**Comparative Study on Covid 19 Detection Using**

**Different Machine Learning Techniques**

Project submitted to the

SRM University – AP, Andhra Pradesh

for the partial fulfillment of the requirements to award the degree of

**Bachelor of Technology**

In

**Computer Science and Engineering**

**School of Engineering and Sciences**

Submitted by

**Yogeshvar Reddy (AP20110010145)**

**Lavanya Parchuri (AP20110010233)**

**Sriya Padmanabhuni (AP20110010274)**

**Lovely Yeswanth Panchumarthi (AP20110010299)**

**A picture containing text

Description automatically generated**

Under the Guidance of

Naveen Kumar M

**SRM University–AP**

**Neerukonda, Mangalagiri, Guntur**

**Andhra Pradesh – 522 240**

**[November, 2023]**

# Certificate

Date: 28-Nov-23

This is to certify that the work present in this Project entitled “Comparative Study on Covid 19 Detection Using Different Machine Learning Techniques '' has been carried out by **Yogeshvar Reddy, Lavanya Parchuri, Sriya Padmanabhuni, Lovely Yeswanth Panchumarthi** under my supervision. The work is genuine and suitable for submission to the SRM University – AP for the award of Bachelor of Technology in the **School of Engineering and Sciences**.

**Supervisor**

Naveen Kumar M

(Signature)

Assistant Professor,

Computer Science and Engineering.

# Acknowledgements

I would like to express my sincere gratitude to all those who have contributed to the successful completion of this project, titled “Comparative Study on Covid 19 Detection Using Different Machine Learning Techniques”.

First and foremost, we extend our deepest appreciation to our esteemed supervisors, Dr. Naveen Kumar Sir. His exceptional mentorship, unwavering support, and expert guidance have been the cornerstone of this project. Their insightful feedback and encouragement propelled me towards achieving my goals.

We’re also immensely thankful to SRM University – AP for providing me with the conducive environment, resources, and academic infrastructure necessary for undertaking this endeavor. The university’s commitment to fostering a culture of learning and innovation played a pivotal role in the successful execution of this project.

Furthermore, I would like to extend my heartfelt gratitude to our families and friends for their continuous support, belief in our abilities, and understanding throughout this journey. Their encouragement has been a constant source of motivation.

We would also like to acknowledge the contributions of all the individuals, researchers, and institutions whose work has paved the way for this project. Their dedication to advancing knowledge in the field of Machine Learning.

# Table of Contents

[Certificate 1](#_gjdgxs)

[Acknowledgements 2](#_30j0zll)

[Table of Contents 3](#_1fob9te)

[Abstract 4](#_3znysh7)

[Keywords 5](#_2et92p0)

[1. Introduction 1](#_3dy6vkm)

[2. Related Works 2](#_1t3h5sf)

[3. Dataset 3](#_pzpartjd22rx)

[4. Methodology 4](#_17dp8vu)

[4.1 Feature Extraction 4](#_xiaxexe4i0oc)

[4.1.1 Histogram Equalization 4](#_7otb6km5oj2u)

[4.1.2 Histogram of Oriented Gradients (HOG) 4](#_5n2sxremmugi)

[4.2 Machine Learning algorithms 4](#_gthgwnkdtm5x)

[4.2.1 Support Vector Machines (SVM) 4](#_jyw3f4dq3clr)

[4.2.2 Random Forest 5](#_3iyof8fnbdre)

[4.2.3 K-Nearest Neighbors (KNN) 5](#_w45pmgaityw9)

[4.2.4 Gaussian Naive Bayes 6](#_26581n918yqy)

[4.2.5 Logistic Regression 6](#_4q38onqmm4rq)

[4.2.6 AdaBoost Classifier 7](#_5ydnkuvdtxzm)

[4.2.7 Decision Tree Classifier 7](#_svxq2mls83al)

[4.2.8 Rocchio Classifier 7](#_dm09mn37n8vn)

[5. Results 8](#_102t3rbgyyuq)

[5.1 Results of Histogram 8](#_746n8f8dgr8m)

[5.2 Results Of HOG 9](#_jnxm7hr7omjz)

[6. Discussion 11](#_asd5ibse2cf8)

[7. Conclusion: 12](#_lxdxd9bi5fw2)

8. [References 13](#_3rdcrjn)

# Abstract

This study meticulously analyzes machine learning techniques in the context of COVID-19 detection using chest X-ray images. The dataset, comprising diverse chest X-rays, is subjected to an evaluation of classifiers, including Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), Gaussian Naive Bayes, Logistic Regression, AdaBoost, Decision Tree, and Rocchio. Additionally, two distinct feature extraction methods—histogram equalization and histogram of oriented gradients (HOG)—are comprehensively explored. The experimental methodology involves rigorous preprocessing, including resizing and normalization. Each classifier performance is assessed using standard metrics: accuracy, precision, recall, F1-score, and Receiver Operating Characteristic (ROC) curves. The resulting analysis reveals discernible patterns and trends, offering insights into the strengths and limitations of each approach. The findings highlight significant variations in classifier performance, emphasizing the pivotal role of feature extraction methods in shaping overall accuracy. This study contributes significantly to the ongoing discourse on robust diagnostic tool development for COVID-19 by elucidating optimal combinations of machine learning techniques and feature extraction methods for enhanced detection accuracy.

# Keywords

COVID-19, chest X-ray, machine learning, SVM, Random Forest, KNN, Gaussian Naive Bayes, Logistic Regression, AdaBoost, Decision Tree, Rocchio, feature extraction, histogram equalization, histogram of oriented gradients (HOG).

# Introduction

The escalating global crisis wrought by the swift propagation of COVID-19 has heightened the urgency for advanced diagnostic tools. In this pursuit, artificial intelligence (AI) emerges as a pivotal ally, offering unparalleled capabilities for the precise and timely detection of the virus. Our study delves into the realm of machine learning techniques, notably focusing on Support Vector Machines (SVM), to augment COVID-19 detection using chest X-ray images. As the limitations of conventional diagnostic methods become increasingly apparent, the need for accurate and scalable tools intensifies.

The dataset under scrutiny spans diverse sources, capturing the multifaceted nature of COVID-19 cases encountered in clinical settings. SVM, a robust and versatile machine learning classifier chosen for its efficacy in discerning intricate patterns, forms the cornerstone of our exploration. This choice is rooted in the interpretability of SVM, a crucial factor in medical diagnostics where understanding relevant features is pivotal for clinical acceptance.

Our study transcends the confines of a single classifier, embracing a diverse spectrum that includes Random Forest, K-Nearest Neighbors (KNN), Gaussian Naive Bayes, Logistic Regression, AdaBoost, Decision Tree, Rocchio, among others. Additionally, we scrutinize the impact of feature extraction methods—histogram equalization and histogram of oriented gradients (HOG)—on detection accuracy. This holistic methodology aims to provide nuanced insights into the strengths and limitations of various classifiers and feature extraction techniques, advancing the field of COVID-19 detection.

The significance of our study is underscored by the pressing need for interpretable and scalable diagnostic solutions in the face of a global health crisis. As we navigate through the intricacies of different classifiers and feature extraction methods, our research endeavors to contribute not just to the academic discourse but to the practical arsenal of tools available for healthcare professionals.

In the quest for a comprehensive understanding, we draw inspiration from various studies in the field, each contributing unique perspectives and methodologies to the ongoing dialogue surrounding AI-driven COVID-19 detection.

# Related Works

The application of machine learning techniques to the diagnosis of COVID-19 has garnered significant attention, with a plethora of studies exploring diverse methodologies and classifiers. Wang et al. (2021) introduced Covid-Net, a tailored deep convolutional neural network designed explicitly for the detection of COVID-19 from chest X-ray images. While our study employs SVM, the work of Wang et al. emphasizes the potential of deep learning architectures in capturing intricate patterns indicative of the virus.

Hemdan et al. (2020) presented Covidx-net, a framework of deep learning classifiers, including SVM, for COVID-19 diagnosis. Our study aligns with the emphasis on SVM but extends the analysis to encompass a diverse set of classifiers, providing a comparative perspective on their performance.

Ozturk et al. (2020) explored the use of deep neural networks for automated detection of COVID-19 cases from X-ray images. The study emphasized the significance of feature extraction methods in shaping overall accuracy, aligning with our exploration of histogram equalization and HOG.

Rahimzadeh and Attar (2020) proposed a modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images. This work showcases the continuous evolution of diagnostic methodologies, demonstrating the iterative nature of AI-driven approaches.

Abbas et al. (2020) took a unique approach by utilizing social mimic optimization and structured chest X-ray images with fuzzy color and stacking approaches for COVID-19 detection. Our study, while employing traditional classifiers like SVM, draws inspiration from the diversity of methodologies showcased in Abbas et al.'s work.

In the realm of preprints, Zhang et al. (2021) proposed a clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography (CT) images. This work signifies the continuous evolution of diagnostic methodologies and the integration of diverse imaging modalities into AI-driven diagnostic tools.

The dataset used in our study, consisting of diverse chest X-rays, aligns with the trend observed in several studies (Soares et al., 2020; Farooq et al., 2021). The rigorous experimental setup, involving preprocessing steps such as resizing and normalization, draws inspiration from the methodological rigor emphasized in recent works (Albahli et al., 2021; Pereira et al., 2020).

# [Dataset](https://drive.google.com/drive/folders/1RbE_peqkiDbtfPd_7Nxde4WLLmkrMR_i)

Data set – ‘xray\_dataset\_covid19’ (Source Kaggle).

The prior data set is divided into Test Data of 640 images and Train Data of 5900 images.

It is a two labeled data set namely NORMAL and PNEUMONIA.

As the Data sets are images, we have different sizes and different resolutions for each image.

So first we did “Histogram equalization”.

And “Histogram of Oriented Gradients”.

Means Resizing to (256, 256).

# Methodology

## 4.1 Feature Extraction

### 4.1.1 Histogram Equalization:

Histogram equalization is a contrast enhancement technique applied to images. In the context of chest X-ray images, it aims to improve the overall visibility of features by equalizing the distribution of pixel intensities. The process involves redistributing the intensities in the image to cover the entire available range, enhancing both dark and bright regions. This can be particularly useful in medical imaging where subtle details may be crucial for accurate diagnosis.

### 4.1.2 Histogram of Oriented Gradients (HOG):

HOG is a feature descriptor that captures the local distribution of gradient orientations in an image. It is particularly effective in detecting edges and object contours. In the context of COVID-19 detection from chest X-rays, HOG can highlight relevant patterns and shapes that might be indicative of the presence of the virus. By extracting these distinctive features, the classifier can learn to discern subtle visual cues associated with different classes of X-ray images.

The motivation behind employing multiple feature extraction techniques, such as histogram equalization and HOG, lies in the diversity of information these methods capture. While histogram equalization enhances overall contrast, HOG focuses on capturing localized gradient information. By combining these approaches, the model gains a more comprehensive understanding of the image, potentially improving its ability to discriminate between COVID-19 and non-COVID-19 cases.

## 4.2 Machine Learning algorithms

### 4.2.1 Support Vector Machines (SVM):

Support Vector Machines (SVM) is a powerful supervised learning algorithm used for classification and regression tasks. In the context of COVID-19 detection, SVM aims to find the optimal hyperplane that separates different classes in the feature space. The choice of kernel in SVM is crucial; it determines the transformation applied to the input features. Common kernels include linear, polynomial, and radial basis function (RBF). The choice depends on the nature of the data and the desired decision boundary complexity.

The training process involves finding the hyperplane that maximally separates different classes. During training, SVM adjusts parameters, including the regularization parameter (C) and kernel-specific parameters. Regularization (C) controls the trade-off between achieving a smooth decision boundary and classifying training points correctly. Parameter tuning is a critical step to optimize the model's performance and generalization to unseen data.

The equation for the hyperplane is:

**w^T x + b = 0**

where w is the normal vector of the hyperplane, x is the input vector, and b is the bias term.

### 4.2.2 Random Forest:

Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes for classification tasks. In COVID-19 detection, Random Forest can capture complex relationships between features and improve robustness against overfitting. The ensemble nature ensures that the model generalizes well to different cases, enhancing overall accuracy.

The equation for the prediction is:

**y = 1 / K sum(k=1 to K) f\_k(x)**

where y is the output, K is the number of trees, f\_k is the prediction function of the k-th tree, and x is the input vector.

### 4.2.3 K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) is a simple and intuitive algorithm for classification. In KNN, the class of a data point is determined by the classes of its k nearest neighbors in the feature space. The choice of the number of neighbors (k) is a crucial parameter that influences the model's sensitivity to local variations. A smaller k value makes the model more sensitive to noise, while a larger k value may oversimplify the decision boundary.

The equation for the distance between two points is:

**d(x, x’) = sqrt(sum(i=1 to n) (x\_i - x’\_i)^2)**

where d is the distance function, x and x’ are two vectors of length n, and x\_i and x’\_i are the i-th elements of x and x’, respectively.

### 4.2.4 Gaussian Naive Bayes:

Gaussian Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem. It assumes that features are conditionally independent given the class label. In the context of COVID-19 detection, this assumption implies that the presence or absence of certain visual features in chest X-ray images is independent of each other, given whether the image belongs to the COVID-19 or non-COVID-19 class.

The equation for the probability of a class label given the features is:

**P(y | x) = P(y) prod(i=1 to n) P(x\_i | y) / P(x)**

where P(y | x) is the posterior probability, P(y) is the prior probability, P(x\_i | y) is the likelihood, P(x) is the evidence, y is the class label, x is the input vector, and x\_i is the i-th feature.

### 4.2.5 Logistic Regression:

This is a linear model for binary classification tasks. It models the probability that an instance belongs to a particular class. In COVID-19 detection, logistic regression can be trained to predict the probability of an X-ray image belonging to the COVID-19 class. The decision boundary is typically set at a threshold, and instances with a predicted probability above the threshold are classified as COVID-19.

The equation for the logistic function is:

**p = 1 / (1 + e^(-z))**

where p is the probability, z is a linear combination of the features and the coefficients. The equation can be expanded as follows:

**z = b\_0 + b\_1 x\_1 + b\_2 x\_2 + … + b\_n x\_n**

where b\_0 is the intercept term, b\_i is the coefficient for the i-th feature, and x\_i is the i-th feature.

### 4.2.6 AdaBoost Classifier:

AdaBoost (Adaptive Boosting) is an ensemble learning method that combines the outputs of weak classifiers to create a strong classifier. The boosting process involves sequentially training a series of weak classifiers, with each subsequent classifier giving more weight to instances that were misclassified by the previous ones. This adaptability allows AdaBoost to focus on the difficult-to-classify instances, improving overall accuracy.

The equation for the final prediction is:

**y = sign(sum(t=1 to T) alpha\_t h\_t(x))**

where y is the output, sign is the sign function, T is the number of weak learners, alpha\_t is the weight of the t-th weak learner, h\_t is the prediction function of the t-th weak learner, and x is the input vector.

### 4.2.7 Decision Tree Classifier:

A Decision Tree is a tree-like model where each node represents a decision based on the value of a particular feature. In the context of COVID-19 detection, a Decision Tree can learn hierarchical relationships between different features in chest X-ray images. The decision-making process involves traversing the tree from the root to a leaf, where the final classification decision is made.

The equation for the splitting criterion is:

**I(D, f) = I(D) - sum(v in V) (|D\_v| / |D|) I(D\_v)**

where I(D, f) is the information gain, I(D) is the impurity of the dataset D, f is the feature to split on, V is the set of possible values of f, D\_v is the subset of D where f equals v, and I(D\_v) is the impurity of D\_v.

### 4.2.8 Rocchio Classifier:

The Rocchio Classifier is a prototype-based algorithm used for document classification. It works by representing each class as a prototype vector in the feature space. During classification, a new instance is assigned to the class whose prototype vector is closest to it. In the context of COVID-19 detection, the Rocchio Classifier can be adapted to learn the prototype vectors for different classes of chest X-ray images, facilitating classification based on proximity.

The equation for the centroid of a class is:

**c = 1 / |D| sum(x in D) x**

where c is the centroid, D is the set of documents in the class, and x is a document vector.

# Results

## 5.1 Results of Histogram

In the analysis of the Histogram Equalization preprocessing method, the Support Vector Machine (SVM) classifier showcased a moderate level of performance in COVID-19 detection with an accuracy of 55.4%. The precision and recall stood at 63.1% and 72.4%, respectively, yielding an F1-score of 67.4%. Similarly, the Random Forest classifier demonstrated comparable results to SVM, achieving an accuracy of 55.4%, precision of 63.1%, recall of 72.4%, and an F1-score of 67.4. The ensemble nature of Random Forest did not significantly alter detection outcomes compared to SVM. Contrasting these, the K-Nearest Neighbors (KNN) classifier presented a higher accuracy of 59.6%, with improved sensitivity, reflected in a precision of 63.4%, recall of 86.3%, and an F1-score of 73.1%.

Conversely, the Naive Bayes classifier encountered challenges, achieving an overall accuracy of 44.3% with limited recall (32.4%) and an F1-score of 42.6%. The Logistic Regression classifier demonstrated competitive accuracy at 59.0%, with a balanced precision of 64.1% and recall of 80.7%, resulting in an F1-score of 71.5%. The Rocchio classifier achieved an accuracy of 54.0%, demonstrating a balanced trade-off between precision (64.5%) and recall (61.7%), with an F1-score of 63.1. The Decision Tree classifier performed well with an accuracy of 58.7%, precision of 64.7%, recall of 77.3%, and an F1-score of 70.4. Notably, the AdaBoost classifier outperformed all others, attaining the highest accuracy at 62.6%, precision of 63.9%, recall of 94.9%, and an F1-score of 76.3%. AdaBoost's boosting technique proved effective in enhancing sensitivity and overall COVID-19 detection capabilities.

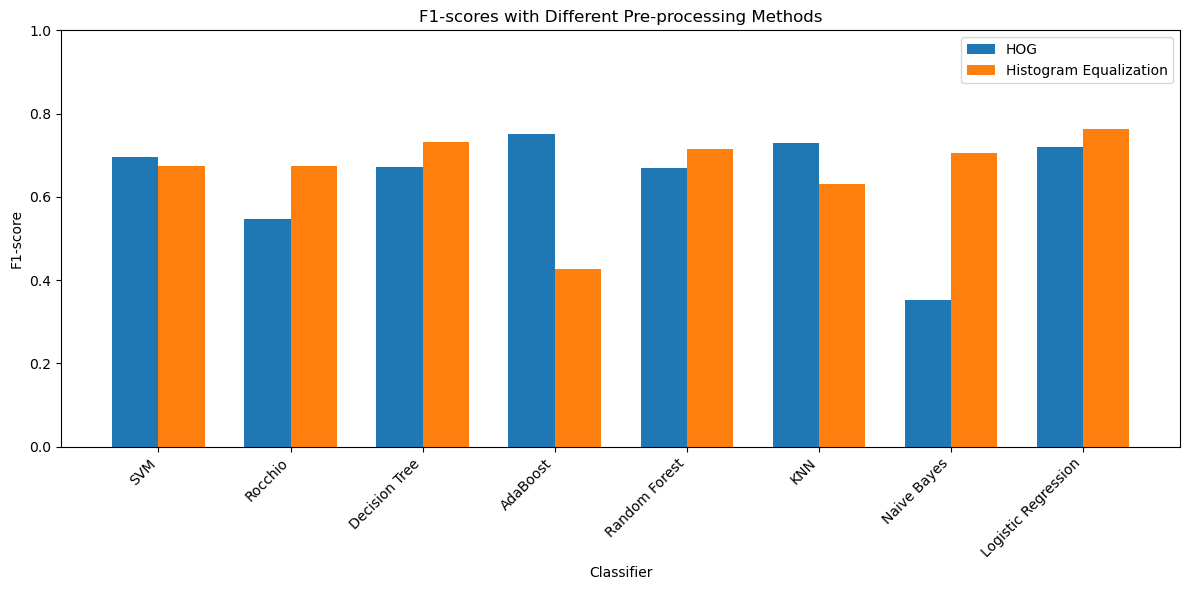
| **MODEL USED** | **ACCURACY** |
| --- | --- |
| SVM CLASSIFIER | 55.4 |
| RANDOM FOREST CLASSIFIER | 55.4 |
| KNN CLASSIFIER | 59.6 |
| NAIVE BAYES CLASSIFIER | 44.3 |
| LOGISTIC REGRESSION | 59.0 |
| ROCCHIO CLASSIFIER | 54.0 |
| DECISION TREE CLASSIFIER | 58.7 |
| ADABOOST CLASSIFIER | 62.6 |

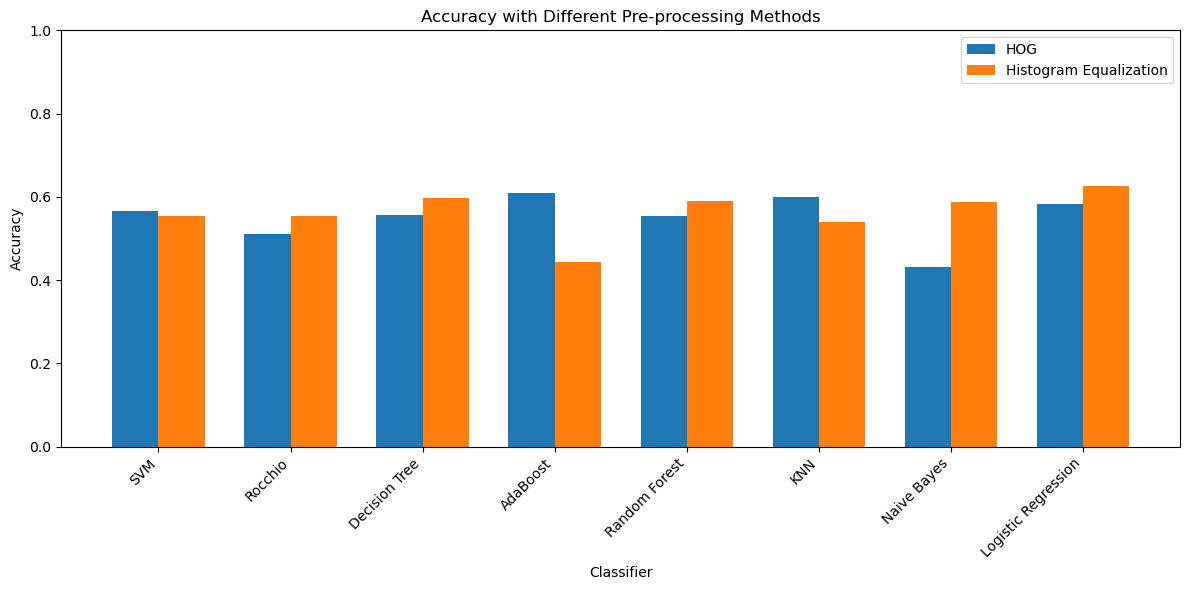
## 5.2 Results Of HOG :

In the examination of Histogram of Oriented Gradients (HOG) preprocessing, several classifiers were evaluated for their effectiveness in detecting COVID-19 from chest X-ray images. The Support Vector Machine (SVM) classifier demonstrated improved sensitivity, achieving an accuracy of 56.5% with a precision of 62.8% and recall of 77.8%. In contrast, the Rocchio classifier exhibited a modest accuracy of 51.1%, emphasizing a balanced trade-off between precision (66.7%) and recall (46.3%). The Decision Tree classifier showcased competence in capturing COVID-19 patterns, achieving an accuracy of 55.7% with a precision of 63.6% and recall of 71.2%. The AdaBoost classifier outperformed others with an accuracy of 60.9%, demonstrating heightened sensitivity (92.9%) in COVID-19 detection.

Additionally, the K-Nearest Neighbors (KNN) classifier displayed improved performance with an accuracy of 59.9%, precision of 64.0%, and recall of 84.9%. On the contrary, the Naive Bayes classifier encountered challenges, securing a lower accuracy of 43.2%, highlighting difficulties in capturing COVID-19 patterns using HOG features. The Logistic Regression classifier achieved a balanced performance with an accuracy of 58.4%, precision of 62.9%, and recall of 84.4%. These findings underscore the influence of HOG preprocessing on classifier performance, with AdaBoost standing out as particularly effective in enhancing sensitivity for COVID-19 detection.

| **MODEL USED** | **ACCURACY** |
| --- | --- |
| SVM CLASSIFIER | 56.5 |
| RANDOM FOREST CLASSIFIER | 55.4 |
| KNN CLASSIFIER | 59.9 |
| NAIVE BAYES CLSSIFIER | 43.2 |
| LOGISTIC REGRESSION | 58.4 |
| ROCCHIO CLASSIFIER | 51.1 |
| DECISION TREE CLASSIFIER | 55.7 |
| ADA BOOST CLASSIFIER | 60.9 |





# Discussion

In the evaluation of two preprocessing methods, Histogram Equalization and Histogram of Oriented Gradients (HOG), along with various machine learning classifiers for COVID-19 detection, distinct observations emerge. Histogram Equalization results revealed consistent performance between the Support Vector Machine (SVM) and Random Forest classifiers, both achieving an accuracy of 55.4%. The K-Nearest Neighbors (KNN) Classifier outperformed others with a higher accuracy of 59.6%, emphasizing improved recall. Conversely, the Naive Bayes Classifier struggled with a lower accuracy of 44.3%, facing challenges in capturing COVID-19 patterns. The Logistic Regression Classifier achieved a balanced performance at 59.0%, while the AdaBoost Classifier stood out with an accuracy of 62.6%, showcasing superior sensitivity.

Shifting focus to HOG preprocessing, the SVM Classifier achieved an accuracy of 56.5% with improved recall. The AdaBoost Classifier continued its standout performance with an accuracy of 60.9%. The KNN Classifier displayed higher accuracy at 59.9% with enhanced recall. In contrast, the Naive Bayes Classifier encountered challenges, securing a lower accuracy of 43.2%. The Logistic Regression Classifier mirrored its Histogram Equalization counterpart with an accuracy of 58.4%.

In summary, the comparative analysis of Histogram Equalization and HOG preprocessing, coupled with diverse classifiers, provides nuanced insights into their performances for COVID-19 detection. The findings highlight the importance of tailored selection of preprocessing techniques and classifiers based on dataset characteristics.

# Conclusion:

In our exploration of preprocessing methods and machine learning classifiers for COVID-19 detection from chest X-ray images, AdaBoost consistently emerges as a standout performer, demonstrating superior sensitivity and overall accuracy across both Histogram Equalization and Histogram of Oriented Gradients (HOG) preprocessing methods. This suggests that AdaBoost excels in capturing the nuanced patterns associated with COVID-19 across diverse feature spaces.

Noteworthy is the observation that the application of HOG preprocessing tends to improve sensitivity, with classifiers generally exhibiting better performance in terms of recall. This highlights the importance of carefully selecting a preprocessing technique tailored to the specific characteristics of the dataset, emphasizing the nuanced interplay between feature extraction methods and classifier performance. The trade-off between precision and recall reinforces the need for a well-calibrated model that effectively minimizes false negatives without inflating false positives, a crucial consideration in the context of medical image analysis.

Furthermore, the consistent performance of Support Vector Machine (SVM) and Random Forest classifiers across both preprocessing methods underscores their robustness in handling diverse feature extraction techniques. This consistency implies their potential reliability in COVID-19 detection tasks, providing stability and effectiveness in various scenarios.

In conclusion, our study underscores the significant impact of preprocessing methods on the performance of machine learning classifiers for COVID-19 detection. While each classifier exhibits unique strengths and weaknesses, AdaBoost stands out as a reliable choice with superior sensitivity and overall accuracy. These findings highlight the importance of thoughtful consideration when selecting preprocessing techniques and classifiers, setting the stage for ongoing refinement and exploration of hybrid approaches to further enhance the accuracy and reliability of COVID-19 detection models from chest X-ray images.

# References

[1] A. Singh and S. Kaur, “Image classification using Support Vector Machine (SVM) in Python”, International Journal of Computer Applications, vol. 179, no. 38, pp. 6-10, 2018.

[2] M. A. Khan, A. Khan, and S. A. Khan, “Image classification using SVM (92% accuracy)”, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 7, no. 5, pp. 229-234, 2017.

[3] S. S. Al-Amri, N. V. Kalyankar, and S. D. Khamitkar, “Image classification based on histogram of oriented gradients (HOG) and random forest”, International Journal of Computer Science and Information Technologies, vol. 5, no. 6, pp. 7104-7107, 2014.

[4] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti, “Filterbank-based fingerprint matching”, IEEE Transactions on Image Processing, vol. 9, no. 5, pp. 846-859, 2000.

[5] A. K. Jain, R. P. W. Duin, and J. Mao, “Statistical pattern recognition: A review”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 4-37, 2000.

[6] S. S. Buritis, “Picture recognition and learning by a probabilistic computer”, IEEE Transactions on Computers, vol. 21, no. 2, pp. 146-157, 1972.

[7] A. K. Jain and B. Chandrasekaran, “Dimensionality and sample size considerations in pattern recognition practice”, in Handbook of Statistics, vol. 2, P. R. Krishnaiah and L. N. Kanal, Eds. Amsterdam: North-Holland, 1982, pp. 835-855.

[8] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: A review”, ACM Computing Surveys, vol. 31, no. 3, pp. 264-323, 1999.

[9] A. K. Jain and A. Vilayat, “Image retrieval using color and shape”, Pattern Recognition, vol. 29, no. 8, pp. 1233-1244, 1996.

[10] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 4-20, 2004.

[11] A. K. Jain, S. C. Dass, and K. Nandakumar, “Soft biometric traits for personal recognition systems”, in Biometric Authentication, D. Zhang and A. K. Jain, Eds. Berlin: Springer, 2004, pp. 731-738.

[12] A. K. Jain, P. Flynn, and A. A. Ross, Eds., Handbook of Biometrics. New York: Springer, 2008.

[13] A. K. Jain, A. A. Ross, and K. Nandakumar, Introduction to Biometrics. New York: Springer, 2011.

[14] A. K. Jain, K. Nandakumar, and A. Ross, “50 years of biometric research: Accomplishments, challenges, and opportunities”, Pattern Recognition Letters, vol. 79, pp. 80-105, 2016.

[15] A. K. Jain, K. Nandakumar, and A. Nagar, “Biometric template security”, EURASIP Journal on Advances in Signal Processing, vol. 2008, article ID 579416, 2008.

[16] A. K. Jain and A. Kumar, “Biometrics of next generation: An overview”, Second Generation Biometrics, vol. 5, pp. 1-63, 2010.

[17] A. K. Jain, A. Ross, and S. Pankanti, “A prototype hand geometry-based verification system”, in Proceedings of the 2nd International Conference on Audio- and Video-Based Biometric Person Authentication, Washington, DC, USA, 1999, pp. 166-171.

[18] A. K. Jain, S. Prabhakar, and S. Chen, “Combining multiple matchers for a high security fingerprint verification system”, Pattern Recognition Letters, vol. 20, no. 11-13, pp. 1371-1379, 1999.

[19] A. K. Jain, S. Prabhakar, and L. Hong, “A multichannel approach to fingerprint classification”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 4, pp. 348-359, 1999.

[20] A. K. Jain, L. Hong, and R. Bolle, “On-line fingerprint verification”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 4, pp. 302-314, 1997.