# Bone Fracture Classification from X-Ray Images using Machine Learning

**Dr.K.S.Kalaivani,** Assistant Professor (SrG), Dept. of Artificial Intelligence. Kongu Engineering College, Erode, Tamil Nadu.

**Mohamed Bilal Z H [21ALR027],** Student, Dept. of Artificial Intelligence. Kongu Engineering College, Erode, Tamil Nadu. [mohamedbilalzh.21aim@kongu.edu](mailto:mohamedbilalzh.21aim@kongu.edu)

**Sukesh G [21ALR048],** Student, Dept. of Artificial Intelligence. Kongu Engineering College, Erode, Tamil Nadu.[sukeshg.21aim@kongu.edu](mailto:sukeshg.21aim@kongu.edu)

**Yokeswaran S [21ALR056],** Student, Dept. of Artificial Intelligence. Kongu Engineering College, Erode, Tamil Nadu.[yokeswarans.21aim@kongu.edu](mailto:yokeswarans.21aim@kongu.edu)

**ABSTRACT**

Bone fracture classification plays a crucial role in orthopedics for accurate diagnosis and treatment planning. Traditional methods rely on manual assessment, which can be time-consuming and subjective. In this study, we propose a machine learning-based approach for automated bone fracture classification from X-ray images. The objective is to develop a reliable and efficient system that can assist clinicians in the classification process. The study begins with a comprehensive literature survey, exploring existing research on bone fracture classification and machine learning in medical image analysis. Based on the findings, a methodology is devised, comprising data pre-processing, Hyperparameter Tuning, and model training using various machine learning algorithms. Performance evaluation is conducted using appropriate metrics such as accuracy, precision, recall, and F1-score. The results are compared across different algorithms to identify the most effective approach for bone fracture classification. Limitations and potential improvements are discussed, along with the impact of the proposed system on orthopedic practice. The study concludes that machine learning-based bone fracture classification shows promising results, offering automation and accuracy in the classification process. The proposed system has the potential to improve efficiency and provide valuable assistance to orthopedic practitioners. Further research and development in this area can lead to enhanced fracture diagnosis and treatment planning.

*Keywords* Bone Fracture, Classification, X-Ray images, Machine Learning, Logistic Regression, Random Forest, K-Nearest Neighbors, Decision Tree, Gaussian NB, SVM, Multinomial NB, MLP, Extra Tree

**I. INTRODUCTION**

Bone fracture classification from X-ray images using machine learning is an emerging field in medical image analysis and orthopedics. X-ray imaging is a widely used diagnostic tool for identifying and assessing bone fractures. Traditionally, fracture classification has relied on manual interpretation by radiologists or orthopedic specialists, which can be subjective and time-consuming. With the advancements in machine learning algorithms and techniques, there is an opportunity to automate and enhance the fracture classification process. Machine learning models can be trained on large datasets of labelled X-ray images to learn patterns and features indicative of different types of fractures. These models can then be used to classify new X-ray images, assisting clinicians in accurate fracture diagnosis and treatment planning. The application of machine learning in bone fracture classification has the potential to improve the efficiency and reliability of fracture assessment. It can reduce the dependence on human expertise and provide consistent and objective results. Additionally, machine learning models can be continuously improved through iterative training on new data, leading to enhanced classification accuracy over time. By leveraging the power of machine learning, bone fracture classification from X-ray images aims to enhance clinical decision-making, improve patient outcomes, and streamline the workflow in orthopedic practice. Ongoing research and development in this field are focused on optimizing the performance of machine learning models, exploring novel algorithms, and addressing challenges such as dataset imbalance and generalization to diverse patient populations.

**II. LITERATURE SURVEY**

Machine Learning is an emerging field nowadays, many medical field problems can be resolved using ML techniques. One of the problems is Bone Fracture Classification, This subject has piqued the interest of researchers, and several people are currently working on it.

[1] Machine learning and artificial intelligence play a vital role in clinical imaging. Image pre-processing tools such as noise removal and contrast enhancement are used to enhance the medical images quality. Gray Level Co-occurrence Matrix (GLCM) texture features are used to classify bone X-ray images into fractures and not fracture categories. Performance evaluation of abnormality detection in X-ray images is performed using five statistical parameters such as Sensitivity, Specificity, Precision, Accuracy and F1 Score.

[2] In this paper introduces the image classification paradigm and compares different feature extraction and classification algorithms. It provides background information on machine learning, the Bag of Features paradigm, the SURF detector, BoW models, and experimentations.

[3] Juan W proposed a hyperspectral remote sensing image classification method based on the improved optimal exponential factor (OIF) and support vector machine (SVM) algorithm. The experimental results showed that the SVM algorithm can effectively obtain the optimal classification band combination with high classification accuracy. Zhao L's experiment shows that improved KNN can achieve higher classification accuracy in high-resolution remote sensing image classification. Deep learning has powerful functions and flexibility, and this paper compares SVM and CNN algorithms from different perspectives. Different data sets have different effects on the experimental results of image classification.

[4] Morphological summaries with subjective restoration and steering MPs are used to classify excessive decision hyperspectral snap shots from city areas, while supervised face extraction is used to reduce dimensionality.

[5] This paper proposes a semi supervised learning based image classification method, which uses a small amount of labeled pathological image data to train the network model and integrates the features extracted by the network to classify the image. The results show that the classification effect of the neural network is better than convolution neural network and other traditional image classification models, and the overall classification accuracy and kappa coefficient are increasing with the increasing number of training samples.

**III. METHODOLOGY**

The Image dataset used for this process is available in Kaggle which contains 9463 Images and of two class labels i.e. Fractured and Not Fractured. Since, the experiments are done using machine learning models, the images are transformed into vectors [or] arrays using their size and samples and then stored. Even though transforming them into vectors it will be uneven, So pre-processing methods i.e. scaling used. The experiments are conducted using MaxAbsScaler since it is the most appropriate scaling technique for this Image Dataset. Next comes the important part Hyperparameter Tuning

The Hyperparameter Tuning is the most important part for model performance, experiments are conducted for several parameters to identify the optimal one. Each algorithm has its own parameter needed to be tuned for the optimal model. For this process the utilization of GridSearchCV is needed for Hyperparmeter Tuning. Since the dataset contains 9463 images each of which 4840 is Fractured class and 4623 is not fractured class, it is a hectic process to use all these data for tuning the model. So the experiments are conducted using evenly distributed counts of each classes for training and testing i.e. 100 images for each, therefore 400 images are used for Hyperparameter Tuning for each model and it significantly increases the model accuracy after Hyperparameter Tuning. The Sample Hyperparameter tuning for Support Vector Machine description is given below.

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Search Space** | **Description** |
| C | float | Regularization parameter. |
| class\_weight | dict , balanced | The “balanced” mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data. |
| gamma | Scale, auto, float | Kernel coefficient for ‘rbf’, ‘poly’ and ‘sigmoid’. |
| decision\_function\_shape | ovo, ovr | Whether to return a one-vs-rest (‘ovr’) decision function of shape (n\_samples, n\_classes) as all other classifiers, or the original one-vs-one (‘ovo’) decision function of libsvm which has shape (n\_samples, n\_classes \* (n\_classes - 1) / 2). |
| kernel | linear, poly, rbf, sigmoid, precomputed, callable | Specifies the kernel type to be used in the algorithm. |

Table 1. Hyperparameters and their search space

Now, visualization of the classes count can be done using the matplotlib module to check if the classes are evenly distributed or not.

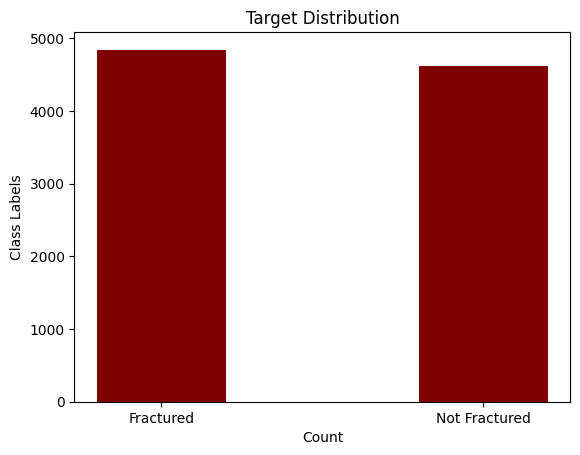


Fig 1. Target Distribution

From Fig 1, it can be visualized that there is some uneven distribution but since it is a small deviation it will not affect the models. The depiction of images with class labels is important for better understand of the dataset, so it is visualized and depicted how it works.



Fig2. Examples of Images from the Dataset

Now, the formulation of proposed model is made that how the model will be trained and evaluated, it contains of three significant steps Image to vector conversion, MaxAbsScaler and Hyperparameter tuning. It is depicted in Fig 3.

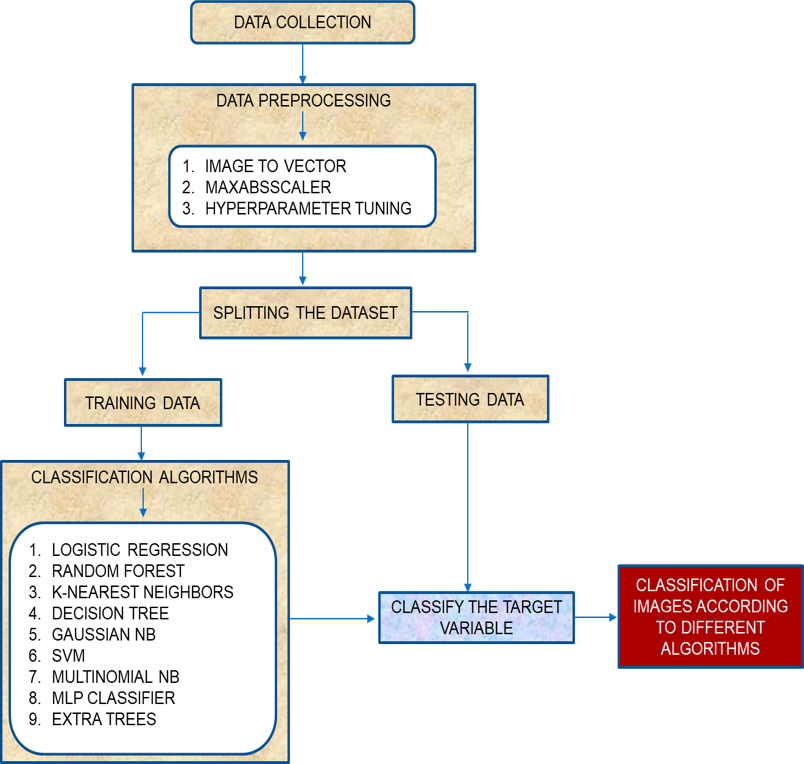


Fig3. Proposed Model

**IV. ERROR METRICS**

For this dataset, to determine the accuracy and error, confusion matrix and classification report has been utilized i.e. precision, recall, F1 score and accuracy. The formula used to determine the mentioned error metrics.

1. Classification Report

It is one of the performance evaluation metrics of a classification-based machine learning model. It displays our model’s precision, recall, F1 score and support. It provides a better understanding of the overall performance of our trained model. To understand the classification report of a machine learning model, we need to know all of the metrics displayed in the report. For a clear understanding, we have explained all of the metrics below so that you can easily understand the classification report of your machine learning model:

| **Metrics** | **Definition** |
| --- | --- |
| **Precision** | Precision is defined as the ratio of true positives to the sum of true and false positives. |
| **Recall** | Recall is defined as the ratio of true positives to the sum of true positives and false negatives. |
| **F1 Score** | The F1 is the weighted harmonic mean of precision and recall. The closer the value of the F1 score is to 1.0, the better the expected performance of the model is. |
| **Support** | Support is the number of actual occurrences of the class in the dataset. It doesn’t vary between models, it just diagnoses the performance evaluation process. |

Table 2. Formulae of different metrics used

2. Confusion Matrix

Classification Models have multiple categorical outputs. Most error measures will calculate the total error in our model, but we cannot find individual instances of errors in our model. The model might misclassify some categories more than others, but we cannot see this using a standard accuracy measure.

Furthermore, suppose there is a significant class imbalance in the given data. In that case, i.e., a class has more instances of data than the other classes, a model might predict the majority class for all cases and have a high accuracy score; when it is not predicting the minority classes. This is where confusion matrices are useful.

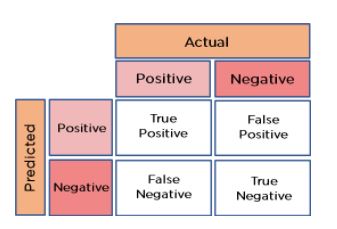


Fig3. Details about Confusion Matrix [2x2]

**•** True Positive: The number of times our actual positive values are equal to the predicted positive. You predicted a positive value, and it is correct.

• False Positive: The number of times our model wrongly predicts negative values as positives. You predicted a negative value, and it is actually positive.

• True Negative: The number of times our actual negative values are equal to predicted negative values. You predicted a negative value, and it is actually negative.

• False Negative: The number of times our model wrongly predicts negative values as positives. You predicted a negative value, and it is actually positive.

**V. PERFORMANCE EVALUATION**

For this dataset, the images are converted into vectors [or] arrays, the data are uneven so scaling is used i.e. MaxAbsScaler is utilized, Since the deployment is using only Machine Learning the models must be at its optimal one. So Hyperparameter tuning for each model is used and the errors and accuracy is calculated is using confusion matrix and classification report.

**A. LOGISTIC REGRESSION**

Logistic regression is a machine learning algorithm used to predict the output of a categorical dependent variable using a given set of independent variables. It is similar to Linear Regression in that it predicts two maximum values (0 or 1). Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets. It can be used to classify observations using different types of data and can easily determine the most effective variables used for the classification.

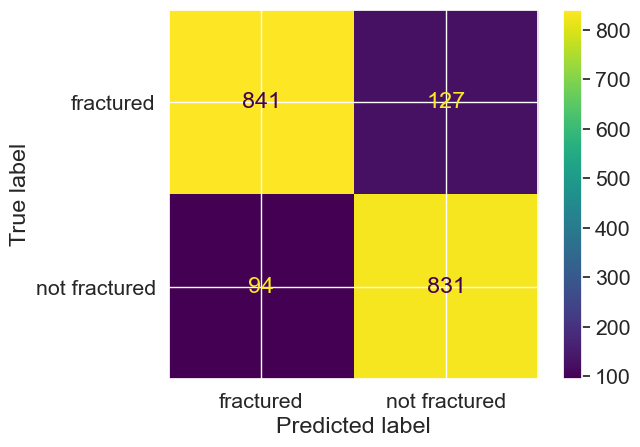


Fig4. Confusion Matrix [Logistic Regression]

From the above Fig4, concluded about the True Positive, True Negative, False Positive, False Negative of Logistic Regression.

**B. RANDOM FOREST**

Random Forest is a popular machine learning algorithm used for both Classification and Regression problems. It is based on ensemble learning, which is a process of combining multiple classifiers to improve the performance of the model. It takes the prediction from each tree and based on the majority votes of predictions, and predicts the final output.

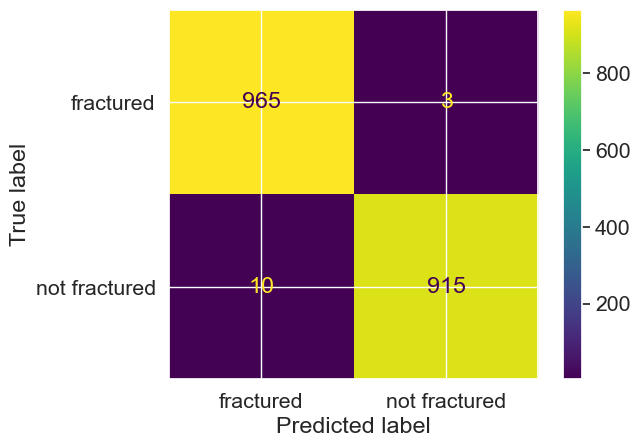


Fig5. Confusion Matrix [Random Forest]

From the above Fig5, concluded about the True Positive, True Negative, False Positive, False Negative of Random Forest.

**C. K-NEAREST NEIGHBORS**

K-Nearest Neighbour is a simple Machine Learning algorithm based on Supervised Learning technique. It assumes the similarity between the new case/data and existing cases and classifies a new data point based on the similarity. It can be used for Regression and Classification, but mostly it is used for Classification problems. K-NN is a non-parametric algorithm that does not make any assumptions on underlying data and is also called a lazy learner algorithm. It stores the dataset and when it gets new data, it classifies that data into a category that is much similar to the new data.

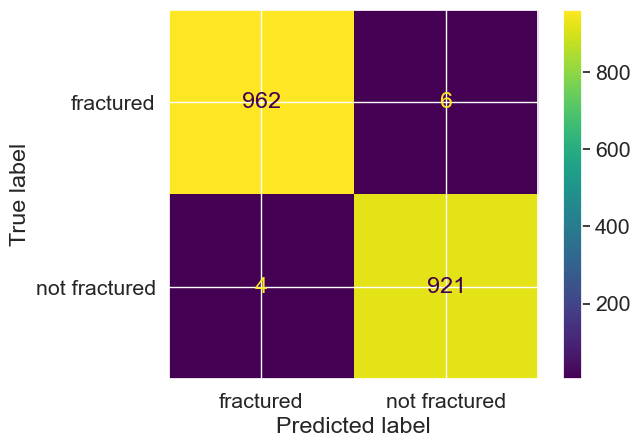


Fig6. Confusion Matrix [K-Nearest Neighbors]

From the above Fig6, concluded about the True Positive, True Negative, False Positive, False Negative of K-Nearest Neighbors.

**D. DECISION TREE**

Decision Tree is a Supervised learning technique used for both classification and Regression problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. There are two nodes, the Decision Node and Leaf Node, which are used to make any decision and have multiple branches. The decisions or the test are performed on the basis of features of the given dataset and it is called a decision tree because it starts with the root node, which expands on further branches and constructs a tree-like structure. To build a tree, the CART algorithm is used.

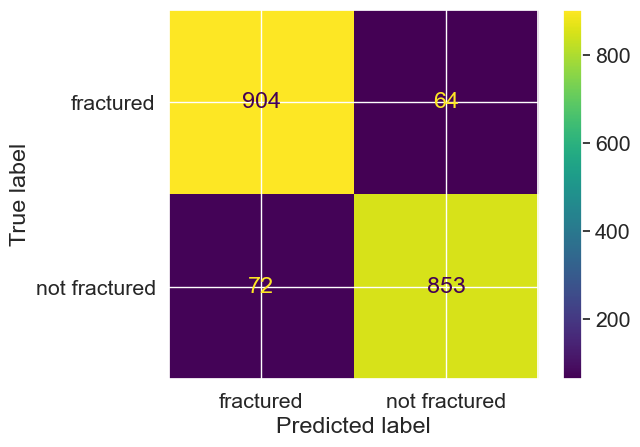


Fig7. Confusion Matrix [Decision Tree]

From the above Fig7, concluded about the True Positive, True Negative, False Positive, False Negative of Decision Tree.

**E. GAUSSIAN NB**

Naive Bayes is a probabilistic machine learning algorithm that can be used in classification tasks such as document classification, filtering spam, and prediction. It is based on the discoveries of Thomas Bayes and is simple yet powerful. It is based on the probabilistic model and can be coded easily and predictions done quickly in real-time. Before diving into Nave Bayes and Gaussian Nave Bayes, it is important to understand conditional probability.

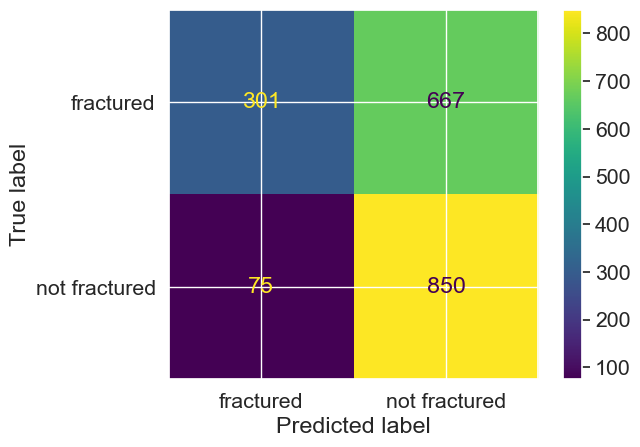


Fig8. Confusion Matrix [Gaussian NB]

From the above Fig8, concluded about the True Positive, True Negative, False Positive, False Negative of Gaussian NB.

**F. SUPPORT VECTOR MACHINE**

SVM is a popular Supervised Learning algorithm used for Classification and Regression problems. It creates the best line or decision boundary to segregate n-dimensional space into classes, known as a hyperplane. SVM chooses extreme points/vectors to create the hyperplane, which are called support vectors.

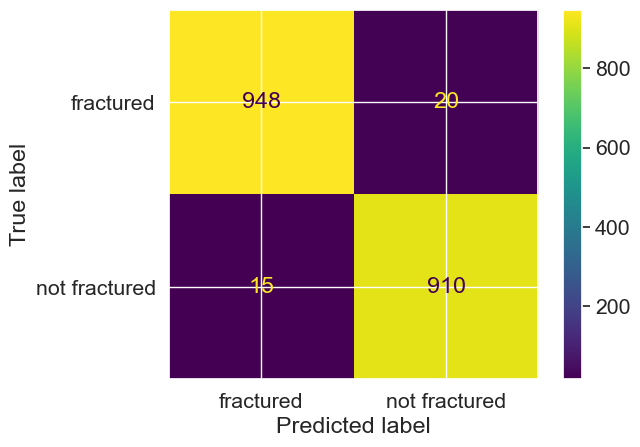


Fig9. Confusion Matrix [Support Vector Machine]

From the above Fig9, concluded about the True Positive, True Negative, False Positive, False Negative of Support Vector Machine.

**G. MULTINOMIAL NB**

Multinomial Naive Bayes algorithm is a probabilistic learning method used in NLP. It predicts the tag of a text based on the Bayes theorem and calculates the probability of each tag for a given sample. Naive Bayes classifier is a collection of many algorithms that share one common principle: each feature being classified is not related to any other feature.

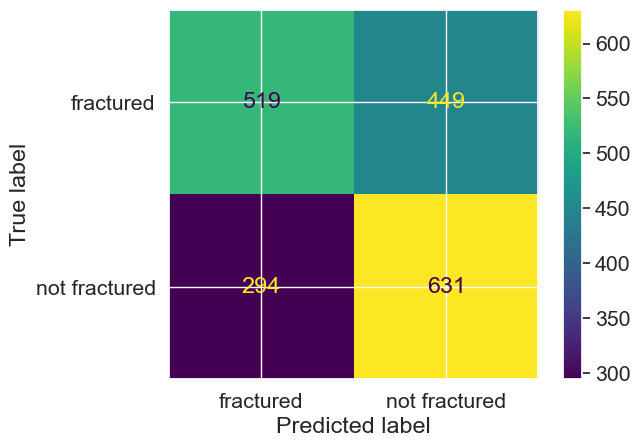


Fig10. Confusion Matrix [Multinomial NB]

From the above Fig10, concluded about the True Positive, True Negative, False Positive, False Negative of Multinomial NB.

**H. MULTI-LAYER PERCEPTRON CLASSIFIER**

The multi-layer perceptron (MLP) classifier is a popular type of artificial neural network used for supervised learning tasks, especially in the field of deep learning. It consists of multiple layers of interconnected nodes, called neurons, arranged in a feed-forward manner. Each neuron applies a non-linear activation function to its input and passes the result to the next layer. The MLP classifier is capable of learning complex patterns and relationships in the data, making it suitable for tasks such as image recognition, natural language processing, and speech recognition.

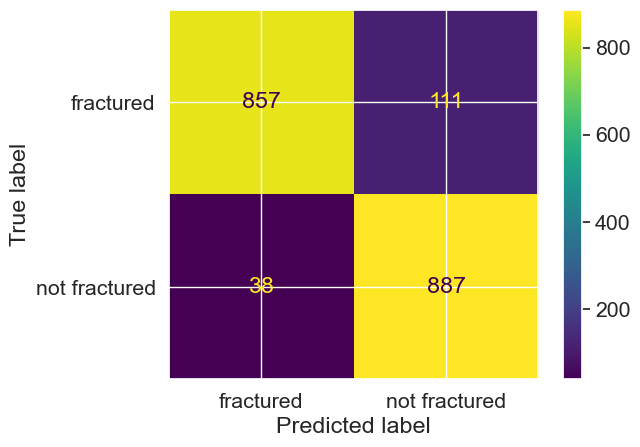


Fig11. Confusion Matrix [Multi-Layer Perceptron Classifier]

From the above Fig11, concluded about the True Positive, True Negative, False Positive, False Negative of Mult-Layer Perceptron Classifier.

**I. EXTRA TREES**

Extra Trees Classifier is an ensemble learning technique that aggregates the results of multiple de-correlated decision trees collected in a "forest" to output its classification result. Each tree is constructed from the original training sample and provided with a random sample of k features from which each tree must select the best feature to split the data. The Gini Importance of each feature is computed and each feature is ordered in descending order according to the Gini Importance.

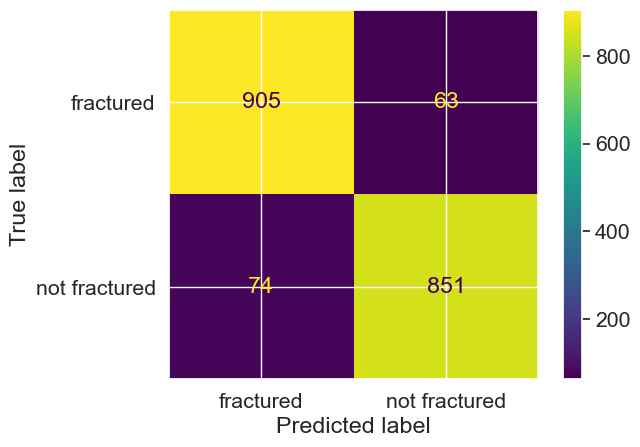


Fig12. Confusion Matrix [Extra Trees]

From the above Fig12, concluded about the True Positive, True Negative, False Positive, False Negative of Extra Trees.

**VI. RESULT AND DISCUSSION**

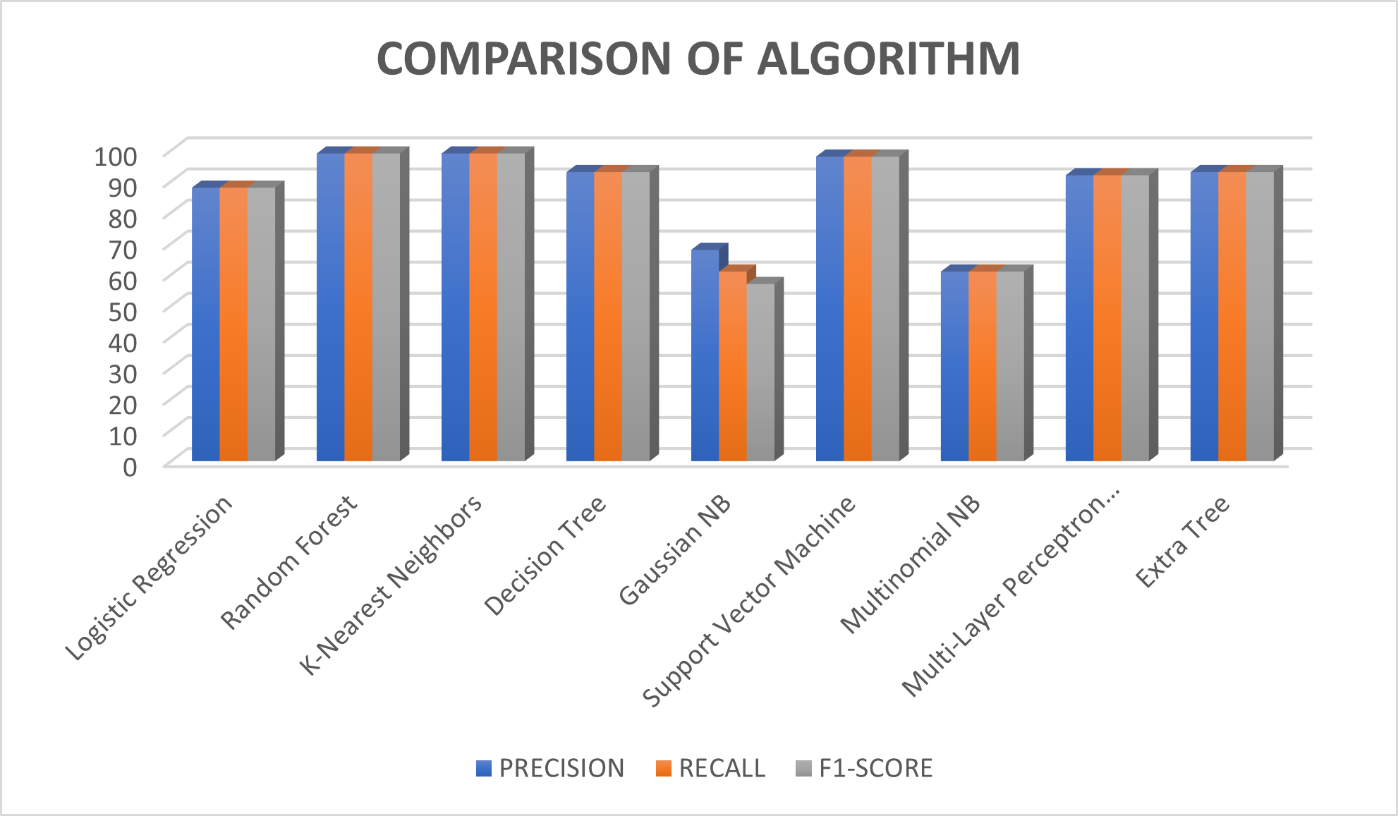
****

Fig13. Visualization of Classification Report

From Fig 13, the errors of various algorithms and their percentage are visualized. The Precision, Recall, F1-score are depicted in the above figure

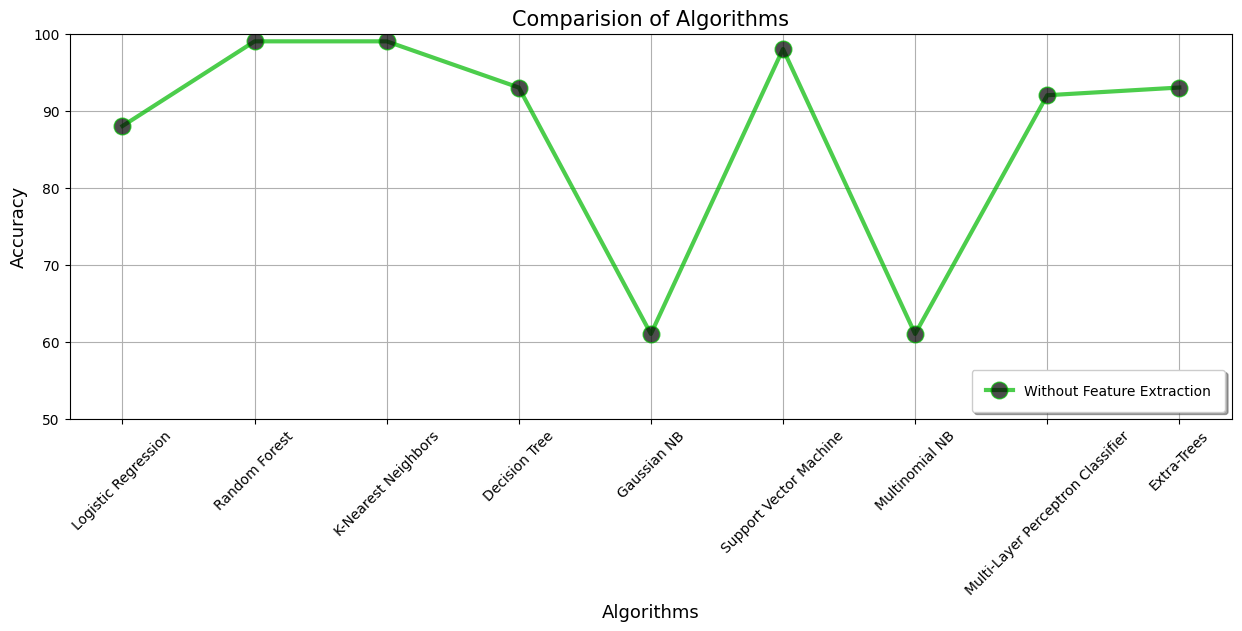


Fig 14. Analysis of accuracy of several algorithms

**V. CONCLUSION**

In conclusion, machine learning has shown great potential in bone fracture classification, revolutionizing the field of medical imaging diagnostics. By utilizing large datasets and advanced algorithms, machine learning models can accurately identify and classify different types of bone fractures. This technology enables faster and more objective diagnoses, aiding healthcare professionals in treatment planning and decision-making. Machine learning algorithms can detect subtle patterns and features in medical images that might be missed by human observers, improving overall diagnostic accuracy. Moreover, the integration of machine learning in fracture classification can help alleviate the burden on radiologists, allowing them to focus on more complex cases. With further advancements and refinement, machine learning has the potential to significantly enhance fracture diagnosis and contribute to better patient care. Utilization of machine learning has shown that the bone fracture classification according to the experiments it can be concluded that the Random Forest and K-Nearest Neighbors are the optimal model which gives an accuracy of 99%, followed by Support Vector Machine which gives an accuracy of 98% respectively.

**REFERENCES**

1. Mall, Pawan Kumar, Pradeep Kumar Singh, and Divakar Yadav. "Glcm based feature extraction and medical x-ray image classification using machine learning techniques." 2019 IEEE Conference on Information and Communication Technology. IEEE, 2019.

2. Loussaief, Sehla, and Afef Abdelkrim. "Machine learning framework for image classification." 2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT). IEEE, 2016.

3. Tang, Hai, and Zhihui Hu. "Research on medical image classification based on machine learning." IEEE access 8 (2020): 93145-93154.

4. Wang, Pin, En Fan, and Peng Wang. "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning." Pattern Recognition Letters 141 (2021): 61-67.

5. Anand, R., et al. "Extended morphological profiles analysis of airborne hyperspectral image classification using machine learning algorithms." International Journal of Intelligent Networks 2 (2021): 1-6.