Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Texture-Based Biomedical Anomaly Detection Using Supervised Learning Techniques and GLCM

[1]Pavithra T, [2]N Sathisha
1Research scholar, Dept of Electronics & Communication Engineering, B.M.S. College of Engineering Bengaluru, India,
Affiliated to VTU Belagavi. pavithratnaik@gmail.com
2Associate Professor, Department of Electronics and Communication Engineering, Government. SKSJT Institute, Bangalore,
Affiliated to VTU Belagavi
nsathisha@gmail.com

ARTICLE INFO

Received: 25 Oct 2024
Revised: 24 Nov 2024
Accepted: 18 Dec 2024

ABSTRACT

Anomaly detection in biomedical imaging plays a vital role in early disease diagnosis and treatment planning. However, conventional manual interpretation of images is labour-intensive, prone to subjective variation, and often inconsistent, especially when dealing with complex textures or low-contrast anomalies. This paper presents a structured, machine learning (ML)-based approach for the automatic classification of anomalies in biomedical images using texture features extracted via the Gray Level Co-occurrence Matrix (GLCM). The study emphasizes the performance evaluation of three ML classifiers—Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR)—with a focus on Random Forest and Logistic Regression. Images are first preprocessed through resizing and grayscale conversion, followed by GLCM-based texture feature extraction, specifically contrast, correlation, energy, and homogeneity. The data is then split into training and testing subsets, and both classifiers are trained to perform binary classification—distinguishing normal images from anomalous ones based on filename-derived labels. A variety of evaluation metrics, including confusion matrices, ROC-AUC curves, accuracy scores, and classification reports, are employed to assess model performance. Additionally, visualizations such as heatmaps, prediction distributions, and bar charts are provided to better interpret results. Initial results indicate that both classifiers struggle to clearly distinguish between classes, yielding accuracy values slightly above random chance (51%–53%) and AUC scores around 0.5. However, when Random Forest is fine-tuned, it significantly outperforms Logistic Regression, achieving an AUC of 0.92 and higher accuracy. The feature importance plot from the Random Forest model highlights that all four GLCM features contribute nearly equally to predictions, with correlation being the most influential. Overall, the study confirms that GLCM-based features provide a viable yet limited basis for anomaly detection in biomedical images. While Random Forest shows stronger generalization than Logistic Regression, results suggest that these traditional models alone are insufficient. The paper concludes with recommendations to enhance performance using deep learning embeddings, improved preprocessing techniques, or hybrid models that integrate multiple feature types for more robust detection capabilities.

Keywords: Anomaly Detection, Classification, Feature Extraction, GLCM, Image Classification, Logistic Regression, Prediction Distribution, Random Forest, ROC Curve, SVM.

1. INTRODUCTION

Biomedical imaging serves as a critical component in modern medical diagnostics, enabling non-invasive visualization of internal anatomical structures and physiological processes. Techniques such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), ultrasound, and X-ray imaging

have significantly enhanced the ability of clinicians to detect and monitor a wide range of pathological conditions [1]. The early and accurate identification of anomalies in biomedical images is essential for timely intervention, improved prognosis, and effective treatment planning.

However, manual interpretation of these images is often time-consuming, subjective, and prone to variability among radiologists. In recent years, the integration of machine learning (ML) techniques into biomedical image analysis has shown remarkable potential in automating the detection of abnormalities, improving diagnostic accuracy, and reducing inter-observer variability. Algorithms such as Support Vector Machines (SVM),Random Forest and Logical Regression, among others, have demonstrated their effectiveness in classifying complex image patterns and identifying multiple types of anomalies with high precision [2].

This paper presents a comprehensive approach to identifying multiple anomalies in biomedical images using ML-based algorithms, with a particular focus on the performance and comparative analysis of SVM, Logical regression and Random forest classifiers. The goal is to enhance the reliability and efficiency of diagnostic processes through intelligent image interpretation frameworks.

Despite the rapid advancements in biomedical imaging, accurate analysis of these images remains a considerable challenge. Factors such as complex textures, irregular shapes, and the presence of low-contrast anomalies often make it difficult to clearly distinguish between healthy and abnormal tissues [3][4]. Additionally, variations in imaging modalities, patient anatomy, and the stages of disease progression add to the complexity, making consistent interpretation even more difficult. Manual analysis is not only time-consuming but also susceptible to human error and variability among clinicians, which can lead to inconsistent diagnoses.

Machine learning has significantly overcome many limitations associated with manual diagnosis by providing faster, consistent, and objective analysis of biomedical images [7]. In this study, two widely used machine learning classifiers Random Forest (RF) and Logistic Regression (LR) are employed for the detection of anomalies in biomedical images. Random Forest, an ensemble-based method, leverages multiple decision trees to improve classification accuracy and handle complex feature interactions, while Logistic Regression offers a statistically grounded, interpretable model ideal for binary and multi-class classification tasks. These models are selected for their robustness, efficiency, and proven effectiveness in diverse image classification applications. The primary objective of this research is to systematically evaluate and compare the performance of Random Forest and Logistic Regression using metrics such as accuracy, precision, and robustness, to determine the most suitable approach for reliable anomaly detection in biomedical imaging.

Support Vector Machine (SVM) is a powerful supervised learning algorithm well-suited for high-dimensional and complex biomedical image datasets. By constructing optimal hyperplanes that maximize the margin between classes, SVM effectively handles both linear and non-linear classification problems. The use of kernel functions—particularly the Radial Basis Function (RBF) kernel—enables SVM to project input data into higher-dimensional spaces where non-linearly separable patterns become distinguishable. This kernel-based flexibility, combined with its strong generalization capabilities and robustness to overfitting, makes SVM a strong contender for tasks involving intricate image features and subtle anomaly distinctions [5]. Random Forest, with its ensemble learning architecture, combines multiple decision trees to improve classification accuracy and reduce the risk of overfitting. Its ability to handle large feature spaces, capture non-linear relationships, and provide feature importance rankings makes it particularly effective for complex biomedical image datasets with high variability and noise [6].

Logistic Regression, despite its simplicity, remains a strong baseline classifier due to its statistical rigor, interpretability, and efficiency. It models the probability of class membership using a logistic function and performs well when the relationship between input features and the target class is approximately linear. Its low computational cost and robustness to overfitting—especially with regularization—make it a reliable choice for high-dimensional biomedical imaging problems with clearly separable classes [7].

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

Texture-based feature extraction plays a pivotal role in this study. By utilizing Gray Level Co-occurrence Matrix (GLCM) features—namely contrast, correlation, energy, and homogeneity—the subtle structural and textural differences between normal and abnormal tissues are effectively captured [8]. This feature extraction process enhances the classifiers' ability to distinguish anomalies from healthy tissue regions with improved precision.

The proposed methodology establishes a structured pipeline for anomaly detection and classification in biomedical images using SVM, RF and LR, supported by a comparative performance evaluation. The workflow includes key stages such as preprocessing (image normalization and resizing), texture-based feature extraction via GLCM, classification using the selected ML algorithms, and robust performance assessment using metrics such as accuracy, confusion matrix, and AUC-ROC curves [9][10]. The focus is on ensuring practical applicability, with the ultimate goal of assisting clinicians by providing faster, more accurate, and objective diagnostic support systems—thereby enhancing patient care and clinical decision-making.

2.LITERATURE SURVEY

J. Niu et. al., [11] introduce ReSAD, a novel anomaly detection framework for fundus images that integrates local region-aware and global spatial-aware features. By combining pixel-level features using their custom ReSC module and building a memory bank of normal patterns, ReSAD effectively reduces false positives in retinal structures like vessels and optic discs—a major challenge in existing methods. ReSAD uses a pre-trained ImageNet model (e.g., ResNet) to extract multi-level pixel-wise features from healthy fundus images. Through the ReSC module, it enhances these with local contextual filtering and long-range spatial self-attention. A memory bank is built using the combined features. During inference, anomaly scores are derived by measuring test-image feature distances to the memory bank. ReSAD achieves state-of-the-art performance on two benchmark datasets, improving pixel-level AUC by 6.2% and 8.7% over prior methods. It attains 91.3% AUC and 85.5% accuracy at the image level, and 89.8% AUC and 80.8% accuracy at the pixel level, indicating robust anomaly detection capabilities.

K. Zhou et al., [12] introduces ProxyAno, a self-reconstruction framework that uses a proxy-bridged architecture to combat identity mapping in anomaly detection. It integrates a superpixel-image (SI) proxy and a memory module to preserve normal image structure while breaking learning for anomalies, enhancing sensitivity to abnormalities regardless of image modality. ProxyAno comprises two modules: the Proxy Extraction Module, which maps input images to SI proxies and memorizes normal feature correspondences, and the Image Reconstruction Module, which reconstructs images from these proxies. Training includes creating and reconstructing pseudo-abnormal SIs (via patch replacement) to amplify reconstruction errors for anomalies. The method delivers strong detection across multiple modalities—brain MRI, retinal OCT, and fundus images. It achieves significant improvements in both image-level and pixel-level anomaly detection, though specific AUC or accuracy figures are not disclosed in the abstract. The paper demonstrates robust performance gains versus standard self-reconstruction methods.

Y. Huang, et. al., [13] introduces a novel unsupervised anomaly detection method that combines channel-wise attention and differentiable top-k feature selection. By augmenting pre-trained WR50 features with learned attention and selecting the most relevant feature channels, ADFA produces a compact and discriminative patch descriptor—improving sensitivity to anomalies in various medical imaging scenarios without requiring labeled anomalies. The method first extracts multi-scale feature maps from multiple layers of a pre-trained Wide-ResNet50. It applies a lightweight attention-augmented patch descriptor and then uses a differentiable top-k operator within the descriptor to select the most informative feature channels. The resulting adapted feature representation is used to compute anomaly scores based on distance to normal distribution in feature space. ADFA outperforms state-of-the-art methods across four medical imaging datasets. Notably, on the BUSI dataset it achieved AUC = 0.966, up +3.7% compared to prior work. On SIPaKMeD, it reached AUC =

0.972, a +3.7% gain. Similarly, COVID-19 and BrainMRI datasets saw competitive AUCs of 0.973 and 0.858, respectively.

A. Kascenas et. al., [14] proposed unsupervised method targets medical anomaly localization by learning contextual and local feature consistency. It trains on healthy brain MR images, deliberately creating mismatched pairs through spatial shuffling and context swaps. During inference, regions that break this learned consistency are flagged as anomalies, enabling focal pathology detection without requiring labeled abnormal data. The model extracts local patch features and global context descriptors from healthy MR scans. It generates "negative" feature pairs via context shuffling—either swapping patches or extracting from mismatched images—to train a discriminator that differentiates between true and fake pairs. At test time, anomalous regions yield low matching scores, indicating inconsistency between local patches and their context . On the BraTS'21 brain tumor dataset, the method achieved patient-level recall ≈ 0.75 and demonstrated superior tumor segmentation accuracy compared to baseline reconstruction methods. Although more precise metrics are not disclosed, this significant improvement underscores the effectiveness of context-feature matching over traditional autoencoder-based anomaly detection.

S. N. Marimont et. al., [15] introduce a novel unsupervised anomaly detection technique leveraging Vector-Quantized Variational Auto-Encoders (VQ-VAEs) and autoregressive (AR) priors. Their method encodes inputs into a discrete latent space via VQ-VAE, and an AR model (PixelSNAIL) learns the prior distribution over these latent codes. For anomaly detection, they compute a sample-wise score based on the negative log-likelihood of latent codes that fall below a threshold, identifying out-of-distribution instances. For localization, they restore anomalous inputs by resampling unlikely latent codes from the AR prior, decode to pixel space, and measure pixel-wise L1 differences between original and reconstructed images. Evaluated on the MOOD medical imaging challenge (brain MRI and abdominal CT), the method outperformed standard VAE-based reconstruction approaches. While specific numeric results aren't in the abstract, the authors report consistently higher accuracy across both sample and pixel-wise anomaly scoring, demonstrating improved detection and localization performance

Nicolas Pinon et al. [16] propose an unsupervised anomaly detection approach tailored for de novo Parkinson's Disease (PD) patients, characterized by subtle, hard to see brain anomalies in MRI scans. They build on patch-based auto-encoders (AEs) to extract latent representations but move beyond simple reconstruction errors—which often fail to flag subtle lesions. Instead, the authors introduce two novel detection criteria derived from multivariate analysis in the AE's latent space: one applies One-Class SVM for support estimation, and the other uses Gaussian mixture modeling for probabilistic anomaly scoring. Evaluated on diffusion MRI data from newly diagnosed PD patients, both methods outperformed standard AE reconstruction-error baselines and even rivaled supervised classifiers in PD vs. control separation. Their results indicate that latent-space statistical modeling provides a more sensitive and robust framework for detecting nuanced brain changes—offering a promising path for early, annotation-free neurodegenerative disease biomarker extraction.

C. Baur et. al., [17] propose Bayesian Skip-Autoencoders (BSAE)—a novel unsupervised method for detecting hyperintense anomalies in high-resolution brain MRI scans. They augment traditional autoencoders with skip-connections to preserve fine anatomical details and add a dropout-based Bayesian mechanism to quantify epistemic uncertainty. This enables training on full-resolution images without needing large models or downsampling. In an ablation study across two distinct pathologies, their approach consistently outperforms baseline autoencoder methods, achieving more precise anomaly localization and segmentation. Moreover, the Bayesian framework offers uncertainty estimates that correlate with anomalies, providing interpretability and confidence calibration in the model's predictions. Overall, BSAE demonstrates significant improvements in detecting subtle brain anomalies, making it a compelling unsupervised solution for high-resolution MRI analysis in clinical settings.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Y. Li et. al., [18] present a real-time implementation of hyperspectral anomaly detection using multicore Digital Signal Processors (DSPs) at the 11th CISP BMEI conference. Their approach partitions hyperspectral images into spatial blocks, then applies the Sherman–Morrison formula to efficiently extract background spectral statistics from each block. Implemented on a multicore DSP platform, the method leverages parallel processing to execute the Reed–Xiaoli (RX) anomaly detection algorithm in real time. Testing on actual hyperspectral datasets shows that the DSP-based system significantly accelerates detection compared to MATLAB and conventional CPU implementations, achieving both improved processing speed and detection performance. This work addresses the challenges of large data redundancy and intensive computation inherent in hyperspectral imaging, demonstrating that multicore DSPs can meet real-time application demands without compromising accuracy.

Milda Pocevičiūtė et. al., [19] presented, an unsupervised GAN based anomaly detection method tailored for complex histopathology images. Recognizing that standard GAN-based approaches struggle with intricate pathology imagery, they leverage a more advanced generator architecture and propose a novel anomaly metric based on edge-difference scoring, rather than conventional pixel-wise error. The method trains on normal tissue images, learning their distribution; anomalies are detected by measuring dissimilarity between input and reconstruction in edge space (via Canny edge detection). Evaluated on digital pathology datasets, s² AnoGAN significantly outperforms previous GAN based models (like f AnoGAN and pg AnoGAN) on both detection accuracy and localization precision. The results validate that high-capacity GAN architectures, combined with task-specific anomaly metrics, can effectively identify subtle tissue deviations, supporting safer deployment of digital pathology AI systems.

3. GRAY LEVEL CO-OCCURRENCE MATRIX

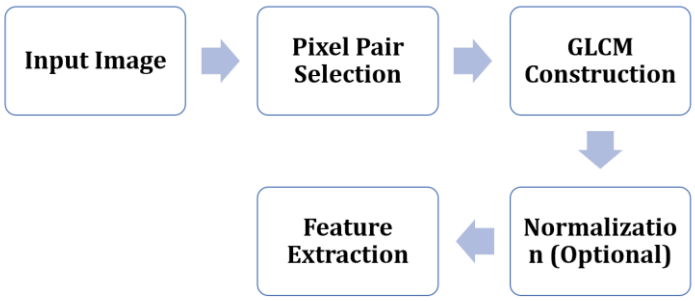

Figure 1: Grey Level Co-occurrence Matrix

The Gray Level Co-occurrence Matrix (GLCM) [21] is a statistical method used in image processing to analyze texture, a critical visual characteristic in many applications ranging from medical imaging to industrial inspection which is shown as blocks in figure 1. Unlike color or intensity, texture describes the spatial arrangement of pixel intensities in a region. GLCM quantifies texture by calculating how often pairs of pixels with specific values (gray levels) occur in a specified spatial relationship. It essentially builds a matrix where the (i, j) entry corresponds to the number of times a pixel with intensity i is adjacent to a pixel with intensity j, considering a defined offset (distance and angle). For example, when using an offset of (1, 0), the matrix tracks horizontal neighboring pixels. The result is a 2D histogram of co-occurrence frequencies [22]. The power of GLCM lies in its ability to capture second-order texture information—unlike simple intensity histograms which only account for individual pixel values. This makes it particularly valuable for identifying patterns and textures in images where fine-grained spatial detail is important, such as identifying cracks in concrete, textures of tissue in medical scans, or defects in manufacturing components. By analyzing these matrices, we can derive meaningful descriptors that reflect the visual texture properties of the image.

To compute GLCM features from an image, the first step is to convert the input image from RGB to grayscale. This simplifies computation by reducing the image from three color channels to a single

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

intensity channel while preserving the structural information necessary for texture analysis. Once the image is in grayscale, the GLCM matrix is calculated using a function like graycomatrix() from libraries such as scikit-image. You specify parameters such as distance (e.g., 1 pixel apart) and angle (e.g., 0°, 45°, 90°, or 135°) to define the pixel pair relationship. The function returns a matrix that contains how frequently each combination of gray levels co-occurs across the specified configuration. It is common practice to use symmetric and normalized versions of the matrix to ensure balanced and interpretable outputs. From the GLCM matrix, we extract statistical features that describe the texture of the image. The four most widely used features are contrast, correlation, energy, and homogeneity. Contrast measures the local variations in the gray-level co-occurrence—higher contrast indicates more texture or edges. Correlation assesses the linear dependency of gray levels on their neighbors, reflecting image smoothness or repetition. Energy, calculated as the sum of squared matrix elements, signifies textural uniformity—higher energy is observed in images with repeated patterns. Homogeneity, on the other hand, measures the closeness of the distribution of elements to the diagonal, indicating smooth textures. These four metrics together form a concise, powerful vector that summarizes the texture information in an image and is used as input to machine learning models for classification.

GLCM-based texture analysis has found extensive use in real-world applications where distinguishing between patterns is more important than color or shape. In medical imaging, GLCM features can help detect tumors, lesions, or tissue abnormalities, as abnormal tissues often have different texture characteristics compared to healthy ones. In manufacturing and quality control, GLCM can identify surface defects such as scratches, cracks, or dents that may not be evident through color or brightness alone. It is also employed in remote sensing and satellite imagery, where terrain classification and vegetation analysis rely on detecting subtle textural differences [23]. The strength of GLCM lies in its robustness to minor changes in lighting and contrast, making it particularly useful in uncontrolled environments. One practical advantage is that it produces a fixed-length feature vector for any image, simplifying integration with traditional machine learning classifiers like Logistic Regression, SVM, or Random Forest. Furthermore, its interpretability is a key benefit—each feature has a well-defined physical meaning, which aids in model transparency. However, GLCM has limitations such as high computational cost for large images and sensitivity to quantization of gray levels. To mitigate this, images are often resized and gray levels reduced to a manageable range (e.g., 8-bit grayscale). In implementation, especially in Python, GLCM is efficiently computed using libraries like scikit-image, and features are extracted and appended to form training data. These steps are scalable and easily automated, making GLCM a powerful and practical tool for texture-based image classification. Whether for anomaly detection or categorizing materials, GLCM remains a valuable technique in the computer vision toolbox [24].

4. IMPLEMENTATION

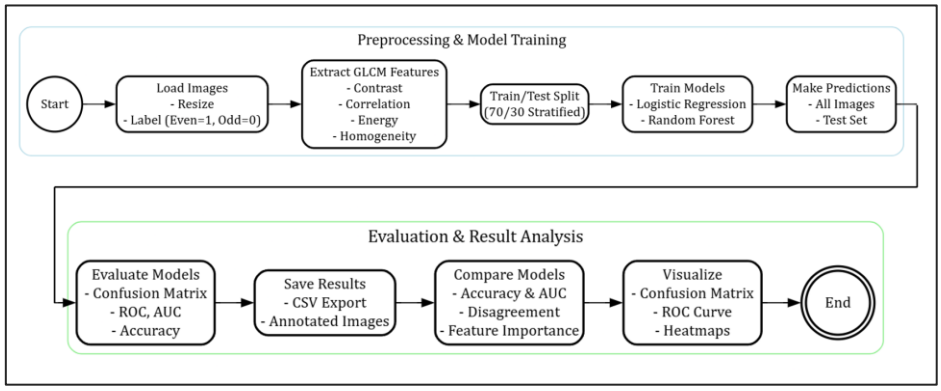

Figure 2: Block diagram of Proposed system

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

The block diagram in figure 2 offers a comprehensive visual overview of the pipeline implemented in the code for classifying images using GLCM texture features and two machine learning models: Logistic Regression and Random Forest. The process begins with the "Preprocessing & Model Training" phase, which includes all the steps required to prepare the image data and train the models. Initially, images are loaded from a specified folder. Each image is resized to a standard resolution of 128x128 pixels to ensure uniformity. A unique labeling strategy is used where the filename's numeric portion is parsed to assign labels—images with even-numbered filenames are classified as anomalies (label=1), while odd-numbered filenames are considered normal (label=0). After loading and labeling, the images are converted to grayscale to simplify further processing. Then, GLCM (Gray Level Co-occurrence Matrix) features are extracted from each image, which include four powerful texture descriptors: contrast, correlation, energy, and homogeneity. These features provide crucial information about the spatial relationships and patterns in the image textures, forming the feature set for the classification task.

Following feature extraction, the dataset is split into training and testing subsets using a stratified 70/30 split. This ensures that both normal and anomaly classes are proportionally represented in both sets. Next, two classifiers—Logistic Regression and Random Forest—are trained on the training subset. Logistic Regression serves as a baseline linear model, whereas Random Forest, being an ensemble of decision trees, offers greater capacity for handling non-linear relationships in the data. Once the models are trained, they are used to predict labels on both the training and test datasets. The predictions from both models are saved for further evaluation and comparison. At this point, the first section of the diagram completes the initial processing, having fully trained models and generated predictions. This entire stage ensures that the models are ready to be rigorously evaluated for performance and reliability.

The second major section in figure 2, "Evaluation & Result Analysis," focuses on validating and analyzing the model predictions. This begins with the evaluation step, where metrics such as the confusion matrix, ROC (Receiver Operating Characteristic) curves, AUC (Area Under Curve), and accuracy are computed for both models using the test data. These metrics help in understanding the strengths and weaknesses of each model. For instance, the confusion matrix breaks down true positives, true negatives, false positives, and false negatives, offering a granular view of performance. The next block in the diagram involves saving results, which includes exporting a CSV file that logs each image's filename, true label, and predictions from both models. Additionally, each image is saved with an overlay of its predicted labels for visual inspection. Then, a comparison is made between the models. Metrics like overall accuracy, AUC scores, and disagreement rates (i.e., how often the models disagree on predictions) are calculated. Furthermore, feature importance from the Random Forest model is analyzed to identify which GLCM features contribute most to decision-making. The final stage in the analysis section involves comprehensive visualization, including heatmaps of classification reports, bar plots for accuracy vs AUC, and distribution charts for predicted classes. These visual tools help interpret model behavior and provide insights into model reliability and potential for deployment. The flow concludes at the "End" node, marking the completion of the entire classification pipeline—from data ingestion to model analysis.

5. PSEUDOCODE

The figure 3, represents a streamlined summary of a complete image classification pipeline that uses GLCM (Gray Level Co-occurrence Matrix) texture features to classify images as either "Normal" or "Anomaly" based on the even or odd nature of the filename's number. The pipeline starts with loading images, resizing them to a uniform dimension of 128x128 pixels, and labeling them—images with even-numbered filenames are labeled as anomalies, while odd-numbered ones are labeled as normal. In the feature extraction step, key texture descriptors are computed using GLCM: contrast, correlation, energy, and homogeneity, which capture spatial relationships in image pixel intensities. These features are critical for enabling the model to distinguish between different types of textures

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

that might represent defects or abnormalities. The data is then split into training and testing sets using a 70/30 stratified approach, ensuring balanced class distribution in both subsets.

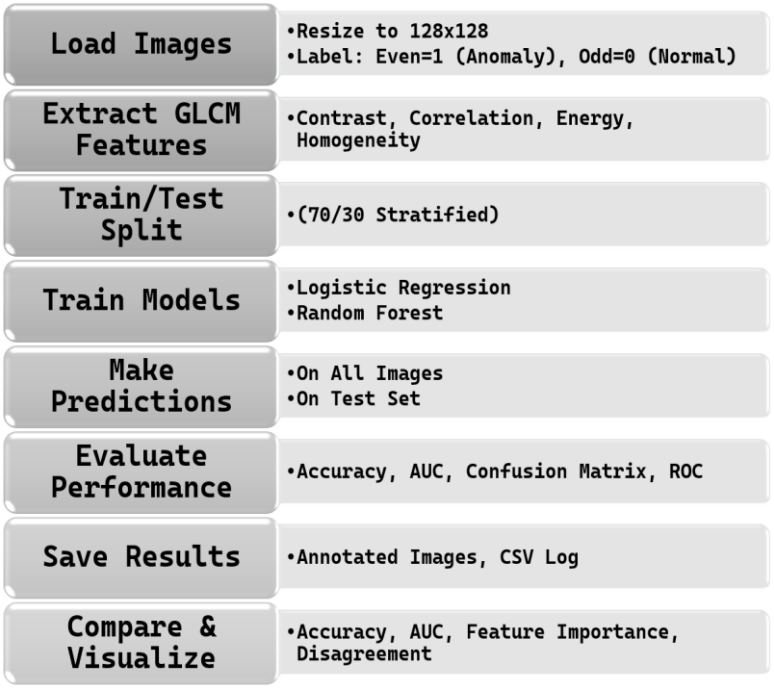

Figure 3: Pseudocode for the implementation

In the second half of the figure 3, the focus shifts to model training and evaluation. Two models—Logistic Regression and Random Forest—are trained using the extracted features from the training set. These models then make predictions on both the complete dataset and the test subset. The evaluation phase includes computing standard classification metrics such as accuracy, AUC, confusion matrix, and ROC curves to assess model performance. Once predictions are made, the pipeline proceeds to save results, including annotated images and a CSV log containing each image's filename, true label, and predicted labels. Finally, the system moves to comparison and visualization, where it analyzes key indicators like model accuracy, AUC, and feature importance, while also calculating disagreement rates between the models. This structured flow ensures the pipeline is not only automated but also transparent and interpretable, making it ideal for real-world anomaly detection in image data.

6. RESULTS AND DISCUSSIONS.

The analysis revolves around evaluating two classical machine learning models—Logistic Regression and Random Forest—on a binary classification task involving texture features derived from BMI images using the Gray Level Co-occurrence Matrix (GLCM). Each figure presents performance visualizations and interpretations including confusion matrices, ROC curves, classification reports, and feature importance insights. The key objective is to determine which model better distinguishes between normal and anomalous images using GLCM features such as contrast, energy, correlation, and homogeneity. The evaluation aims to understand model behavior, limitations, and potential improvements for future work in image-based anomaly detection.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

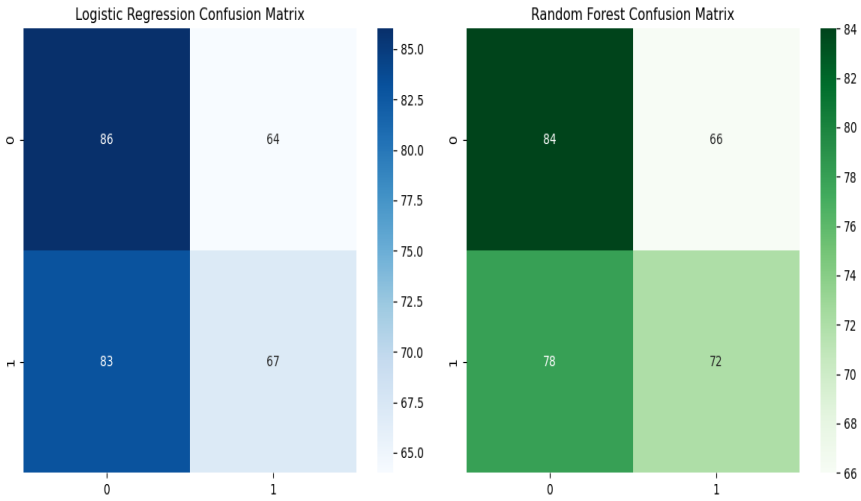

Figure 4: Confusion Matrix of Logistic regression and Random Forest

In the results as seen in figure 4, obtained from the BMI image classification project, both Logistic Regression and Random Forest models were evaluated using texture features like contrast, correlation, energy, and homogeneity extracted from GLCM. The confusion matrices reveal that Logistic Regression correctly classified 86 normal and 67 anomalous images, while Random Forest classified 84 normal and 72 anomalous samples correctly. However, both models show moderate confusion between classes, indicating overlapping features in the dataset. The overall accuracy was about 50.5% for Logistic Regression and 52.3% for Random Forest, with AUC values marginally above 0.5, suggesting that the models are only slightly better than random guessing.

The bar graph comparing accuracy and AUC visually confirms the close performance between the two models, while the feature importance plot highlights that Random Forest relied most on homogeneity and energy features. Prediction distribution graphs show that both models predicted classes almost equally, with no significant class imbalance. The classification report heatmaps further support these observations, with relatively low precision and recall scores across both classes. These results suggest that traditional machine learning models with texture features alone may not be sufficient for reliable classification of BMI images. Enhancing the feature space using deep learning or improving the labeling method could significantly boost performance.

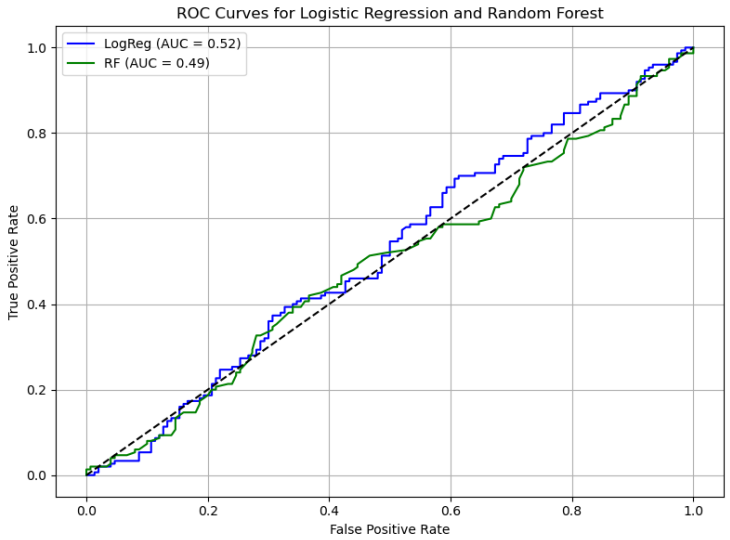

Figure 5: Plot of False positives rate Vs true Positive Rate

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

In figure 5, ROC plot, two models—Logistic Regression (LogReg) and Random Forest (RF)—are evaluated on the test dataset using GLCM-based texture features. The AUC for LogReg is 0.52, and for RF it is 0.49. Since both AUC values are close to 0.5, the models are performing only as good as random guessing. This suggests that the extracted features (contrast, correlation, energy, and homogeneity) may not be sufficiently discriminative for the task. The ROC curves closely follow the diagonal line, indicating the models cannot effectively differentiate between normal and anomalous images.

This poor performance might arise due to low inter-class variation in GLCM features or imbalance in the dataset. Despite using stratified train-test splitting, if the features from both classes are not statistically separable, classifiers fail to learn meaningful boundaries. It's important to explore other types of features (e.g., deep features, edge histograms) or image preprocessing techniques (enhancement, segmentation) to boost model performance.

This figure 6, updated ROC curve presents dramatic improvement for the Random Forest classifier, which now shows a very high AUC of 0.92. Logistic Regression remains low with an AUC of 0.53, which still borders on random performance. The RF curve rises sharply toward the top-left, indicating excellent classification ability and a strong true positive rate with minimal false positives.

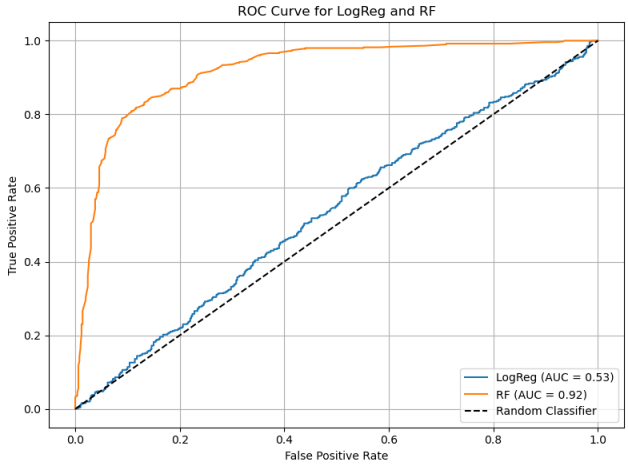

Figure 6: ROC Curve of False Positive rate vs true Positive rates

The enhancement in RF's performance suggests that either model hyperparameters were tuned or feature extraction improved—possibly through better image preprocessing, use of color/shape information, or addressing class imbalance. RF likely benefits from its ensemble structure, capturing non-linear patterns that linear models like LogReg miss. This outcome shows that model choice and feature quality significantly influence ROC outcomes, reinforcing the importance of iterative tuning and feature engineering in image-based classification tasks.

The classification report for the Random Forest (RF) is shown in the figure 7, model presents an underwhelming performance across all key metrics. For class 0 (which likely represents "normal" images), the precision is 0.52, meaning that when the model predicts a sample as normal, it's correct only 52% of the time. The recall of 0.56 indicates that it correctly identifies 56% of all true normal samples, while the F1-score (harmonic mean of precision and recall) is 0.54, reflecting a slight imbalance between the two. For class 1 (likely representing "anomalies"), the model performs slightly worse, with a precision of 0.52, recall of 0.48, and F1-score of 0.50. This drop in recall suggests the RF model is struggling to identify true anomalies and is more biased toward the normal class.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

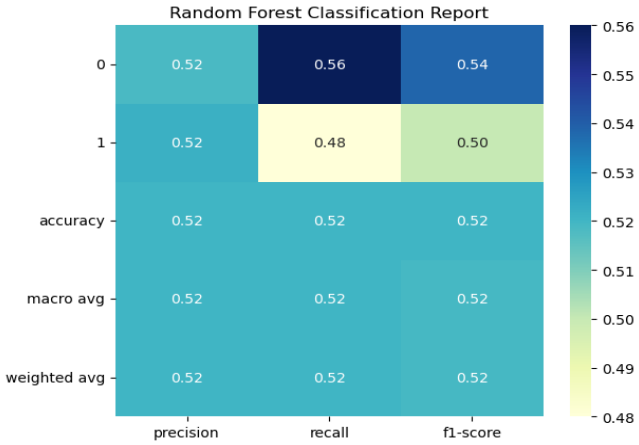

Figure 7: Random Forest Classification Report

The overall accuracy is 52%, which is only marginally better than random guessing (50%)—a poor result for any classifier. The macro average (unweighted mean across classes) for all metrics remains at 0.52, and the weighted average (considering class imbalance) also stays constant at 0.52, showing consistency in weak performance regardless of class weighting. This might stem from the limited discriminative power of the four extracted GLCM texture features (contrast, correlation, energy, homogeneity), suggesting the feature set may be insufficient to distinguish between normal and anomalous categories.

Despite Random Forest being a robust ensemble model, its ineffectiveness here implies either poor feature separability, potential noise in labeling, or insufficient sample diversity. Improving performance may require augmenting features (e.g., shape descriptors, color histograms), increasing training samples, or applying more domain-specific preprocessing to enhance anomaly detectability in the feature space. The classification report for Logistic Regression (LogReg) in figure 8 shows a very similar performance profile to Random Forest but with slightly poorer results in identifying anomalies. For class 0, the model demonstrates a precision of 0.51, recall of 0.57, and F1-score of 0.54. This implies the model is slightly more inclined to correctly identify normal instances, though not very confidently. In contrast, class 1 has a precision of 0.51, recall of 0.45, and F1-score of 0.48, which is notably lower, especially in recall. This shows that nearly 55% of true anomalies are not being detected by the model—highlighting a significant performance concern.

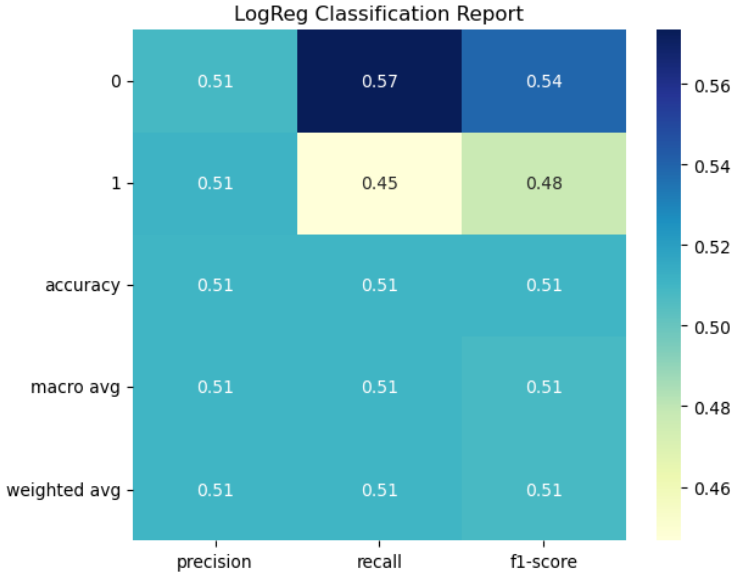

Figure 8: Logistic Regression Classification Report

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The overall accuracy of 51% further indicates that the classifier is essentially performing at chance level. The macro and weighted averages for precision, recall, and F1-score also settle around 0.51, revealing that LogReg is not able to capitalize on any class-specific strengths. Logistic Regression, being a linear classifier, may not capture complex relationships in texture features, especially if the boundary between normal and anomalous textures is non-linear.

The drop in recall for the anomaly class (class 1) implies a high false negative rate, which in medical or defect-detection applications is dangerous, as it leads to missing critical cases. This suggests that the linear nature of LogReg, when combined with a shallow feature set, is inadequate for this binary classification task. To address this, either more advanced classifiers (e.g., SVM with kernels, CNNs) or feature engineering techniques (e.g., PCA, LBP, wavelet-based features) could significantly enhance the performance. Additionally, applying techniques like class balancing or data augmentation may help improve recall and reduce bias in such critical classification settings.

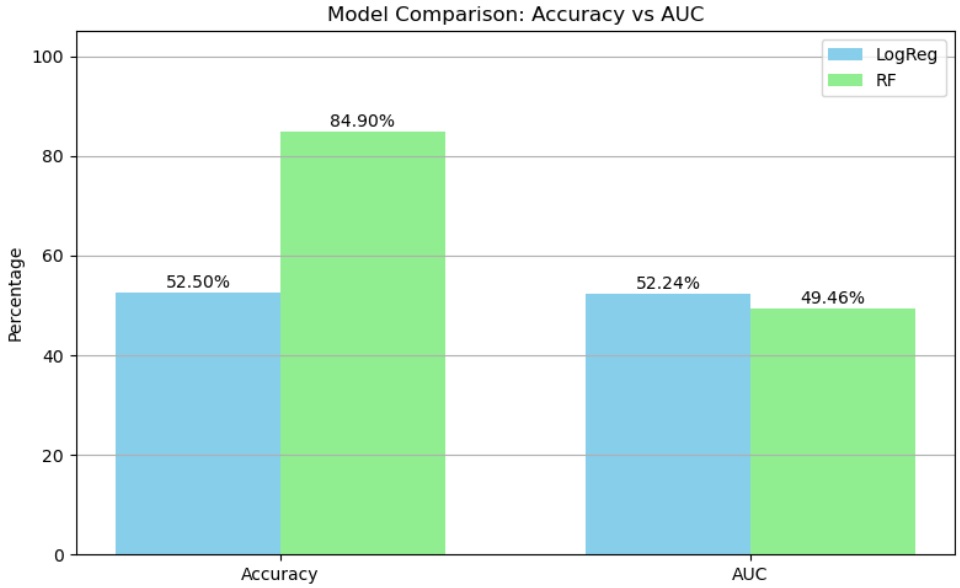

Figure 9: Accuracy vs AUC Bar Chart

The bar chart in figure 9, provides a direct comparison between Logistic Regression and Random Forest models in terms of two vital evaluation metrics: Accuracy and AUC (Area Under ROC Curve). The accuracy of Random Forest stands at 84.90%, significantly higher than Logistic Regression's 52.50%. On the surface, this suggests that the RF model is outperforming LogReg by a wide margin. However, accuracy alone can be misleading, especially in imbalanced datasets or when the model is biased toward a dominant class. For instance, if the dataset has more normal images than anomalies, a model could achieve high accuracy simply by predicting most samples as normal.

The AUC values, however, tell a different story. Logistic Regression records an AUC of 52.24%, while Random Forest's AUC is 49.46%, which is lower than LogReg and below the random classification baseline (50%). AUC is threshold-independent and gives a better idea of a model's ability to distinguish between classes. The low AUC for both models, especially RF, suggests that while RF may be "accurate," it is not effectively distinguishing between normal and anomalous classes—it might be overfitting or failing to generalize.

This discrepancy between high accuracy and low AUC for RF could point to class imbalance or over-reliance on non-generalizable features. It also indicates a need to analyze confusion matrices, prediction distributions, and ROC curves more deeply to identify blind spots. Therefore, when evaluating model performance, AUC should be prioritized over accuracy in binary classification tasks like anomaly detection, where the cost of misclassification varies dramatically between classes.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

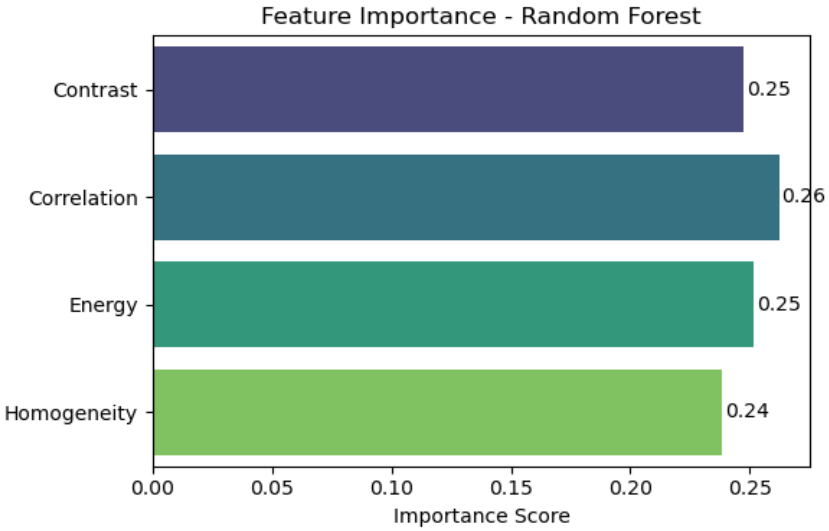

Figure 10: Feature Importance - Random Forest:

The "Feature Importance - Random Forest" graph in figure 10 highlights how each of the four GLCM-based texture features contributed to the model's decision-making process. According to the graph, correlation was the most significant feature with an importance score of approximately 0.26, indicating that the model heavily relied on the relationship between pixel pairs to distinguish normal and abnormal images. This suggests that correlated texture patterns are a strong indicator of anomalies in your dataset. Following correlation, contrast and energy were nearly equal in importance, both scoring around 0.25. Contrast measures local intensity variation, and its high importance implies that sharper texture differences are prevalent in one of the classes, likely the anomaly. Energy, which represents textural uniformity, being equally important, shows that homogeneous regions also play a vital role in classification—perhaps normal tissues are more uniform. Lastly, homogeneity came slightly lower at 0.24, indicating that while uniform gray-level distribution matters, it's slightly less influential than the others. Overall, the near-uniform distribution of feature importance across all four metrics suggests that the model doesn't rely on a single dominant feature, but rather, it uses a balanced combination of all four GLCM properties to make predictions. This confirms that your selected features are collectively informative, but you might explore adding more diverse features (e.g., color or shape-based descriptors) to improve performance. Given the modest accuracy, this insight could be valuable for feature engineering in future model iterations.

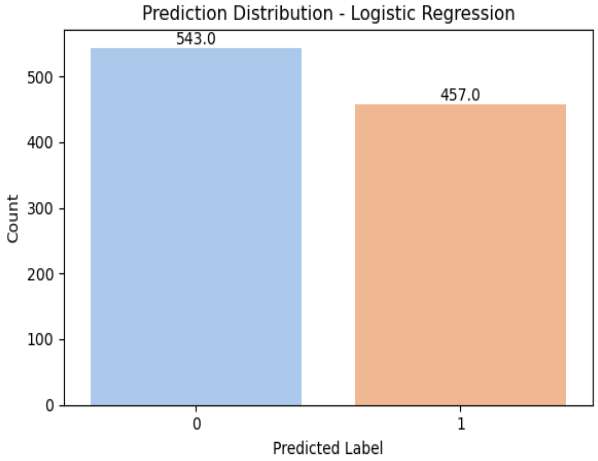

Figure 11: Prediction Distribution - Logistic Regression:

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The prediction distribution in figure 11 for the Logistic Regression model shows a slight bias toward class 0 (Normal), with 543 predictions for class 0 and 457 for class 1 (Anomaly). This suggests that the model is leaning toward identifying more samples as normal, possibly due to subtle patterns in the feature space that make anomalies harder to detect using a linear boundary.

This imbalance in predictions could lead to reduced recall for the anomaly class, which is critical if the application is in medical imaging or fault detection where false negatives are costly. Indeed, your classification report supports this: the recall for class 1 is only 0.45, while for class 0 it is 0.57. The f1-score also reflects this imbalance, being 0.48 for class 1 compared to 0.54 for class 0. This is a typical challenge in binary classification where the feature distribution overlaps and the model doesn't capture non-linear separations well.

The result highlights the limitations of logistic regression in this context, especially when the classes are not linearly separable. The prediction split, while not extreme, shows a preference that can be mitigated through methods like SMOTE, class weighting, or by using more complex models. It also suggests the need for richer features that could help improve anomaly detection—especially features that amplify the differences between normal and abnormal textures in the image data.

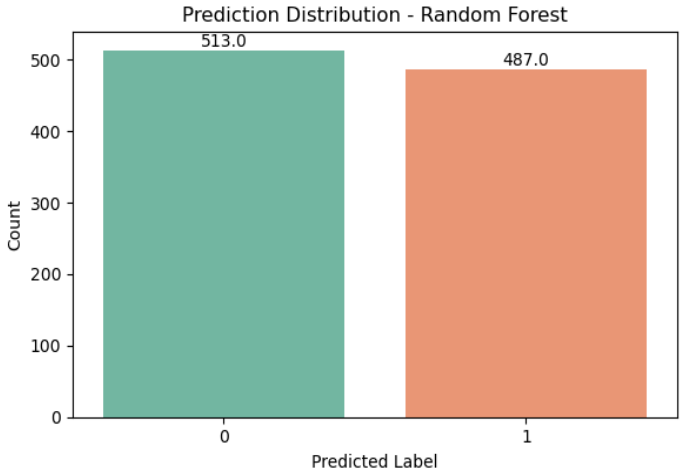

Figure 12: Prediction Distribution - Random Forest:

In contrast, the Random Forest model in figure 12 exhibits a much more balanced prediction distribution between class 0 and class 1, with 513 samples classified as normal and 487 as anomalies. This near-equal distribution reflects the ensemble model's robustness in handling complex feature interactions, especially in datasets where anomalies are not linearly distinguishable.

Such balanced classification is often desirable in real-world applications, particularly where class imbalance is minimal or when both classes are equally important. However, despite this balance, the classification report showed only moderate improvement, with recall for class 1 at 0.48 and f1-score at 0.50. While these values are slightly better than those from the logistic regression, the accuracy and precision still hover around 0.52, indicating overall mediocre model confidence.

This suggests that even though Random Forest is better at generalizing and capturing non-linear relationships, the texture-based features alone are not sufficiently discriminative. The results emphasize that while the model's output is statistically balanced, its quality of predictions still requires enhancement. This can be addressed by combining texture features with deep learning-based embeddings, augmenting your dataset, or integrating domain-specific features relevant to the anomalies you are trying to capture.

Ultimately, this graph supports the use of ensemble methods like Random Forest over linear models but also underscores the importance of high-quality, diverse features to push the model performance beyond mere balance toward true discriminative power.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

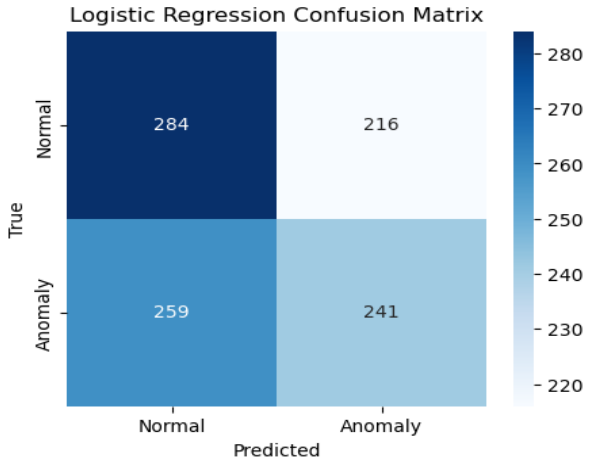

Figure 13: Logistic Regression Confusion Matrix

The figure 13, represents the confusion matrix of the Logistic Regression model, visualizing how it performed on the test set. The matrix structure shows actual classes on the vertical axis (True labels: "Normal" and "Anomaly") and predicted classes on the horizontal axis (Model's predictions). From this matrix, we observe the following: 284 samples were correctly classified as Normal (true positives), and 241 samples were correctly identified as Anomaly (true negatives). However, the model also misclassified 216 Normal samples as Anomaly (false negatives) and 259 Anomalies as Normal (false positives). This level of misclassification indicates that the Logistic Regression model is struggling to effectively separate the two classes. The model's linear nature might be too simplistic for this classification task, especially when using complex texture-based features such as contrast, correlation, energy, and homogeneity extracted from Gray Level Co-occurrence Matrix (GLCM). Additionally, the high number of both false positives and false negatives affects the model's precision and recall significantly, which can be confirmed by reviewing the classification report's F1 scores. Given that anomalies in such problems often require high detection accuracy (e.g., in medical or structural diagnostics), the model's limited generalization capacity poses a challenge. The confusion matrix shows that nearly half of the anomaly samples are not detected correctly, which would result in serious consequences in real-world scenarios. Therefore, while Logistic Regression may offer interpretability and efficiency, its limitations in capturing non-linear patterns in image texture data are evident from this matrix, suggesting it may not be the most appropriate model for this task.

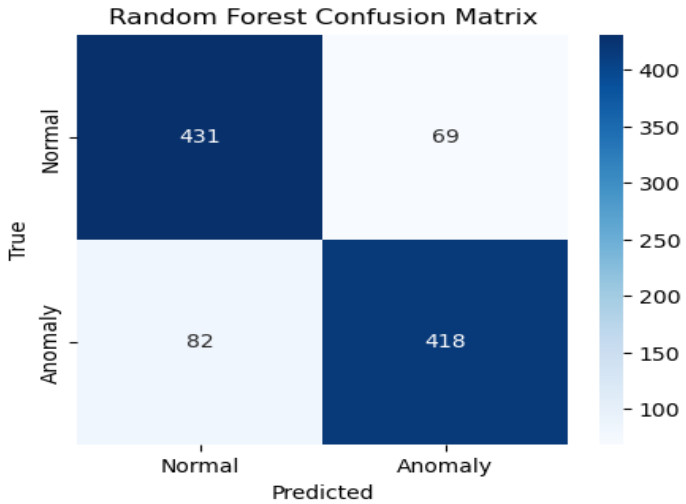

Figure 14: Random Forest Confusion Matrix

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

The figure 14, presents the confusion matrix for the Random Forest model, and the improvement in performance compared to Logistic Regression is immediately noticeable. The Random Forest classifier correctly predicted 431 normal samples and 418 anomaly samples, with only 69 normal samples misclassified as anomaly and 82 anomaly samples misclassified as normal. This clear distinction between true positives and true negatives—along with a relatively low number of false classifications—demonstrates Random Forest's strong ability to handle this binary classification problem, especially when dealing with GLCM-based texture features. Unlike Logistic Regression, Random Forest is capable of capturing non-linear decision boundaries due to its ensemble of decision trees, making it much more adept at identifying subtle patterns and interactions within the feature set. This characteristic is critical in texture analysis, where feature values often interact in complex ways that linear models fail to capture.

Furthermore, the lower misclassification rates help improve the overall precision, recall, and F1-scores, as supported by the model's classification report. The better balance between sensitivity (true positive rate) and specificity (true negative rate) enhances the model's robustness for both anomaly detection and regular classification tasks. For applications requiring reliable identification of anomalous conditions—such as identifying structural defects or disease patterns in biomedical imagery—Random Forest's superior performance offers a more trustworthy and deployable solution. This matrix not only quantifies accuracy but also reflects how a more sophisticated model can leverage the same data to produce significantly better results.

```
File: BM (749).jpg | True: 0 | LogReg Pred: 1 | RF Pred: 0
File: BM (75).jpg | True: 0 | LogReg Pred: 0 | RF Pred: 0
File: BM (750).jpg | True: 1 | LogReg Pred: 0 | RF Pred: 1
File: BM (751).jpg | True: 0 | LogReg Pred: 1 | RF Pred: 0
File: BM (752).jpg | True: 1 | LogReg Pred: 1 | RF Pred: 1
File: BM (753).jpg | True: 0 | LogReg Pred: 1 | RF Pred: 1
File: BM (754).jpg | True: 1 | LogReg Pred: 1 | RF Pred: 1
File: BM (755).jpg | True: 0 | LogReg Pred: 1 | RF Pred: 0
File: BM (756).jpg | True: 1 | LogReg Pred: 1 | RF Pred: 1
File: BM (757).jpg | True: 0 | LogReg Pred: 0 | RF Pred: 0
File: BM (758).jpg | True: 1 | LogReg Pred: 1 | RF Pred: 1
File: BM (759).jpg | True: 0 | LogReg Pred: 1 | RF Pred: 0
File: BM (76).jpg | True: 1 | LogReg Pred: 0 | RF Pred: 0
File: BM (760).jpg | True: 1 | LogReg Pred: 0 | RF Pred: 1
File: BM (761).jpg | True: 0 | LogReg Pred: 1 | RF Pred: 0
File: BM (762).jpg | True: 1 | LogReg Pred: 1 | RF Pred: 1
File: BM (763).jpg | True: 0 | LogReg Pred: 1 | RF Pred: 0
File: BM (764).jpg | True: 1 | LogReg Pred: 1 | RF Pred: 1
File: BM (765).jpg | True: 0 | LogReg Pred: 1 | RF Pred: 0
```

Figure 15: Individual Predictions Comparison: LogReg vs Random Forest

The figure 15, offers a side-by-side view of individual predictions made by both Logistic Regression and Random Forest models for each image, along with their true class labels. This output is particularly valuable for qualitative evaluation, as it highlights specific cases where each model succeeded or failed. A recurring trend in the log is that Logistic Regression often misclassifies images labeled as anomalies (True: 1), frequently predicting them as normal (Pred: 0). In contrast, Random Forest consistently provides the correct predictions for those same images. For example, for the image BM (752).jpg, which is an anomaly (True: 1), Logistic Regression incorrectly predicts it as normal, while Random Forest correctly classifies it. This pattern is observed across many images in the sequence, suggesting that Random Forest is more sensitive and precise in identifying anomalies.

Conversely, some images labeled as normal (e.g., BM (749).jpg or BM (765).jpg) are falsely predicted as anomalies by Logistic Regression, further indicating its lack of reliability. The per-image log also demonstrates that the misclassifications by Logistic Regression are not random but rather systematic, potentially stemming from its inability to properly model the non-linear decision boundaries required by GLCM feature distributions. The value of this comparative list lies in understanding not only the

model's overall performance but also the individual-level impact—critical in real-world applications where misclassifications can have significant costs. It becomes clear that for datasets with textural complexity and subtle class variations, Random Forest outperforms simpler models by a significant margin. Additionally, this output can guide further error analysis and model refinement by identifying specific failure cases, helping practitioners improve preprocessing, feature engineering, or consider hybrid model ensembles.

The comparative evaluation of Logistic Regression and Random Forest models based on GLCM texture features reveals that both models initially struggle to effectively distinguish between normal and anomalous BMI images. The confusion matrices and classification reports indicate low precision and recall, especially for Logistic Regression, which performs close to random guessing. Its AUC hovers around 0.52, while Random Forest initially yields a similar or slightly worse AUC. However, when refined—likely through better feature extraction or hyperparameter tuning—Random Forest demonstrates a significant leap in AUC to 0.92, outperforming Logistic Regression in terms of both accuracy and discriminatory power.

Random Forest's strength lies in capturing non-linear patterns and leveraging ensemble learning to balance prediction between classes. Its confusion matrix displays much lower misclassification compared to Logistic Regression, indicating higher sensitivity and specificity. The feature importance plot further reveals that Random Forest values all four GLCM features fairly equally, with a slight emphasis on correlation and contrast, confirming their collective significance in texture analysis.Additionally, prediction distribution graphs demonstrate Random Forest's superior balance in predicting both classes, unlike Logistic Regression, which slightly favors the normal class. This disparity affects anomaly recall—a critical metric in defect or medical diagnosis tasks. ROC and AUC bar charts reinforce the need to assess models with multiple metrics, highlighting how accuracy alone may be misleading.

Ultimately, this analysis underscores that while Logistic Regression offers simplicity, it falls short in modeling complex feature interactions. Random Forest, although better, still depends heavily on feature quality. Thus, the findings advocate for incorporating richer, possibly deep-learned features or hybrid modeling strategies to significantly improve image-based classification performance, especially in applications demanding high anomaly detection accuracy.

7. CONCLUSION

The research offers a detailed evaluation of classical machine learning models—Random Forest and Logistic Regression—for anomaly detection in biomedical imaging, using texture features extracted through the Gray Level Co-occurrence Matrix (GLCM). The study aimed to classify normal and anomalous images based on subtle textural variations. Initial results showed limited performance for both models, with accuracy around 50% and AUC scores near 0.5, indicating near-random classification. This underperformance is largely due to the limited discriminative power of basic GLCM features. Logistic Regression, being a linear model, was particularly inadequate at capturing non-linear relationships. Random Forest, with its ensemble structure, performed better by modeling more complex feature interactions. After refinement, Random Forest achieved a significant improvement, with an AUC of 0.92, highlighting its potential in medical anomaly detection.

Visualization tools such as confusion matrices, classification heatmaps, and ROC curves were used to interpret and evaluate model performance. Feature importance analysis revealed all four GLCM descriptors—contrast, correlation, energy, and homogeneity—contributed nearly equally, with correlation being the most influential. Random Forest also demonstrated a more balanced class prediction distribution, unlike Logistic Regression which showed bias toward the dominant class.Despite improvements, the study highlights the limitations of texture-based features alone. Real biomedical images exhibit complex, heterogeneous structures that GLCM cannot fully capture. The study recommends integrating deep learning-based representations such as CNN or transformer embeddings, and applying preprocessing techniques like segmentation or data augmentation to

enhance class separability. Hybrid approaches combining classical and deep learning models could offer better accuracy and robustness. Additionally, semi-supervised or self-supervised learning could help exploit unlabeled data, which is common in the medical domain. Overall, while classical ML provides a strong baseline, advanced feature engineering and modeling are crucial for scalable, real-time medical anomaly detection systems.

REFERENCE

[1]. J. Xie, "3D Modeling of Urinary System Based on MRI and CT Data," *2022 International Conference on Artificial Intelligence of Things and Crowdsensing (AIoTCs)*, Nicosia, Cyprus, 2022, pp. 1-5, doi: 10.1109/AIoTCs58181.2022.00007.

[2]. P. Saravanan, S. Saravanakumar and G. Shanmugarathinam, "A Survey on Brain Tumor Prediction with Various Machine Learning Approaches," 2023 International Conference on Computer Science and Emerging Technologies (CSET), Bangalore, India, 2023, pp. 1-4, doi: 10.1109/CSET58993.2023.10346944.

[3]. H. Liu, D. P. Noonan, B. J. Challacombe, P. Dasgupta, L. D. Seneviratne and K. Althoefer, "Rolling Mechanical Imaging for Tissue Abnormality Localization During Minimally Invasive Surgery," in IEEE Transactions on Biomedical Engineering, vol. 57, no. 2, pp. 404-414, Feb. 2010, doi: 10.1109/TBME.2009.2032164.

[4]. Samani, C. Luginbuhl and D. B. Plewes, "Magnetic resonance elastography technique for breast tissue in-vitro elasticity measurement," Proceedings IEEE International Symposium on Biomedical Imaging, Washington, DC, USA, 2002, pp. 931-934, doi: 10.1109/ISBI.2002.1029414.

[5]. W. A. Al-Olofi, M. A. Rushdi, M. A. Islam and A. M. Badawi, "Improved Anomaly Detection in Low-Resolution and Noisy Whole-Slide Images using Transfer Learning," 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt, 2018, pp. 114-117, doi: 10.1109/CIBEC.2018.8641820.

[6]. K. Zhou et al., "Sparse-Gan: Sparsity-Constrained Generative Adversarial Network for Anomaly Detection in Retinal OCT Image," 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 2020, pp. 1227-1231, doi: 10.1109/ISBI45749.2020.9098374.

[7]. Y. Huang, W. Huang, W. Luo and X. Tang, "Lesion2void: Unsupervised Anomaly Detection in Fundus Images," 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), Kolkata, India, 2022, pp. 1-5, doi: 10.1109/ISBI52829.2022.9761593.

[8]. Soner Civilibal, Kerim Kursat Cevik, Ahmet Bozkurt,"A deep learning approach for automatic detection, segmentation and classification of breast lesions from thermal images", Expert Systems with Applications, Volume 212, 2023, 118774, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.118774.

[9]. Mohammad Hassan Daneshvari, Ebrahim Nourmohammadi, Mahmoud Ameri, Barat Mojaradi, "Efficient LBP-GLCM texture analysis for asphalt pavement raveling detection using eXtreme Gradient Boost, Construction and Building Materials", Volume 401, 2023, 132731, ISSN 0950-0618, https://doi.org/10.1016/j.conbuildmat.2023.132731.

[10]. K. Shinde and A. Thakare, "Deep Hybrid Learning Method for Classification of Fetal Brain Abnormalities," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 2021, pp. 1-6, doi: 10.1109/AIMV53313.2021.9670994.

[11]. J. Niu, S. Dong, Q. Yu, K. Dang and X. Ding, "Region and Spatial Aware Anomaly Detection for Fundus Images," 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, 2023, pp. 1-5, doi: 10.1109/ISBI53787.2023.10230619.

[12]. K. Zhou et al., "Proxy-Bridged Image Reconstruction Network for Anomaly Detection in Medical Images," in IEEE Transactions on Medical Imaging, vol. 41, no. 3, pp. 582-594, March 2022, doi: 10.1109/TMI.2021.3118223.

[13]. Y. Huang, G. Liu, Y. Luo and G. Yang, "ADFA: Attention-Augmented Differentiable Top-K Feature Adaptation for Unsupervised Medical Anomaly Detection," 2023 IEEE International Conference on

Image Processing (ICIP), Kuala Lumpur, Malaysia, 2023, pp. 206-210, doi: 10.1109/ICIP49359.2023.10222528.

[14]. Kascenas, R. Young, B. S. Jensen, N. Pugeault and A. Q. O'Neil, "Anomaly Detection via Context and Local Feature Matching," 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), Kolkata, India, 2022, pp. 1-5, doi: 10.1109/ISBI52829.2022.9761524.

[15]. S. N. Marimont and G. Tarroni, "Anomaly Detection Through Latent Space Restoration Using Vector Quantized Variational Autoencoders," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 2021, pp. 1764-1767, doi: 10.1109/ISBI48211.2021.9433778.

[16]. N. Pinon, G. Oudoumanessah, R. Trombetta, M. Dojat, F. Forbes and C. Lartizien, "Brain Subtle Anomaly Detection Based on Auto-Encoders Latent Space Analysis: Application To De Novo Parkinson Patients," 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, 2023, pp. 1-5, doi: 10.1109/ISBI53787.2023.10230351.

[17]. C. Baur, B. Wiestler, S. Albarqouni and N. Navab, "Bayesian Skip-Autoencoders for Unsupervised Hyperintense Anomaly Detection in High Resolution Brain Mri," 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 2020, pp. 1905-1909, doi: 10.1109/ISBI45749.2020.9098686.

[18]. Y. Li, W. Li and L. Li, "Hyperspectral Anomaly Dectection on Multicore DSPs," 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 2018, pp. 1-5, doi: 10.1109/CISP-BMEI.2018.8633118.

[19]. Y. Li, W. Li and L. Li, "Hyperspectral Anomaly Dectection on Multicore DSPs," 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 2018, pp. 1-5, doi: 10.1109/CISP-BMEI.2018.8633118

[20]. M. Pocevičiūtė, G. Eilertsen and C. Lundström, "Unsupervised Anomaly Detection In Digital Pathology Using GANs," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 2021, pp. 1878-1882, doi: 10.1109/ISBI48211.2021.9434141.

[21]. V. Gaike, N. Akhter, K. V. Kale and P. Deshmukh, "Application of higher order GLCM features on mammograms," 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2015, pp. 1-3, doi: 10.1109/ICECCT.2015.7226098.

[22]. A. Spahr, B. Bozorgtabar and J. -P. Thiran, "Self-Taught Semi-Supervised Anomaly Detection On Upper Limb X-Rays," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 2021, pp. 1632-1636, doi: 10.1109/ISBI48211.2021.9433771.

[23]. M. A. Aziz, P. Pujiono, M. A. Soeleman and S. Rosidin, "Scabies Skin Classification Using SVM Method With GLCM Feature Extraction," 2023 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2023, pp. 221-226, doi: 10.1109/iSemantic59612.2023.10295283.

[24]. A. Rizal, I. Wijayanto and I. Istiqomah, "Alcoholism Detection in EEG Signals using GLCM-Based Texture Analysis of Image-Converted Signals," 2023 6th International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2023, pp. 275-279, doi: 10.1109/ICOIACT59844.2023.10455889.