JOURNAL MEDICAL INFORMATICS TECHNOLOGY
medinftech.org ISSN 2988-7003 (online)

Identification of Cell Images in Pap Smear Using GLCM and Classification Methods in Machine Learning

Rano Agustino[1*], *Prima Nanda Fauziah*[2]

[1,2] *Universitas Mohammad Husni Thamrin, Indonesia*

ARTICLE HISTORY

Received: 30 September 24
Final Revision: 20 September 25
Accepted: 23 September 25
Online Publication: 30 September 25

KEYWORDS

Cervical Cancer, Gray-Level Co-occurrence Matrix, Machine Learning, Pap Smear Classification, Texture Feature Extraction

CORRESPONDING AUTHOR

rano.agustino@gmail.com

DOI

10.37034/medinftech.v3i3.85

A B S T R A C T

Early detection of cervical cancer is critical for improving patient outcomes, and accurate classification of Pap smear images supports clinical decision-making. This study aimed to improve cervical cancer diagnosis by classifying Pap smear images using texture features. A dataset of 250 images across five classes underwent preprocessing including grayscale conversion and noise removal. Texture features such as contrast, dissimilarity, homogeneity, energy, correlation, and Angular Second Moment (ASM) were extracted using the Gray-Level Co-occurrence Matrix (GLCM). These features were then used to train and evaluate machine learning algorithms: Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), and Neural Networks (NN). The Decision Tree model achieved the highest accuracy of 95%, outperforming Neural Networks which reached 74%. Ensemble methods like RF and GB showed robust performance across classes. These results demonstrate the effectiveness of GLCM-based feature extraction combined with Decision Tree classification for accurate and reliable Pap smear image analysis. This approach offers valuable insights for enhancing clinical decision support in cervical cancer diagnosis.

1. Introduction

Early detection of cervical cancer is crucial for effective treatment and improved patient outcomes. Accurate classification of Pap smear images plays a key role in supporting clinical decision-making and reducing diagnostic errors. However, medical image classification, particularly for Pap smear slides, remains a challenging task due to variations in cell morphology and image quality.

Random Forest is an ensemble method generating multiple decision trees during training and employing class mode for classification. It has been shown to outperform traditional algorithms like logistic regression and support vector machines, especially on complex, high-dimensional datasets. Previous studies have also demonstrated higher AUC values for RF compared to other models [1], [2], [3]. RF's ability to reduce overfitting through bagging and its robustness to noisy data make it a preferred choice in clinical applications [4], [5].

Gradient Boosting builds models sequentially, with each new model correcting errors made by the previous one. It is known for its high predictive accuracy across various domains, including healthcare. For instance, gradient boosting has proven effective in predicting cardiac surgery outcomes and other medical conditions, solidifying its reliability in clinical settings [6], [3]. GB's performance comparable to RF in this study is consistent with previous research showing that both algorithms can yield high accuracy and AUC scores in complex classification tasks [7], [8].

Decision Trees are machine learning algorithms used for classification and regression. They create tree-like models resembling flowcharts. Each branch represents a data attribute, and each leaf represents a class or predicted value. Pap smear image classification is crucial for early cervical cancer detection. Decision Trees, as part of machine learning techniques, have demonstrated significant potential in improving accuracy and efficiency in classifying cervical cells [9].

Neural Networks (NN), while highly capable of modeling complex and non-linear relationships in unstructured data such as images, typically require large, diverse, and balanced datasets to achieve optimal performance. These models rely on extensive amounts of training data to effectively learn feature

representations and generalize well to unseen data. When datasets are limited in size or imbalanced, as is often the case in medical image analysis, the learning process can be significantly constrained, leading to suboptimal accuracy and potential overfitting. In the present study, NN achieved only 74% accuracy, which reflects these inherent limitations. The relatively small size of the dataset, combined with the high variability among Pap smear images, posed substantial challenges in training the deep learning models. These challenges underscore the importance of dataset quality and quantity when applying NN to medical diagnostics and highlight the need for careful consideration when interpreting the performance of such models in clinical applications [10], [11], [12].

The dataset comprised 250 Pap smear images categorized into five classes: Normal, H-Sil, Koilocyte, Normal Non-ThinPrep, and L-Sil Non-ThinPrep. Preprocessing included grayscale conversion and noise removal to enhance data quality for machine learning [13]. The Gray-Level Co-occurrence Matrix (GLCM) was used to extract texture features, including Contrast, Dissimilarity, Homogeneity, Energy, Correlation, and Angular Second Moment (ASM) [14], which were then used to train RF, GB, DT, and NN models.

Results showed that DT achieved 95% accuracy, while RF and GB reached 92% with AUC scores of 0.94 and 0.95, respectively. This highlights DT's effectiveness in distinguishing classes. The AUC metric is highly relevant in medical diagnostics as it reflects overall model performance across different classification thresholds [15], [16]. These findings support previous research consistently reporting high AUC scores for RF and GB in various medical prediction tasks [17].

Previous research [18], [19] primarily focused on fewer classes and a smaller set of texture features, often using deep learning models like VGG16, VGG19, AlexNet, and ResNet50. This study addresses these gaps by adding the ASM feature to the GLCM feature set and expanding the classification task to five classes. These enhancements improve the representation of cervical cell variations, while the comparative evaluation of four algorithms highlights the strengths and weaknesses of each approach. The findings, particularly the high accuracy achieved by Decision Tree, demonstrate the novelty and potential contribution of this study to the development of automated cervical cancer detection methods, which could inform future clinical decision-making processes.

Previous studies on Pap smear image classification were generally limited to fewer classes and a small set of texture features. This research addresses these gaps by adding the Angular Second Moment (ASM) feature to the GLCM feature set and expanding the classification task to five classes. These enhancements improve the representativeness of cervical cell variations, while the comparative evaluation of four algorithms highlights the strengths and weaknesses of each approach [20]. The findings, particularly the high accuracy achieved by the Decision Tree model, demonstrate the novelty and potential contribution of this study to automated cervical cancer detection.

In summary, this study compares four machine learning algorithms: Random Forest, Gradient Boosting, Decision Tree, and Neural Networks for Pap smear image classification. The results indicate that DT achieved the highest accuracy (95%), followed by RF and GB (92%), while NN reached 74%, thereby emphasizing the differences in performance among these approaches.

2. Research Method

2.1. Research Methodology

The methodology employed in this research involved a systematic approach to classify Pap smear images using four machine learning algorithms: Random Forest (RF), Gradient Boosting (GB), Decision Tree (DT), and Neural Networks (NN). The research was structured into several sequential phases: dataset preparation, data preprocessing, feature extraction, model training, and evaluation. Each phase was designed to ensure data quality and reliable model development. The following diagram illustrates the overall research workflow:

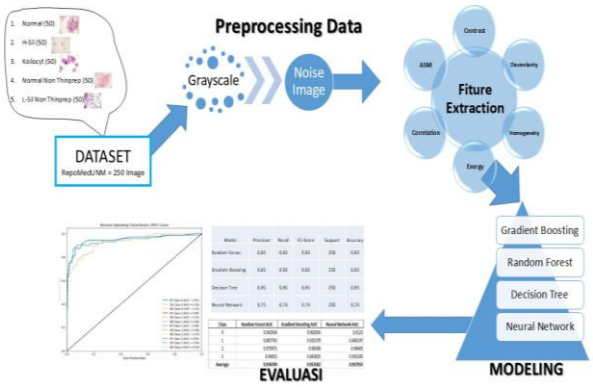

Figure 1. Business Process

2.2. Dataset Preparation

As depicted in the Business Process diagram, the initial step involves dataset preparation. The dataset utilized in this study consists of 250 Pap smear images sourced from the RepoMedUNM database. These images are categorized into five distinct classes: Normal, H-Sil, Koilocyt, Normal Non-ThinPrep, and L-Sil Non-ThinPrep, with each class containing 50 images. The selection of this dataset is crucial as it provides a balanced representation of various cervical cell types, which is essential for training robust classification models [21]. Figure 2. Presented below are images of the various classes or categories present in the Pap smear dataset.

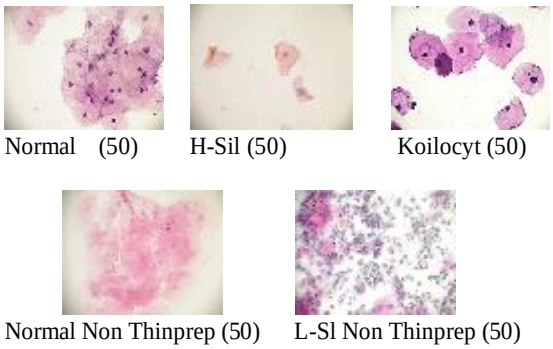

Figure 2. Dataset Image Pap Smear

2.3. Data Preprocessing

Data preprocessing is a critical step in preparing images for analysis. In this study, preprocessing was performed using Python, involving grayscale conversion and noise removal. Grayscale conversion simplifies image data, reducing input complexity while preserving essential features necessary for classification. Noise removal enhances image quality, which is crucial for accurate feature extraction and subsequent processing stages [22], [23]. These preprocessing steps align with best practices in medical image analysis, as they help to mitigate the impact of artifacts and improve the overall quality of the dataset [24]. Figure 3 presented are visualizations of the data preprocessing stages.

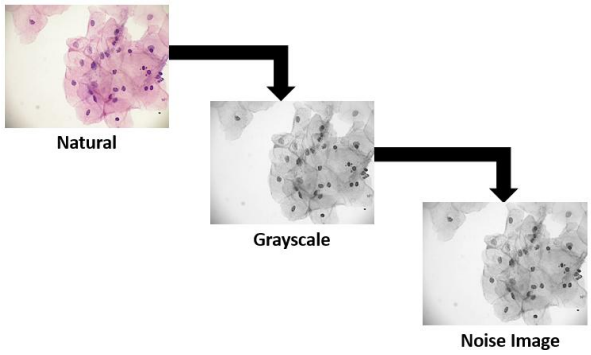

Figure 3. Preprocessing Data Image

2.4. Feature Extraction

Feature extraction was conducted using the Gray-Level Co-occurrence Matrix (GLCM) technique, which is well-suited for image processing to capture texture features. GLCM generates a matrix describing the spatial relationship between pixels, allowing for the extraction of important features such as Contrast, Dissimilarity, Homogeneity, Energy, Correlation, and Second Angular Moment (ASM) [25], [26]. This structured approach to feature extraction is essential for enabling algorithms to learn and differentiate between various classes effectively [27], [28].

2.5. Modeling

The extracted features were used to train four machine learning algorithms: Random Forest (RF), Gradient Boosting (GB), Decision Tree (DT), and Neural Networks (NN). Random Forest builds multiple decision trees and aggregates their outputs to improve generalization and reduce overfitting. Gradient Boosting constructs models sequentially, where each model corrects the errors of its predecessor. Decision Tree creates a hierarchical structure of nodes and branches to classify data based on attribute values. Neural Networks, inspired by the structure of the human brain, consist of layers of interconnected nodes that can model non-linear relationships in the data. The training process included hyper parameter optimization and evaluation procedures to ensure each algorithm was fitted appropriately for the dataset [29], [30]. Neural Networks, particularly deep learning architectures, were also employed in this study, although they generally require larger datasets to achieve optimal performance [31], [32]. The training process involved hyperparameter optimization and ensuring that each model adequately fit the training data.

2.6. Evaluation

Model performance was evaluated using the Confusion Matrix and the Area Under the Curve – Receiver Operating Characteristic Curve (AUC-ROC). The Confusion Matrix provides detailed insights into class-level performance through metrics such as precision, recall, F1-score, and overall accuracy, allowing evaluation of how well the model classifies each Pap smear category. Meanwhile, the AUC-ROC is a threshold-independent metric that reflects the model's discriminative ability across all classes, offering a more comprehensive perspective compared to accuracy alone. These evaluation methods are particularly crucial in medical image analysis, where distinguishing between normal and abnormal samples has direct clinical implications [31], [32].

3. Result and Discussion

3.1. Feature Extraction Results

The feature extraction process was carried out using the Gray Level Co-occurrence Matrix (GLCM) method to obtain texture-based characteristics from Pap smear images. The extracted features were then compiled into a structured dataset and stored in a CSV file named glcm_features_all.csv. This dataset consists of six texture descriptors, namely Contrast, Dissimilarity, Homogeneity, Energy, Correlation, and Angular Second Moment (ASM), along with their respective class labels. An example of the extracted features is presented in Table 1, which demonstrates the numerical values of each descriptor across sample images labeled as "Normal."

Table 1. Example of Feature Extraction Dataset

Contrast	Dissimilarity	Homogeneity	Energy	Correlation	ASM	Class
1.982389322	0.6129 13479	0.78150 4427	0.1195 0127	0.9990 7345	0.01 42806	Normal
0.966626205	0.3623 80543	0.85912 9834	0.1539 7578	0.9987 4062	0.02 37085	Normal
0.727323236	0.3051 46683	0.87870 1147	0.1987 3074	0.9991 6057	0.03 94939	Normal
0.696247814	0.3042 69724	0.87823 6646	0.1951 0198	0.9991 9274	0.03 80648	Normal
1.067771899	0.3861 13164	0.85293 7005	0.1846 604	0.9992 1606	0.03 40995	Normal

To facilitate interpretation, the distributions and pairwise relationships among the extracted features are visualized in Figure 2. These figures provide insight into how each feature behaves across different classes (Normal, H-SIL, L-SIL, Koilocyte, and Normal-NT).

Based on the pair plot visualization in Figure 2, several important observations can be made regarding the distribution and relationships of the extracted GLCM features:

1. Patterns and Clusters Between Classes

Normal (Green): This class exhibits a distinct pattern compared to others, particularly in features like Contrast and Homogeneity, where Contrast values are higher and Homogeneity values are lower.

H-SIL (Red) and L-SIL (Purple): These classes appear to share some similarities but still exhibit differences in feature distributions such as ASM and Energy.

Koilocyte (Orange): This class has a unique distribution pattern in features like Correlation and Dissimilarity, but its data points are relatively fewer.

Normal-NT (Blue): Generally, it has a feature distribution similar to Normal but spreads slightly wider in some features.

2. Feature Value Distribution

Contrast: This feature shows a wider range of values compared to other features, with the Normal class dominating at higher values.

Homogeneity and Energy: These features tend to cluster at high values for most classes, but with some variation for abnormal classes like H-SIL.

ASM and Correlation: These features have a very narrow range of values, which may be less informative for class discrimination when used alone.

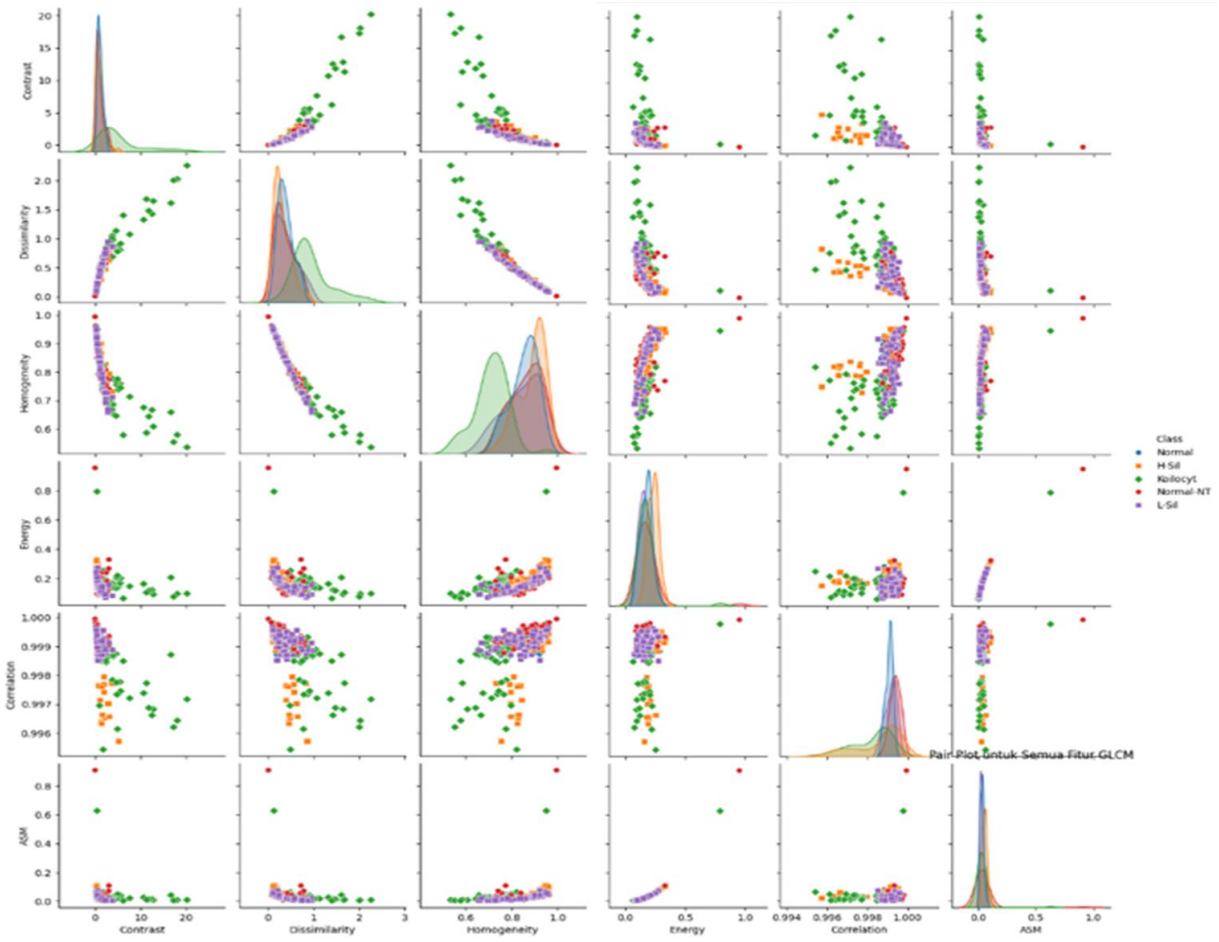

Figure 2. Pair Plot of Extracted Features

3. Feature Correlations

There is a clear negative correlation between Homogeneity and Contrast: high Homogeneity values tend to correspond to low Contrast values, especially for the Normal class.

Energy and ASM seem to have a strong positive correlation, where an increase in one feature is closely related to an increase in the other for all classes.

Some features, like Correlation, do not show strong correlations with other features, indicating that they might be less informative.

4. Feature Discriminative Power

Features like Contrast, Homogeneity, and Energy seem to have good discriminative power in separating the Normal class from abnormal classes (H-SIL, L-SIL, Koilocyte).

Features like Correlation and ASM, due to their narrow value range, may have a smaller contribution to class discrimination when used independently.

3.2. Model Performance Evaluation

As shown in Table 2, both Random Forest (RF) and Gradient Boosting (GB) achieved an accuracy of 92%, while Neural Network (NN) obtained the lowest accuracy at 74%. In contrast, Decision Tree (DT) recorded the highest accuracy of 95%.

Table 2. Confussion Matrix

Model	Class	Precision	Recall	F1-Score	Support	Accuracy
Random Forest	0	0.96	0.92	0.94	50	0.92
	1	0.96	0.98	0.97	50	
	2	0.86	0.88	0.87	50	
	3	0.94	0.94	0.94	50	
	4	0.90	0.90	0.90	50	
	Macro Avg	0.92	0.92	0.92	250	
	Weighted Avg	0.92	0.92	0.92	250	
Gradient Boosting	0	0.96	0.96	0.96	50	0.92
	1	0.96	0.98	0.97	50	
	2	0.88	0.86	0.87	50	
	3	0.92	0.94	0.93	50	
	4	0.90	0.88	0.89	50	
	Macro Avg	0.92	0.92	0.92	250	
	Weighted Avg	0.92	0.92	0.92	250	
Decision Tree	0	0,94	0,96	0,95	50	0.95
	1	0,98	0,98	0,98	50	
	2	0,92	0,92	0,92	50	
	3	0,94	0,96	0,95	50	
	4	0,96	0,92	0,94	50	
	Macro Avg	0,95	0,95	0,95	250	
	Weighted Avg	0,95	0,95	0,95	250	
Neural Network	0	0.78	0.78	0.78	50	0.74
	1	0.81	0.94	0.87	50	
	2	0.60	0.66	0.63	50	
	3	0.78	0.76	0.77	50	
	4	0.76	0.58	0.66	50	
	Macro Avg	0.75	0.74	0.74	250	
	Weighted Avg	0.75	0.74	0.74	250	

Furthermore, the AUC scores indicate the discriminative power of the models, with RF achieving 0.90, GB reaching 0.88, DT scoring 0.76, and NN obtaining 0.83. The following Table 3 and Figure 3 presents the AUC-ROC values for this study.

Table 3. AUC-ROC

Class	Random Forest AUC	Gradient Boosting AUC	Decision Tree	Neural Network AUC
0	0.9100	0.8700	0.7600	0.8200
1	0.9600	0.9400	0.8900	0.9300
2	0.8700	0.8500	0.6600	0.8300
3	0.8900	0.8900	0.7400	0.7600
Average	0.9075	0.8875	0.7625	0.8350

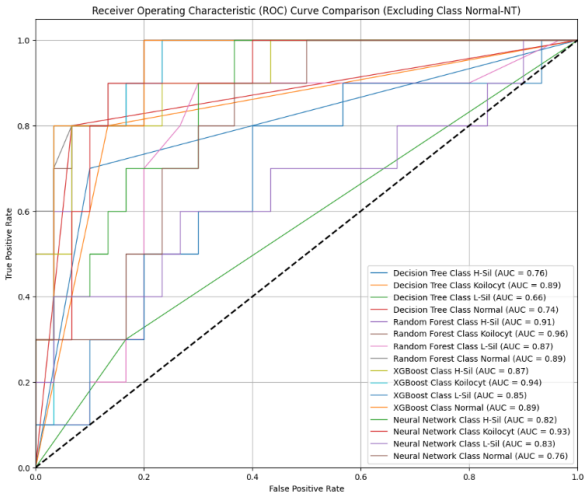

Figure 3. AUC-ROC Curve

An interesting finding from this study is the performance discrepancy of the Decision Tree (DT) model. While DT achieved the highest overall accuracy (95%), it also obtained the lowest average AUC score (0.76) compared to Random Forest (0.90), Gradient Boosting (0.88), and Neural Networks (0.83). This indicates that DT classified most samples correctly, but its ability to consistently distinguish between all classes was weaker than ensemble methods such as RF and GB.

Further analysis of the confusion matrix (Table 2) shows that DT reached very high precision and recall for certain classes (e.g., 0.98 for Class 1), which contributed to its high accuracy, but its performance dropped for others (e.g., 0.92 for Class 2). The AUC-ROC visualization (Figure 3) reinforces this finding by showing that DT's ROC curves are less consistent across classes compared to RF and GB, whose curves

lie closer to the top-left corner, indicating stronger class separability.

In the context of medical diagnostics, this discrepancy highlights the importance of evaluating models not only by accuracy but also by AUC. Accuracy reflects the overall proportion of correct classifications, while AUC provides a more comprehensive measure of a model's discriminative power across all classes, which is especially critical when distinguishing between normal and abnormal Pap smear samples. Beyond the performance of DT, it is also important to examine how other models behaved in terms of Accuracy and AUC, as these metrics provide a broader perspective on model effectiveness. The AUC scores for RF and GB were 0.90 and 0.88, respectively, indicating strong discriminative power. Conversely, the NN model, with an AUC of 0.83, while still reasonably good, highlights the challenges faced by deep learning models when trained on smaller datasets [33]. These findings align with previous research that has reported similar performance levels for RF and GB in various classification tasks, reinforcing their status as robust algorithms for medical image analysis [34], [35].

3.2. Decision Tree Performance

The Decision Tree (DT) algorithm demonstrated the highest classification accuracy in this study, achieving 95% as shown in Table 2. This strong performance can be attributed to its straightforward rule-based structure, which enables the model to capture clear decision boundaries in the dataset. In particular, DT exhibited very high precision and recall for certain classes (e.g., 0.98 for Class 1), contributing significantly to its overall accuracy.

However, despite its high accuracy, DT achieved the lowest average AUC score (0.76) compared to other algorithms (Table 3 and Figure 3). This discrepancy suggests that while DT can correctly classify a large proportion of samples, its ability to consistently discriminate between all classes is weaker than ensemble-based methods such as Random Forest and Gradient Boosting. For instance, DT tended to perform well on classes with distinct feature patterns but showed performance drops in classes with overlapping feature distributions (e.g., 0.92 for Class 2).

This result highlights an important characteristic of DT models: they are prone to overfitting when dealing with complex datasets, which can lead to high accuracy but reduced generalization in terms of discriminative ability. Nevertheless, the simplicity and interpretability of DT make it attractive for medical diagnostics, as its rule-based structure allows clinicians to trace decision paths and understand classification outcomes.

3.2. Random Forest and Gradient Boosting Performance

The superior performance of RF and GB can be attributed to their ensemble learning techniques, which combine multiple weak predictions to produce a strong predictive model. RF builds multiple decision trees and aggregates their outputs, effectively reducing overfitting and improving generalization [36]. GB, on the other hand, builds models sequentially, focusing on correcting errors from previous iterations, allowing it to adaptively improve performance with each additional tree [37]. This adaptability is particularly beneficial in medical diagnostics, where the complexity and variability of data can significantly impact classification accuracy. The results of this study support findings from other research that has demonstrated the effectiveness of RF and GB in medical image classification tasks. For instance, studies have shown that RF consistently outperforms traditional classification, achieving higher accuracy and AUC scores in various applications, including medical imaging [38], [39]. Similarly, GB has been recognized for its high predictive accuracy in various domains, including healthcare, where it has been successfully applied to predict patient outcomes and disease progression [40].

3.3. Neural Network Performance

The lower accuracy of NN in this study highlights the challenges associated with training deep learning models on limited datasets. Although NNs have demonstrated remarkable success in various image classification tasks, especially in scenarios with large amounts of labeled data, their performance can be hindered when datasets are small or imbalanced [41], [42]. This observation is consistent with the existing literature, which suggests that deep learning models often require extensive training data to achieve optimal performance. Despite these limitations, the potential of NNs in medical image classification remains significant, especially as larger and more diverse datasets become available for training.

3.4. Implications for Medical Diagnostics

The findings of this study have significant implications for the field of medical diagnostics, particularly in the context of cervical cancer detection. The high accuracy achieved by DT indicates that this algorithm can be effectively integrated into clinical decision support systems to assist pathologists in analyzing Pap smear images. By providing accurate and reliable classifications, this model can improve the efficiency of cervical cancer screening programs and enhance patient outcomes. Moreover, these results highlight the importance of selecting the appropriate classification algorithm based on the characteristics of the dataset and the specific requirements of the classification task. As the field of medical image analysis continues to evolve, the integration of ensemble methods like DT with more recent techniques such as hybrid models may further improve accuracy.

4. Conclusion

This study demonstrates that the Decision Tree (DT) algorithm achieved the highest accuracy (95%) in classifying Pap smear images, although its discriminative ability, reflected by the AUC score, was lower than that of ensemble methods such as Random Forest and Gradient Boosting. These findings suggest that while DT offers simplicity and interpretability, ensemble models provide more consistent class separability. Neural Networks, in contrast, exhibited lower performance due to the limited dataset but remain promising for medical image classification when larger datasets are available.

This study has several limitations. The dataset size was relatively small (250 images), which may limit generalizability. The Neural Network model did not perform optimally due to the limited data size, and the study only relied on GLCM-based texture features without exploring other descriptors such as morphological or deep learning-based features. These limitations indicate that the reported results should be interpreted with caution, particularly for broader clinical applications.

Future research should therefore focus on addressing these limitations. Expanding the dataset with larger and more diverse Pap smear images, integrating deep learning approaches (e.g., CNN and transfer learning) with handcrafted features, and developing hybrid or ensemble models that combine interpretable algorithms (such as Decision Tree) with more powerful learners (such as Neural Networks) are promising directions. Validating the proposed methods using real-world clinical data will also be essential to ensure reliability and applicability in practical cervical cancer screening.

References

[1] H. Lee, S. Yoon, S. Yang, W. Kim, H. Ryu, C. Jung, *et al*., "Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model," *Journal of Clinical Medicine*, vol. 7, no. 11, p. 428, 2018. doi: 10.3390/jcm7110428.

[2] K. Ellis, J. Kerr, S. Godbole, G. Lanckriet, D. Wing, and S. Marshall, "A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers," *Physiological Measurement*, vol. 35, no. 11, pp. 2191-2203, 2014. doi: 10.1088/0967-3334/35/11/2191.

[3] Y. Fan, J. Dong, Y. Wu, M. Shen, S. Zhu, X. He, *et al*., "Development of machine learning models for mortality risk prediction after cardiac surgery," *Cardiovascular Diagnosis and Therapy*, vol. 12, no. 1, pp. 12-23, 2022. doi: 10.21037/cdt-21-648.

[4] S. Park, C. Kim, and X. Wu, "Development and validation of an insulin resistance predicting model using a machine-learning approach in a population-based cohort in korea," *Diagnostics*, vol. 12, no. 1, p. 212, 2022. doi: 10.3390/diagnostics12010212.

[5] S. Weng, J. Reps, J. Kai, J. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," *Plos One*, vol. 12, no. 4, p. e0174944, 2017. doi: 10.1371/journal.pone.0174944.

[6] U. Benedetto, S. Sinha, M. Lyon, A. Dimagli, T. Gaunt, G. Angelini, *et al*., "Can machine learning improve mortality prediction following cardiac surgery?," *European Journal of Cardio-Thoracic Surgery*, vol. 58, no. 6, pp. 1130-1136, 2020. doi: 10.1093/ejcts/ezaa229.

[7] Y. Shen, L. Wang, W. Jian, J. Shang, X. Wang, L. Ju, *et al*., "Big-data and artificial-intelligence-assisted vault prediction and evo-icl size selection for myopia correction," *British Journal of Ophthalmology*, vol. 107, no. 2, pp. 201-206, 2021. doi: 10.1136/bjophthalmol-2021-319618.

[8] J. Feng, J. Ye, G. Qi, L. Hong, F. Wang, S. Liu, *et al*., "A comparative analysis of eight machine learning models for the prediction of lateral lymph node metastasis in patients with papillary thyroid carcinoma," *Frontiers in Endocrinology*, vol. 13, 2022. doi: 10.3389/fendo.2022.1004913.

[9] D. Na, M. Zhai, L. Zhao, and C.-H. Wu, "Cervical cell classification based on the CART feature selection algorithm," *Journal of Ambient Intelligence and Humanized Computing*, 2021. doi: 10.1007/S12652-020-02256-9.

[10] A. Romagnoni, S. Jégou, K. Steen, G. Wainrib, J. Hugot, L. Peyrin–Biroulet, *et al*., "Comparative performances of machine learning methods for classifying crohn disease patients using genome-wide genotyping data," *Scientific Reports*, vol. 9, no. 1, 2019. doi: 10.1038/s41598-019-46649-z.

[11] D. Wang, W. Xu, S. Wang, S. Wang, W. Leng, L. Fu, *et al*., "Lupus nephritis or not? a simple and clinically friendly machine learning pipeline to help diagnosis of lupus nephritis," *Inflammation Research*, vol. 72, no. 6, pp. 1315-1324, 2023. doi: 10.1007/s00011-023-01755-7.

[12] H. Salah and S. Srinivas, "Explainable machine learning framework for predicting long-term cardiovascular disease risk among adolescents," *Scientific Reports*, vol. 12, no. 1, 2022. doi: 10.1038/s41598-022-25933-5.

[13] Y. Du, A. Rafferty, F. McAuliffe, J. Mehegan, and C. Mooney, "Towards an explainable clinical decision support system for large-for-gestational-age births," *Plos One*, vol. 18, no. 2, p. e0281821, 2023. doi: 10.1371/journal.pone.0281821.

[14] J. Zhu, J. Zheng, L. Li, R. Huang, H. Ren, D. Wang, *et al*., "Application of machine learning algorithms to predict central lymph node metastasis in t1-t2, non-invasive, and clinically node negative papillary thyroid carcinoma," *Frontiers in Medicine*, vol. 8, 2021. doi: 10.3389/fmed.2021.635771.

[15] J. Park, T. Hsu, J. Hu, C. Chen, W. Hsu, M. Lee, *et al*., "Predicting sepsis mortality in a population-based national database: machine learning approach," *Journal of Medical Internet Research*, vol. 24, no. 4, p. e29982, 2022. doi: 10.2196/29982.

[16] S. Ou, K. Lee, M. Tsai, W. Tseng, F. Lee, and D. Tarng, "Artificial intelligence for risk prediction of rehospitalization with acute kidney injury in sepsis survivors," *Journal of Personalized Medicine*, vol. 12, no. 1, p. 43, 2022. doi: 10.3390/jpm12010043.

[17] H. Wang, C. Liu, and L. Deng, "Enhanced prediction of hot spots at protein-protein interfaces using extreme gradient boosting," *Scientific Reports*, vol. 8, no. 1, 2018. doi: 10.1038/s41598-018-32511-1.

[18] D. Riana, S. Hadianti, S. Rahayu, Frieyadie, M. Hasan, I. N. Karimah, and R. Pratama, "Repomedunm: A new dataset for feature extraction and training of deep learning network for classification of pap smear images," in *International Conference on Neural Information Processing*, Cham: Springer International Publishing, 2021, pp. 317–325.

[19] D. Riana, S. Hadianti, S. Rahayu, F. Aziz, and O. Kalsoem, "DEEPREPOMEDUNM: A Train Deep Learning Network and Extraction Feature for the Classification of Pap Smear Images," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 19, pp. 5787–5800, 2022.

[20] M. Mahendra, J. Jumadi, and D. Riana, "Cervical cancer papsmear classification through meta-learning technique using convolution neural networks," *Journal Medical Informatics Technology*, vol. 1, no. 4, pp. 105–108, 2023. doi: 10.37034/medinftech.v1i4.23.

[21] A. D. Purwanto, K. Wikantika, A. Deliar, and S. Darmawan, "Decision tree and random forest classification algorithms for mangrove forest mapping in Sembilang National Park, Indonesia," *Remote Sensing*, vol. 15, no. 1, p. 16, 2023. doi: 10.3390/rs15010016.

[22] X. Fu, Y. Chen, J. Yan, Y. Chen, and F. Xu, "BGRF: A broad granular random forest algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 44, no. 5, pp. 8103-8117, 2023.

[23] S. Amini, S. Homayouni, A. Safari, and A. A. Darvishsefat, "Object-based classification of hyperspectral data using Random Forest algorithm," *Geo-spatial information science*, vol. 21, no. 2, pp. 127-138, 2018.

[24] B. Balnarsaiah, T. Prasad, and P. Laxminarayana, "Pixel-Based SAR Image Classification Using Random Forest Algorithm," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 10, pp. 4351-4356, 2019.

[25] D. Jollyta, G. Gusrianty, and D. Sukrianto, "Analysis of Slow Moving Goods Classification Technique: Random Forest and Naïve Bayes," *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, vol. 5, no. 2, pp. 134-139, 2019.

[26] K. Maurya, S. Mahajan, and N. Chaube, "Remote sensing techniques: Mapping and monitoring of mangrove ecosystem—A review," *Complex & Intelligent Systems*, vol. 7, no. 6, pp. 2797-2818, 2021.

[27] C. Ma, B. Ai, J. Zhao, X. Xu, and W. Huang, "Change detection of mangrove forests in coastal guangdong during the past three decades based on remote sensing data," *Remote Sensing*, vol. 11, no. 8, p. 921, 2019. doi: 10.3390/rs11080921.

[28] A. N. Oo and T. Naing, "Decision tree models for medical diagnosis," *International Journal of Trend in Scientific Research and Development*, vol. 3, no. 3, pp. 1697-1699, 2019. doi: 10.31142/ijtsrd23510.

[29] K. Maurya, S. Mahajan, and N. Chaube, "Remote sensing techniques: mapping and monitoring of mangrove ecosystem—a review," *Complex & Intelligent Systems*, vol. 7, no. 6, pp. 2797-2818, 2021. doi: 10.1007/s40747-021-00457-z.

[30] R. K. Gupta, J. Manhas, and M. Kour, "Hybrid feature extraction based ensemble classification model to diagnose oral carcinoma using histopathological images," *Journal of Scientific Research*, vol. 66, no. 03, pp. 219-226, 2022. doi: 10.37398/jsr.2022.660327.

[31] F. Arnia, K. Saddami, and K. Munadi, "Dcnet: noise-robust convolutional neural networks for degradation classification on ancient documents," *Journal of Imaging*, vol. 7, no. 7, p. 114, 2021. doi: 10.3390/jimaging7070114.

[32] Z. Rustam, F. Zhafarina, G. S. Saragih, and S. Hartini, "Pancreatic cancer classification using logistic regression and random forest," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 2, p. 476, 2021. doi: 10.11591/ijai.v10.i2.pp476-481.

[33] Z. Yu, C. Zhang, N. Xiong, and F. Chen, "A new random forest applied to heavy metal risk assessment," *Computer Systems Science and Engineering*, vol. 40, no. 1, pp. 207-221, 2022. doi: 10.32604/csse.2022.018301.

[34] S. Shafei, H. Vahdati, T. Sedghi, and A. Charmin, "Novel high level retrieval system based on mathematic algorithm & technique for mri medical imaging and classification," *Journal of Instrumentation*, vol. 16, no. 07, p. P07055, 2021. doi: 10.1088/1748-0221/16/07/p07055.

[35] N. H. Alkurdy, H. K. Aljobouri, and Z. K. Wadi, "Ultrasound renal stone diagnosis based on convolutional neural network and vgg16 features," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 3, p. 3440, 2023. doi: 10.11591/ijece.v13i3.pp3440-3448.

[36] A. Wang, Y. Wang, and Y. Chen, "Hyperspectral image classification based on convolutional neural network and random forest," *Remote Sensing Letters*, vol. 10, no. 11, pp. 1086-1094, 2019. doi: 10.1080/2150704x.2019.1649736.

[37] N. T. Dinh, N. T. U. Nhi, T. M. Le, and T. T. Van, "A model of image retrieval based on kd-tree random forest," *Data Technologies and Applications*, vol. 57, no. 4, pp. 514-536, 2023. doi: 10.1108/dta-06-2022-0247.

[38] D. K. Prasad, L. Vibha, and K. R. Venugopal, "Early detection and multistage classification of diabetic retinopathy using random forest classifier," *International Journal on Computer Science and Engineering*, vol. 10, no. 3, pp. 77-84, 2018. doi: 10.21817/ijcse/2018/v10i3/181003012.

[39] Y. Suganya, S. Ganesan, and P. Valarmathi, "Comparative analysis of ovarian images classification for identification of cyst using ensemble method machine learning approach," *ECS Transactions*, vol. 107, no. 1, pp. 7407-7415, 2022. doi: 10.1149/10701.7407ecst.

[40] S. Perumal and T. Velmurugan, "Lung cancer detection and classification on ct scan images using enhanced artificial bee colony optimization," *International Journal of Engineering & Technology*, vol. 7, no. 2.26, p. 74, 2018. doi: 10.14419/ijet.v7i2.26.12538.

[41] W. Feng, G. Dauphin, W. Huang, Y. Quan, and W. Liao, "New margin-based subsampling iterative technique in modified random forests for classification," *Knowledge-Based Systems*, vol. 182, p. 104845, 2019. doi: 10.1016/j.knosys.2019.07.016.

[42] W. Feng, W. Huang, and W. Bao, "Imbalanced hyperspectral image classification with an adaptive ensemble method based on smote and rotation forest with differentiated sampling rates," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 12, pp. 1879-1883, 2019. doi: 10.1109/lgrs.2019.2913387.