Performance Evaluation of Classification Algorithms For Defect Detection in Patterned Fabric Images

Y.E.MERAL[1] and O. GOKALP[2]

[1] Izmir Katip Celebi University, Izmir/Turkiye, yunusemremeral.cs@gmail.com
[2] Izmir Katip Celebi University, Izmir/Turkiye, osman.gokalp@ikcu.edu.tr

***Abstract* - Defect identification holds substantial relevance across diverse industries, encompassing the textile sector, as well as materials like marble and metal. Automatic defect detection technology, by minimizing human factors influenced by employee distraction and limited concentration time, offers consistent and accurate identification of defects. This study aims to evaluate the performance of various classification algorithms on automatic defect detection in patterned fabric images. The feature extraction stage incorporates Grey-Level Co-occurrence Matrix utilizing seven distinct statistics, as well as Local Binary Pattern techniques. The obtained results on 4 public datasets indicate that Random Forest and XGBoost consistently outperform other classifiers in terms of accuracy, precision, recall, and F1 score across all datasets. This suggests that ensemble methods are well-suited for fabric image classification tasks.**

***Keywords* – automatic defect detection, feature extraction, GLCM, fabric images, classification.**

I. INTRODUCTION

Defect detection is a field with significant applications in various industries including the textile sector, as well as materials such as marble and metal. Its primary objective is to detect and identify defects on material surfaces. This has the potential to improve product quality, enhance manufacturing processes, and reduce costs. Fabric defect identification is a quality control procedure designed to detect and find defects on fabrics [1]. Fabric defects can lead to losses of approximately 45-65% in selling prices. Furthermore, around 85% of rejected products in the ready-to-wear sector are attributed to fabric defects [2].

Manual defect detection can lead to missing or incorrect results due to factors such as employee distraction and limited concentration time. Automatic defect detection technology offers the ability to consistently identify defects by minimizing human factors. Furthermore, artificial intelligence and image processing techniques allow automatic systems to rapidly analyze large datasets, enabling continuous quality control on the production line. The limited concentration time of quality control personnel and the difficulty of recognizing only clearly visible defects demonstrate the limitations of the manual approach. According to research, even under test conditions, approximately 30% of defects may go undetected, and in some well-performing companies considered as good, the repeatability of defect assessments by quality control personnel has not exceeded 50%. Therefore, the automation of the fabric inspection process is a heavily researched topic [2, 3].

Feature extraction is a process utilized to reduce the dimensionality of input data and to transform it into meaningful information that represents its main characteristics. It renders the data more comprehensible and manageable while emphasizing the fundamental attributes of the data to enhance the effectiveness of classification or analysis procedures. In tasks such as fabric defect detection, where the aim is to reduce the dimensionality of input data, the relevant information, referred to as features, is chosen or transformed for accurate classification. This process simplifies the data and enhances the efficiency of data processing algorithms by focusing on the most distinguishing aspects of the data. In the context of fabric defect detection, feature extraction may encompass the identification and quantification of specific visual attributes or patterns within fabric images that signify defects, thereby facilitating a machine learning algorithm's ability to discern differences between normal and defective samples.

GLCM (Grey-Level Co-occurrence Matrix) is a highly effective technique for extracting features from textured images. This matrix systematically evaluates the relationships between pixel values within an image. Specifically, this method holds substantial potential for examining the textural characteristics of images in detail. One of the most noteworthy attributes of GLCM is its position-invariant property, which ensures consistent and reliable results when extracting features from images of an object in different positions.

In this study, feature extraction from images was carried out using GLCM. In particular, the features computed by GLCM were employed for the purpose of analyzing and classifying images to better understand and identify patterns, textures, and objects within them. This demonstrates how GLCM serves as an effective tool in applications such as object recognition and image analysis.

The remainder of the paper is organized as follows. Section 2 gives the literature review and gap. Section 3 explains materials and methods used in this study. Section 4 shares the results obtained after experimental work. Finally, section 5 discusses the findings and concludes the paper.

II. LITERATURE REVIEW

Ngan et al. [4] utilized the direct thresholding approach, based on the wavelet transform of detailed sub-images, as an automated visual inspection technique for discovering defects in patterned fabric. The authors reported a success rate of 96.7% in detecting defects on a common class of patterned jacquard fabric using the wavelet pre-processed golden image subtraction method. The authors used 30 defect-free images and 30 defective patterned images in their study. This method segments out defective regions effectively by employing wavelet-based techniques to extract detailed and approximation sub-images from a histogram equalized defective image. These sub-images were subsequently processed for defect detection.

Kang et al. [5] used a genetic algorithm (GA) to optimize Gabor filters for defect detection in patterned fabrics. Their method, implemented in MATLAB, aimed to reduce noise in printed patterned fabrics and achieve efficient defect detection. The research employed distance matching and the Regular Band to determine fabric units. It enabled fast and effective detection of both random and regular patterned fabrics. The Gabor filter enhanced defect regions by removing background interference, but it faced challenges in cases where defects were not clearly distinguishable from the fabric background. The study also highlighted the importance of color differences between fabric defects and backgrounds for Gabor filtering. Training with non-defective fabrics facilitated real-time defect detection with minimal computational complexity. The proposed methods offered swift and real-time defect detection, especially for regular patterns.

Reddy et al. [6] developed a texture classification system that is robust to changes in illumination. They used a combination of the Gray Level Co-occurrence Matrix (GLCM) and different binary patterns. Their model included adaptive histogram equalization normalization to enhance local contrast. The binary patterns were used to extract texture features, and a rotation-invariant, scale-invariant steerable decomposition filter was applied at different angles. Statistical measures such as mean, standard deviation, contrast, correlation, angular second moment, entropy, and difference entropy were used to capture image content. They employed three classifiers (k-NN, PNN, and SVM) for texture similarity and defect classification. The proposed method, combining binary patterns and GLCM, achieved a 100% defect classification accuracy for both SVM and K-NN classifiers. Additionally, it outperformed individual binary pattern methods in similarity classification, achieving a high rate of 93.75% with SVM and K-NN and 31.25% with PNN.

Despite numerous studies focusing on automatic defect detection in patterned fabric images, there is a lack of comprehensive performance evaluations that simultaneously consider a wide range of classification algorithms. Previous research has not addressed this gap, and there is a need to conduct a comparative evaluation of multiple algorithms to gain insights into their effectiveness. By considering the "no free lunch" theorem in machine learning [7], which suggests the absence of a universally superior algorithm, this study aims to provide a thorough performance evaluation of diverse classification algorithms for defect detection in patterned fabric images. The findings will assist in guiding the selection of appropriate algorithms for real-world applications in the textile industry.

III. Materials and Methods

A. *Feature Extraction for Patterned Images Using GLCM*

In our study, we employed a feature extraction technique for the analysis of patterned images. This technique primarily revolves around the utilization of the Gray-Level Co-occurrence Matrix (GLCM). We calculated GLCM matrices from the input images to capture their inherent textural characteristics. The matrices were further subjected to various statistical calculations to obtain specific textural properties, which include contrast, dissimilarity, homogeneity, energy, correlation, angular second moment, and mean. These properties collectively constitute an essential feature set that enables a comprehensive characterization of the image textures.

B. *Classification Algorithms*

To classify the image data and evaluate its patterns, we employed a range of distinct classification algorithms. These algorithms were specifically chosen to cater to the complexities of the problem at hand. The set of algorithms used in our analysis comprises the Random Forest, Bagging, Logistic Regression, Support Vector Classifier, K-Nearest Neighbors, Naive Bayes, Decision Tree, Gradient Boosting, and XGBoost. In our study, each of these algorithms underwent a training process on the feature vectors derived from our training dataset. Subsequently, they were assessed and validated using a separate test dataset.

C. *Data Acquisition and Processing*

The data utilized in this study was obtained from the TILDA dataset, specifically focusing on the c3-r1, c3-r3, c4-r1, and c4-r3 datasets, which contain patterned image samples. Each dataset classifies images into 7 classes; 6 different kinds of defects labels and 1 defect-free label. To prepare the data, the Grey-Level Co-occurrence Matrix was computed for each image, capturing the relationships among pixel values in various orientations and distances. The GLCM matrices were normalized and contained 256 levels to ensure consistency in feature extraction across all images.

In the context of data preparation and preprocessing, these GLCM matrices served as a fundamental step. They were employed for subsequent feature extraction, specifically to calculate contrast, dissimilarity, homogeneity, energy, correlation, angular second moment, and mean values. Additionally, local binary pattern features were computed from the images, further enriching the feature set. The combination of these features laid the foundation for the classification tasks.

D. *Evaluation of Classification Results*

Our study meticulously evaluated the classification results using various well-defined metrics, including accuracy, precision, recall, and F1 score. These metrics allowed us to thoroughly assess the performance of the classification algorithms in classifying patterned images.

Accuracy scores were computed for each classification algorithm, providing a quantitative measure of how effectively the algorithms classified the images. The accuracy was calculated using the formula (1):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

However, accuracy alone does not provide a complete picture of performance.

In addition to accuracy metric, precision, recall, and F1 score metrics were also analyzed.

Precision measures the proportion of true positive predictions out of all positive predictions, indicating the algorithm's ability to avoid false positives. It is calculated using the formula (2):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall, on the other hand, assesses the proportion of true positive predictions out of all actual positive instances, demonstrating the algorithm's capability to detect all relevant patterns. The formula for recall is (3):

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

Furthermore, we calculated the F1 score, which is a harmonic mean of precision and recall. It offers a balanced evaluation of a classification algorithm's performance by considering both false positives and false negatives. The F1 score is computed as in (4):

$$\text{F1Score} = \frac{2(precision-recall)}{precision+recall} \quad (4)$$

IV. Experimental Work

In this section, we detail the experimental setup used for our analysis. The experiments were conducted using Python version 3.10.12 and the scikit-learn library version 1.2.2. The analysis aims to evaluate the performance of classification algorithms in the context of fabric defect detection. Default hyper-parameters were used for the classification algorithms except the following settings:

- Random Forest: A Random Forest classifier with 100 estimators was used, and the random state was set to 42 for reproducibility.
- Bagging: A Bagging classifier, primarily utilizing Decision Tree classifiers, was employed with 100 estimators. The random state was set to 42.
- Logistic Regression: Logistic Regression was performed with a maximum of 1000 iterations.
- Gradient Boosting: A Gradient Boosting classifier with 100 estimators was employed, and the random state was set to 42 for reproducibility.

The dataset used in the experiments contains patterned images obtained from the TILDA database, specifically including four different datasets: c3-r1, c3-r3, c4-r1, and c4-r3. When splitting the dataset into training and test data, a test data split ratio of 20% was applied. The results of the experiments are presented in the next section.

The performance metrics used in the paper include Accuracy, precision, recall and F1 Score evaluations.

The results for four different data sets are shown in the tables below (Tables 1-4).

Table 1: The Results obtained with the C3-R1 dataset.

Classifier	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Random Forest	0.864	0.878	0.864	0.859
Bagging	0.778	0.804	0.778	0.782
Logistic Regression	0.457	0.454	0.457	0.416
Support Vector Classifier	0.346	0.304	0.346	0.248
K-Nearest Neighbors	0.531	0.557	0.531	0.531
Naïve Bayes	0.519	0.611	0.519	0.505
Decision Tree	0.630	0.688	0.630	0.645
Gradient Boosting	0.790	0.808	0.790	0.789
XGBoost	0.864	0.878	0.864	0.859

Table 2: The Results obtained with the C3-R3 dataset.

Classifier	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Random Forest	0.667	0.731	0.667	0.663
Bagging	0.691	0.736	0.691	0.694
Logistic Regression	0.370	0.393	0.370	0.374
Support Vector Classifier	0.333	0.350	0.333	0.328
K-Nearest Neighbors	0.420	0.460	0.420	0.426
Naïve Bayes	0.457	0.457	0.457	0.449
Decision Tree	0.580	0.587	0.580	0.572
Gradient Boosting	0.667	0.703	0.667	0.670
XGBoost	0.704	0.722	0.704	0.699

Table 3: The results obtained with the C4-R1 dataset.

Classifier	Accuracy	Precision	Recall	F1 Score
Random Forest	0.642	0.691	0.642	0.647
Bagging	0.704	0.716	0.704	0.706
Logistic Regression	0.407	0.497	0.407	0.411
Support Vector Classifier	0.173	0.174	0.173	0.142
K-Nearest Neighbors	0.345	0.392	0.345	0.354
Naïve Bayes	0.481	0.543	0.481	0.470
Decision Tree	0.556	0.589	0.556	0.556
Gradient Boosting	0.654	0.684	0.654	0.656
XGBoost	0.704	0.725	0.704	0.706

Table 4: The results obtained with the C4-R3 dataset.

Classifier	Accuracy	Precision	Recall	F1 Score
Random Forest	0.675	0.699	0.675	0.647
Bagging	0.700	0.720	0.700	0.689
Logistic Regression	0.375	0.418	0.375	0.373
Support Vector Classifier	0.312	0.323	0.312	0.274
K-Nearest Neighbors	0.375	0.455	0.375	0.397
Naïve Bayes	0.450	0.417	0.450	0.412
Decision Tree	0.475	0.525	0.475	0.489
Gradient Boosting	0.688	0.731	0.688	0.699
XGBoost	0.675	0.688	0.675	0.675

V. Discussion and Conclusion

A. Performance Comparison Across Datasets

Experiments on various datasets consistently demonstrate the superior performance of Random Forest and XGBoost. Particularly, in the C3-R1 and C3-R3 datasets, these two algorithms exhibit a significant advantage over other classifiers (See Table 1 and Table 2). In the case of the C4-R1 and C4-R3 datasets, Random Forest and XGBoost also stand out with high accuracy, precision, recall, and F1 scores (See Table 3 and Table 4).

B. Ensemble Methods' Superiority

The obtained results indicate that Random Forest and XGBoost consistently outperform other classifiers in terms of accuracy, precision, recall, and F1 score across all datasets. This suggests that ensemble methods are well-suited for fabric image classification tasks. Conversely, Logistic Regression, Support Vector Classifier, and Naive Bayes yield lower performance scores, underscoring their limitations in handling complex fabric patterns.

C. Impact of Dataset Selection

The results underscore the substantial influence of dataset selection on classifier performance. Depending on the diversity of fabric textures, classification methods yield better or worse predictions. This highlights the critical importance of selecting the most appropriate dataset for specific applications.

D. Future Directions and Significance

In conclusion, our study offers valuable insights into fabric image classification. We emphasize the pivotal role of classifier selection in achieving accurate results. Random Forest and XGBoost have proven to be the top choices for this task, emphasizing the importance of considering dataset characteristics when choosing the most suitable model. Our research contributes to the fields of computer vision and textile industry applications. These findings underscore the potential of machine learning in enhancing automated quality control and product identification processes in textile manufacturing. Future studies may encompass improving the dataset collection process and exploring advanced deep learning models for enhanced accuracy.

References

[1] M. S. Priyanka, P. D. Manish, & R. S. Shekhar, (2012),"Overview: Methods of Automatic Fabric Defect Detection", Global Journal of Engineering, Design and Engineering., Education, published by Global Institute for Research and Vol. 1(2), pp 42-46, ISSN 2319 – 7293, 2019.

[2] Arıkan C.O., "Developing an Intelligent Automation and Reporting System for Fabric Inspection Machines" Tekstil ve Konfeksiyon, 29 (1), pp:86-93.

[3] Büyükkabasakal, K, "Kumaş dokuma hatalarının tespiti ve sınıflandırılması", İzmir, Ege Üniversitesi, 2010.

[4] H. Y. T. Ngan, G. K. H. Pang, S. P Yung, and M. K Ng,. "Wavelet-based Methods on Patterned Fabric Defect Detection", Pattern Recognition, Vol. 38, No.4, pp. 559–576, 2005.

[5] X. Kang, P. Yang, and J. Jing , "Defect Detection on Printed Fabrics Via Gabor Filter and Regular Band", Journal of Fiber Bioengineering and Informatics, Volume 8, Issue 1 pp 195–206, 2015.

[6] R. O. K Reddy, B. E. Reddy, and E. K Reddy, "Classifying Similarity and Defect Fabric Textures based on GLCM and Binary Pattern Schemes", International Journal of Information Engineering and Electronic Business, 5, pp 25-33, 2013.

[7] Adam, S. P., Alexandropoulos, S. A. N., Pardalos, P. M., & Vrahatis, M. N. No free lunch theorem: A review. Approximation and optimization: Algorithms, complexity and applications, 57-82, 2019.