Cervical Cancer Detection using Traditional Computer Vision

Sriram Sunderrajan
CSE (AI&ML)
MIT Manipal
Roll Number: 77
Reg No: 220962444

Medha Chawla
CSE (AI&ML)
MIT Manipal
Roll Number: 55
Reg No: 220962340

Pratik Chakraborty
CSE (AI&ML)
MIT Manipal
Roll Number: 57
Reg No: 220962350

***Abstract*—This research explores the automated detection of leukemia cells in Pap smear images through advanced computer vision techniques. Using the Sipakmed dataset, we investigated the effects of preprocessing strategies—specifically Otsu's thresholding and watershed segmentation—on classification accuracy. Feature extraction was performed using five methods: Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Local Binary Patterns (LBP), Oriented FAST and Rotated BRIEF (ORB), and Gabor filters. These extracted features were classified using four machine learning models: Logistic Regression, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Additionally, we applied three deep learning architectures—AlexNet, VGG19, and ResNet—to the original images to further investigate classification performance. Results were evaluated across multiple metrics, with confusion matrices highlighting the performance of the deep learning models. The comparative analysis provides insight into the impact of preprocessing and feature extraction methods on the accurate detection of leukemia cells in cervical images. It also shows the utility and benefits of traditional Computer Vision techniques in today's Deep Learning landscape.**

I. Introduction

Leukemia, a type of blood cancer that primarily affects white blood cells, poses a significant global health burden. Early and accurate diagnosis is essential for effective treatment and improved patient outcomes. Traditionally, leukemia detection relies on the manual examination of blood smear or pap smear images by trained pathologists, a process that is not only time-consuming but also prone to human error. With recent advancements in computer vision and artificial intelligence, automated detection methods have gained traction as a promising solution to streamline leukemia diagnostics [1], [2].

Computer vision approaches for leukemia cell detection involve preprocessing, feature extraction, and classification steps. Preprocessing methods such as Otsu thresholding [4] and watershed segmentation are commonly used to enhance image quality, enabling better delineation of cell boundaries. Feature extraction methods, including Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Local Binary Patterns (LBP), and Gabor filters, help capture essential morphological and textural characteristics of leukemia cells [2], [3]. Machine learning (ML) classifiers, such as support vector machines (SVM), k-nearest neighbors (KNN), and random forests, further enhance classification accuracy by learning these extracted features [1].

The rise of deep learning (DL) has introduced more sophisticated models, such as convolutional neural networks (CNNs), capable of learning complex, hierarchical features directly from images. Prominent CNN architectures like AlexNet, VGG19, and ResNet have demonstrated state-of-the-art accuracy in cell classification tasks, minimizing the need for manual feature engineering [6], [7]. While these models offer high performance, their computational requirements often limit their application in resource-constrained settings, motivating further research to optimize these methods.

In this study, we utilize the Sipakmed dataset [7], a well-annotated collection of pap smear images, to explore leukemia cell classification. We investigate the impact of preprocessing and feature extraction methods on classification accuracy by using both ML and DL models. Specifically, we examine the performance of five feature extraction techniques—HOG, SIFT, LBP, ORB, and Gabor filters—paired with four ML classifiers, including logistic regression, SVM, random forest, and KNN. Additionally, we benchmark the performance of three CNN architectures: AlexNet, VGG19, and ResNet.

II. Literature Survey

Automated detection of leukemia cells in pap smear images has received significant attention in recent years due to its potential to improve diagnostic accuracy and efficiency. The Sipakmed dataset, extensively used in research, provides a structured and labeled collection of pap smear images, aiding in the training and evaluation of machine learning and deep learning models for cell classification tasks [7]. The following survey highlights recent advancements in this domain, focusing on preprocessing, feature extraction, and classifier techniques.

A. Leukemia Cell Detection in Microscopic Images

Automating the detection of leukemia cells has the potential to streamline diagnostics by reducing reliance on manual observation. Ghaderzadeh et al. [1] provide a systematic review on the use of machine learning in the detection and classification of leukemia cells, illustrating the effectiveness of ML algorithms in distinguishing between healthy and malignant

cells in blood smear images. Baig et al. [2] demonstrated that deep learning approaches, such as convolutional neural networks, can achieve high accuracy in identifying malignant leukemia cells from microscopic images, thus reducing diagnostic variability.

B. Preprocessing Techniques for Enhanced Detection

Preprocessing plays a critical role in preparing images for feature extraction and classification, as it enhances cell boundaries and reduces noise. The Otsu thresholding technique, introduced by Otsu [4], has been widely adopted for binarizing images by determining optimal threshold values based on image histograms. Additionally, watershed segmentation is frequently employed alongside thresholding to improve the delineation of cell boundaries, which is essential for accurate feature extraction in medical images. Edge detection methods, such as the Canny edge detector [3], are also commonly used to highlight structural features in cell images. These preprocessing steps are foundational in achieving reliable feature extraction results, particularly when dealing with microscopic images where cell morphology varies.

C. Feature Extraction Techniques

Feature extraction methods are employed to capture morphological and textural information necessary for cell classification:

- Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT) are popular techniques for capturing shape and gradient-based features, essential in distinguishing leukemia cells in complex backgrounds [1].
- Local Binary Patterns (LBP) effectively capture texture information, which is valuable for differentiating between normal and malignant cells based on subtle textural changes [2].
- Oriented FAST and Rotated BRIEF (ORB), a computationally efficient alternative to SIFT, is often applied in real-time applications for capturing key points [7].
- Gabor Filters are known for their effectiveness in texture analysis, crucial in capturing multi-scale and multi-orientation features for medical imaging applications [1].

D. Machine Learning Classifiers for Cell Classification

Machine learning classifiers, such as k-nearest neighbors (KNN) [5], logistic regression, random forests, and support vector machines (SVM), have been extensively employed for cell classification tasks. These classifiers, when combined with robust feature extraction techniques, have shown competitive accuracy in leukemia cell detection. For instance, SVM is widely used for its effectiveness in handling high-dimensional data, making it well-suited for image classification tasks [1]. Random forests are effective in dealing with imbalanced datasets, which is a common challenge in medical imaging. KNN, introduced by Cover and Hart [5], is noted for its simplicity and effectiveness, especially when paired with texture-based features such as LBP.

E. Deep Learning Models in Cell Classification

Deep learning models, particularly convolutional neural networks (CNNs), have emerged as powerful tools for automated cell classification due to their ability to learn hierarchical features directly from raw images. AlexNet, VGG19, and ResNet, widely recognized CNN architectures, have been applied in medical imaging for leukemia detection:

- VGG19, proposed by Simonyan and Zisserman [6], utilizes a deep architecture to capture fine-grained structural details, making it highly effective for image classification tasks, including cell detection.
- ResNet, known for its residual learning framework, addresses the vanishing gradient problem, allowing for deeper architectures that are advantageous in medical image analysis [2].

These deep learning approaches often outperform traditional ML methods in terms of accuracy but come with increased computational requirements, which may be challenging in resource-limited settings [2].

III. Experiment Setup

This study utilizes the Sipakmed dataset for cervical cell classification across five categories, sourced from Kaggle. All experiments were conducted in Google Colab with GPU support to handle computational tasks such as feature extraction, segmentation, and model training [7].

A. Dataset Preparation

The dataset includes images categorized into five cell types: **im_Dyskeratotic**, **im_Koilocytotic**, **im_Metaplastic**, **im_Parabasal**, and **im_Superficial-Intermediate**. Images were resized to 128×128 pixels for feature extraction and to the required input sizes for specific deep learning models.

B. Preprocessing and Segmentation

To assess the impact of preprocessing on classification performance, two sets of experiments were conducted: one using the original images and another using preprocessed images.

The preprocessing steps were as follows:

- **Otsu Thresholding**: Applied to grayscale images to create binary masks, highlighting cell structures against the background [4].
- **Watershed Segmentation**: Following Otsu thresholding, watershed segmentation was applied to define cell boundaries and separate cells from the background, creating segmented images.

Only the preprocessed (segmented) images were used for ORB feature extraction, whereas the original (non-preprocessed) images were used for deep learning models.

C. Feature Extraction

Feature extraction methods were applied to both the original and preprocessed images, except for ORB, which was used only on the preprocessed images. Five feature extraction

methods were employed to capture important morphological and textural characteristics:

- **Histogram of Oriented Gradients (HOG)**: Extracted gradient orientations with pixels per cell = 8 × 8 and cells per block = 2 × 2, used for both original and preprocessed images.
- **Local Binary Patterns (LBP)**: Configured with a radius of 1 and 8 sample points to capture texture information, used for both original and preprocessed images.
- **Scale-Invariant Feature Transform (SIFT)**: Extracted keypoints and descriptors, which were flattened into fixed-length feature vectors, used for both original and preprocessed images.
- **Oriented FAST and Rotated BRIEF (ORB)**: Applied exclusively to the preprocessed (segmented) images, ORB captured texture-based keypoints, which were then flattened to create standardized feature vectors.
- **Gabor Filters**: Applied to both original and preprocessed images, using a kernel size of 21, sigma of 5.0, orientation of pi/4, wavelength of 10.0, and aspect ratio of 0.5 to capture texture information across multiple orientations.

Feature vectors from each extraction method were stored separately for original and preprocessed images, depending on applicability, and categorized by cell type.

D. Machine Learning Classifiers

Feature sets from both original and preprocessed images (with ORB applied only to preprocessed images) were used to train four machine learning classifiers, allowing for a comparative analysis:

- **Logistic Regression**: Configured with a maximum iteration of 1000.
- **Random Forest**: Selected for its robustness with imbalanced datasets.
- **Support Vector Machine (SVM)**: Known for effective performance with high-dimensional features.
- **k-Nearest Neighbors (KNN)**: Effective for texture-based classification tasks, due to its simplicity and performance.

Each classifier was trained independently on each feature set (HOG, LBP, SIFT, ORB for preprocessed images only, and Gabor), and results were recorded for comparison between original and preprocessed datasets.

E. Deep Learning Models

Deep learning models were applied exclusively to the original (non-preprocessed) images to capture high-level patterns directly from raw data. Three architectures were utilized:

- **AlexNet**: Images were resized to 227×227 and normalized before being input into the network, with the output layer modified for five classes.
- **VGG19**: Input images were resized to 224×224, and the output layer was adjusted for multi-class classification.
- **ResNet50**: Images were resized to 256×256, and the final classification layer was modified for five-class output.

Category	Number of Images	Number of Cells
Superficial/Intermediate	126	813
Parabasal	108	787
Koilocytotic	238	825
Metaplastic	271	793
Dyskeratotic	223	813
Total	966	4049

Table I: Distribution of Cell Categories and Counts

Each model was evaluated on the original dataset, and metrics such as accuracy, precision, recall, F1 score, and confusion matrices were computed.

IV. Dataset

The Sipakmed dataset used in this study consists of **4,049 manually annotated images of isolated cells** cropped from **966 cell cluster images** obtained from Pap smear slides. Cells are categorized into five classes, classified based on annotations by expert cytopathologists who defined the cytoplasm and nuclear areas. This dataset includes both normal and abnormal cell types, enabling comprehensive model training for cell classification.

A. Cell Categories and Counts

The distribution of cell categories and their respective counts in terms of images and individual cells is provided in **Table I**.

B. Cell Types

1) **Normal Cells**:
   - **Superficial-Intermediate Cells**: Typically flat, round, oval, or polygonal with large, well-defined cytoplasm and a central pycnotic nucleus. They may show nuclear abnormalities in severe lesions.
   - **Parabasal Cells**: Small epithelial cells with a large vesicular nucleus and cyanophilic cytoplasm, similar in appearance to metaplastic cells, making differentiation challenging.
2) **Abnormal Cells**:
   - **Koilocytotic Cells**: Found in mature squamous cells, these cells are characterized by perinuclear cavities, enlarged hyperchromatic nuclei, and irregular nuclear membranes, often associated with HPV infections.
   - **Dyskeratotic Cells**: Display premature keratinization, orange-stained cytoplasm, and vesicular nuclei, indicative of HPV infection.
3) **Benign Cells / Metaplastic Cells**: Characterized by round cytoplasm with well-defined cellular borders and occasionally large vacuoles, associated with increased detection rates of pre-cancerous lesions.

The dataset's diversity in cell types provides a robust foundation for training and evaluating models in cervical cell classification tasks, encompassing both benign and abnormal cellular presentations.

V. Methodology

This study employs a structured pipeline to classify cervical cells from the Sipakmed dataset using both machine learning and deep learning models, comparing the effects of preprocessing on classification accuracy. The methodology is organized into data preprocessing, feature extraction, model training, and evaluation.

a) Data Preprocessing: Two parallel processing paths were established: one using original images and another using preprocessed images to examine the impact of segmentation techniques on classification performance.

- **Otsu Thresholding**: Each image was first converted to grayscale and resized to 128×128. Otsu's thresholding was applied using `cv2.threshold` to automatically determine the optimal threshold for binary segmentation, enhancing the contrast between cells and the background.
- **Watershed Segmentation**: The binary masks obtained from Otsu's thresholding were further processed using watershed segmentation. Distance transform was applied with `cv2.distanceTransform` to identify distinct cell regions, and markers were created by thresholding the transformed image. These markers were fed to the `cv2.watershed` function to finalize the segmentation boundaries, producing labeled images with isolated cell structures.

b) Feature Extraction: Feature extraction methods were applied independently to both the original and preprocessed images (except ORB, which was used only on preprocessed images) to capture diverse characteristics from each dataset:

1) **Histogram of Oriented Gradients (HOG)**: HOG features were extracted using the `hog` function from `skimage.feature`, with `pixels_per_cell` set to 8×8 and `cells_per_block` set to 2×2. This approach captured gradient orientation distributions, producing feature vectors that represent cell shapes and boundary orientations.
2) **Local Binary Patterns (LBP)**: Using `local_binary_pattern` from `skimage.feature`, LBP was configured with 8 points in a radius of 1. Histograms of uniform patterns were created by binning the LBP values, then normalized to form feature vectors representing the cell's texture.
3) **Scale-Invariant Feature Transform (SIFT)**: Keypoints and descriptors were extracted with OpenCV's `SIFT_create()` function. For dimensionality reduction, Principal Component Analysis (PCA) was applied to reduce descriptor dimensionality to 100 components, using `PCA(n_components=100)`. Descriptors were averaged across keypoints to create a consistent feature vector for each image.
4) **Oriented FAST and Rotated BRIEF (ORB)**: ORB keypoints and descriptors, extracted exclusively from the preprocessed images using `cv2.ORB_create()`, provided a computationally efficient alternative for texture-based keypoints. Descriptors were flattened and standardized to a fixed length (500 elements) using padding, ensuring compatibility across samples.
5) **Gabor Filters**: For texture analysis, Gabor filters were applied with `cv2.getGaborKernel`, using parameters (ksize=21, sigma=5.0, theta= pi/4, lambd=10.0, gamma=0.5). Each filtered response was flattened to form feature vectors, capturing multi-scale and multi-orientation textural details.

The resulting feature vectors for each method were organized separately for both the original and preprocessed images, providing two distinct feature datasets.

c) Machine Learning Classification: Four machine learning classifiers were trained on the extracted features from both original and preprocessed images. The classifiers were initialized and configured as follows:

- **Logistic Regression**: Configured with a maximum iteration parameter of 1000 using `LogisticRegression(max_iter=1000)` to ensure convergence across larger feature sets.
- **Random Forest**: Implemented using `RandomForestClassifier()`, this ensemble method was particularly useful for handling feature variability across the original and preprocessed datasets.
- **Support Vector Machine (SVM)**: A linear SVM, initialized with `SVC()`, was used for its high-dimensional handling capabilities. The SVM was trained on each feature set, helping to evaluate the separability of cell categories.
- **k-Nearest Neighbors (KNN)**: Configured with `KNeighborsClassifier()` to determine class labels based on the closest neighboring points. This approach proved effective for texture-rich descriptors like LBP and ORB.

Each classifier was trained on each feature set independently (HOG, LBP, SIFT, ORB for preprocessed only, and Gabor), and accuracy results were stored for comparison across models and feature sets.

d) Deep Learning Classification: Deep learning models were applied exclusively to the original images. Three pre-trained architectures were fine-tuned using PyTorch to handle the five-class cervical cell classification task:

- **AlexNet**: Initialized with `models.alexnet(pretrained=True)`, AlexNet was modified to match the five classes by replacing the output layer with a fully connected layer using `nn.Linear(4096, 5)`. The model was trained with a batch size of 32 and learning rate of 0.001, optimized using Adam (`optim.Adam`).
- **VGG19**: Loaded with `models.vgg19(pretrained=True)`, the VGG19 model was adapted similarly by replacing the output layer for multi-class classification. Input images were resized to 224×224 to match VGG's requirements, and training parameters mirrored those of AlexNet.

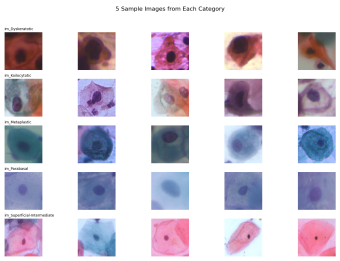

Figure 1: 5 Sample Images from Each Category

- **ResNet50**: Initialized with `models.resnet50(pretrained=True)`, ResNet50 was modified with a fully connected layer for five-class output. Images were resized to 256×256 and the network was optimized with Adam, tracking the weighted average of metrics over epochs.

Each model was evaluated on accuracy, precision, recall, and F1 score. Confusion matrices were generated to visualize class prediction performance.

e) Evaluation Metrics: To comprehensively compare the models, performance metrics were calculated for each classifier and deep learning model:

- **Accuracy**: Calculated as the ratio of correctly predicted instances to total instances.
- **Precision, Recall, and F1 Score**: For each model, precision, recall, and F1 scores were computed using `precision_score`, `recall_score`, and `f1_score` with `average='weighted'` to account for class imbalances.
- **Confusion Matrix**: The confusion matrix, generated with `confusion_matrix`, provided a detailed view of correct and incorrect predictions across cell types. Heatmaps of these matrices were plotted using Seaborn (`sns.heatmap`) to visualize classification performance per class.

This methodology facilitated a structured comparison between traditional machine learning models and deep learning architectures, highlighting the impact of preprocessing on model performance for cervical cell classification.

VI. Results

Figure 1 shows five sample images from each of the five cell categories in the dataset: **Dyskeratotic**, **Koilocytotic**, **Metaplastic**, **Parabasal**, and **Superficial-Intermediate**. These examples illustrate the morphological diversity within the dataset, providing a foundational view of the data used for classification.

a) Feature Extraction Results:

1) **HOG Features**: In *Figure 2*, the original grayscale image of a dyskeratotic cell is shown alongside its HOG (Histogram of Oriented Gradients) features. The HOG representation captures gradient orientations, effectively highlighting cell shape and boundary structure.

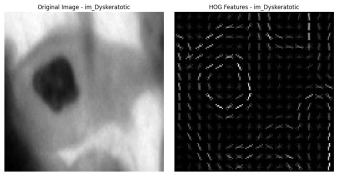

Figure 2: Image with HOG features

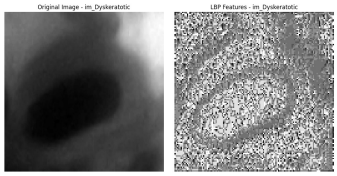

Figure 3: Image with LBP features

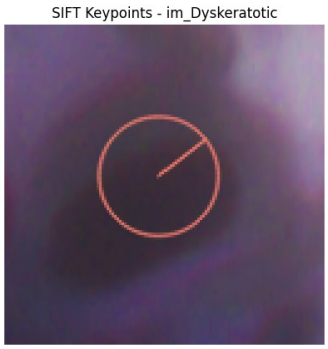

Figure 4: Image with SIFT features

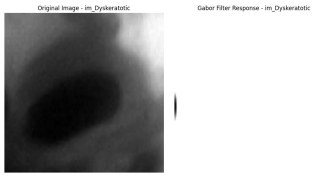

Figure 5: Image with Gabor filters

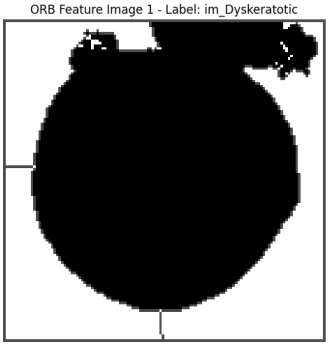

Figure 6: Image with ORB features

Feature Detector	Logistic Regression	Random Forest	SVM	k-NN
HOG	63.21	62.72	69.26	28.02
LBP	54.69	66.42	62.96	63.58
SIFT	47.25	53.91	53.39	46.74
Gabor Filters	37.04	40.49	34.32	32.47

Table II: Comparison of accuracy percentages with different ML Models and Feature Extraction Methods without preprocessing

Feature Detector	Logistic Regression	Random Forest	SVM	k-NN
HOG	51.73	54.81	64.57	34.20
LBP	44.69	43.95	45.93	40.12
SIFT	28.64	35.43	27.41	25.68
ORB	34.73	43.26	34.42	36.12
Gabor Filters	31.36	39.63	29.75	34.69

Table III: Comparison of accuracy percentages with different ML Models and Feature Extraction Methods with preprocessing

2) **LBP Features**: *Figure 3* presents the original grayscale image of a dyskeratotic cell and its corresponding Local Binary Patterns (LBP) feature map. LBP captures texture details by encoding local intensity differences, making it useful for distinguishing texture-rich areas within the cell.
3) **SIFT Features**: *Figure 4* displays the original image of a dyskeratotic cell with detected keypoints using Scale-Invariant Feature Transform (SIFT). SIFT identifies distinct keypoints, which are rotation- and scale-invariant, allowing for robust matching based on the cell's structure.
4) **Gabor Filter Features**: *Figure 5* shows the application of Gabor filters to the original dyskeratotic cell image. The Gabor filter response highlights frequency and orientation information, which can capture detailed texture patterns within the cell structure.
5) **ORB Features**: In *Figure 6*, the ORB (Oriented FAST and Rotated BRIEF) features extracted from a dyskeratotic cell in a segmented, preprocessed image are displayed. ORB efficiently captures keypoints and descriptors, which are particularly useful for texture and pattern recognition in preprocessed images.

A. ML Models without Preprocessing Results

Table II shows the performance metrics (accuracy scores) of four classifiers—Logistic Regression, Random Forest, SVM, and k-NN—applied to feature vectors generated from the original images without any preprocessing. The HOG feature extraction method, when combined with the SVM classifier, achieved the highest accuracy at **69.26%**, indicating that HOG features were particularly effective at capturing distinguishing patterns in the unprocessed images. The LBP method also performed reasonably well, with SVM achieving an accuracy of **62.96%**. In general, SVM consistently outperformed other classifiers across all feature extraction methods, while k-NN had the lowest performance, particularly on HOG and SIFT features.

B. ML Models with Preprocessing Results

Table III presents the accuracy percentages achieved by the same classifiers when applied to feature vectors generated from preprocessed images (using Otsu's thresholding and watershed segmentation). The preprocessing steps appear to enhance performance across multiple feature extraction methods, particularly for the LBP and ORB features. In this setup, the HOG and ORB methods with SVM yielded some of the highest accuracies, with HOG and SVM reaching **64.57%** accuracy, and ORB achieving an accuracy of **43.26%** with Random Forest. Preprocessing allowed for improved differentiation between classes, especially for the LBP and ORB methods, where texture-based information became more prominent and distinguishable.

C. Comparison of Results with and without preprocessing

The comparison between *Table II* and *Table III* reveals the impact of preprocessing on classification performance across feature extraction methods:

- **HOG Features**: HOG with SVM achieved a slightly higher accuracy without preprocessing (69.26%) compared to with preprocessing (64.57%). This suggests that HOG features effectively capture cell shape information from original images and may not require additional preprocessing for optimal performance.
- **LBP Features**: LBP showed stable performance across both setups, with minor differences across classifiers. While preprocessing did not drastically improve accuracy, it led to marginal gains for some classifiers like k-NN, highlighting how texture-based features may respond variably to segmentation techniques.
- **SIFT and ORB Features**: Preprocessing had a notable positive impact on ORB, which was only applied to preprocessed images, with Random Forest achieving 43.26% accuracy. SIFT also saw some improvement with Random Forest after preprocessing, indicating that segmentation and contrast enhancement can help highlight keypoints and descriptors, especially for models sensitive to structure.
- **Gabor Features**: Gabor filters showed limited improvement with preprocessing, with a slight increase in accuracy for Random Forest (39.63%) and k-NN (34.69%) but overall remained less effective than other feature types. This suggests that while Gabor filters capture multi-scale and orientation-based textures, the preprocessing steps did not significantly enhance its effectiveness for this dataset.

In summary, the results indicate that preprocessing with Otsu's thresholding and watershed segmentation can enhance performance for certain feature extraction methods, particularly ORB and, to a lesser extent, SIFT. However, for HOG, which performs well on original images, preprocessing does not necessarily yield improved performance, suggesting that HOG features are inherently robust to variations in preprocessing.

D. DL Models Results

This section presents the performance of three deep learning models—AlexNet, VGG19, and ResNet—applied directly to the original images. Each model's performance was evaluated using accuracy, precision, recall, and F1 score. The confusion matrices for each model provide additional insight into the classification performance across the five cell categories.

1) AlexNet:

- **Accuracy**: 80.99%
- **Precision**: 83.03%
- **Recall**: 80.99%
- **F1 Score**: 80.33%

Figure 7 shows the confusion matrix for AlexNet. The model demonstrates strong classification capability for most categories, with the highest misclassifications occurring between classes with similar morphological features, such as **Koilocytotic** and **Metaplastic** cells. AlexNet achieves a balanced performance, with high precision indicating its effectiveness in reducing false positives.

2) VGG19:

- **Accuracy**: 72.84%
- **Precision**: 75.41%
- **Recall**: 72.84%
- **F1 Score**: 71.50%

Figure 8 displays the confusion matrix for VGG19, which shows a lower overall accuracy compared to AlexNet. VGG19 tends to have more misclassifications, particularly in distinguishing **Koilocytotic** and **Superficial-Intermediate** cells, which likely share textural similarities. The model's moderate precision and F1 score reflect its challenges in differentiating closely related categories, highlighting that VGG19 may require further tuning or pretraining on similar histological data for improved performance.

3) ResNet:

- **Accuracy**: 93.08%
- **Precision**: 93.08%
- **Recall**: 93.08%
- **F1 Score**: 93.08%

As shown in *Figure 9*, ResNet significantly outperforms both AlexNet and VGG19, achieving an impressive accuracy of 93.08%. The confusion matrix reveals minimal misclassification, with high accuracy across all categories. ResNet's high F1 score and precision indicate a balanced capability to handle both false positives and false negatives effectively. This superior performance can be attributed to ResNet's deeper architecture, which allows it to capture intricate details and hierarchical features within cell images, making it highly suitable for this classification task.

4) Comparative Analysis of Deep Learning Models: In summary, ResNet exhibits the best performance across all metrics, followed by AlexNet, with VGG19 showing the lowest performance among the three. The deeper architecture of ResNet allows it to capture complex patterns within the images, contributing to its higher accuracy. AlexNet, while simpler, also shows reliable performance, making it a viable option where computational resources are limited. VGG19, while effective in general image classification tasks, may require further optimization for medical images to reach the same level of performance as ResNet.

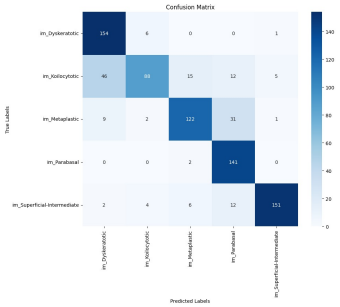

Figure 7: AlexNet Confusion Matrix

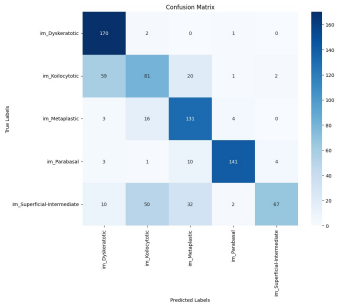

Figure 8: VGG19 Confusion Matrix

The high performance of ResNet underscores the potential of deep learning, especially deep architectures, in medical image analysis for cell classification. Further tuning and domain-specific pretraining could further enhance the effectiveness of these models for clinical applications in identifying cellular abnormalities.

5) Comparasion between best Computer Vision Method and Deep Learning Benchmark: The best Computer Vision technique was determined to be the HOG feature extractor, with the Support Vector Machine Classifier. *Table IV* shows its comparison with the Deep Learning Benchmark (ResNet50), in terms of accuracy and floating-point operations(FLOP's).

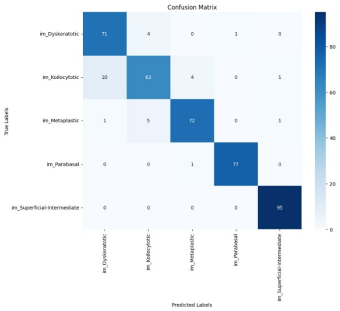

Figure 9: ResNet Confusion Matrix

FLOP's and Accuracy Analysis	HOG extractor + SVM classifier	ResNet50 Benchmark	Fraction of proposed method compared to Benchmark
FLOP's	14,106,245	5,413,388,608	0.26%
Accuracy	69.26%	93.08%	74.41%

Table IV: Comparison of FLOP's and Accuracy between the best Computer Vision Method and the Deep Learning Benchmark

VII. Conclusion

This research investigates the automated classification of leukemia cells in Pap smear images using both traditional machine learning methods and advanced deep learning models. By leveraging the Sipakmed dataset, we applied a comprehensive approach that included preprocessing techniques (Otsu's thresholding and watershed segmentation), multiple feature extraction methods (HOG, SIFT, LBP, ORB, and Gabor filters), and various machine learning classifiers (Logistic Regression, Random Forest, SVM, and k-NN). Additionally, deep learning models (AlexNet, VGG19, and ResNet) were utilized directly on the original images to evaluate their effectiveness in cell classification.

Our results show that preprocessing generally enhances classification performance for certain feature extraction techniques, particularly ORB, which yielded the highest improvements with Random Forest and k-NN classifiers. Among feature extraction methods, HOG with SVM achieved the highest accuracy without preprocessing, indicating the robustness of HOG features on original images. In contrast, other methods like LBP and Gabor showed varied responses to preprocessing, highlighting the importance of selecting feature extraction methods based on specific classifier compatibility.

The deep learning models further demonstrated the power of convolutional neural networks (CNNs) in this domain. ResNet achieved the highest overall performance, with an accuracy of 93.08%, outperforming both AlexNet and VGG19. This result underscores the potential of deep architectures in capturing complex cellular patterns and textures that are crucial for accurate classification in medical imaging tasks. AlexNet, while simpler, also showed promising results with competitive accuracy, suggesting its applicability where computational efficiency is prioritized.

This study provides a thorough analysis of machine learning and deep learning approaches for leukemia cell classification in Pap smear images. Our findings highlight the importance of selecting appropriate preprocessing, feature extraction, and model architectures to achieve optimal performance. Future work could explore the benefits of transfer learning on pre-trained medical models, further optimization of deep learning models for medical imaging, and the integration of ensemble methods to improve classification accuracy. This research contributes to the ongoing efforts in developing automated diagnostic tools that can assist pathologists in early detection and diagnosis of leukemia, ultimately enhancing clinical decision-making and patient outcomes.

When comparing the Computer Vision approach with the Deep Learning Benchmark, we are able to achieve almost 75% of the benchmark's performance, while being nearly 400 times faster in inference (383.76). This is not by any means a replacement for deep learning algorithms, but can serve its purpose when classifying a large amount of samples at once. This could be a valid proposal for automating medical database creation and saving on time, space and resources. It is also more accessible, since it comes without the need of GPU machines or cloud services.

REFERENCES

[1] M. Ghaderzadeh, A. Asadi, and A. Shalbaf, "Machine learning in detection and classification of leukemia using smear blood images: a systematic review," *Scientific Programming*, vol. 2021, Article ID 9933481, 2021.

[2] R. Baig, J. Muhammad, S. Rasheed, et al., "Detecting malignant leukemia cells using microscopic blood smear images: a deep learning approach," *Applied Sciences* , vol. 12, no. 13, p. 6317, 2022.

[3] J. Canny, "A computational approach to edge detection,"*IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[4] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics* , vol. 9, no. 1, pp. 62–66, Jan. 1979.

[5] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory* , vol. 13, no. 1, pp. 21–27, Jan. 1967.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. on Learning Representations (ICLR)* , 2015. [Online]. Available: https://arxiv.org/abs/1409.1556

[7] M. E. Plissiti, A. Dimitrakopoulou-Strauss, and A. Charchanti, "SIPaKMeD: A new dataset for feature and image based classification of normal and pathological cervical cells in Pap smear images," *IEEE Transactions on Medical Imaging* , vol. 37, no. 3, pp. 799–805, Mar. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8451588

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* , vol. 25, pp. 1097–1105, 2012.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* , 2016, pp. 770–778.

[10] P. Mehandiratta, "Cervical Cancer Largest Dataset - Sipakmed," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/prahladmehandiratta/cervical-cancer-largest-dataset-sipakmed