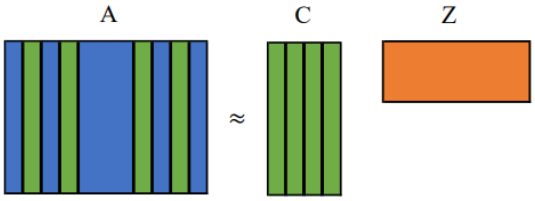


Figure 1. Illustration of ID.

Figure 1 illustrates the Interpolative Decomposition (ID) process, where the original $m \times n$ matrix A is decomposed into a skeleton matrix C of size $m \times k$ and an interpolation matrix Z of size $k \times n$, for any real number k satisfying $k \leq m$ and $k \leq n$.

Singular Value Decomposition (SVD)

Let A be an $m \times n$ matrix. SVD decomposes A into a product of three matrices, $A = U\Sigma V^T$, where U is an $m \times m$ orthogonal matrix, where its columns represent the left singular vectors, Σ is an $m \times n$ diagonal matrix with decreasing singular values, and V^T is an $n \times n$ orthogonal matrix, whose rows represent the right singular vectors.

Both U and V^T are unitary matrices, i.e., $U^T = U^{-1}$ and $V^T = V^{-1}$. The singular values in Σ indicate the importance of the corresponding singular vectors in the decomposition, with higher values showing greater importance (Strang, 2006). With a higher k value, more singular values and singular vectors from the original matrix are retained, which increases the accuracy of the approximation and reduces matrix norms (e.g., Frobenius norm or spectral norm). Consequently, the compression ratio is lower to store more information of the original data, and the compressed representation is less compact. SVD and its variants are commonly applied to perform data compression (Bentbib et al., 2022) and dimensionality reduction (Libal et al., 2020).

The randomised SVD can better deal with the standard SVD because it embeds the original large $m \times n$ matrix A to subspaces for memory-efficient low rank- k approximation $A \approx U\Sigma V$.

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

Algorithm 2. An outline of the randomised SVD for a matrix A (Halko et al, 2010):

Step 1. Draw an $n \times k$ Gaussian random matrix Ω and Ψ.

Step 2. Form the $m \times k$ sample matrix $Y = A\Omega$ and $Z = A\Psi$.

Step 3. Find orthogonal matrices Q and W such that $Y = QQ^*Y$ and $Z = WW^*Z$.

Step 4. Solve for T, the linear system $QY = T(W^*\Omega)$ and $W^*Z = T^*(Q^*\Psi)$.

Step 5. Compute the SVD of the small matrix T: $T = \hat{U}\Sigma\hat{V}^*$.

Step 6. Form $U = Q\hat{U}$ and $V = W\hat{V}$, and return the triple (U, Σ, V).

2. Methodology

To enable image compression via matrix approximation, we represent a grayscale image as an $m \times n$ matrix A, or a colour image as $A = [R, G, B]$, where R, G, B denote the red, green, and blue channels, respectively. Once the image is expressed in matrix form, matrix decomposition techniques, such as Interpolative Decomposition (ID) or Singular Value Decomposition (SVD), are employed to determine A with a lower-rank matrix $\tilde{A}$. The decomposition rank k, which determines the degree of compression, is varied from $\min(m, n)$ down to 1 to determine the smallest value of k that preserves visual quality above a defined threshold. The visual similarity between the original and compressed image is measured using the Structural Similarity Index Measure (SSIM).

Figure 2 presents the procedure for determining the optimal compression parameters using ID. The process begins with image loading and matrix extraction, followed by the definition of candidate lists for the decomposition rank k and a sparsity threshold. An empty result list is initialised to store the compression outcomes. For each combination of k and threshold value, the following steps are executed:

1. Apply ID to the matrix A;
2. Apply thresholding to the resulting coefficient matrix Z to induce sparsity;
3. Reconstruct the compressed matrix;
4. Compute key metrics, including reconstruction error, compression ratio, and SSIM; and
5. Record the results.

After evaluating all possible combinations, the results with SSIM ≥ 0.8 are retained to ensure acceptable image quality. The configuration that yields the maximum compression ratio among the filtered results is selected as the optimal setting. This approach facilitates a comprehensive exploration of the parameter space, while maintaining a balance between compression efficiency and visual fidelity.

Figure 2. Flowchart to obtain the optimal compression parameters of an image using ID.

Image Pre-processing

Images are loaded into Python using libraries like OpenCV or Pillow as pixel grids. Grayscale images form a 2D matrix with intensity values $(0 - 255)$, RGB images have three channels storing intensity data. This matrix representation enables detailed analysis of pixel colour intensities for several image-processing tasks. After loading, the image is converted to grayscale, and its matrix A is extracted. If the number of rows m exceeds the number of columns n, the matrix is transposed. The prepared image is then ready for reduction and compression.

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

Image Reduction and Compression

An iterative process determines optimal parameters k (rank) and a threshold for the Interpolative Decomposition (ID). The matrix A is approximated as $A \approx CZ$, where C is an $m \times k$ matrix of selected columns from A, and Z is the $k \times n$ interpolation matrix. Elements in Z below the threshold are set to their minimum value after normalisation using a min-max scaler. This adjustment reduces less impactful weights in Z to simplify reconstruction without greatly affecting image quality. The compressed matrix $\tilde{A}$ is then reconstructed as $CZ \approx \tilde{A} = CZ'$, where Z' is the modified interpolation matrix.

Performance Metrics

The compression error, denoted as ϵ_c, of the compressed matrix is calculated using the Frobenius norm, $\left\|A - \tilde{A}\right\|_F^2 = \epsilon_c$. Mean Squared Error (MSE) is defined as

$$\text{MSE} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\tilde{A}_{i,j} - A_{i,j}\right)^2, \quad (1)$$

where $\tilde{A}_{i,j}$ is the (i,j)-entry of the compressed matrix $\tilde{A}$, and $A_{i,j}$ is the (i,j)-entry of the original matrix A, while Peak Signal-to-Noise Ratio (PSNR) is defined as

$$\text{PSNR} = 10 \times \log_{10} \frac{(2^N - 1)^2}{\text{MSE}}, \quad (2)$$

where N is the bit depth per channel of the image. MSE and PSNR are the two metrics applied to evaluate the quality of image reconstruction tasks. MSE quantifies the average squared difference between corresponding pixels in the original image and the reconstructed image following compression. A lower MSE value indicates a smaller overall difference and potentially improved reconstruction quality. In contrast, PSNR offers a measure of the signal-to-noise ratio expressed in decibels (dB). A higher PSNR value signifies a higher ratio of the original signal power (image content) to the noise power (introduced by compression artifacts).

The Structural Similarity Index Measure (SSIM) evaluates image quality, with iterative adjustments applied to refine parameters (Wang et al., 2003). The process is adapted for RGB images with modified error computation. ID's compression performance is then compared to SVD using the same image set.

4. Results and Discussion

For a fixed k, SVD is always the best approximation for an image matrix A, theoretically. ID is a poorer approximation of A. However, for the miscellaneous set (the baboon, peppers, and other favourites) of USC-SIPI Image Database, we obtain Table 1. The table consists of image label, image size (gray-scale images are pairs, while colour images are triplets with the third position 3 indicating red, green, and blue colours), the chosen rank k with the accuracy parameter $\text{eps} = 0.05$, and the MSE calculated by using Equation (1).

According to Table 1, we can observe that the majority of k values are over 20% of the image size. However, further compression may be possible by removing visually similar or redundant images. Only images with strong horizontal similarity (e.g., gray21.512.tiff) achieve low MSE. A diverse set of images from the USC-SIPI Image Database was considered for analysis, encompassing grayscale and colour images with varying levels of detail and texture. For experimental evaluations, four representative images, "Baboon", "Sailboat on Lake", "Peppers", and "Airplane F-16" were selected to perform detailed compression analysis. These images were selected for their distinct characteristics, which provide a comprehensive understanding of how Interpolative Decomposition (ID) and Singular Value Decomposition (SVD) handle varying image complexities. However, the generalised parameters derived are intended to extend beyond these four examples, as demonstrated in the "Splash" (image test) and shown across the complete dataset in Table 1.

Table 1. k and Mean Squared Errors for UCS-SIPI misc images.

Image label	Image size	k with $\text{eps} = 0.05$	Mean Squared Error
4.1.01.tiff	(256, 256, 3)	128	14.76
4.1.02.tiff	(256, 256, 3)	123	9.32
4.1.03.tiff	(256, 256, 3)	41	24.46
4.1.04.tiff	(256, 256, 3)	32	51.83
4.1.05.tiff	(256, 256, 3)	56	48.93
4.1.06.tiff	(256, 256, 3)	118	42.43
4.1.07.tiff	(256, 256, 3)	30	30.17
4.1.08.tiff	(256, 256, 3)	44	34.56
4.2.01.tiff (Splash (Test Image))	(512, 512, 3)	51	36.98
4.2.03.tiff (Baboon)	(512, 512, 3)	329	52.23
4.2.05.tiff (Airplane (F-16))	(512, 512, 3)	103	45.58
4.2.06.tiff (Sailboat on lake)	(512, 512, 3)	217	47.24
4.2.07.tiff (Peppers)	(512, 512, 3)	156	43.33

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

5.1.09.tiff	(256, 256)	77	31.66
5.1.10.tiff	(256, 256)	132	20.24
5.1.11.tiff	(256, 256)	46	32.4
5.1.12.tiff	(256, 256)	47	45.71
5.1.13.tiff	(256, 256)	50	44.45
5.1.14.tiff	(256, 256)	120	11.02
5.2.08.tiff	(512, 512)	131	26.11
5.2.09.tiff	(512, 512)	211	31.09
5.2.10.tiff	(512, 512)	258	29.43
5.3.01.tiff	(1024, 1024)	380	19.11
5.3.02.tiff	(1024, 1024)	538	14.17
7.1.01.tiff	(512, 512)	176	13.45
7.1.02.tiff	(512, 512)	52	38.8
7.1.03.tiff	(512, 512)	134	29.13
7.1.04.tiff	(512, 512)	119	21.63
7.1.05.tiff	(512, 512)	271	13.75
7.1.06.tiff	(512, 512)	296	9.67
7.1.07.tiff	(512, 512)	258	13.57
7.1.08.tiff	(512, 512)	99	26.34
7.1.09.tiff	(512, 512)	191	18.71
7.1.10.tiff	(512, 512)	127	20.91
7.2.01.tiff	(1024, 1024)	449	7.83
boat.512.tiff	(512, 512)	156	31.3
gray21.512.tiff	(512, 512)	2	0.19
house.tiff	(512, 512, 3)	160	44.35
ruler.512.tiff	(512, 512)	44	141.35

Optimal Parameters for ID

The optimal compression parameters for the "Baboon" image were determined through 160 iterations, testing 20 k-values (from 1 to the image's rank) with eight threshold values (0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, and 0.5). Metrics, including MSE (refer to Equation (1)), PSNR (refer to Equation (2)), Structural Similarity Index Measure (SSIM), and compression ratio, were recorded. SSIM and the compression ratio were used to identify the optimal combinations of k and the threshold for effective image compression.

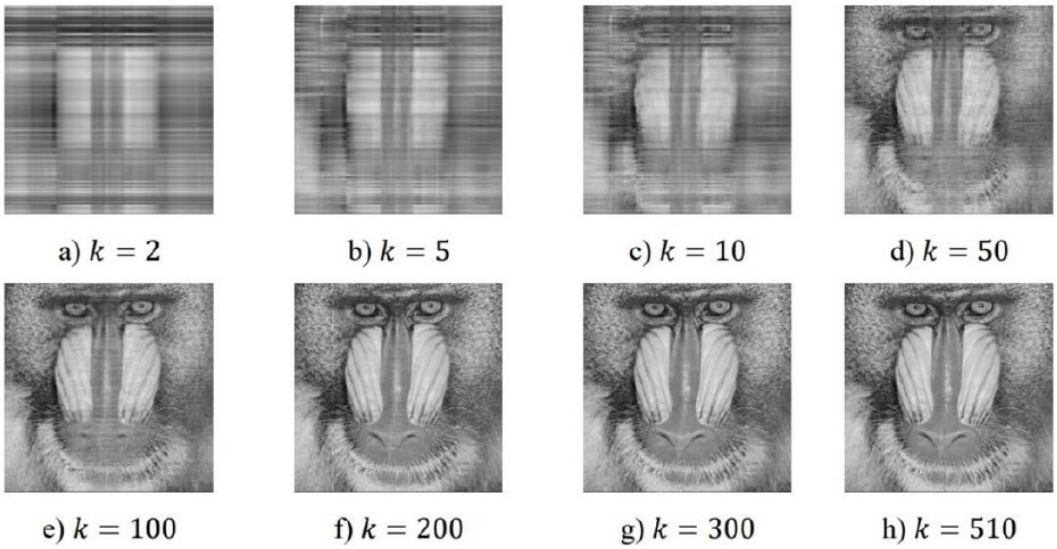


Figure 3. ID compression of different k values with a threshold of 0.01.

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

Table 2. Reconstruction Errors and Compression Ratios for Various k Values in ID Compression of the "Baboon" Image.

k	MSE	PSNR	SSIM	Compression Ratio
2	2010.716	15.0973	0.150149	5.803884
5	1463.848	16.4758	0.176223	4.852780
10	1241.537	17.1912	0.226701	4.352396
50	651.644	19.9907	0.477479	4.183527
100	373.117	22.4001	0.658740	3.584975
200	142.651	26.5842	0.853026	2.589475
300	42.50984	31.8460	0.940425	2.534132
510	25.88132	34.0090	0.989844	2.536124

Further experiments were conducted to investigate the impact of varying rank parameters (k) on image compression and quality. As illustrated in Figure 3 and Table 2, ID decomposition was applied to the "Baboon" image using k values ranging from 2 to 510. The analysis reveals that adjusting the rank parameter has a great effect on reconstruction quality and compression efficiency. As k increases, the MSE consistently decreases, showing improved image fidelity. Simultaneously, PSNR and SSIM values rise, reflecting better preservation of image details. However, this occurs at the cost of a reduced compression ratio, highlighting the trade-off between image quality and storage efficiency. These findings demonstrate that ID can be effectively fine-tuned to balance compression and quality based on specific conditions. While these results exhibit ID's ability to generalize across various parameter settings within the USC-SIPI Image Database, extending this evaluation to other datasets would further confirm its robustness and applicability for diverse image types.

Numerical analysis established that a threshold value of 0.01 is optimal for all four images using ID across most k-values. Figure 4 illustrates a comparison between uncompressed and compressed images using optimal compression parameters in ID for both grayscale and colour images. To generalize a k-value for compressing four different images, the average k-value was calculated and is shown at the bottom of the figure.

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

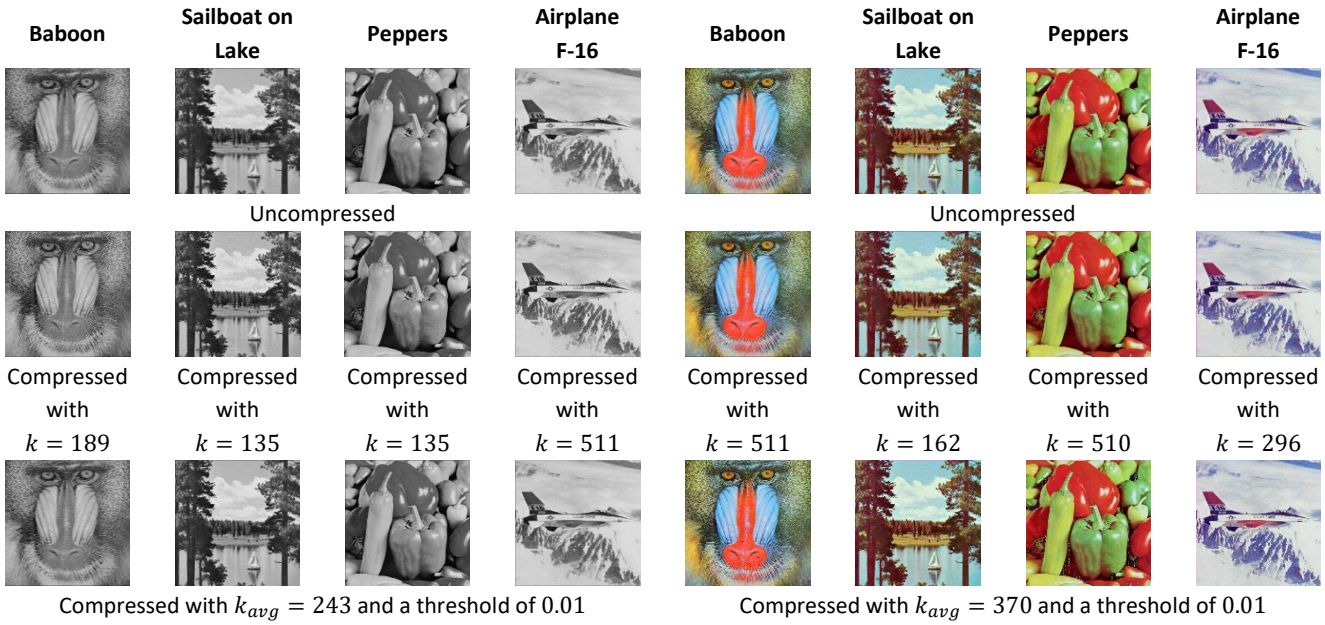


Figure 4. Comparison of uncompressed and compressed images using optimal compression parameters in ID: grayscale images on the left and colour images on the right.

Table 3. Compression results for the four colour images using ID with $k_{avg} = 370$ with a threshold of 0.01.

Image	Baboon	Sailboat on Lake	Peppers	Airplane F-16
MSE	34.57612	7.379738	200.8264	0.687366
PSNR (dB)	33.70509	40.14759	30.47195	49.79675
SSIM	0.966854	0.977783	0.947861	0.992076
File Size (bytes)	123864	90560	82346	63464
Compression Ratio	6.350287	8.685645	9.552037	12.39399

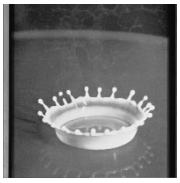
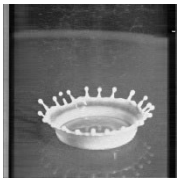
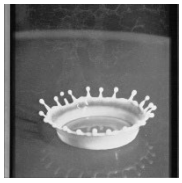
Table 3 shows that all four colour images achieved high SSIM values, indicating good quality. Compression ratios were generally acceptable, except for “Sailboat on Lake” and “Peppers”. For “Sailboat on Lake”, the difference was from the gap between the average k-value and its optimal k-value. For “Peppers”, the larger compression ratio difference was due to lines of dots in darker areas (visible in Figure 4), caused by low-intensity regions being discarded with a small k-value, resulting in a higher MSE and a need for more information to accurately store the image.

To evaluate performance, compression parameters were tested on new images, using the Splash image for both grayscale and colour. The results, compared with their optimal values, are presented in Table 4. Table 4 and Figure 5 demonstrate that grayscale image compression results were acceptable, with improved quality and a tolerable compression ratio decrease despite a larger k-value. However, colour image compression faced issues, similar to those seen with the "Peppers" image, where lines of dots appeared in darker areas. This highlights the challenges of compressing images with dark regions using low-rank approximations. For 512×512 grayscale images, $k = 243$ and a threshold of 0.01 are recommended, while for colour images, $k = 370$ with the same threshold is recommended, although caution is needed for images with dark areas.

Table 4. Comparison results between compressing using optimal and generalised parameters on both grayscale and colour images using ID.

Image Mode	Grayscale		Colour	
Compression Parameters Used	Optimal	Generalised	Optimal	Generalised
k	55	243	484	370
Threshold	0.01	0.01	0.01	0.01
MSE	30.07837	1.971176	0.027419	159.4838
PSNR (dB)	33.34826	45.18355	68.61479	42.46367
SSIM	0.844145	0.978776	0.999614	0.974872
File Size (bytes)	42574	45729	55700	66448
Compression Ratio	3.922676	3.652037	14.12158	11.83741

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

Grayscale		
Original	**Optimal**	**Using General Parameters**
		
Uncompressed	$k = 55$, threshold of 0.01	$k_{avg} = 243$, threshold of 0.01

Colour		
Original	**Optimal**	**Using General Parameters**
		
Uncompressed	$k = 484$, threshold of 0.01	$k_{avg} = 370$, threshold of 0.01

Figure 5. Comparative images of uncompressed, optimally compressed, and generally compressed versions using ID.

Optimal Parameters for SVD

The optimal compression parameters for SVD were determined similarly to ID, but without a threshold parameter. In total, 20 equidistant k-values were tested, and configurations with SSIM values below 0.8 were excluded. The optimal k-value was selected based on the highest compression ratio from the remaining results. Figure 6 presents a comparison of uncompressed and compressed images using optimal SVD compression parameters, with grayscale images shown on the left and colour images on the right.

The compression parameters were tested on the same image to assess their performance. Splash image was employed to evaluate the parameter for both grayscale and colour images. The results were compared with their optimal values and were listed in Table 5. The table shows that using a higher k-value improves image quality for both grayscale and colour images, despite a considerable decrease in the compression ratio. For colour images, the anomaly lines became less noticeable, resulting in significant visual improvements at the cost of larger file sizes.

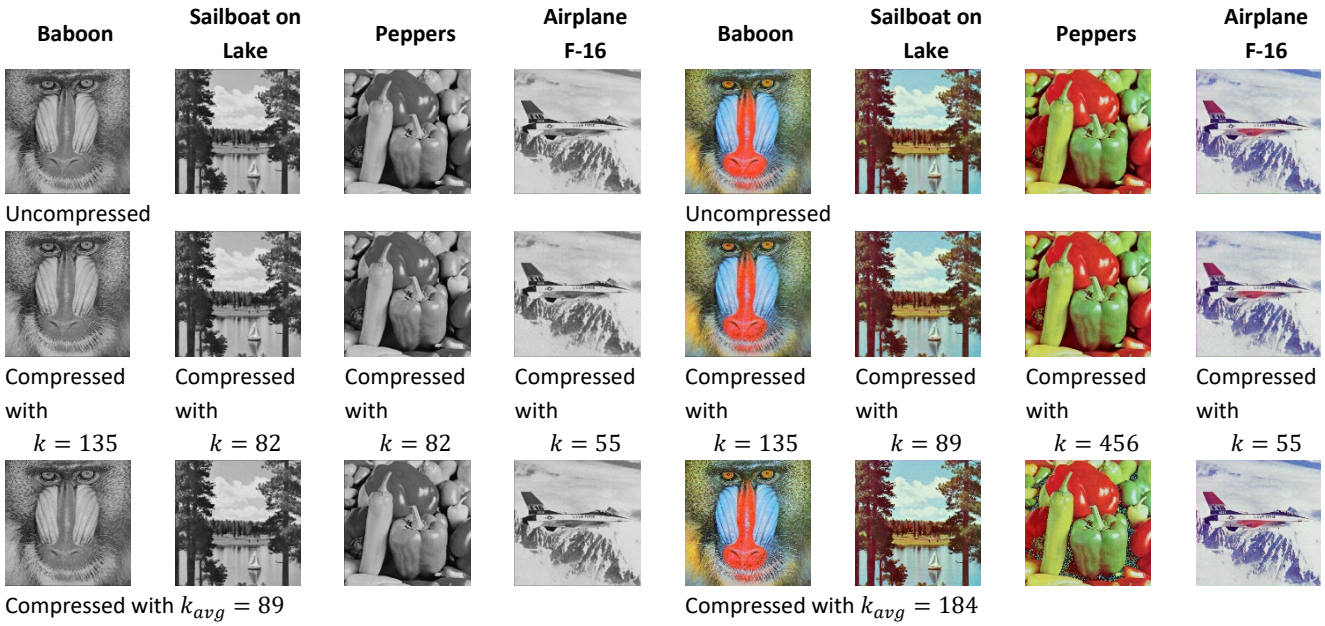


Figure 6. Comparison of uncompressed and compressed images using optimal SVD compression parameters: grayscale images on the left and colour images on the right.

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

Table 5. Comparison results between compressing using optimal and the general parameters for both grayscale and colour images using SVD.

Image Mode	Grayscale		Colour	
Compression Parameters Used	Optimal	Generalised	Optimal	Generalised
k	28	89	28	184
MSE	31.92671	6.080673	451.7643	24.06033
PSNR (dB)	33.08926	40.29129	29.71018	40.92835
SSIM	0.836819	0.937818	0.835039	0.971149
File Size (bytes)	33648	41180	51226	58242
Compression Ratio	4.963267	4.055464	15.35494	13.50524

Comparative Analysis of ID-method and SVD-method

The performance of ID and SVD in image compression highlights a trade-off between image quality (SSIM) and file size (compression ratio), as they are inversely related. Figure 7 shows that ID achieves higher SSIM values than SVD for the tested images when the rank k exceeds the rank of SVD, as determined by our methodology. Grayscale images exhibit lower SSIM values compared to colour images, as colour images retain more information across their three channels.

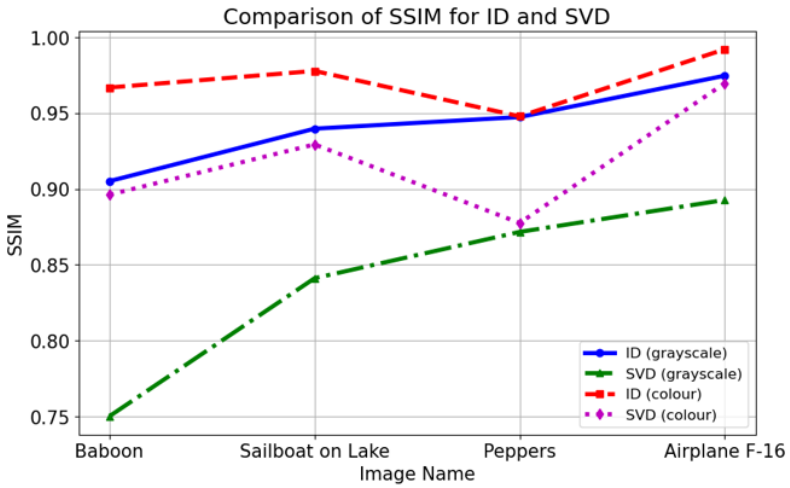


Figure 7. Comparison of SSIM using the generalised parameters for ID and SVD.

5. Computational Performance Comparison

The efficiency of matrix decomposition techniques is typically evaluated not only by their ability to reconstruct data accurately but also by their computational efficiency. This section offers a comparative analysis of the runtime and memory consumption of Interpolative Decomposition (ID), emphasizing its practical implications in image compression tasks.

Experimental Setup

The experiments were conducted on a machine equipped with an Intel Core i7-11700 CPU, 16 GB of RAM, and running Python 3.10 on Windows 11. ID computations were performed using the scipy.linalg.interpolative module. Runtime measurements were recorded with the time module, and peak memory consumption was monitored using the memory profiler package. ID was evaluated on a set of 512×512 colour images from the USC-SIPI Image Database, ensuring consistent matrix rank (k) values based on predefined compression parameters.

Results and Analysis

Table 6 presents the runtime results for ID on selected 512×512 colour images with a rank of $k = 370$. The experimental data demonstrate that Interpolative Decomposition (ID) consistently achieved efficient runtimes for these image decompositions. ID completed matrix decompositions faster for images with detailed features such as textures and edges, as seen with images like Baboon and Peppers.

It is vital to highlight that on standard CPUs, computing the full SVD of a 512×512 matrix typically takes between 1.5 to 3 seconds. This is consistent with theoretical expectations, as the computational complexity of SVD is $O(mn \cdot \min(m,n))$, which is generally more demanding compared to the $O(mk)$ complexity of ID.

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

Table 6. Runtime of ID on the selected 512×512 colour images for $k = 370$.

Image	Runtime (seconds)
Baboon	1.303
Sailboat on Lake	1.088
Peppers	0.958
Airplane F-16	0.841
Splash (Test Image)	0.721

These results are consistent with theoretical expectations, as the computational complexity of ID is $O(mk)$, where m is the number of rows and k is the rank of the approximation. The results suggest that ID not only achieves faster runtimes but also maintains stable performance across different types of image content, even at high resolutions (512×512).

Practical Implications

The computational benefits of ID have significant real-world implications. It is beneficial in applications where speed and memory efficiency are essential, such as real-time image processing, mobile computing, or large-scale data management. ID performs well because it uses less computing power and memory, while still delivering good image quality. This makes it a practical choice for devices with limited resources.

Our study demonstrates that ID achieves faster runtimes compared to typical SVD performance on the same image resolutions while maintaining consistent performance across different types of image content. This highlights its suitability for applications requiring rapid decompositions with minimal resource consumption, especially in real-time processing and edge computing environments.

Interpolative Decomposition (ID) and Singular Value Decomposition (SVD) have demonstrated their effectiveness in various real-world applications. ID is applied in quantum mechanics for solving nonlinear eigenvalue problems, particularly within Kohn-Sham density functional theory (DFT), where it efficiently approximates the density matrix through selected columns (Damle et al., 2017). Additionally, ID is examined in signal processing for musical instrument source separation, although its randomised column selection occasionally introduces redundancy (Tang et al., 2023). In contrast, SVD is extensively employed in medical image compression, video streaming optimization, and AI-based image classification. Its ability to retain key image features while reducing dimensionality makes it suitable for applications requiring both storage efficiency and image fidelity (Parameshachari et al., 2023; Libal et al., 2020). Furthermore, SVD has been adapted for hexagonal grid image compression, inspired by the structure of the human retina, demonstrating enhanced storage efficiency (Varghese & Saroja, 2021). These examples demonstrate the versatile applications of ID and SVD across various domains, underscoring their significance in data reduction tasks.

6. Conclusion

This study examined the effectiveness of Interpolative Decomposition (ID) and Singular Value Decomposition (SVD) for image data reduction through low-rank approximation. ID was shown to maintain matrix sparsity and reusing original columns, resulting in reduced storage and computational cost. However, it requires a higher rank k to achieve acceptable quality for complex or dense images.

Experimental results demonstrated that both ID and SVD yield comparable performance in image compression. Images with high variability or dark regions, such as “Baboon” and “Peppers,” required larger values of k to retain visual quality. SVD consistently achieved better compression ratios, making it suitable where storage efficiency is a priority. In contrast, ID better preserved image fidelity, especially under constrained computational resources. Generalised compression parameters for grayscale and colour images were proposed and tested across various images. For grayscale images, ID with a threshold of 0.01 and SVD with moderate k values produced effective results. For colour images, a similar thresholding was applied, although caution is needed when compressing images with dark regions. For reproducibility, generalised compression parameters were proposed: for 512×512 grayscale images, $k = 243$ with a threshold of 0.01 for ID and $k = 89$ for SVD; for colour images, $k = 370$ with a threshold of 0.01 for ID and $k = 184$ for SVD.

The findings obtained in this study indicate that method selection should be guided by specific application goals, whether prioritizing speed and memory usage (ID) or optimal storage and accuracy (SVD). Future work may explore the application of these methods on real-world health tabular data, such as the CDC Diabetes Health Indicators. By using ID on health data, we can analyse redundant information and highlight important health indicators.

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

7. References

Achlioptas, D., McSherry, F. (2001). Fast computation of low rank matrix approximations. *In Proceedings of the thirty-third annual ACM symposium on Theory of computing (STOC '01). Association for Computing Machinery*, New York, NY, USA, 611-618. https://doi.org/10.1145/380752.380858

Advani, R., O'Hagan, S. (2022). Efficient algorithms for constructing an interpolative decomposition. arXiv.

Aruchunan, E., Siri, Z., Noor Aziz, M. H. B., Ab Wahab, M. H. B., Muthuvalu, M. S., Sulaiman, J. (2022). Solution of peak junction temperature with Crank-Nicolson and SOR approach. *In Intelligent Systems Modeling and Simulation II: Machine Learning, Neural Networks, Efficient Numerical Algorithm and Statistical Models* (pp. 225-234). Springer. https://doi.org/10.1007/978-3-031-04028-3_15

Bentbib, A. H., Kreit, K., Labaali, I. (2022). Randomized tensor singular value decomposition for multidimensional data compression. *In 2022 11th International Symposium on Signal, Image, Video and Communications (ISIVC)* (pp. 1-6). El Jadida, Morocco: IEEE.

Bhaskara, A., Lattanzi, S., Vassilvitskii, S., Zadimoghaddam, M. (2020). Residual based sampling for online low rank approximation. In *2020 Information Theory and Applications Workshop (ITA)*. San Diego, CA, USA: IEEE. 1-19. https://doi.org/10.1109/ITA50056.2020.9244974.

Damle, A., Lin, L., Ying, L. (2017). SCDM-k: Localized orbitals for solids via selected columns of the density matrix. *Journal of Computational Physics*, 334, 1-15. https://doi.org/10.1016/j.jcp.2016.12.053.

Golub, G. (1965). Numerical methods for solving linear least squares problems. *Numerische Mathematik*, 7(3), 206-216.

Halko, N., Martinsson, P. G., Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review,* 53(2), 217-288. https://doi.org/10.1137/090771806.

Kawamura, H., Suda, R. (2021). Least upper bound of truncation error of low-rank matrix approximation algorithm using QR decomposition with pivoting. *Japan Journal of Industrial and Applied Mathematics*, 38(3), 757-779. https://doi.org/10.1007/s13160-021-00459-x.

Khoei, T. T., Singh, A. (2024). Data reduction in big data: a survey of methods, challenges and future directions, *International Journal of Data Science and Analytics*, 2364-4168.

Libal, U., Baras, J. S., Johansson, K. H. (2020). Dimensionality reduction of volterra kernels by tensor decomposition using higher-order SVD. In *2020 59th IEEE Conference on Decision and Control (CDC)*. Jeju, Korea (South): IEEE. 5935-5941. https://doi.org/10.1109/CDC42340.2020.9303951.

Li, S., Lu, J., Hu, Y., Mattos, L. S., Li, Z. (2025). Towards scalable medical image compression using hybrid model analysis. *Journal of Big Data*, 12(45). https://doi.org/10.1186/s40537-025-01073-1

Liberty, E., Woolfe, F., Martinsson, P. G., Rokhlin, V., Tygert, M. (2007). Randomized algorithms for the low-rank approximation of matrices. *Proceedings of the National Academy of Sciences* 104(51), 20167-20172. https://doi.org/10.1073/pnas.0709640104(2007).

Liu, W., He, M. (2019). Accelerating solution of volume-surface integral equations with multiple right-hand sides by improved skeletonization techniques. *IEEE Antennas and Wireless Propagation Letters* 18(10), 2006-2010.

Lu, J., Ying, L. (2015). Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost. *Journal of Computational Physics* 302, 329-335. https://doi.org/10.1016/j.jcp.2015.09.014.

Mersereau, R. M. (1979). The processing of hexagonally sampled two-dimensional signals. *Proceedings of the IEEE* 67(6), 930-949. https://doi.org/10.1109/PROC.1979.11356.

Muravev, A., Tran, D. T., Iosifidis, A., Kiranyaz, S., Gabbouj, M. (2018). Acceleration approaches for big data analysis. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. Athens: IEEE, 311-315. https://doi.org/10.1109/ICIP.2018.8451082.

Parameshachari, B. D., Kumar, D. S., Prafulla, P. S., Yashwanth, J. (2023). Singular Value Decomposition (SVD) based optimal image compression technique. *In 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*. Bengaluru, India: IEEE, 1-6.

Strang, G., 2006. Linear algebra and its applications. 4th ed. Boston: Cengage Learning.

Su, Q., Wang, G., Zhang, X., Lv, G., Chen, B. (2018). A new algorithm of blind color image watermarking based on LU decomposition. *Multidimensional Systems and Signal Processing*, 29(3), 1055-1074. https://doi.org/10.1007/s11045-017-0487-7.

Tang, W. K. A., Ng, W. S., Liew, H. H. (2023). Separation of two musical instruments using matrix factorisation techniques. *International Journal of Applied Mathematics* 36(3), 425. http://dx.doi.org/10.12732/ijam.v36i3.8.

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

Varghese, P., Saroja, G. A. S. (2021). Hexagonal image compression using singular value decomposition in Python. In *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)* IEEE. 211-215. https://doi.org/10.1109/ACCESS51619.2021.9563312.

Wang, Z., Simoncelli, E. P., Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. *In The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers 2003*, 2, 1398-1402. IEEE. https://doi.org/10.1109/ACSSC.2003.1292216.

Wüthrich, C. A., Stucki, P. (1991). An algorithmic comparison between square-and hexagonal-based grids. *CVGIP: Graphical Models and Image Processing*, 53(4), 324-339. https://doi.org/10.1016/1049-9652(91)90036-J.

DOI:https//doi.org/10.22452/mjs.vol45no2.8
Malaysian Journal of Science 45(2): 75-86 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 45(2): 87-95 (June 2026) https://mjs.um.edu.my

First Report and Molecular Identification of *Sergentomyia babu babu* Annandale (1910) (Diptera: Psychodidae: Phlebotominae) in Iraq

Younis Ibrahim Mohammed Ali Al-Joary[1a]* and Muneef Abid Mustafa Al-Hamdani[2a]

Abstract: This study aims to investigate the morphology and molecular identification of sandflies, one of the most significant disease vectors in Iraq, across the provinces of Nineveh Governorate (northern Iraq) during the period between May 2023 and June 2024. The male genitalia, female pharynx and spermatheca of the subfamily Phlebotominae were used as the basis for the morphological identification. For molecular identification, the conventional polymerase chain reaction (PCR) was used to amplify the mitochondrial cytochrome c oxidase gene subunit 1 gene, and the DNA sequences were analyzed and compared to isolates listed in the National Center for Biotechnology Information (NCBI). The findings showed that 8,355 sandflies (3,779 males and 4,575 females) and seven species were collected, belonging to the two genera: *Phlebotomus* and *Sergentomyia*. The genus *Phlebotomus* includes five species: *Phlebotomus papatasi*, *Phlebotomus sergenti*, *Phlebotomus alexandri*, *Phlebotomus kazeruni*, and *Phlebotomus tobbi*. As for the genus *Sergentomyia*, it includes two species: *Sergentomyia dentata* and *Sergentomyia babu babu*. The results of phenotypic and molecular identification showed that the *S. babu babu* was recorded for the first time in Iraq, as three isolates were documented in NCBI. The evolutionary tree and genetic divergence analysis confirmed the great genetic similarity between *S. babu babu* and the Indian and Pakistani species, which confirms the correctness of the identification. This study is the second in the country to address the molecular identification of the *Sergentomyia* species. So, our emphasis is on the necessity of coordinating efforts to perform additional morphological and molecular taxonomic investigations in order to accurately identify the sandflies in Iraq, particularly those belonging to the genus *Sergentomyia*.

Keywords: *Phlebotominae, Sandflies, leishmaniasis, Sergentomyia spp.*

1. Introduction

Insect-borne diseases represent a major public health problem worldwide. One of the most important vectors of these diseases is the sand fly, which belongs to the subfamily Phlebotominae and comprises blood-feeding insects associated with the transmission of many pathogens, such as *Leishmania*, viruses, and bacteria. Recent studies have indicated the possible role of sand fly species in the transmission of *Trypanosoma* parasites and *Onchocerca* worms. Given their substantial importance, sand fly species have been identified in endemic areas of both the Old and New Worlds, as these insects are associated with these diseases (Srisuton *et al*., 2019; Brilhante *et al*., 2020; Cecílio *et al*., 2022).

Sand fly taxonomy is a cornerstone of scientific research aimed at identifying groups of disease vectors and the pathogens transmitted by these insects. Traditionally, the use of phenotypic characters in sand fly taxonomy, which began in 1786, has been the main guide for species identification, but the use of genetic sequences provides an integrative and accurate approach for demarcating these species (Rodrigues & Galati, 2023; Galati & Rodrigues, 2023).

Species of *Sergentomyia* are found throughout the Old World but are particularly abundant in tropical Africa, the Middle East, Australia, and the Indian subcontinent (Theodor, 1948). Although the genus *Phlebotomus* in the Old World and the genus *Lutzomyia* in the New World are responsible for transmitting cutaneous and visceral leishmaniasis in endemic areas, recent evidence has emerged that members of the genus *Sergentomyia* may also be involved as vectors of *Leishmania* species to mammals (Maia & Depaquit, 2016). Experimentally, *S. baghdadi* has been shown to transmit both *Leishmania tropica* and *Leishmania major* in Iraq (Al-Mashhadani, 2006). In molecular studies, the DNA of *Leishmania* has been identified in *S. dentata,* but its species has not been identified in Turkey (Özbel *et al*., 2016).

Iraq is considered one of the endemic areas for leishmaniasis in the Middle East, with the disease being widespread across all governorates and more prevalent in warm and hot areas, especially in the southern regions of the country. For this reason, studies on the identification of sandflies have been numerous in those governorates, while fewer have been conducted in the northern governorates. Previously, three significant phenotypic surveys were conducted in Iraq, resulting in the identification of several sand fly species, including species from the genus *Sergentomyia*. However, the *S. babu babu* was not identified in any of these investigations (Adler & Theodor, 1929; Pringle, 1953; Abu Al-Hab, 1979). Regarding the molecular identification of the genus *Sergentomyia*, there are only two previous studies, the first conducted in Babylon Governorate, but without confirming the species (Al-Obaidi, 2021). The second is our previous study, during which we identified *S. dentata* at the molecular level for the first time in Iraq (Al Joary & Al Hamdani, 2024). Given the above and the health and veterinary importance of sand fly species, the present study was conducted in Nineveh Governorate, northern Iraq.

Authors information:
[a]Department of Biology, College of Science, University of Mosul, Mosul, Nineveh, IRAQ. Email: yonsbio127@uomosul.edu.iq[1]; munsbio6@uomosul.edu.iq[2]
*Corresponding Author: yonsbio127@uomosul.edu.iq

Received: December, 2024
Accepted: August, 2025
Published: June, 2026

DOI:https//doi.org/10.22452/mjs.vol45no2.9
Malaysian Journal of Science 45(2): 87-95 (June 2026)

2. Methods

Study Area, Collection of Sandflies, and Morphological Identification

Sand fly Specimens were collected during the period between May 2023 until June 2024 in nine provinces of Nineveh Governorate (northern Iraq), with a focus on areas known for leishmaniasis outbreaks: Mosul (43.159 ◦E, 36.349◦N), Alhamdanyia (43.399 ◦E, 36.190◦N), Al-Baaj (41.716◦E, 36.046◦N), Al-Hadar (42.734◦E, 35.575◦N), Sinjar (41.863◦E, 36.314◦N), Tel Kaif (43.120◦E, 36.490◦N), Tel Afar (42.402◦E, 36.362◦N), Al-shikhan (43.440◦E, 36.785◦N) and Makhmur (43.576◦E, 35.774◦N). The latitude and longitude of each study site were determined using Google Maps (Figure 1).

Photocatalysis-based Mosquito Killer light traps (China) were used to collect specimens, and the traps were installed at the study sites from 5:00 p.m. until 9:00 a.m. the next day, Baghdad time. Then, the specimens of sand fly were isolated from other insects using a dissecting microscope) Wild M7A, Switzerland), magnifying lenses, and at times with the naked eye, based on the initial characteristics of the sandflies, including their small size, gray color, large black eyes, raised wings, and long legs. When dead, the insect can also be distinguished by an inverted (L) letter (Pringle, 1953). Then, the specimens were separated by sex (male and female). The isolated specimens were preserved in test tubes containing 70% ethyl alcohol.

Prior to morphological identification, the method described by Dantas-Torres *et al*. (2014) was followed to prepare the specimens for microscopic examination, with the specimens being placed in potassium hydroxide solution for about 24 hours to make the internal structures more visible. The head and abdomen regions were then separated from the rest of the body and mounted on slides using Canada balsam. The morphological keys of the *S. babu babu* mentioned by Kakarsulemankhel (2004) were used for identification. The slides were examined using a Navite light microscope (China) at various magnifications, and then photographed using an RS500W digital camera (China).

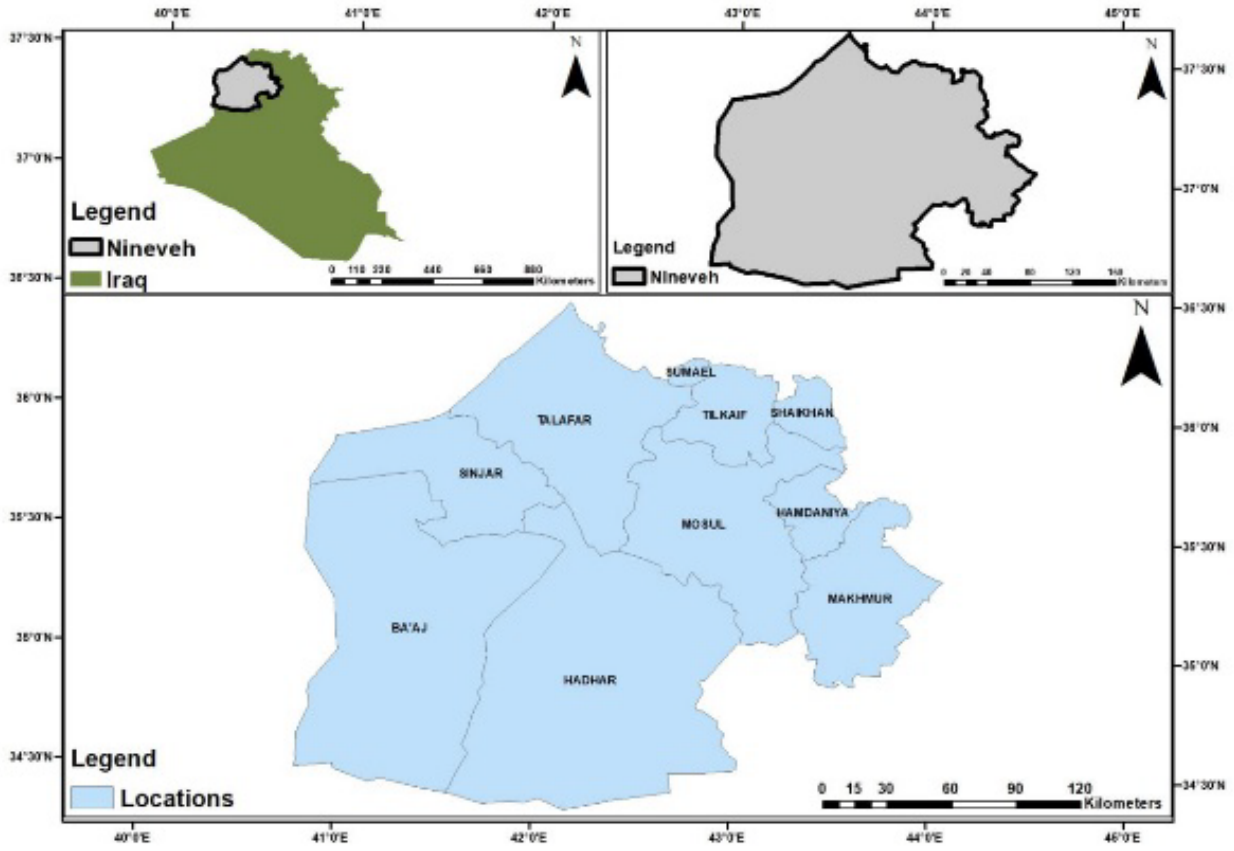


Figure 1. Map of study regions, designed using ArcGIS Pro3.0.2 program

Molecular Identification

The genomic DNA of sandflies was extracted following the manufacturer's instructions of Geneaid DNA Isolation Kit (Taiwan). The concentration and purity of the extracted DNA were measured using a BioDrop device (Finland) at a wavelength of 260 nm, and the purity of the extract was estimated from the 260/280 nm absorbance ratio. PCR with specific primers for amplification of the mitochondrial cytochrome c oxidase gene subunit 1 (*COX)* was performed using Taq Master Mix 2X (PROMEGA, USA) and the primers LCO 1490 (5-GGTCAACAAATCATAAAGATATTGG-3), HCO 2198 5-(TAAACTTCAGGGTGACCAAAAAATCA-3) designed by Macrogen (Korea) (Folmer *et al*., 1994). After preparing the PCR tubes, they were placed in a thermal cycler (BioRad, USA). The reactions were performed using the following program: 1 cycle of

DOI:https//doi.org/10.22452/mjs.vol45no2.9
Malaysian Journal of Science 45(2): 87-95 (June 2026)

primary denaturation for 5 min and 94°C, followed by denaturation (40 cycles of 40 seconds at 94°C), annealing (40 cycles of 4 seconds at 52°C), and extension (40 cycles of 1 min and 72°C). Then, final extension (1 cycle of 5 min and 72°C), followed by cooling at 4°C. Gel electrophoresis (1% concentration) was conducted for PCR products and then examined under a gel documentation system (Bio-Rad, USA).

DNA Sequencing and Phylogenetic Tree

The positive PCR products that showed primer binding at 700 bp were sent to Psomagene company (USA) for sequencing of the *COX* gene by the Sanger method. After receiving the sequencing results, the high-quality sequences were matched with the global isolates documented in NCBI using the Blast program and recorded. For the purpose of constructing the phylogenetic tree of sandflies under study, the *COX* sequences were aligned with the corresponding species sequences documented in the NCBI of Old World sand flies, considering that Iraq is located within the Old World, using the MEGA 11 program and according to the following steps: alignment method (ClustalW), Test (Neighbor-Joining (NJ) tree, Model (Kimura-2 parameter) and Bootstrap *P* values (1000). Finally, evolutionary divergence was estimated using the Kimura-2 parameter model (K2P) and Bootstrap *P* values (1000) (Tamura *et al*., 2021).

3. Results

Sandflies Collection and Morphological Identification

In total, 8,355 sandflies were collected during the current study period, belonging to two genera: *Phlebotomus* and *Sergentomyia*. The genus *Phlebotomus* included five species: *Ph. papatasi* (males: 2471; females: 3111), *Ph. sergenti* (males: 1,227; females: 1,252), *Ph. alexandri* (males: 70; females: 203), *Ph. kazeruni* (Female: 5) and *Ph. tobbi* (males: 4; females: 2), While the genus *Sergentomyia* includes two species: *S. dentata* (1 male) and *S. babu babu* (males: 7; females: 2). Further details about molecular identification of the recorded species (except for *S. babu babu*) were addressed during our study (Al Joary & Al Hamdani, 2024). *Sergentomyia babu babu* (Annandale, 1910) was recorded for the first time in Iraq and in Nineveh Governorate. Nine specimens were found in three regions (Alhamdanyia, Al-Hadar, and Makhmur) during a two-month period (July 2023 and May 2024). The molecular identification confirmed the phenotypic identification of this species, which is briefly characterized as follows:

Male Genitalia

The stylets have four spines, two terminals and two subterminal. The paramere is long, with the basal and middle parts of its body extending downwards, with a ventral tubercle bearing three short spines; the body of the paramere has 8 to 10 long setae, gradually narrows anteriorly, and finally curves laterally like a bird's head. The aedeagus is long, straight, and highly pigmented, tapering to an almost colorless, rounded, and blunt tip; the genital pump is filiform with transverse striations.

Female Spermatheca

An oblong, oval, thick-walled capsule in which the sperm ducts communicate through a common opening in the reproductive atrium.

Female Pharynx

The pharyngeal armature consists of 4 to 5 rows of scales with pointed posterior ends and longer basal spicules. Most of the pharyngeal armature is confined to the basal part of the pharyngeal bulb, and the lateral walls of the pharynx on the side of the bulb bear 5 to 8 long, sparsely arranged spines anteriorly (Figure 2,3) (Kakarsulemankhel, 2004).

DOI:https//doi.org/10.22452/mjs.vol45no2.9
Malaysian Journal of Science 45(2): 87-95 (June 2026)

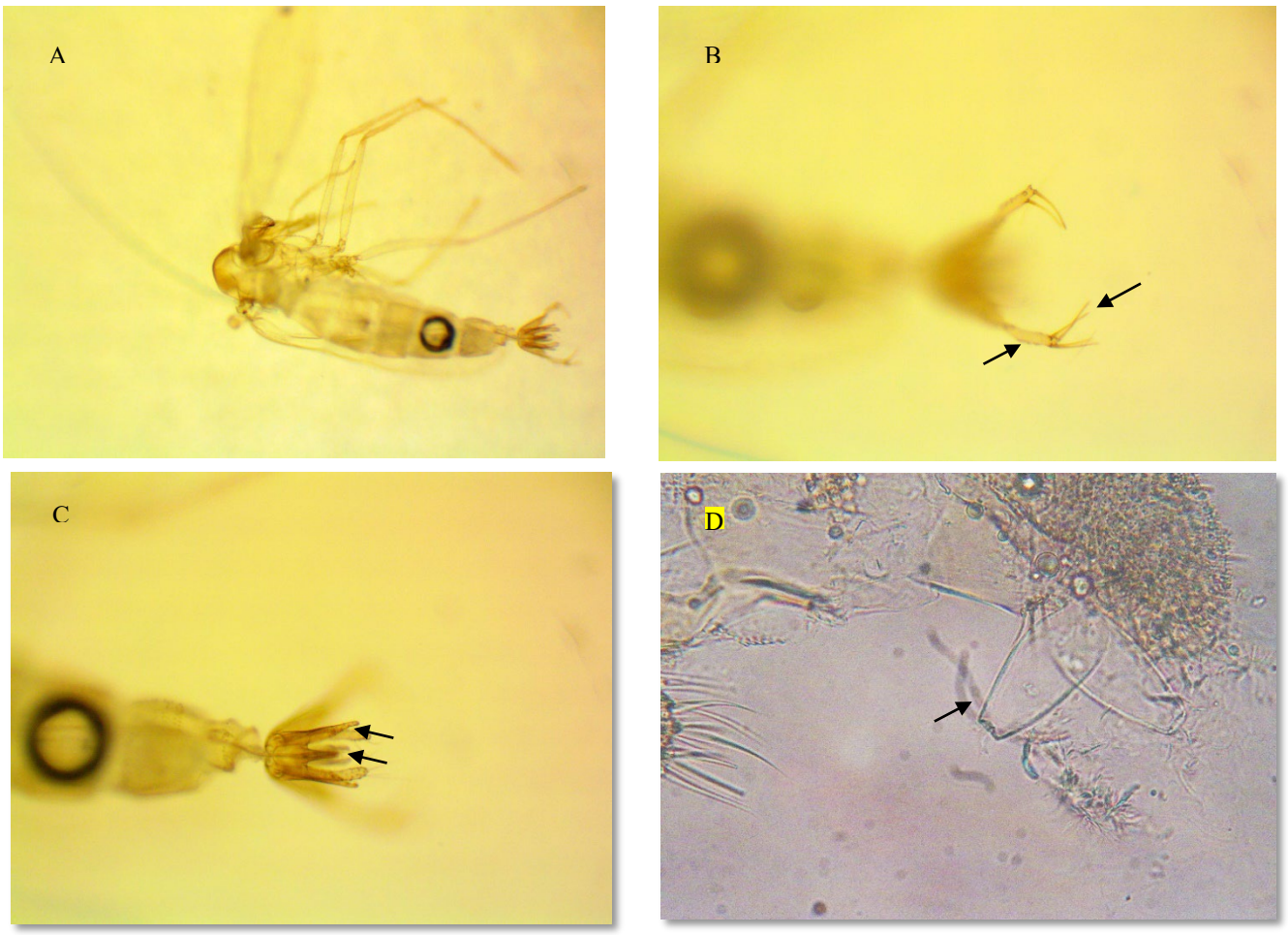


Figure 2. Phenotypic characters of male and female S. babu babu: A- Male (40X), B- Male genitalia (stylets with spines) (100X), C- Male genitalia (Aedeagus and paramere) (100X), D- Female spermatheca (640X).

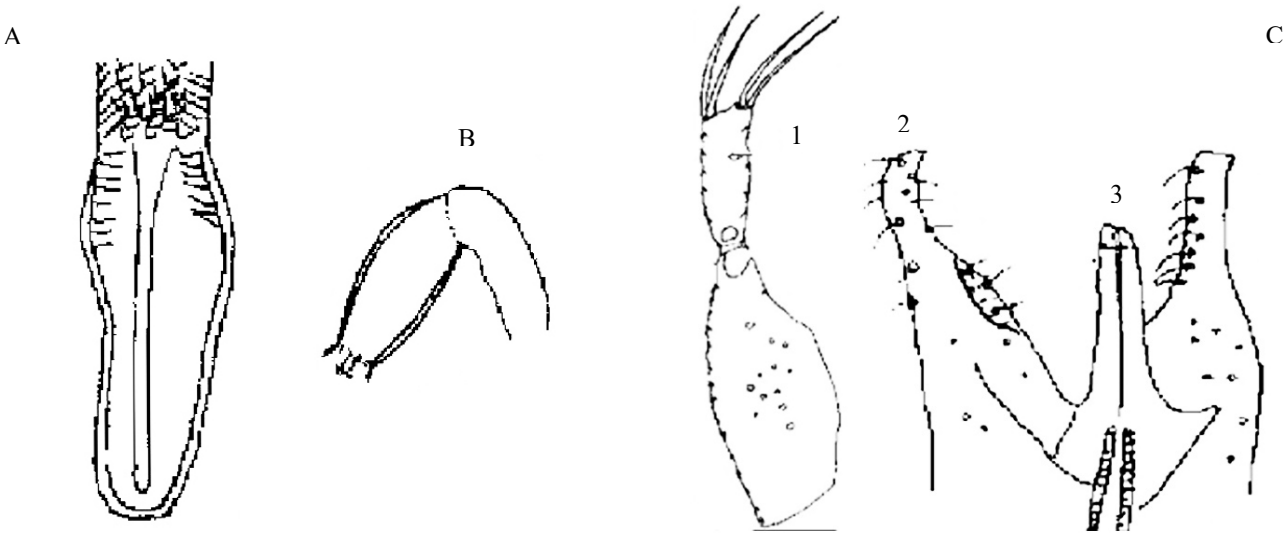


Figure 3. Diagram of phenotypic characters of male and female S. babu babu: A- Female pharynx, B- Female spermatheca, C- Male genitalia:1- stylets with spines, 2- paramere, 3-Aedeagus (Kakarsulemankhel, 2004).

DOI:https//doi.org/10.22452/mjs.vol45no2.9
Malaysian Journal of Science 45(2): 87-95 (June 2026)

Molecular Identification

DNA Extraction and PCR Results

The genome of the sand fly specimens was extracted at high concentration (53.62 ng/µl) and purity (1.7 ng/µl), as shown in Figure 4. These were measured using the BioDrop device, which is highly efficient in estimating the concentration and purity of DNA, and is easy to use, inexpensive, and provides rapid results (Hameed & Hamed, 2023).

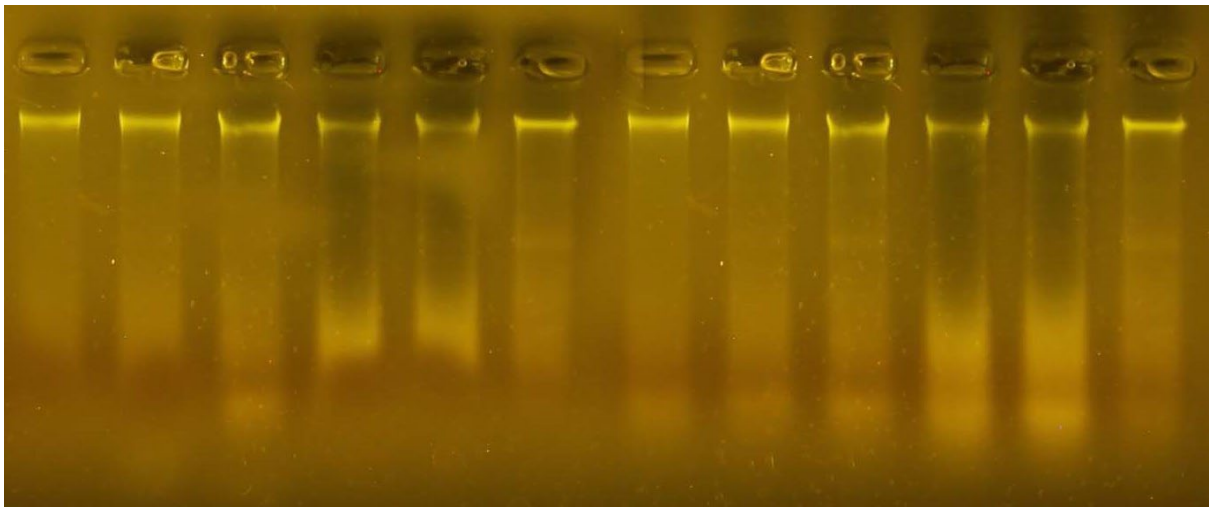


Figure 4. Whole genome extraction from sand fly specimens electrophoresed on 1% agarose gel.

In addition, the results in Figure 5 show the electrophoresis of the PCR products, where the primer used in this study bound to the 700 bp band of the *COX* gene region in the mitochondrial sand fly genome, which confirms the validity of using this primer in the current study.

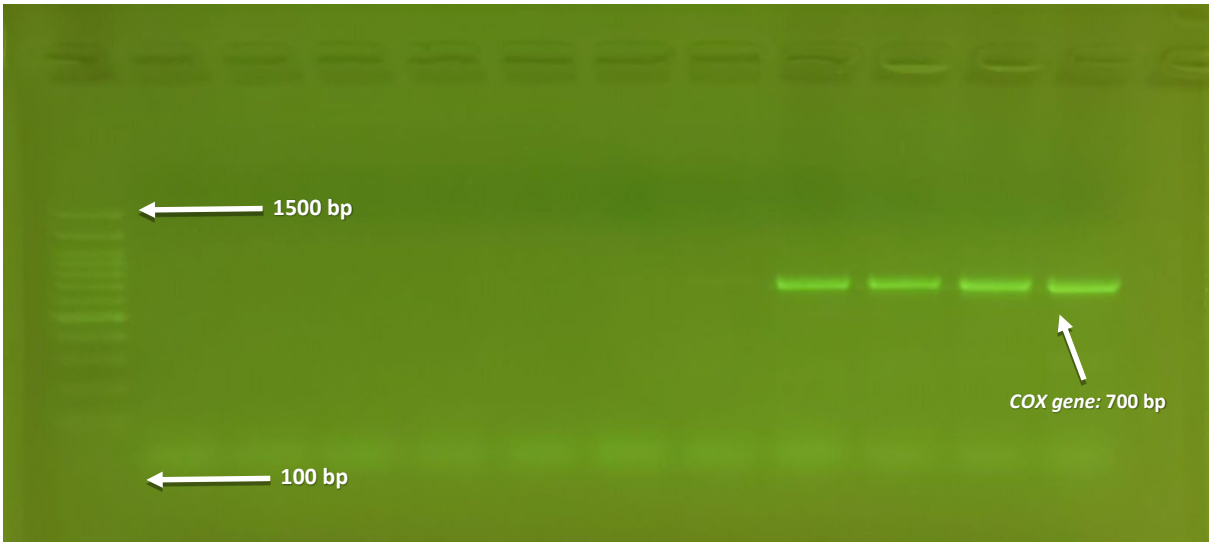


Figure 5. PCR amplification of the mitochondrial COX gene of sand fly specimens and primer binding with 700 bp PCR product separated by 2% gel electrophoresis.

Matching the Sequences of the Species Under Study and Recording Them in The NCBI.

The aim of conducting sequencing and matching the results with the isolates documented in the NCBI was to identify the species under study accurately and to determine the genetic variations and variability relative to global isolates, especially for species located within the Old World (Figure 6).

DOI:https//doi.org/10.22452/mjs.vol45no2.9
Malaysian Journal of Science 45(2): 87-95 (June 2026)

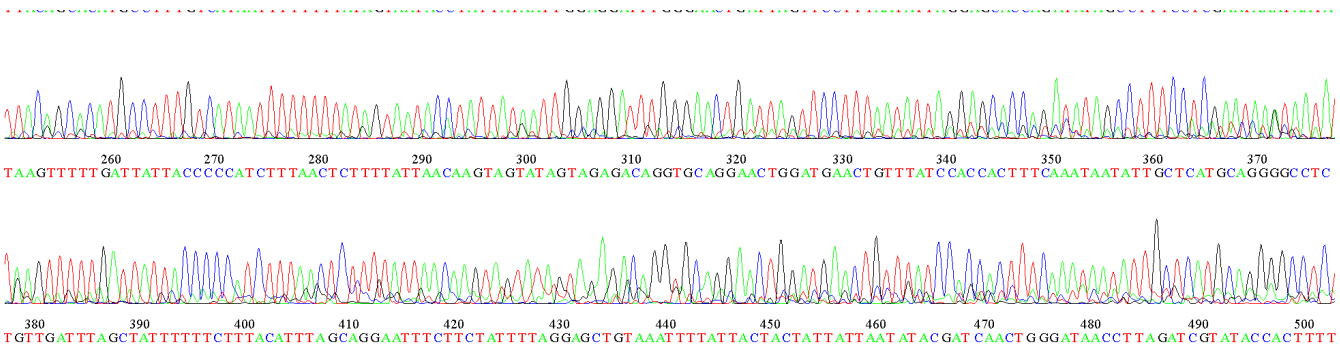


Figure 6. COX gene sequence analysis in the DNA-Sequencing test.

Overall, the high-quality sequences showed matching identity and identity with global isolates ranging from 94.42% to 98.59% (Table 1), and this resulted in the recording of three isolates in NCBI (accession numbers: PP856257, PQ056783, PQ063273) belonging to the species that were phenotypically identified, named YIMA in reference to the name of the authors.

Table 1. The accession numbers of *S. babu babu* isolates that were recorded and matched with different global isolates in NCBI.

No.	Isolates	Accession Numbers	Release Date	Matching with NCBI			
				Query Cover	Identity	Acc. Numbers	Country
1	*S. babu babu* isolate YIM32	PP856257	6-6-2024	100%	94.42	HQ585362	India
				100%	95.94	KY834515	Pakistan
				100%	95.43	MT472523	India
2	*S. babu babu* isolate YIM33	PQ056783	28-7-2024	100%	97.66	HQ585362	India
				100%	98.59	KY834515	Pakistan
				100%	98.36	MT472523	India
3	*S. babu babu* isolate YIM34	PQ063273	29-7-2024	100%	96.48	HQ585359	India
				98%	96.76	HQ585362	India
				99%	96.47	HQ585356	India

Phylogenetic tree and Evolutionary divergence

To determine the genetic proximity between the isolates recorded in the NCBI and to confirm the results of the molecular identification of the *COX* gene, an evolutionary tree was conducted with several closely related global isolates using the NJ tree and K2P model (Figure 7). The clades reflect the extent to which the isolates are related to each other. In general, the bootstrap values, shown as percentages at the key clade nodes, indicate the level of confidence or stability of the nodes observed in the tree. The group of isolates that are closely related are clustered together and supported by high bootstrap values (Aljubouri, 2024).

The isolates recorded during the current study were related to each other with a bootstrap value of 96%. Their association with the Pakistani and Indian isolates (Accession Numbers: KY834515, MT472523) with bootstrap values of 95, 98% respectively, indicates a very strong phylogenetic relationship. On the other hand, the three isolates showed evolutionary divergence values with global isolates ranging from 0.004 to 0.064, and the overall mean distance was determined to be 0.02 (Table 2). This low value indicates little evolutionary divergence between the Iraqi and global isolates, and also supports the results of the phylogenetic relationship.

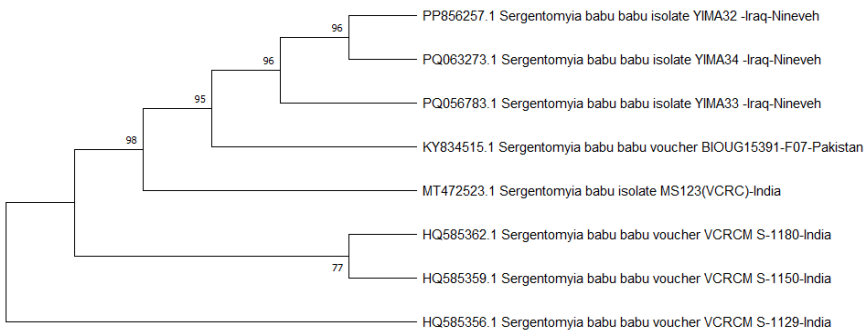


Figure 7. Phylogenetic tree of *S. babu babu* isolates and the Indian and Pakistani isolates documented in NCBI.

DOI:https//doi.org/10.22452/mjs.vol45no2.9
Malaysian Journal of Science 45(2): 87-95 (June 2026)

Table 2. Evolutionary divergence (K2P) between *S. babu babu* isolates and the Indian and Pakistani isolates documented in NCBI.

No.	Isolates	1	2	3	4	5	6	7	8
1	*Sergentomyia babu babu* Isolate YIMA32-Iraq-Nineveh								
2	*Sergentomyia babu babu* Isolate YIMA33-Iraq-Nineveh	0.026							
3	*Sergentomyia babu babu* Isolate YIMA34-Iraq-Nineveh	0.021	0.005						
4	*Sergentomyia babu babu* Voucher VCRCM S-1180-India	0.058	0.024	0.030					
5	*Sergentomyia babu* Isolate MS123(VCRC)-India	0.047	0.017	0.026	0.011				
6	*Sergentomyia babu babu* Voucher BIOUG15391-F07-Pakistan	0.042	0.014	0.013	0.018	0.014			
7	*Sergentomyia babu babu* Voucher VCRCMS-1150-India	0.064	0.029	0.033	0.004	0.016	0.023		
8	*Sergentomyia babu babu* Voucher VCRCM S-1129-India	0.058	0.024	0.035	0.001	0.013	0.018	0.009	

4. Discussion

The sandfly fauna of Iraq has received some attention from taxonomists. Newstead (1920) described specimens collected in southern Iraq during World War I (Newstead, 1920). Adler and Theodor (1929) constructed an extensive epidemiological survey of sandflies and Oriental sore (leishmaniasis) in Palestine, Syria, and Iraq; in Iraq, their study was limited to the urban areas of Baghdad, Basra, and Mosul, where they were able to record six species of the genus *Phlebotomus*. Pringle (1953) recorded 12 species of sandflies during his survey of most governorates of Iraq in the fall of 1949 and throughout 1950, four of which belonged to the genus *Phlebotomus* and eight to the genus *Sergentomyia* (Pringle, 1953). Abu al-Hab in 1979 recorded sixteen species across all governorates of Iraq, six of which belonged to the genus *Phlebotomus* and ten to the genus *Sergentomyia* (Abu al-Hab, 1979).

The genus *Sergentomyia*, a genus within the subfamily Phlebotominae (Theodor, 1948), has a variable distribution in Iraq, being more prevalent in the central and southern regions of the country (Al-Mayali & Al-Hassani, 2017), while its distribution is lower in the northern regions, as observed in our previous study in Nineveh Governorate (Al Joary & Al Hamdani, 2024). It was not identified in Erbil Governorate in 2022 during the study carried out by Rasool *et al*. (2024). Therefore, our current study also observed a low distribution of *Sergentomyia* species in general and the species *S. babu babu* in particular. The *S. babu babu* belongs to the subgenus *Parratomyia*, which includes species previously recorded in Iraq, such as *S. baghdadis* and *S. palestinensis* (Pringle, 1953; Abu Al-Hab, 1979). It is also more common in countries such as India (Shah *et al*., 2023) and Pakistan (Rasheed *et al*., 2023).

Given that sandflies are significant disease vectors, understanding the genetic diversity of these species is important for ecological monitoring. Many species may become resistant to insecticides as a result of gene flow between them, which could affect control programs (Pathirage *et al*., 2021). Overall, there are few studies on the molecular identification of *Sergentomyia* species throughout the Old World (Rodrigues & Galati, 2023). The *COX* gene is the most widely used molecular marker for studying sandflies despite their diversity, as it offers a substantial and rapid benefit in identifying species. It can also be used in cases where phenotypic identification is difficult, in addition to its benefit in suggesting and analyzing the population structure of the sandfly's community (Grant *et al*., 2021) and to correctly identify females from several genera, including the genus *Sergentomyia* (Srinivasan *et al*., 2014).

The primer used in this study (LCO1490 and HC02198) was first described by Folmer and used to amplify the *COX* gene in the mitochondrial DNA of 11 invertebrate phyla. Initial comparisons of the use of the *COX* primer showed that it generates informative sequences for phylogenetic analyses at the species and higher taxonomic levels. Since then, this primer has been used to detect the *COX* gene in sand flies and is among the most widely used primers for this purpose, which is why it was used in the current study (Folmer *et al*., 1994).

Regarding the results of evolutionary divergence, the findings of the current study are consistent with many previous studies. In India, Kumar found that the evolutionary divergence values among eight sand fly species, seven of which belonged to the genus *Sergentomyia* and one to the genus *Phlebotomus,* ranged between 0.031 and 0.207 (Kumar *et al*., 2012). In Brazil, Rodrigues found that the genetic divergence values between the sand fly species recorded in their study ranged from 0.008 to 0.177 (Rodrigues *et al*., 2018). In Jordan (west of Iraq), the genetic divergence between sand fly species belonging to the genus *Phlebotomus* ranged from 0.0853 to 0.2692 (Mukbel *et al*., 2024).

DOI:https//doi.org/10.22452/mjs.vol45no2.9
Malaysian Journal of Science 45(2): 87-95 (June 2026)

5. Recommendations

The current study recommends the re-identification of the species of the genus *Sergentomyia* in Iraq using both phenotypic and molecular methods rather than relying on either approach alone. This is deferred as future work using a dual approach to take into account both methods simultaneously.

6. Conclusion

The current study addressed an important disease vector, the sand fly, in Iraq, a country known for its outbreaks of leishmaniasis, and recorded for the first time in Iraq a new species belonging to the genus *Sergentomyia*, confirming this through molecular identification, which represents an integrative approach in the classification of species. The results of phenotypic and molecular identification demonstrate that the *S. babu babu* was recorded for the first time in Iraq, as three isolates were documented in NCBI. The evolutionary tree and genetic divergence analysis confirmed high genetic similarity between *S. babu babu* and the Indian and Pakistani species, which confirms the correctness of the identification. This study is the second in the country to address the molecular identification of the *Sergentomyia* species. Our emphasis is on the necessity of coordinating efforts to perform additional morphological and molecular taxonomic investigations to accurately identify the sandflies in Iraq, particularly those belonging to the genus *Sergentomyia*.

7. Acknowledgements

The authors would like to extend their sincere thanks and appreciation to the College of Science, University of Mosul, Iraq, for its support of this work.

8. References

Abu Al-Hab, J.K. (1979). Medical and veterinary insects in Iraq, Theoretical Department, Baghdad University Press, University of Baghdad, Baghdad, Iraq, 451p.

Adler, S., Theodor, O. (1929). The distribution of sandflies and leishmaniasis in Palestine, Syria and Mesopotamia. Annals of Tropical Medicine & Parasitology, 23 (2): 269-306.

Al-Joary, Y.I., Al-Hamdani, M.A. (2024). First molecular identification of phlebotomine sandflies (Diptera: Psychodidae) in Nineveh Governorate, Northern of Iraq. Biomedical and Biotechnology Research Journal, 8 (2):187-93.

Aljubouri, M.M.S. (2024). Detection of *Eikenella corrodens* from Patients with Periodontitis and a Novel Local Study of its Whole Genome Sequence. PHD thesis,University of Mosul, Nineveh, Iraq.

Al-Mashhadani, W.J.H. (2006). Growth and development of the *Leishmania* parasite in the sandfly. Umm Salamah Journal Science, 4 (3): 556-564.

Al-Mayali, H.M.H., Al-Hassani, M.K. (2017). Molecular identification of Phlebotominae sandflies (Diptera: Psychodidae) in Al-Diwaniyah Region/Iraq. Journal of Global Pharma Technology, 9 (10): 424–430.

Al-Obaidi, M.J.L.(2021). Environmental Suitability of the Occurrence of Cutaneous Leishmaniasis and its Vectors in Iraq. PHD thesis, University of Baghdad, Baghdad, Iraq.

Brilhante A.F., de Albuquerque A.L., Rocham,A.C.B., Ayres, C.F.J., Paiva, M.H.S., de Ávila, M.M., Cardoso, C.O., Mauricio, I.L., Galati, E.A.B. (2020). First report of an Onchocercidae worm infecting *Psychodopygus carrerai carrerai* sandfly, a putative vector of *Leishmania braziliensis* in the Amazon. Scientific Reports, 10(1):15246.

Cecílio, P., Cordeiro-da-Silva, A,, Oliveira, F. (2022). Sand flies: Basic information on the vectors of leishmaniasis and their interactions with Leishmania parasites. Communications biology, 5(1): 305.

Dantas-Torres, F., Tarallo, V.D., Otranto, D. (2014). Morphological keys for the identification of Italian phlebotomine sand flies (Diptera: Psychodidae: Phlebotominae). Parasites & Vectors, 7:479.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology, 3 (5):294-9.

Galati, E.A.B., Rodrigues, B.L. (2023). A Review of Historical Phlebotominae Taxonomy (Diptera:Psychodidae). Neotropical Entomology,52:539–559.

Grant, D.M., Brodnicke, O.B., Evankow, A.M., Ferreira, A.O., Fontes, J.T., Hansen, A.K., Jensen, M.R., Kalaycı, T.E., Leeper, A., Patil, S.K., Prati, S. (2021). The future of DNA barcoding: reflections from early career researchers. Diversity, 13 (7): 313.

Hameed, M.A., Hamed, O.M. (2023). Detection of P53 suppressor gene mutation in women with breast cancer in Mosul city. AIP Conference Proceedings, 2834(1), 020007.

Kakarsulemankhel, J.K. (2004). Taxonomic Morphology of *Sergentomyia* (*Sergentomyia*) *babu babu* Annandale (1910), *Sergentomyia palestinensis* Adler and Theodor (1927) and *Sergentomyia baghdadis* Adler and Theodor (1929) (Diptera, Psychodidae) of Pakistan. Pakistan Journal of Biological Sciences,7 (5):721-733.

DOI:https//doi.org/10.22452/mjs.vol45no2.9
Malaysian Journal of Science 45(2): 87-95 (June 2026)

Kumar, N.P., Srinivasan, R., Jambulingam, P. (2012). DNA barcoding for identification of sand flies (Diptera: Psychodidae) in India. Molecular Ecology Resources, 12 (3): 414-420.

Maia, C, Depaquit, J. (2016). Can *Sergentomyia* (Diptera, Psychodidae) play a role in the transmission of mammal-infecting *Leishmania*?. Parasite, 23:55.

Mukbel, R. M., Kanaan, D.A., Abu Halaweh, M.M., Amr, Z.S. (2024). Morphological, distributional, and molecular study of the Phlebotomus sand flies of Jordan. Journal of Vector Ecology, 49 (1):64-79.

Newstead, R. (1920). On the genus *Phlebotomus*. Part IV. Bulletin of Entomological Research, 11(3): 305-311.

Özbel ,Y., Karakuş, M., Arserim, S.K., Kalkan, Ş.O., Töz, S. (2016). Molecular detection and identification of *Leishmania* spp. in naturally infected *Phlebotomus tobbi* and *Sergentomyia dentata* in a focus of human and canine leishmaniasis in western Turkey. Acta Tropica, 155:89-94.

Pathirage, D.R.K., Weeraratne, T.C., Senanayake, S.C., Karunaratne, S.H., Karunaweera, N.D. (2021). Genetic diversity and population structure of *Phlebotomus argentipes*: Vector of *Leishmania donovani* in Sri Lanka. PLOS One, 16 (9): e0256819.

Pringle, G. (1953). The Sandflies (Phlebotominae) of Iraq. Bulletin of Entomological Research, 43(4): 707-734.

Rasheed, S.B., Farrah Zaidi, S., Jan, T. (2023). Sand Fly (Psychodidae: Phlebotominae) Ecology of an Emerging Cutaneous Leishmaniasis Focus in Khyber Pakhtunkhwa with an Updated Species Checklist from Pakistan. Pakistan Journal of Zoology, 56 (4): 1589-1598.

Rasool, S.H., Toprak, S., Karim, A.Y., Parmaksiz, A., Yildiz Zeyrek, F., Hamad, S.W. (2024) Morphological and molecular identification of sandfly species (diptera: psychodidae) and the bio-ecology of cutaneous lesheimaniasis vectors in Erbil province, Iraq. Applied Ecology and Environmental Research, 22 (2):1543-1562.

Rodrigues, B.L., Carvalho-Costa, L.F., Pinto, I.D., Rebˆelo, J.M.(2018). DNA barcoding reveals hidden diversity of sand flies (Diptera: Psychodidae) at fine and broad spatial scales in Brazilian endemic regions for leishmaniasis. Journal of Medical Entomology 55 (4):893-901.

Rodrigues, B.L., Galati, E.A.B. (2023). Molecular taxonomy of phlebotomine sand flies (Diptera, Psychodidae) with emphasis on DNA barcoding: A review. Acta Tropica, 238:106778.

Shah, H., Fathima, P., Kumar, N.P., Kumar, A., Saini, P. (2023). Faunal richness and checklist of sandflies (Diptera: Pyschodidae) in India. Asian Pacific Journal of Tropical Medicine, 16 (5): 193.

Srinivasan, R., Jambulingam, P., Kumar, N.P. (2014). *Sergentomyia* (Neophlebotomus) *monticola*, a new species of sand fly (Diptera: Psychodidae) from the Western Ghats, Thiruvananthapuram District, Kerala, India. Acta Tropica,137:74-79.

Srisuton, P., Phumee, A., Sunantaraporn, S., Boonserm, R., Sor-Suwan, S., Brownell, N., Pengsakul, T., Siriyasatien, P. (2019). Detection of *Leishmania* and *Trypanosoma* DNA in Field-Caught Sand Flies from Endemic and Non-Endemic Areas of Leishmaniasis in Southern Thailand. Insects, 10 (8):238.

Tamura, K., Stecher, G., Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution,38 (7):3022-3027.

Theodor, O. (1948). Classification of the Old World species of the subfamily Phlebotominae (Diptera, Psychodidae). Bulletin of Entomological Research, 39 (1) : 85–115.

DOI:https//doi.org/10.22452/mjs.vol45no2.9
Malaysian Journal of Science 45(2): 87-95 (June 2026)