AUTOMATIC IDENTIFICATION OF COVID-19 FROM CHEST X-RAY IMAGES USING ENHANCED MACHINE LEARNING TECHNIQUES

Ankit Ghosh[1], Purbita Kole[2], Alok Kole[3]

[1]*Student, School of Nuclear Studies and Application, Jadavpur University, Kolkata, India*
[2]*Student, Department of Physics, Christ University, Bangalore, India*
[3]*Professor, Department of Electrical Engineering, RCC Institute of Information Technology, Kolkata, India*

Abstract - *The Covid-19 pandemic has led to the loss of millions of human lives across the globe. The public health care system has been facing an unprecedented challenge since the outbreak of Covid-19. The accurate and the timely diagnosis of Covid-19 are extremely important. Artificial Intelligence (AI) and Machine Learning (ML) can play a crucial role in the fight of humanity against the Covid-19 pandemic. This paper presents a detailed overview of how different ML algorithms can be implemented for the automatic identification of Covid-19 infected patients using chest X-ray images. The dataset that has been used comprises of chest X-ray images of both Covid and non-Covid patients obtained from various sources. 9 ML algorithms namely, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbour (KNN), Naïve Bayes (NB), Decision Tree (DT) Classifier, Random Forest Classifier, Stochastic Gradient Descent (SGD) Classifier, XGBoost Classifier and Gradient Boosting Classifier have been implemented to perform the classification of the images. A performance comparative analysis of the different ML algorithms has been conducted based on a few metrics such as accuracy, recall, precision, F1-score and the AUC-ROC curve. XGBoost has surpassed all the other classifiers with an accuracy as high as 93.9%, recall, precision and F1-score of 91.3% respectively and an AUC of 93.3%.*

***Key Words*: artificial intelligence, automatic identification, classification, comparative analysis, diagnosis, machine learning**

1. INTRODUCTION

Artificial Intelligence (AI) and Machine Learning (ML) have proven to be a valuable ally for medical practitioners [1] [2] [3]. Medical imaging data is a rich but complex source of information about the patient [4]. Although medical images like computed tomography (CT), magnetic resonance imaging (MRI), mammograms, X-Ray types such as fluoroscopy and angiography, and ultrasounds are a valuable collection of potentially life-saving data for healthcare researchers, providers and their patients, these kind of images present unique challenges that have traditionally limited radiologists' effectiveness in the diagnosis and treatment of different ailments [5][6]. Medical image analysis is performed by radiologists and clinical doctors. Due to the increasing complexities of the images and overload of work the interpretations of medical images are at times prone to human error. There have been instances where one of four patients has experienced false positives from their medical image review. Undoubtedly traditional manual review and diagnosis of medical images have saved countless lives over the years. However, these methods can be improved. The ability of ML to analyse and learn from vast quantities of data has made them a favourite choice among medical and technology researchers [7] [8] [9] [10]. Image scanning using ML techniques have accomplished more reliable results. AI and ML also promise speed and consistency [11] [12]. Therefore, time sensitive cases can be addressed first and the diagnosis is also faster. AI and ML algorithms also eliminate errors introduced due to natural cognitive bias from clinical diagnoses [13] [14]. Therefore, by merging AI and ML techniques with the competency of medical practitioners, it is feasible to bring about significant advances in the field of medicine and healthcare [15] [16] [17].

2. RELATED WORK

A lot of research has been carried out in the field of AI and ML for medical image analysis. (**Mamlook** et al., 2020) have used a Convolutional Neural Network (CNN) model to perform the classification of chest X-ray images into health vs. sick for the detection of Pneumonia. Their proposed model has achieved an extremely high overall accuracy. (**Sorić** et al., 2020) have proposed the implementation of a CNN for the classification of chest X-ray images. The dataset that they have used contains both pneumonia and non-pneumonia images. Their model has produced satisfactory results. (**Ohata** et al., 2021) have applied the concept of transfer learning for the automatic identification of Covid-19 based on chest X-ray images. They have used different architectures of CNNs trained on ImageNet and then modified them in order to extract the features from the X-ray images. The CNNs have then been combined with other ML algorithms like K-Nearest Neighbours, Bayes, Random Forest, Multi-layer Perceptron (MLP) and Support Vector Machine (SVM). The test results that they have obtained have been quite encouraging. (**Karhan** and **Akal**, 2020) have used ResNet50 model, which is a convolutional neural network architecture for the detection of Covid-19 using chest X-ray images. The experimental results have been quite promising. (**Asif** et al., 2020) have proposed the implementation of deep convolutional neural network (DCNN) based model Inception V3 with transfer learning for the automatic identification of

Covid-19 pneumonia patients using digital chest x-ray images. Their proposed model has achieved very high classification accuracy. (**Arias-Londoño** et al., 2020) have presented an evaluation of different methods based on a deep neural network for the automatic identification of COVID-19 using chest X-ray images to differentiate between controls, pneumonia, or COVID-19 groups. They have described the steps involved in training a CNN with an enormous dataset of X-ray images compiled from various sources. The test results that they have obtained have been quite promising. (**Alghamdi** et al., 2021) have reviewed and analysed the preprint and published reports between March and May 2020 for the diagnosis of COVID-19 via chest X-ray images using CNNs and other deep learning architectures.

3. DATASETS

The dataset that has been used consists of the posteroanterior (PA) view of chest X-ray images. There are 1155 chest X-ray images of patients infected with Covid-19 and 668 chest X-ray images of patients not infected with Covid-19. Therefore, the dataset comprises of 1823 images. The chest X-ray image of a Covid patient has been shown in Fig-1. Fig-2 shows the chest X-ray image of a non-Covid patient.

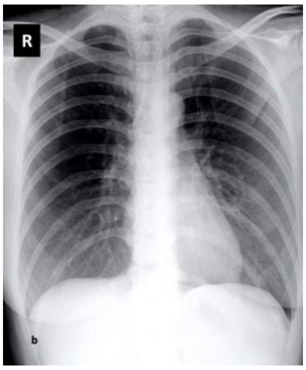

Fig -1: Chest X-ray image of a Covid patient

Fig -2: Chest X-ray image of a non-Covid patient

The chest X-ray images of patients who have been infected with Covid-19 have been labeled as 1 while the chest X-ray images of patients who have not been infected with Covid-19 have been labeled as 0.

4. Methodology

The dataset has been pre-processed by labeling the Chest X-ray images of Covid infected patients as 1 while the Chest X-ray images of non-Covid patients have been labeled as 0. This has been shown in Fig-3.

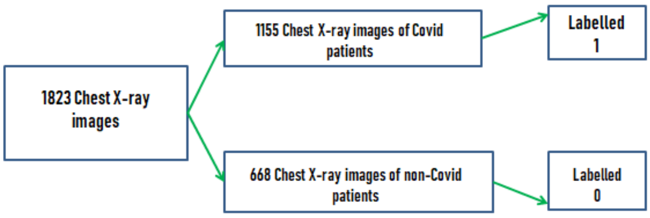

Fig -3: Data pre-processing

The proposed workflow has been illustrated in Fig -4.

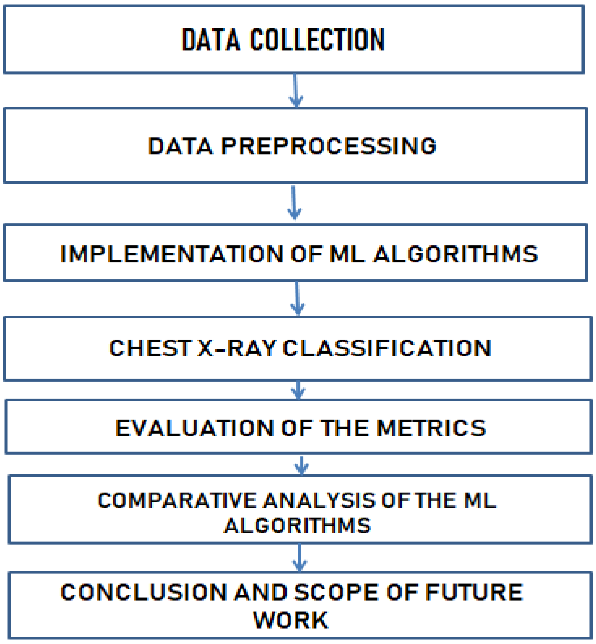

Fig -4: Workflow

4.1 Implementation of the ML algorithms for automatic identification Covid-19

The following ML algorithms have been used to predict whether a patient is Covid infected or not:

- Support Vector Machine
- Logistic Regression
- K-Nearest Neighbour
- Naïve Bayes
- Decision Tree Classifier
- Random Forest Classifier
- Stochastic Gradient Descent Classifier
- XGBoost Classifier
- Gradient Boosting Classifier

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 09 | Sep 2021 www.irjet.net p-ISSN: 2395-0072

The methodology has been described in Fig-5.

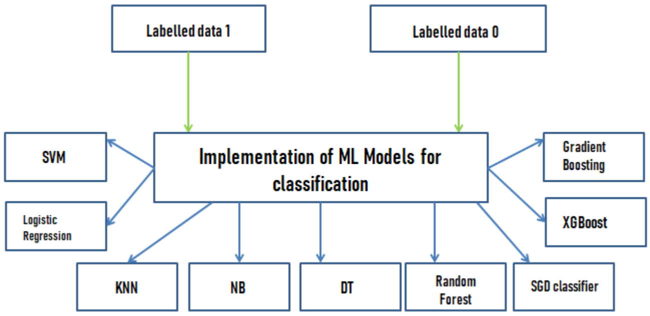

Fig -5: Methodology

The above mentioned state-of-the-art ML algorithms have been used to address the image classification problem in hand. After successful implementation of all the ML algorithms, test scores have been evaluated to perform a comparative analysis of all the algorithms. This has been discussed in the following section.

5. EXPERIMENTAL RESULTS AND COMPARISON

A collection of Chest X-ray images of patients diagnosed with Covid-19 has been shown in Fig-6.

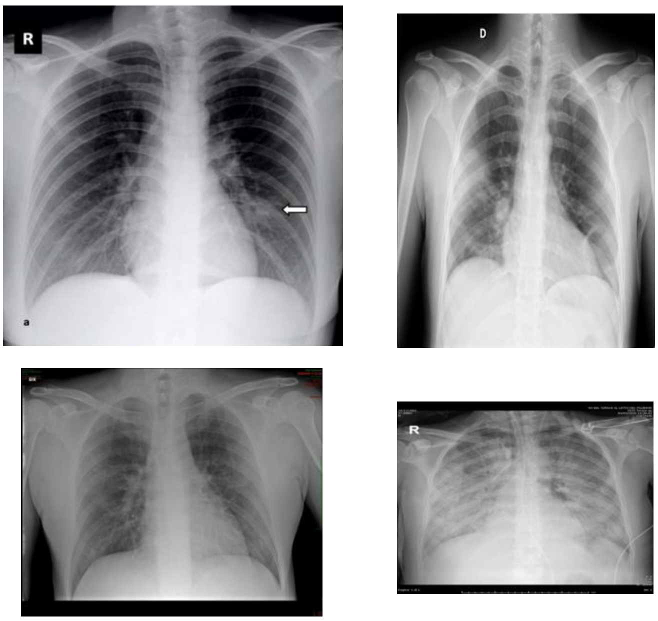

Fig -6: Chest X-ray images of Covid-19 patients

The presence of white spots in the lungs called infiltrates as shown in Fig-6 is an indication of the infection. Detection of infection from these images is not an easy task. So, ML algorithms have been implemented for the automatic identification of Covid-19 from these Chest X-ray images.

A collection of the Chest X-ray images of non-Covid patients has been shown in Fig-7.

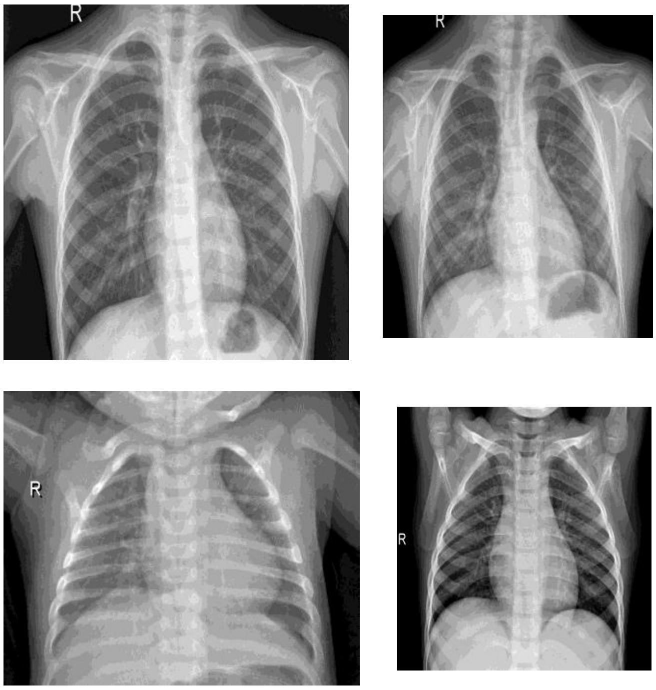

Fig -7: Chest X-ray images of non-Covid patients

Manual classification of these images is quite difficult. So, ML algorithms have been used for the accurate classification of these images so as to predict whether a person has been infected with Covid-19 or not.
A comparative analysis of the 9 ML algorithms that have been used has been done based on the following metrics:

- Accuracy
- Recall
- Precision
- F1-Score
- Area Under Curve – Receiver Operating Characteristics (AUC-ROC) Curve

5.1 Performance analysis and comparison of the different ML algorithms

The confusion matrices of the ML algorithms have been shown in Fig-8.

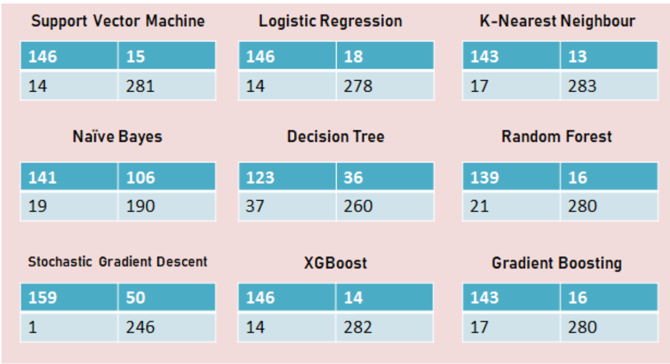

Fig -8: Confusion matrices of the different ML algorithms

The performance of the different ML algorithms has been compared based on accuracy, recall, precision and F1-Score and the test results that have been obtained have been enlisted in Table-1 and Table-2.

Table-1: Performance comparison of the proposed algorithms

Metric	SVM	Logistic Regression	KNN	NB	DT
Accuracy	0.936	0.929	0.934	0.726	0.839
Recall	0.913	0.913	0.894	0.881	0.768
Precision	0.907	0.890	0.917	0.571	0.774
F1-Score	0.909	0.901	0.905	0.693	0.771

Table-2: Performance comparison of the proposed algorithms

Metric	Random Forest	SGD	XGBoost	Gradient Boosting
Accuracy	0.919	0.888	0.939	0.928
Recall	0.869	0.994	0.913	0.894
Precision	0.877	0.761	0.913	0.899
F1-Score	0.873	0.862	0.913	0.896

From the test results it can be inferred that XGBoost has outperformed the other ML algorithms with an accuracy of 0.939, recall, precision and F1-Score of 0.913 respectively.

5.2 AUC-ROC curves of the different algorithms

The AUC-ROC curves corresponding to SVM, Logistic Regression, KNN, NB, DT Classifier, Random Forest Classifier, SGD Classifier, XGBoost Classifier and Gradient Boosting Classifier have been shown in Fig-9, 10, 11, 12, 13, 14, 15, 16 and 17 respectively.

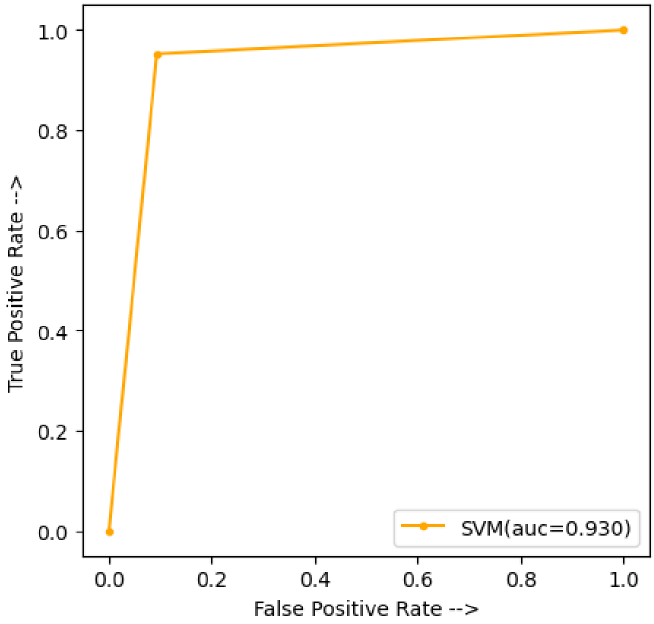

Fig-9: AUC-ROC curve of SVM

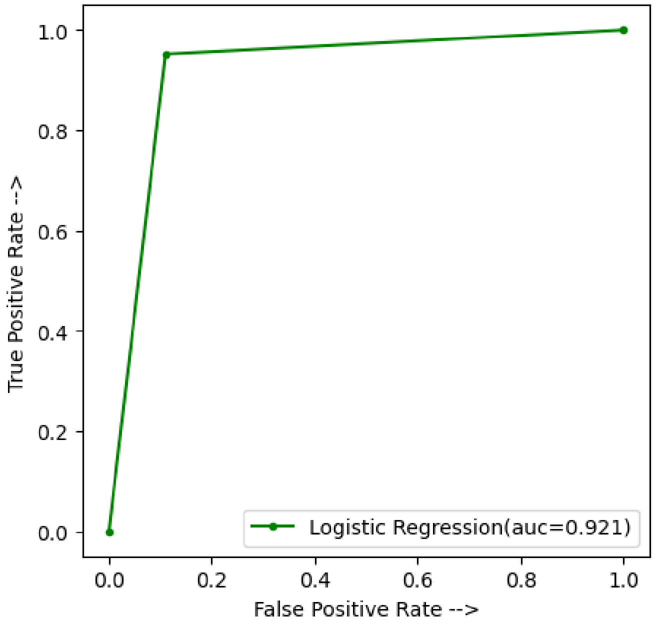

Fig-10: AUC-ROC curve of Logistic Regression

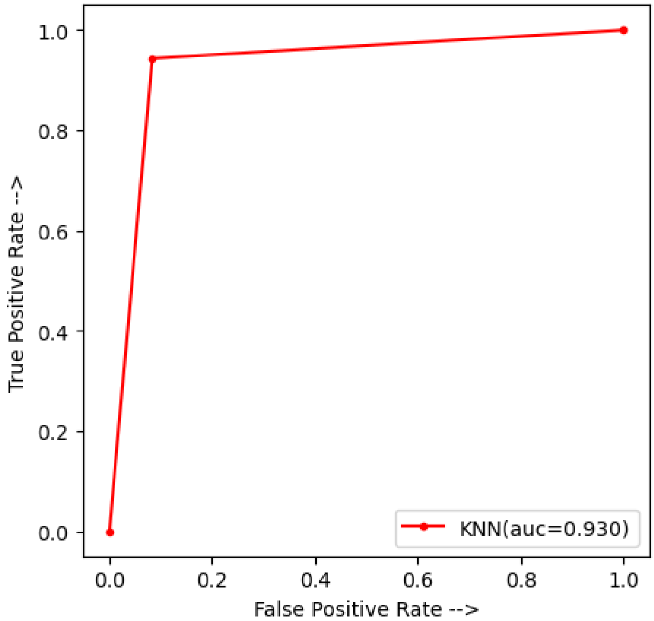

Fig-11: AUC-ROC curve of KNN

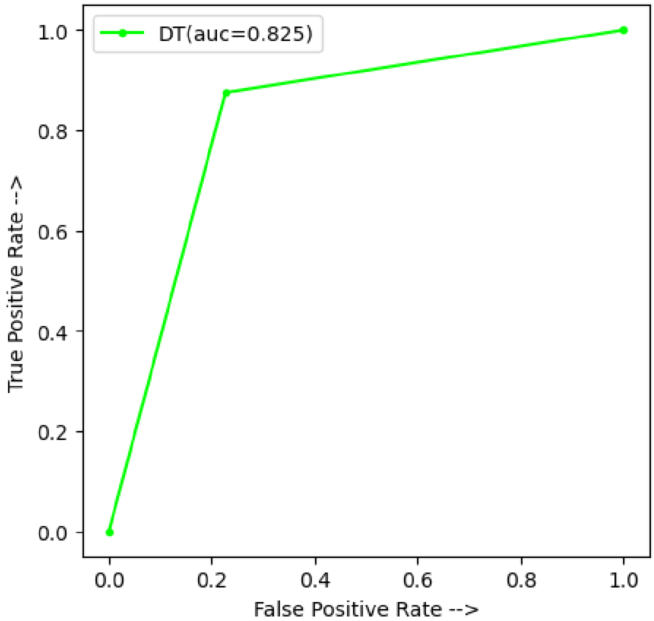

Fig-13: AUC-ROC curve of DT

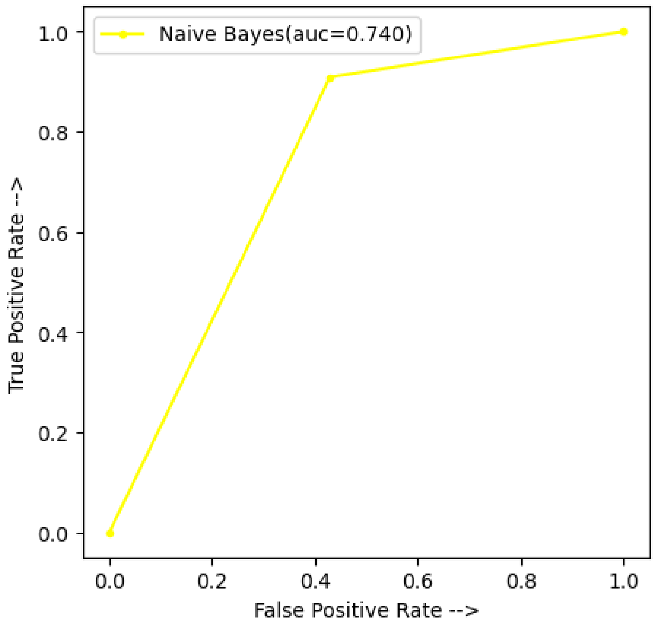

Fig-12: AUC-ROC curve of NB

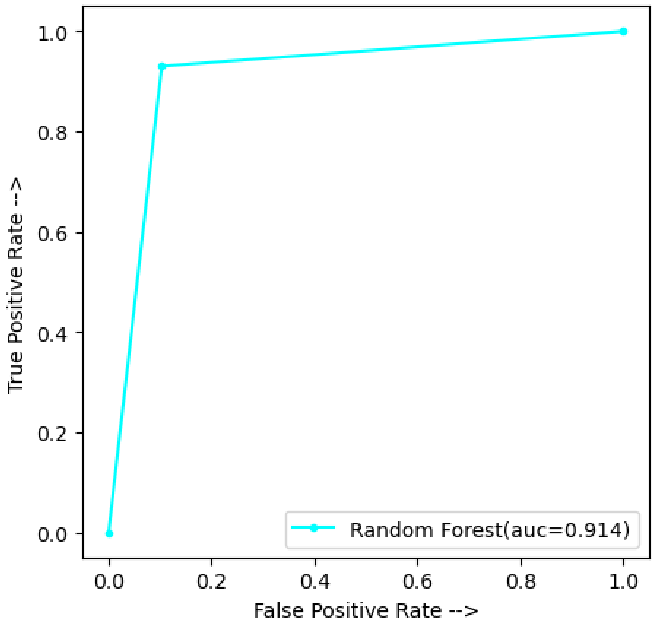

Fig-14: AUC-ROC curve of Random Forest

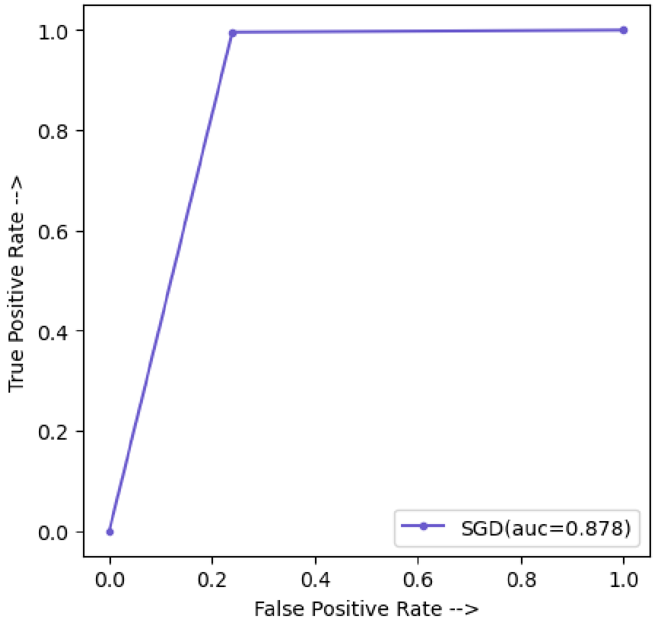

Fig-15: AUC-ROC curve of SGD

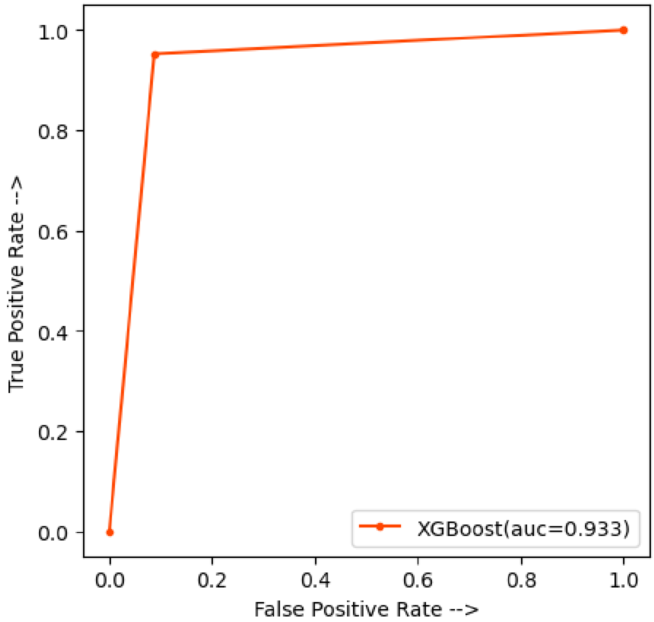

Fig-16: AUC-ROC curve of XGBoost

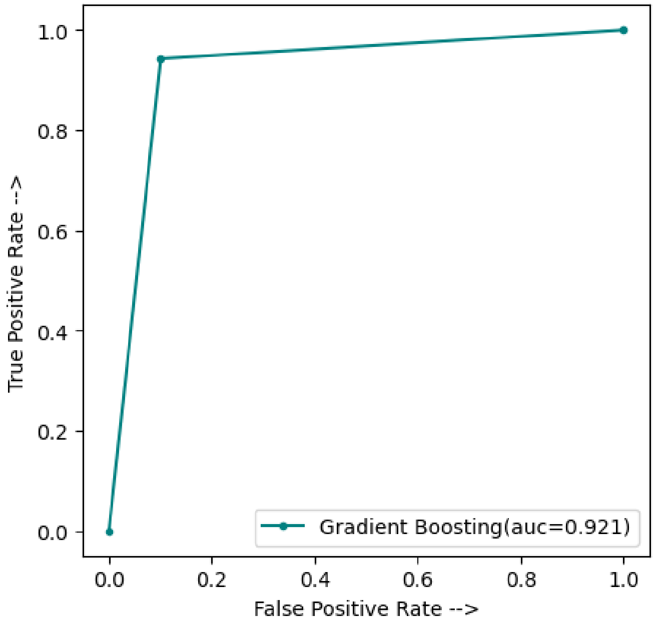

Fig-17: AUC-ROC curve of Gradient Boosting

The AUC of all the ML algorithms has been enlisted in Table-3 and Table-4.

Table-3: AUC of the ML algorithms

Metric	SVM	Logistic Regression	KNN	NB	DT
AUC	0.930	0.921	0.930	0.740	0.825

Table-4: AUC of the ML algorithms

Metric	Random Forest	SGD	XGBoost	Gradient Boosting
AUC	0.914	0.878	0.933	0.921

As it can be observed from the test results, the AUC of XGBoost classifier is 0.933 which is very near to 1. AUC being the measure of ability of a classifier to segregate between the two classes, it can be concluded that XGBoost with an AUC of 0.933 and an accuracy of 0.939 is the best classifier in this case.

6. CONCLUSION

This paper presents a detailed overview of how with the help of enhanced machine learning tools like Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbour (KNN), Naïve Bayes (NB), Decision Tree (DT) Classifier, Random Forest Classifier, Stochastic Gradient Descent (SGD) Classifier, XGBoost Classifier and Gradient Boosting Classifier, automatic identification of Covid-19 can be conducted using Chest X-ray images of patients. A performance comparative analysis of all the machine learning algorithms has been done based on a few metrics such as accuracy, recall, precision, F1-Score and the AUC-ROC curve. The test results that have been obtained have been quite encouraging paving the way for the implementation of other machine learning and deep learning algorithms as well in future. Therefore, it can be concluded that in the fight against Covid-19, Artificial Intelligence, Machine Learning and Deep Learning will undoubtedly play a major role.

REFERENCES

[1] Aryan Sagar Methil, 'Brain Tumor Detection using Deep Learning and Image Processing', 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 25-27 March 2021, Coimbatore, India

[2] Chinmaya Kumar Pradhan, Shariar Rahaman, Md. Abdul Alim Sheikh, Alok Kole and Tanmoy Maity, 'EEG Signal Analysis Using Different Clustering Techniques', International Conference on Emerging Technologies in Data Mining and Information Security, 23- 25 February, 2018, Kolkata, West Bengal

[3] Tushar Kant Mudgal, Aditya Gupta, Siddhant Jain and Kunal Gusain, 'Automated system for Brain tumour detection and classification using eXtreme Gradient Boosted decision trees', 2017 International Conference on Soft Computing and its Engineering Applications (icSoftComp), 1-2 Dec. 2017, Changa, India

[4] Gajendra Raut, Aditya Raut, Jeevan Bhagade, Jyoti Bhagade and Sachin Gavhane, 'Deep Learning Approach for Brain Tumor Detection and Segmentation', 2020 International Conference on Convergence to Digital World - Quo Vadis (ICCDW), 18-20 Feb. 2020, Mumbai, India

[5] Masoumeh Siar and Mohammad Teshnehlab, 'Brain Tumor Detection Using Deep Neural Network and Machine Learning Algorithm', 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE), 24-25 Oct. 2019, Mashhad, Iran

[6] Sneha Grampurohit, Venkamma Shalavadi, Vaishnavi R. Dhotargavi, Megha Kudari and Soumya Jolad, 'Brain Tumor Detection Using Deep Learning Models', 2020 IEEE India Council International Subsections Conference (INDISCON), 3-4 Oct. 2020, Visakhapatnam, India

[7] Alok Kole, "Design and Stability Analysis of Adaptive Fuzzy Feedback Controller for Nonlinear Systems by Takagi-Sugeno Model based Adaptation Scheme", International Journal of Soft Computing, Vol.19, Issue.6, pp.1747-1763, Springer SCI Indexed, 2015

[8] Sanjeevani Bhardwaj and Alok Kole, 'Review and Study of Internet of Things: It's the Future', IEEE International Conference on Intelligent Control, Power and Instrumentation (ICICPI-2016), 2016, Kolkata, India

[9] Mahima Rai and H. L. Mandoria, 'Network Intrusion Detection: A comparative study using state-of-the-art machine learning methods', 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), 27-28 Sept. 2019, Ghaziabad, India

[10] Ankit Ghosh, Purbita Kole and Alok Kole, 'Fault Detection and Diagnosis of various Mechanical Components in a Nuclear Power Plant combining Noise Analysis and Machine Learning Techniques', International Research Journal of Engineering and Technology (IRJET), Vol.8, Issue.8, pp.3930-3936, 2021

[11] Rabia Emhamed Al Mamlook, Shengfeng Chen and Hanin Fawzi Bzizi, 'Investigation of the performance of Machine Learning Classifiers for Pneumonia Detection in Chest X-ray Images', 2020 IEEE International Conference on Electro Information Technology (EIT), 31 July-1 Aug. 2020, Chicago, IL, USA

[12] Matija Sorić, Danijela Pongrac and Iñaki Inza, 'Using Convolutional Neural Network for Chest X-ray Image classification', 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 28 Sept.-2 Oct. 2020, Opatija, Croatia

[13] Elene Firmeza Ohata, Gabriel Maia Bezerra, João Victor Souza das Chagas, Aloísio Vieira Lira Neto, Adriano Bessa Albuquerque, Victor Hugo C. de Albuquerque and Pedro Pedrosa Rebouças Filho, 'Automatic Detection of COVID-19 Infection Using Chest X-Ray Images Through Transfer Learning', IEEE/CAA JOURNAL OF AUTOMATICA SINICA, Vol. 8, No.1, pp. 239-248, 2021

[14] Zehra Karhan and Fuat Akal, 'Covid-19 Classification Using Deep Learning in Chest X-Ray Images', 2020 Medical Technologies Congress (TIPTEKNO), 19-20 Nov. 2020, Antalya, Turkey

[15] Sohaib Asif, Yi Wenhui, Hou Jin and Si Jinhai, 'Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Network', 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 11-14 Dec. 2020, Chengdu, China

[16] JULIÁN D. ARIAS-LONDOÑO, JORGE A. GÓMEZ-GARCÍA, LAUREANO MORO-VELÁZQUEZ3, and JUAN I. GODINO-LLORENTE, 'Artificial Intelligence Applied to Chest X-Ray

Images for the Automatic Detection of COVID-19.A Thoughtful Evaluation Approach, IEEE Transactions, Vol.8, pp. 226811-226827, 2020

[17] Hanan S. Alghamdi, Ghada Amoudi, Salma Elhag, Kawther Saeedi and Jomanah Nasser, 'Deep Learning Approaches for Detecting COVID-19 From Chest X-Ray Images: A Survey', IEEE Engineering in Medicine and Biology, IEEE Access, Vol.9, pp. 20235 – 20254, 2021