Geetha K/Afr.J.Bio.Sc. 6(6) (2024) 7045-7061 ISSN: 2663-2187

https://doi.org/10.33472/AFJBS.6.6.2024.7045-7061

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Lung Cancer Detection in CT Images Using Auto Color Correlogram Features and Multiple Machine Learning Classifiers

Geetha K[1], Dr.Karthikeyan Elangovan[2]

[1]Research Scholar (Part time Internal), Department of Computer and Information Science, Annamalai University, Chidambaram, India.

[2]Research Supervisor, Assistant Professor/Programmer, (Deputed as Assistant Professor and Head in Government Arts and Science College,Gingee,Villupuram), Department of Computer and Information Science ,Faculty of Science, Annamalai University, Annamalai Nagar, Tamilnadu, India.

Email: geetha2kumar@gmail.com, kanchikarthi2010@gmail.com

Article Info

Volume 6, Issue 6, July 2024

Received: 24 May 2024

Accepted: 21 June 2024

Published: 10 July 2024

doi: 10.33472/AFJBS.6.6.2024.7045-7061

ABSTRACT:

Early detection is crucial in improving lung cancer outcomes. However, lung cancer often remains asymptomatic in its early stages, leading to late-stage diagnoses in many cases. Computed Tomography (CT) scanning has emerged as a powerful tool for lung cancer screening and diagnosis, offering detailed, three-dimensional images of the lungs that can reveal small nodules or tumors before they become symptomatic. The interpretation of CT images, however, is a complex and time-consuming task that requires significant expertise. Radiologists must carefully analyze numerous images to identify potential malignancies, a process that is susceptible to human error and fatigue. This challenge has spurred the development of computer-aided detection (CAD) systems, which aim to assist radiologists by automatically identifying and classifying suspicious areas in CT scans. Recent advancements in artificial intelligence and machine learning have opened new avenues for improving the accuracy and efficiency of lung cancer detection and classification. These technologies offer the potential to analyze vast amounts of imaging data, recognize subtle patterns, and provide rapid, consistent results. By augmenting human expertise with machine learning algorithms, we can potentially enhance early detection rates, reduce false positives and negatives, and ultimately improve patient outcomes. This study investigates the effectiveness of various machine learning models for lung cancer detection and classification using CT images. The research employs Auto Color Correlogram (ACC) features and compares the performance of six classifiers: Additive Regression (AR), Naive Bayes (NB), Linear Regression (LR), Attribute Selected Classifier (ASC), Naive Bayes Multinomial (NBM), and Logistic Regression. The results demonstrate that the AR model outperforms other classifiers across multiple evaluation metrics. AR achieves the highest accuracy at 90.10%, precision of 0.90, recall of 0.89, ROC of 0.97, and PRC of 0.97. It also exhibits superior performance in terms of kappa statistic (0.65), F-Measure (0.89), and Matthews Correlation Coefficient (0.65). In contrast, the ASC model generally shows the lowest performance, with an accuracy of 84.10%, precision of 0.81, recall of 0.83, ROC of 0.84, and PRC of 0.78. The ASC model also has the lowest kappa value (0.53) and Matthews Correlation Coefficient (0.53) among the compared models. Notably, the Logistic Regression model matches the AR model in precision (0.90), while the NBM model shows the lowest F-Measure (0.34) among all classifiers. These findings suggest that the Additive Regression model, when combined with ACC features, offers a promising approach for automated lung cancer detection and classification in CT images. This research contributes to the ongoing efforts to enhance computer-aided diagnosis systems in oncology, potentially improving early detection and classification of lung cancer.

Keyword: Lung cancer,Additive Regression (AR), Naive Bayes (NB), Linear Regression (LR), Attribute Selected Classifier (ASC), Naive Bayes Multinomial (NBM), and Logistic Regression

1. Introduction

Cancer is an uncontrollably growing, aberrant tissue that swiftly spreads throughout the body. If not treated appropriately in the beginning, it can also spread to other organs. Modern technologies have greatly benefited the medical industry, particularly lung cancer detection. It helps doctors diagnose and properly treat medical conditions. Because lung cancer claims so many lives, it's one of the most terrifying illnesses in the world. According to the study, there were 2.21 million instances of lung cancer detected in 2020, and 1.8 million of those cases resulted in death [1]. Lung cancer has the highest annual death rate of any cancer type (1.80 million), according to a 2020 World Health Organization (WHO) study. It illustrates the number of species that will become extinct due to cancer by 2020, according to the WHO. When it comes to lung cancer, early detection and effective management are critical to receiving the best possible care. Lung cancer is one of the leading causes of cancer-related deaths worldwide, and early detection is critical to improving patient outcomes. Thanks to computer-aided diagnostic (CAD) systems, radiologists can now more effectively detect and categorize lung cancer with the aid of medical images. This project aims to develop a more sophisticated computer-aided detection (CAD) system that can detect and categorize lung cancer using computed tomography (CT) scans.

The recommended approach detects lung cancer quickly and precisely by utilizing multiple machine learning models in conjunction with auto-color correlogram (ACC) feature extraction. Physicians frequently utilize CT scans to detect lung cancer because they can produce precise, three-dimensional images of lung tissue. However, interpreting these images can be challenging for physicians, particularly when dealing with minor issues or tumors that are still in the early stages of the disease.

Our technique uses the ACC filter, which has shown promising results in color-based image retrieval tasks, to extract meaningful color and texture information from CT images. Then, you can teach and test various machine learning models, including Logistic Regression, Attribute Selected Classifier, Naive Bayes Multinomial, Additive Regression, Linear Regression, and Naive Bayes Multinomial, using these features.

This study's primary objective is to develop robust software that can distinguish between healthy lung tissue and the three forms of lung cancer—adenocarcinoma, squamous cell carcinoma, and big cell carcinoma. Our objective is to compare the performance of several algorithms in order to determine which approach, or combination of methods, is optimal for precisely identifying and categorizing lung cancer.

This study contributes to the growing body of research on CAD systems for lung cancer detection by:

- Exploring the effectiveness of ACC features in capturing relevant visual characteristics of lung CT images.
- Evaluating and comparing the performance of multiple machine learning classifiers in the context of lung cancer detection.
- Providing insights into the strengths and limitations of different classification approaches for this specific medical imaging task.

The successful development of such a system could potentially assist radiologists in their diagnostic work, leading to earlier detection of lung cancer and improved patient outcomes. Additionally, the findings from this research may inform future studies in medical image analysis and contribute to the ongoing efforts to enhance the accuracy and efficiency of CAD systems in oncology.

This paper organizes section 2 focuses on literature survey; in section 3 presents materials and methods; in section 4 shows results and interpretations, and finally section 5 has conclusion of this research work.

2. Literature Survey

The use of deep learning methods, particularly Deep Convolutional Neural Networks (DCNN), to automate the diagnosis and categorization of lung cancer is examined in this review paper. It offers an overview of methodology, advances, quality assessments, and tailored deep learning algorithms and covers a variety of medical imaging modalities, including MRIs, CT scans, WSI, and X-rays. The study emphasizes how DCNN helps with lung cancer categorization and detection [1].This study presented a DL-based Lung Cell Cancer Detection (DL-LCCD) technique that reliably detects and classifies malignant cells in lung tissue. The methodology uses a hybrid CNN model and digital image processing techniques to accurately and precisely diagnose cancer from CT scanned pictures.[2].The use of machine learning techniques for lung cancer prediction and detection utilizing medical imaging data was covered in this overview study. In order to compare different classifiers and image processing processes for accurately diagnosing malignant and normal lung cancers, it reviews a number of proposed systems [3]. Other methods have been developed to read and learn data representation from the disorganized (raw) data using a deep learning algorithm. Details of the inner body are inspected, and useful information is gleaned from this data. The use of deep learning models, algorithms, and techniques has been shown to significantly improve classification accuracy and reduce error in cases of lung cancer. In many ways, automatic segmentation based on deep learning is superior to manual segmentation [4].Deep learning produces high-quality images, lowers error rates, prevents misclassification, and effectively diagnoses cancer. Various classifiers are employed to eliminate false positive nodules [5]. The radiologist's ability to diagnose patients quickly and accurately is directly correlated with accurate and high-quality images. Additionally, deep learning techniques are used to forecast lung cancer. [6]. Features are automatically derived from training photos. Deep learning is less expensive than traditional CAD frameworks in comparison. The radiologist benefits from deep learning's HD representation of the input data, which streamlines and expedites the process of detection and identification. Since pixels are used to distinguish between cancerous and non-cancerous areas, each pixel in the image immediately aids in the identification of cancer. Thus, medical practitioners may better serve the healthcare system by using deep learning to help with accurate diagnosis and disease classification. Making appropriate decisions about the illness is aided by it. Convolutional Layers are one of the several stages that make up the CNN architecture (CLs). The CL layers are able to extract specific information from the provided images of cancer cells by utilizing several types of convolution filters [7]. The layers of a convolutional neural network (CNN) are numerous. Convolutional layer that chooses the feature and extracts its features from the image pooling layer. Combining the collected features is the task of the third layer, often known as the fully connected or FC layer. Recurrent neural networks (RNNs) are mostly utilized for text, audio, and video and are appropriate for sequential data. A Deep Belief Network (DBN) is made up of several RBM. These generative models are probabilistic in nature. DBN comes in a wide variety. An method based on statistical theory is called Support Vector Machine (SVM). Artificial Neural Networks (ANNs) are referred to as biologically inspired networks because of their structure, which is similar to that of human brains with neurons. In the realm of artificial intelligence, the Deep Neural Network (DNN) is a novel and sophisticated method that may also be applied to complex nonlinear relationships. Asthma, cancer, and AIDS are among the human diseases closely linked to DNA-binding proteins [8].Early illness identification and real-time patient monitoring are made possible by integrating CC technology with wireless body area networks (WBANs) devices to develop sensor-cloud infrastructure (S-CI), which benefits the healthcare sector [9] while protecting patient privacy. A well-developed deep learning model may assist avoid time wastage and incorrect diagnoses [10]. Deep machine learning could be used for image preprocessing, image segmentation to highlight the diagnostic objects under examination, and object classification to determine if the objects are benign or cancerous [11,25,26]. Predicting human diseases, particularly cancer, in order to provide more efficient and timely care is difficult. Numerous organs and systems in the human body are impacted by the potentially fatal condition known as cancer [12,27]. This study employed the Convolutional Neural Network (CNN) Deep Learning algorithm to identify a lung nodule, which has the potential to be malignant, based on a variety of CT scan pictures provided to the model. To solve the problem of lung nodule detection, an ensemble approach has been devised for this work. To improve performance and outcome prediction accuracy, we pooled the output of two or more CNNs rather of utilizing a single Deep Learning model. The dataset used for the LUNA 16 Grand Challenge is accessible online at their website. The dataset comprises of a CT scan annotated with information to help comprehend the data and specifics of each scan. Artificial Neural Networks are the foundation of deep learning because they function similarly to brain neurons. The deep learning model is trained on a large dataset of CT scans. CNNs are trained on the data set to distinguish between photos with and without cancer. For our Deep Ensemble 2D CNN, a set of training, validation, and testing datasets is created. The Deep Ensemble 2D CNN is made up of three distinct CNNs with various pooling strategies, layers, and kernels. With a total accuracy of 95%, our Deep Ensemble 2D CNN outperformed the baseline technique.[13] Since it might be difficult to distinguish between a lung nodule and lung tissue, an ensemble technique has been devised to aid in the detection of lung nodules. To achieve this, a more precise model for differentiating between a lung nodule candidate and a lung nodule should be created. Instead of the availability of picture data, the primary challenge for any researcher is obtaining pertinent annotations and labeled image data. All free-text reports that are based on the conclusions of radiologists are kept in PACS format. Therefore, turning all of these reports into data that is more appropriately and accurately labeled and into structural results can be a difficult task that calls for text-mining techniques. The study of these text-mining techniques alone is crucial. These days, text mining is another common use for deep learning. In this sense, machine and deep learning goals will profit from the creation of an organized reporting system. This advancement has the potential to improve radiologic findings, and radiologists may be able to handle several doctors' responsibilities with the use of the patient care computerized aid system. A thorough examination of both Nodule Candidates and True Nodules is part of the lung nodule identification procedure. actual and fake nodules that mimic actual nodules make up lung nodule candidates. In order to choose genuine nodules from among all potential candidate nodules, a categorization system needs to be created. In order to identify actual nodules, two issues must be given additional consideration in order to establish such nodules. Lung nodules in the CT image have been identified using a two-dimensional CNN. CNN only considers two dimensions in 2D CNN. Deep learning and ensemble learning techniques were applied to classification issues in numerous studies[14,24]. The purpose of this work is quite near to the present CAD applications for lung cancer classification of lung nodules. As a result, we investigated the most recent and advanced methods for classifying lung nodules. CNN with a transfer learning strategy was created with Multiresolution CNN [15] and Knowledge Transfer for Candidate Classification in Lung Nodule Detection in order to extract the primary characteristics from the picture and learn these features. CNN image-wise computation using various depth layers employed for Luna lung nodule classification 16 Data Set to increase lung nodule detection accuracy to 0.9733 accuracy. [16] created a multi-view convolutional network CAD system for lung nodules in order to reduce false positives. A deep residual learning method utilizing CT scan data for cancer detection is Multiview-KBC[17], which is based on Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest[18]. This approach uses the ResNet14 and UNet models for feature extraction. algorithms for machine learning Random forest and XGBoost are used to classify photos that are malignant. This model's accuracy was 84%. the study that suggests using ensemble learning and machine learning techniques to forecast lung cancer based on early symptoms.[19] In order to classify lung cancer, this study used a variety of machine learning methods, such as MLP (multilayer perceptron)[20], SVM (support vector machine)[21], Naïve Bayes[22], and neural networks. The UCI repository provided the dataset that was used in this investigation. For the suggested investigation, the ensemble learning method's accuracy was 90%[23].

3. Materials and Methods

The materials and techniques used in the research investigation are particularly covered in this section. The public data repository, namely the Kaggle data repository, is where the Chest CT-Scan pictures Dataset was acquired [24]. It was an experiment on applying deep learning and machine learning to identify chest cancer (CNN). Utilizing an AI model, here categorize and determine if the patient has cancer or not. Here enlighten them on the kind of cancer and its course of treatment. We made an effort to gather all the information required for the model to quickly classify the photos. Thus, in order to begin the research, this work had to gather data from numerous sources. This work conducted extensive investigation to gather all the information from various sources and prepared it for CNN.

Data

To fit the model, the photos are in jpg or png format instead of dcm format. Three kinds of chest cancer—adenocarcinoma, large cell carcinoma, and squamous cell carcinoma—as well as one folder containing normal cells are contained in the data. The main folder is called "Data," and it contains all of the step folders: "test," "train," "valid test represent testing set," "train represent training set," "valid represent validation set," with training comprising 70%, testing comprising 20%, and validation comprising 10%.

Adenocarcinoma

This type of lung cancer accounts for roughly 40% of non-small cell lung cancer cases and 30% of all cases overall. It is the most frequent type of lung cancer. Adenocarcinomas can be detected in the prostate, breast, and colorectal malignancies, among other prevalent cancers. Lung adenocarcinomas are located in the glands that release mucus and aid in breathing in the outer part of the lung. Coughing, hoarseness, weight loss, and weakness are some of the symptoms.

Large cell carcinoma

Anywhere in the lung can develop large-cell undifferentiated carcinoma lung cancer, which spreads swiftly and expands. Ten to fifteen percent of all instances of non-small cell lung cancer (NSCLC) are typically of this kind. Undifferentiated large-cell carcinoma typically grows and spreads swiftly.

Squamous cell carcinoma

This kind of lung cancer is located in one of the major airway branches or centrally in the lung, where the bigger bronchi connect the trachea to the lung. About 30% of non-small cell lung cancers are squamous cell lung cancers, which are typically associated with smoking. and the standard CT-Scan pictures are in the final folder.

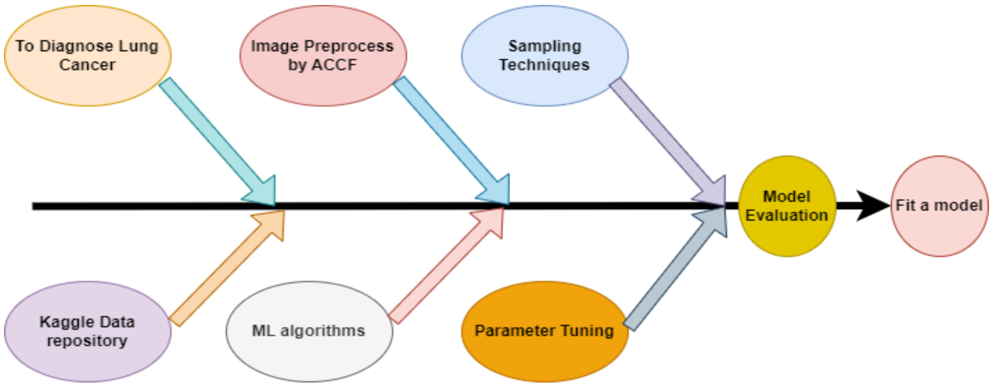

Figure 1: Proposed Architecture

The architecture shows the flow process of this research work. The collected dataset to be applied image filtering and features selection through learning models in weka 3.9.5 open-source tool by 10:90 sampling techniques.
This work considers following algorithms:

- Naïve Bayes (NB) is a method that calculates the posterior probability of each class based on the observable data. The predicted class is determined by selecting the class with the highest probability.
- Linear Regression (LR) is a statistical technique that models the connection between a dependent variable and one or more independent variables by fitting a linear equation to observed data.
- Additive Regression (AR) is an advanced technique that builds upon linear regression to predict non-linear interactions. It achieves this by mixing numerous additive components. Additive regression models the link between predictors and the response by considering the total of smooth functions of each individual predictor, rather than assuming a linear relationship.
- Naïve Bayes Multinomial (NBM) is a specialized version of the Naïve Bayes method that is tailored for text classification applications. It is especially suitable for situations where the characteristics are distinct and indicate the frequencies of words or occurrences of terms in a document.
- Logistic Regression is a statistical technique employed for binary classification tasks, where it predicts the likelihood that an instance belongs to a specific class. Logistic regression, despite its misleading name, is actually used for solving classification problems rather than regression difficulties.
- The Attribute Selected Classifier (ASC) is a method used to choose a subset of important characteristics from the original set. The goal is to enhance the efficiency and performance of a classifier.

Algorithm: ACCF with Hybrid ML Techniques
The ACC is an effective method for color-based image retrieval and can be useful for extracting color and texture features from CT images.

Input: Large cell carcinoma, Squamous cell carcinoma, Adenocarcinoma, Normal CT images
Output: Fit an efficient model for diagnosing lung cancer
Here's the updated algorithm with mathematical notation, including the ACC filter:
1. Data Representation:

Let $I = \{I_1, I_2, ..., I_□\}$ be the set of input CT images Let $Y = \{y_1, y_2, ..., y_□\}$ be the set of corresponding labels where $y_i \in$ {Large cell carcinoma, Squamous cell carcinoma, Adenocarcinoma, normal}

2. Auto Color Correlogram Filter: For each image I in the dataset:

ACC(I) = {γ^(k)_c(I)}(c∈C, k∈K) where:

- C is the set of quantized colors
- K is the set of distance values
- γ^(k)_c(I) is the probability of finding a pixel of color c at distance k from a pixel of the same color

Mathematically, γ^(k)_c(I) is defined as: γ^(k)_c(I) = Pr(p_2 ∈ I_c | p_1 ∈ I_c, $\|p_1 - p_2\|$ = k) where:

- I_c is the set of pixels with color c in image I
- p_1 and p_2 are pixels in I
- $\|p_1 - p_2\|$ is the distance between p_1 and p_2

3.Feature Extraction: X = ACC(I) = {ACC(I_1), ACC(I_2), ..., ACC($I_□$)}

4.Feature Selection (optional): X' = F(X), where F is the feature selection function

5.Data Split: (X_train, y_train), (X_test, y_test) = split(X', Y)

6.For each classifier:

a) Naive Bayes: P(y|x) = P(x|y)P(y) / P(x)

b) Linear Regression: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_□ x_□ + \varepsilon$

c) Additive Regression: $f(x) = f_0(x) + \beta_1 f_1(x) + \beta_2 f_2(x) + ... + \beta_□ f_□(x)$

d) Naive Bayes Multinomial: $P(y|x) = P(y) \prod_i P(x_i|y) / P(x)$

e) Logistic Regression: P(y=1|x) = 1 / (1 + e^(-z)) where $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_□ x_□$

f) Attribute Selected Classifier: X" = S(X'), y = C(X") where S is the attribute selection function and C is the chosen classifier.

7. Model Evaluation: Accuracy = (TP + TN) / (TP + TN + FP + FN) Precision = TP / (TP + FP) Recall = TP / (TP + FN) F1-score = 2 * (Precision * Recall) / (Precision + Recall)

8.Model Selection: $M *= argmax_M Evaluation_M etric(M)$

IV Outcome and Interpretations

This section focuses the outcome of ACCF + AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models. The above table 2 shows the accuracy, precision, recall, receiver operating characteristic curve (ROC) and precision recall curve (PRC) value of ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models.

Table 1: Classifiers Vs Classification Outcomes

S.No	Classifier	Accuracy	Precision	Recall	ROC	PRC
1	ACCF + AR	90.10%	0.90	0.89	0.97	0.97
2	ACCF + NB	86.52%	0.89	0.85	0.93	0.93
3	ACCF + LR	87.21%	0.88	0.86	0.94	0.93
4	ACCF + ASC	84.10%	0.85	0.83	0.84	0.78
5	ACCF + NBM	87.10%	0.84	0.86	0.94	0.93
6	ACCF + Logistic	84.87%	0.90	0.84	0.93	0.92

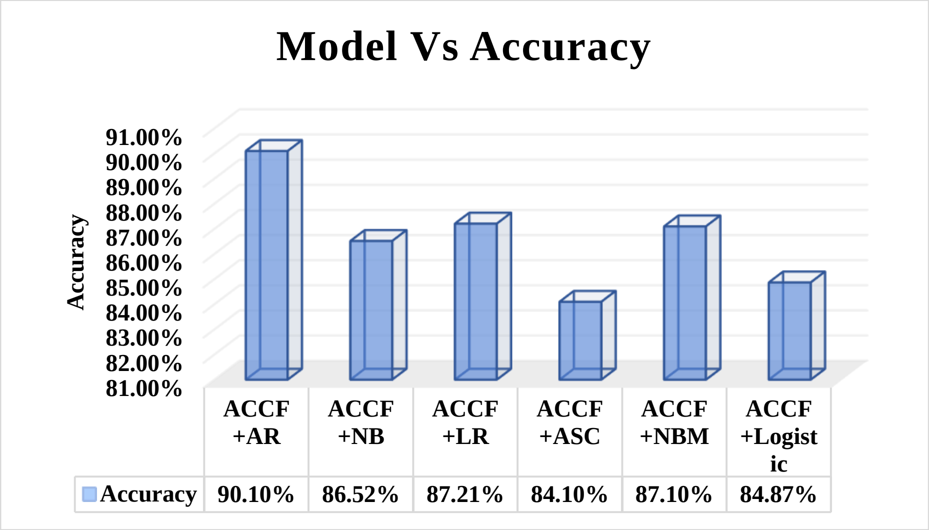

Figure 3: Model Vs Accuracy

The ACCF+AR model achieved the highest outcome with an accuracy of 90.10%, while the ACCF+ASC model had the lowest accuracy at 84.10%. The ACCF+Naive Bayes (NB) model achieves an accuracy of 86.52%, the ACCF+Logistic Regression (LR) model achieves an accuracy of 87.21%, the ACCF+Naive Bayes Multinomial (NBM) model achieves an accuracy of 87.10%, and the ACCF+logistic model has an accuracy of 84.87%.

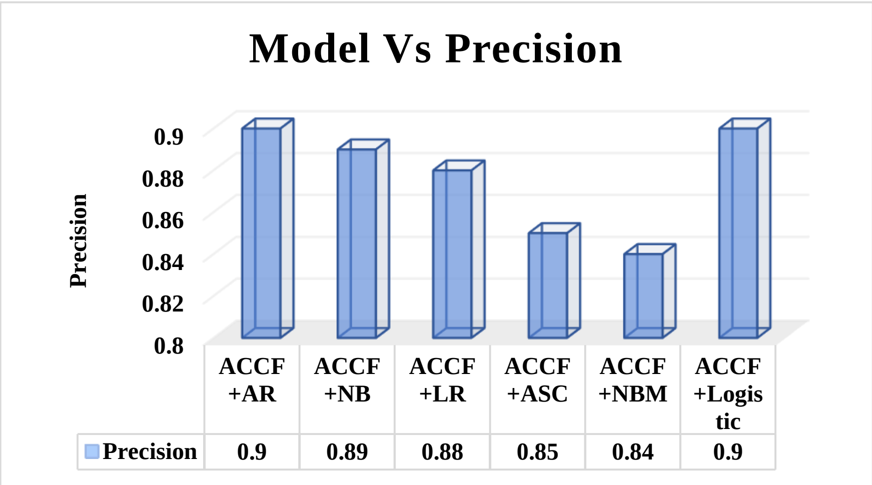

Figure 4: Model Vs Precision

The graphic above, labeled as graphic 4, illustrates the different levels of precision associated with ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models. The ACCF+AR and ACCF+Logistic models yield identical maximum outcomes with a precision of 0.87, while the ACCF+ASC model has the lowest precision at 0.81. The ACCF+Naive Bayes (NB) model has a precision of 0.86, the ACCF+Logistic Regression (LR) model has a precision of 0.85, the ACCF+Naive Bayes Multinomial (NBM) model has a precision of 0.83, and the logistic model has a precision of 0.87.

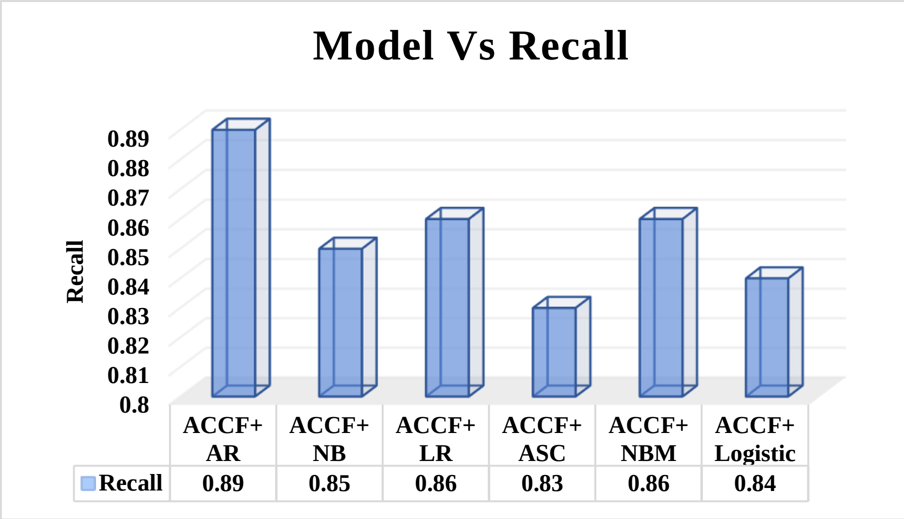

Figure 5: Model Vs Recall

The graphic above, labeled as graphic 5, illustrates the different levels of recall for the ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models. The ACCF+AR model achieves the highest memory rate of 0.89, while the ACCF+ASC model has the lowest remember rate of 0.80. The ACCF+Naive Bayes (NB) model has a recall of 0.85, the ACCF+Logistic Regression (LR) model has a recall of 0.86, the ACCF+Naive Bayes Multinomial (NBM) model has a recall of 0.83, and the logistic model has a recall of 0.84.

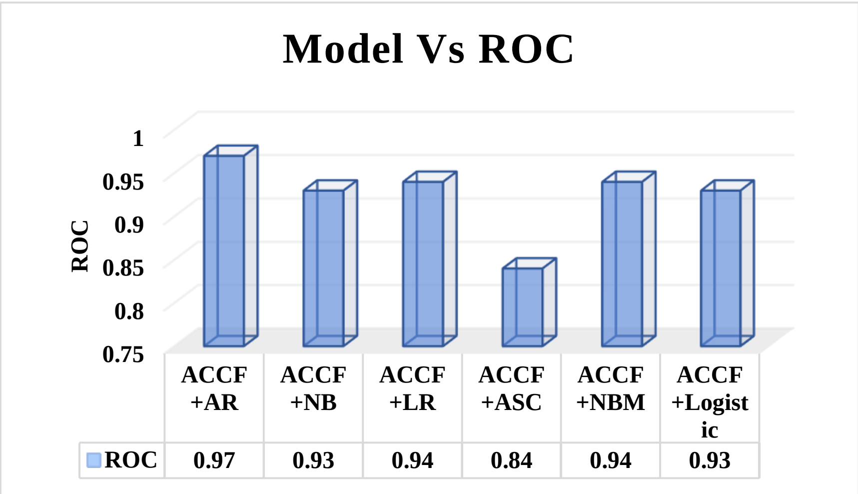

Figure 6: Model Vs ROC

The figure 6 above illustrates the ROC values of ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models. The ACCF+AR model has the highest outcome with a maximum ROC of 0.97, while the ACCF+ASC model has the lowest ROC of 0.84. The ACCF+NB and ACCF+Logistic models have an identical ROC value of 0.93. Similarly, the ACCF+LR model and the ACCF+NBM models both have a ROC value of 0.94 for recall.

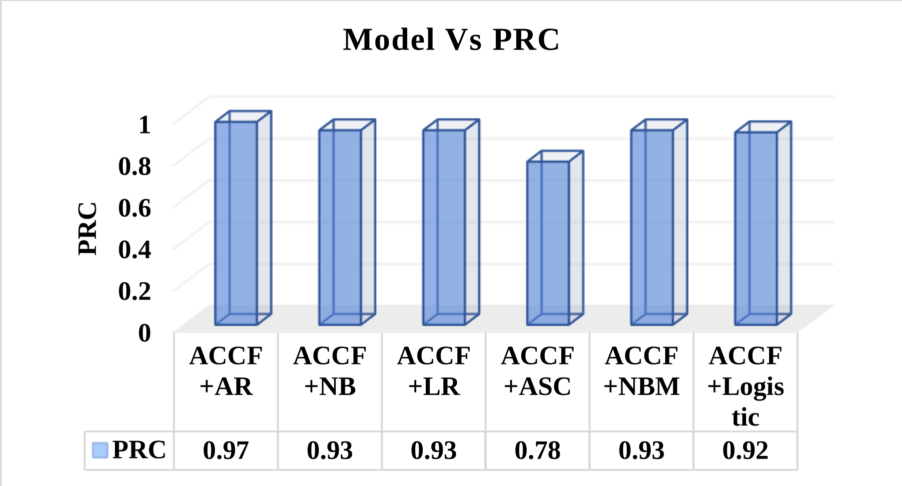

Figure 7: Model VsPRC

Figure 7 illustrates the different degrees of PRC for ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models. The ACCF+AR model achieves the highest outcome with a PRC of 0.97, while the ACCF+ASC model has the lowest PRC of 0.78. The ACCF+NB, ACCF+LR, and ACCF+NBM models all have a PRC value of 0.93, whereas the ACCF+Logistic model has a PRC value of 0.92.

Table 2: Classifiers Vs Statistical outcome

S.No	Classifier	Time	Kappa	F-Measure	MCC
1	ACCF + AR	0.21	0.65	0.89	0.65
2	ACCF + NB	0.15	0.61	0.85	0.61
3	ACCF + LR	0.14	0.61	0.86	0.61
4	ACCF + ASC	1.07	0.53	0.83	0.53
5	ACCF + NBM	3.17	0.58	0.34	0.58
6	ACCF + Logistic	0.14	0.61	0.83	0.6

The table 2 above illustrates the time consumption, Kappa, F-Measure, and Matthews Correlation Coefficient (MCC) values for the ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models.

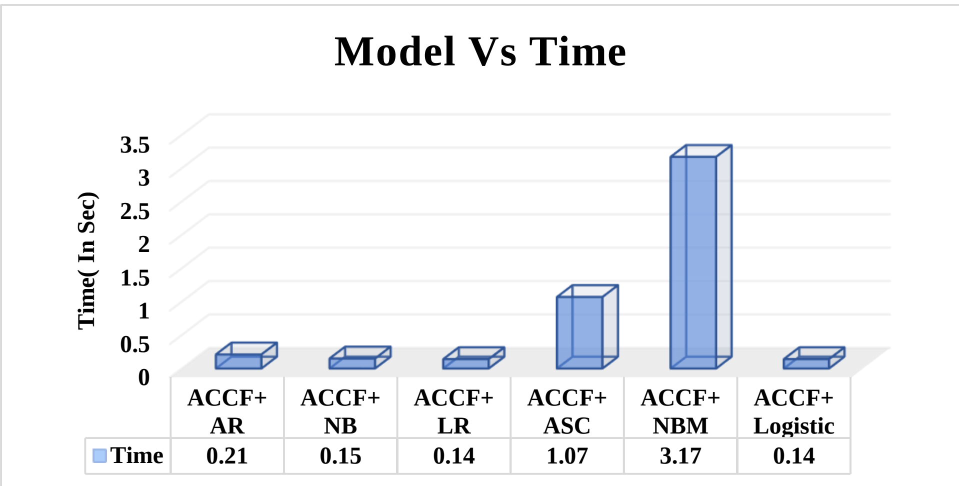

Figure 8: Model Vs Time

The chart above illustrates the time required, measured in seconds, to create models of ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic. Both the ACCF+LR and ACCF+Logistic models have a time consumption of 0.14 seconds for model creation, making them equally efficient in terms of time. The ACCF+NBM requires a longer time for constructing its model(3.17 seconds). The ACCF+NB model requires 0.15 seconds, the ACCF+AR model requires 0.21 seconds, and the ACCF+ASC model requires 0.21 seconds to generate their respective models.

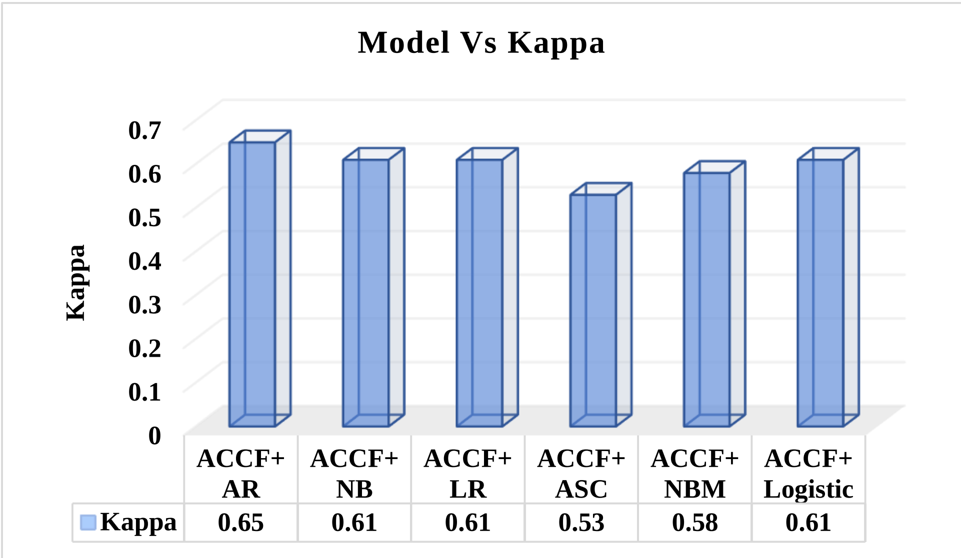

Figure 9: Model Vs Kappa

Figure 9 illustrates the different kappa levels of ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models. The ACCF+AR model exhibits the highest kappa value, with a kappa of 0.65, while the ACCF+ASC model has the lowest value of 0.53 compared to the other models. The ACCF+NB and ACCF+LR models both have a kappa value of 0.61, while the ACCF+NBM model has a kappa value of 0.58 and the ACCF+Logistic model has a kappa value of 0.61.

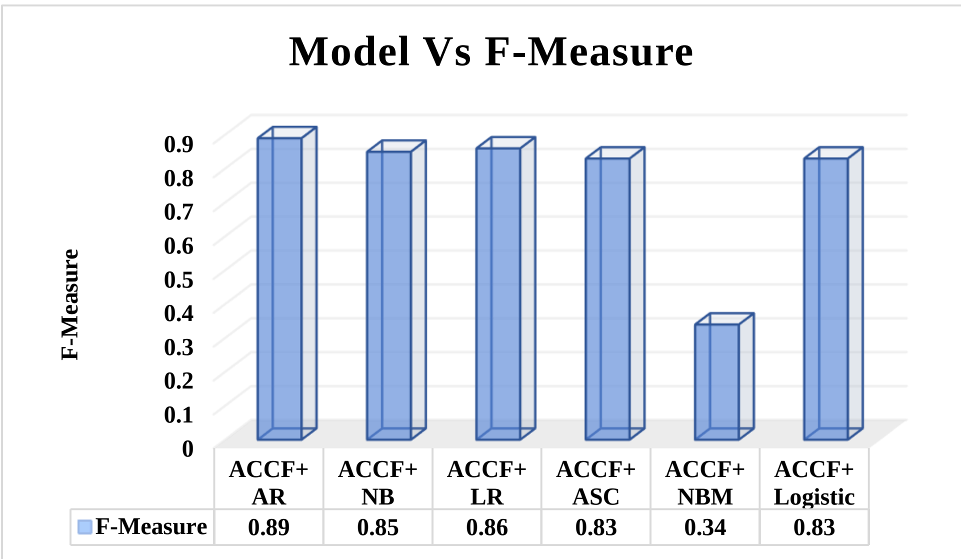

Figure 10: Model Vs F-Measure

The figure 10 above illustrates the different degrees of F-Measure for the ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models. The ACCF+AR model exhibits the highest F-Measure value among the other models, with a value of 0.89. In contrast, the ACCF+NBM model has the lowest F-Measure value of 0.34, which is the least among the other models. The ACCF+Naive Bayes (NB) classifier has an F-Measure of 0.85, while the ACCF+Logistic Regression (LR) classifier has an F-Measure of 0.86. Both the ACCF+ASC and ACCF+Logistic classifiers have an F-Measure value of 0.83.

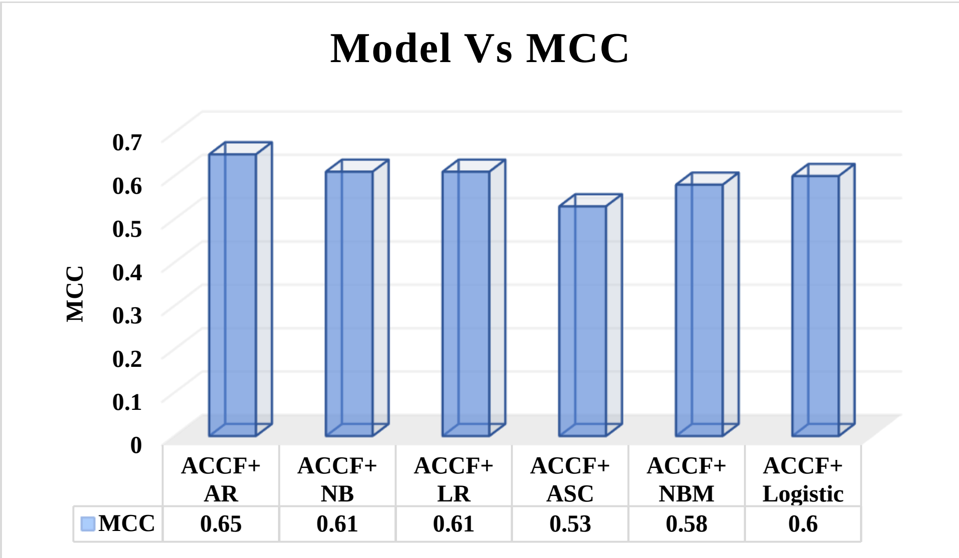

Figure 11: Model Vs MCC

The figure 11 above illustrates the different levels of MCC (Matthews Correlation Coefficient) for ACCF+AR, ACCF+NB (Naive Bayes), ACCF+LR (Logistic Regression), ACCF+ASC, ACCF+NBM (Naive Bayes Multinomial), and ACCF+Logistic models. The ACCF+AR model exhibits the greatest Matthews Correlation Coefficient (MCC) value among the other models, with a score of 0.65. On the other hand, the ACCF+ASC model has the lowest MCC value of 0.53, which is the least among all the models. The ACCF+NB and ACCF+LR models both have an MCC value of 0.61, while the ACCF+NBM model has an MCC value of 0.58 and the ACCF+Logistic model has an MCC value of 0.60.

Table 3: Classifiers Vs Errors

S.No	Classifier	MAE	RMSE	RAE	RRSE
1	ACCF + AR	0.24	0.33	53.00%	73.37%
2	ACCF + NB	0.40	0.4	88.86%	88.92%
3	ACCF + LR	0.32	0.35	71.80%	78.71%
4	ACCF + ASC	0.23	0.47	54.00%	103.28%
5	ACCF + NBM	0.25	0.37	58.00%	79.22%
6	ACCF + Logistic	0.23	0.45	52.74%	98.14%

The above table 3 depicts the Mean Absolute Error (MAE), Relative Absolute Error (RAE), Root Measure Squared Error (RMSE), and Relative Root Squared Error (RRSE) of ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models.

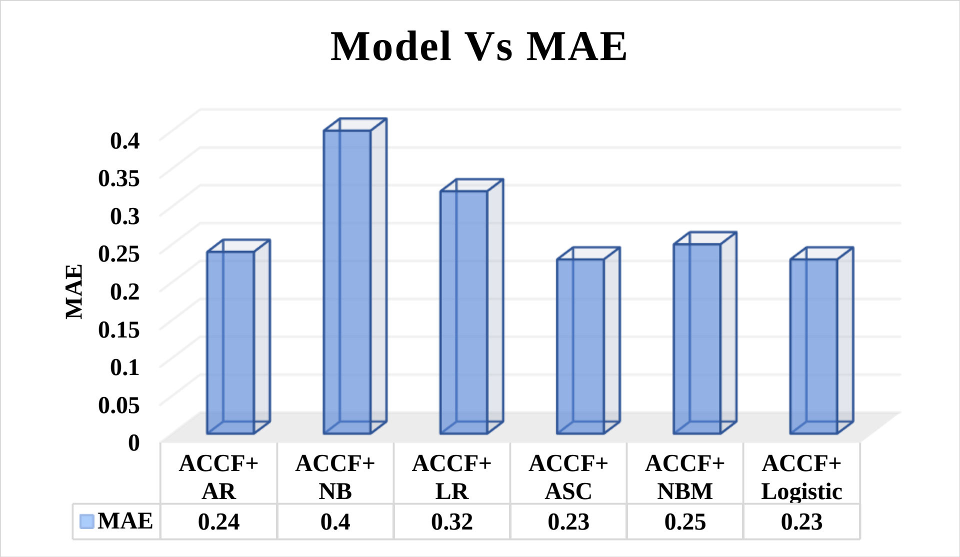

Figure 12: Model Vs MAE

The image 12 above illustrates the different Mean Absolute Error (MAE) levels of the ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models. The logistic model achieved a Mean Absolute Error (MAE) of 0.22, indicating the best performance. The ACCF+AR and ACCF+ASC have similar MAE values, with each having an MAE of approximately 0.24 and 0.23, respectively. The ACCF+NB exhibits the poorest performance, with a Mean Absolute Error (MAE) of 0.40. The ACCF+Linear Regression (LR) model has a Mean Absolute Error (MAE) of 0.32, whereas the ACCF+Naive Bayes Model (NBM) has an MAE of 0.25.

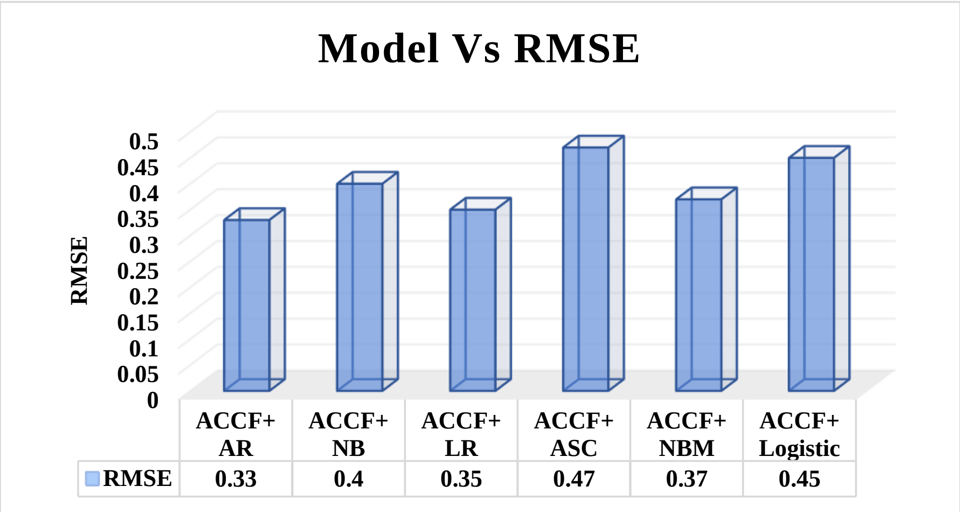

Figure 13: Classifiers Vs RMSE

Figure 13 illustrates the different degrees of root mean square error (RMSE) for the ACCF+AR, ACCF+NB, ACCF+ LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models. The AR model exhibits an RMSE of 0.33, indicating the lowest level of deviations when compared to other models. The ACCF+Naive Bayes (NB) model has a root mean square error (RMSE) of 0.40, the ACCF+Logistic Regression (LR) model has an RMSE of 0.35, and the ACCF+Naive Bayes Multinomial (NBM) model has an RMSE of 0.37. The ACCF+ASC exhibits the poorest performance with an RMSE of 0.47, while the ACCF+logistic model achieves an RMSE of 0.45.

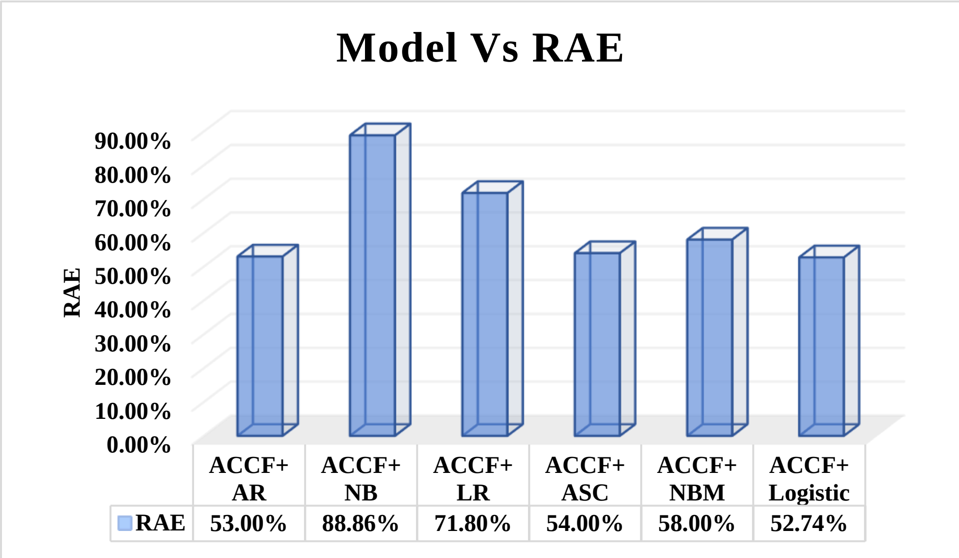

Figure 14: Classifiers Vs RAE

The picture 14 above illustrates the different levels of ACCF+RAE, ACCF+NB, ACCF+LR, ACCF+ASC, ACCF+NBM, and ACCF+Logistic models. The ACCF+NB model exhibits a RAE of 88.86%, indicating the poorest performance. The ACCF+Logistic model had a RAE (Relative Absolute Error) of 52.74%, indicating the best performance. On the other hand, the ACCF+LR (Linear Regression) model achieved a RAE of 71.80%. The ACCF+AR, ACCF+ASC, and ACCF+NBM have RAE percentages of 53%, 54%, and 58% respectively.

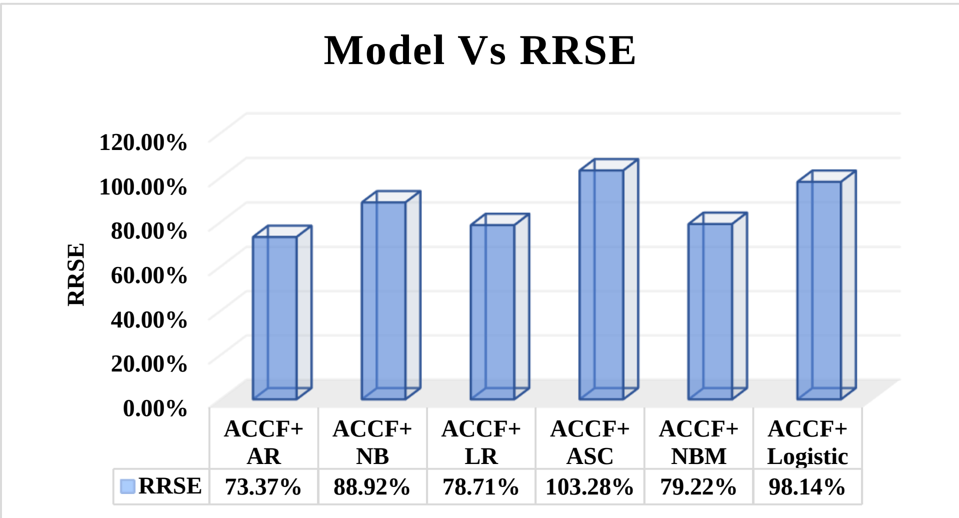

Figure 15: Classifiers Vs RRSE

The image 15 above illustrates the different degrees of RRSE (Root Relative Squared Error) for ACCF+AR, ACCF+NB, ACCF+LR, ACCF+ASC , ACCF+NBM, and ACCF+Logistic models. The ACCF+AR model exhibits an RRSE of 73.37%, indicating minimal variations. The ACCF+ASC exhibits an RRSE of 103.28%, which is the largest level of deviations when compared to other models. The ACCF+Naive Bayes (NB) model has an 88.92% Root Relative Squared Error (RRSE), the ACCF+Linear Regression (LR) model has a 78.71% RRSE, the

ACCF+Naive Bayes Multinomial (NBM) model has a 79.22% RRSE, and the ACCF+Logistic model has a 98.14% RRSE.

4. Conclusion

This research work shows that the ACCF+logistic has 0.19 MAE which is best performance. The ACCF+AR and ACCF+ASC has more or less same MAE value 0.21 MAE and 0.20 MAE. The ACCF+AR model has 0.30 RMSE which is least deviations compare with other models. The ACCF+ASC has worst performance which is 0.44 RMSE and the ACCF+logistic has 0.42 RMSE. The ACCF+NB model has 85.86% RAE which is worst performance. The ACCF+Logistic has 49.74% RAE which is best performance and the ACCF+LR model has 68% RAE. The ACCF+AR model has 70.37% RRSE which is having least deviations. The ACCF+ASC has 100.28% RRSE which is highest deviations compare with other models. The ACCF+AR model has least deviation and it has best efficiency compare with other models. So that this work explores that the ACCF+AR deductive learning model is best for diagnosing lung cancer.

5. References

1. Javed, R., Abbas, T., Khan, A.H. et al. Deep learning for lungs cancer detection: a review. Artif Intell Rev 57, 197 (2024). https://doi.org/10.1007/s10462-024-10807-1
2. Wankhade, S., Vigneshwari, S. Lung cell cancer identification mechanism using deep learning approach. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08661-4
3. Ahmed, Bakhan. (2019). Lung Cancer Prediction and Detection Using Image Processing Mechanisms: An Overview. Signal and Image Processing Letters. 1. 10.31763/simple.v1i3.11.
4. Liu X, Li K-W, Yang R, Geng L-S (2021) Review of deep learning based automatic segmentation for lung cancer Radiotherapy. Front Oncol 11. https://doi.org/10.3389/fonc.2021.717039
5. Jiang W, Zeng G, Wang S et al (2022) Application of Deep Learning in Lung Cancer Imaging Diagnosis. J Healthc Eng 2022:1–12. https://doi.org/10.1155/2022/6107940
6. Chae KJ, Jin GY, Ko SB et al (2020) Deep Learning for the Classification of Small (≤2 cm) Pulmonary Nodules on CT Imaging: A Preliminary Study. Acad Radiol 27:e55–e63. https://doi.org/10.1016/j. acra.2019.05.018
7. Manjula Devi R, Dhanaraj RK, Pani SK et al (2023) An improved deep convolutionary neural network for bone marrow cancer detection using image processing. Inf Med Unlocked 101233. https://doi.org/10. 1016/j.imu.2023.101233
8. Ali F, Kumar H, Patil S et al (2022) Target-DBPPred: An intelligent model for prediction of DNA-binding proteins using discrete wavelet transform based compression and light eXtreme gradient boosting. Comput Biol Med 145:105533–105533. https://doi.org/10. 1016/j.compbiomed.2022.105533
9. 10.Masood I, Wang Y, Daud A et al (2018) Towards Smart Healthcare: Patient Data Privacy and Security in Sensor Cloud Infrastructure. Wirel Commun Mob Comput 2018:1–23. https://doi.org/10.1155/2018/2143897
10. Javed R, Abbas T, Jamshaid Iqbal Janjua et al (2023) wrist fracture prediction using transfer learning, a case study. J Popul Ther Clin Pharmacol 30. https://doi.org/10.53555/jptcp.v30i18.3161
11. Jamshaid Iqbal Janjua, Tahir Abbas Khan, Nadeem M (2022) Chest x-ray anomalous object detection and classification framework for medical diagnosis. 2022 International

conference on information networking (ICOIN). https://doi.org/10.1109/icoin53446.2022.9687110
12. Abbas S, Issa GF, Fatima A et al (2023) Fused Weighted Federated Deep Extreme Machine Learning Based on Intelligent Lung Cancer Disease Prediction Model for Healthcare 5.0. Int J Intell Syst 2023:1–14. https://doi.org/10.1155/2023/2599161
13. Shah, A.A., Malik, H.A.M., Muhammad, A. et al. Deep learning ensemble 2D CNN approach towards the detection of lung cancer. Sci Rep 13, 2987 (2023). https://doi.org/10.1038/s41598-023-29656-z
14. Ghasemi Darehnaei, Z., Shokouhifar, M., Yazdanjouei, H. & Rastegar Fatemi, S. M. J. SI-EDTL: Swarm intelligence ensemble deep
15. transfer learning for multiple vehicle detection in UAV images. Int. J. Commun. Syst. https://doi.org/10.1002/cpe.6726 (2022).
16. Zuo, W., Zhou, F., Li, Z. & Wang, L. Multi-resolution cnn and knowledge transfer for candidate classifcation in lung nodule detection. IEEE Access 7, 32510–32521 (2019).
17. Setio, A. A. A. et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks.IEEE Trans. Med. Imaging 35, 1160–1169 (2016).
18. 12. Xie, Y. et al. Knowledge-based collaborative deep learning for benign-malignant lung nodule classifcation on chest CT. IEEETrans. Med. Imaging 38, 991–1004 (2019).
19. Rao, G. S., Kumari, G. V., & Rao, B. P. Network for biomedical applications. vol. 2 (Springer Singapore, 2019).
20. Wang, W. et al. Exploring cross-image pixel contrast for semantic segmentation. In Proceedings of the. IEEE Int. Conf. Comput. Vis.7283–7293 (2021) doi:https://doi.org/10.1109/ICCV48922.2021.00721.
21. Ramchoun, H., Amine, M., Idrissi, J., Ghanou, Y. & Ettaouil, M. Multilayer perceptron: Architecture optimization and training.Int. J. Interact. Multimed. Artif. Intell. 4, 26 (2016).
22. Berwick, R. An Idiot's Guide to Support vector machines (SVMs): A New Generation of Learning Algorithms Key Ideas. Village Idiot 1–28 (2003).
23. Faisal, M. I., Bashir, S., Khan, Z. S. & Hassan Khan, F. An evaluation of machine learning classifers and ensembles for early stage prediction of lung cancer. In 2018 3rd International Conference on Emerging Trends Engineering Science Technology. ICEEST 2018 1–4 (2019). https://doi.org/10.1109/ICEEST.2018.8643311.
24. N. Bharanidharan et al., "Local Entropy Based Remora Optimization and Sparse Autoencoders for Cancer Diagnosis Through Microarray Gene Expression Analysis," in IEEE Access, vol. 12, pp. 39285-39299, 2024, doi: 10.1109/ACCESS.2024.3371887.
25. R.Sugumar , Dr.E.Mohan. "Magnetic Resonance Imaging Segmentation for Brain Tumor Detection Using New Robust Global Kernel Fuzzy C-Means Clustering Algorithm (NRGKFCM-F)," International Journal of Applied Engineering Research – Volume 9 ,issue 21, 10889-10908, 2014, (ISSN: 0973-4562).
26. Thambu Gladstan , Dr.E.Mohan. "Object Recognition Based on Wave Atom Transform," Research Journal of Applied Sciences, Engineering and Technology – Volume 8 ,issue 13, 1613-1617, 2014, ISSN: 2040-7459; e-ISSN: 2040-7467
27. Dr.E.Mohan, Dr.A.Annamalai Giri,S.V.AswinKumer "A Novel Image Segmentation Approach for Brain Tumor Detection Using Dual Clustering Approach" International Journal of Applied Engineering Research, Volume 13 ,issue 11, 9807-9810, 2018, (ISSN: 0973-4562).
28. https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images