

RESEARCH ARTICLE



An Intelligent Diagnosis of Adenovirus Disease for Child Healthcare and Prognosis

 OPEN ACCESS

Received: 18-03-2023

Accepted: 31-05-2023

Published: 09-06-2023

Sumit Das^{1*}, Amarta Kundu¹, Abhishek Kumar¹, Biswajit Karmakar¹, Anirban Saha¹¹ JIS College of Engineering, Information Technology, Kalyani, 741235, India

Citation: Das S, Kundu A, Kumar A, Karmakar B, Saha A (2023) An Intelligent Diagnosis of Adenovirus Disease for Child Healthcare and Prognosis . Indian Journal of Science and Technology 16(23): 1716-1725. <https://doi.org/10.17485/IJST/v16i23.447>

* Corresponding author.

sumit.das@jiscollege.ac.in

Funding: None

Competing Interests: None

Copyright: © 2023 Das et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN

Print: 0974-6846

Electronic: 0974-5645

Abstract

Objective: This proposed study is based upon the recent disease Adenovirus classification of several physical activities based upon different Machine Learning algorithms. We have chosen this topic because Adenovirus mostly infected children, and we are trying to handle this virus at an early stage so that it doesn't affect a large population like COVID-19. The social impact of this proposed study is that each and every individual can use this model free of cost to find the Adenovirus. **Methods:** The dataset contains 5434 physical samples with 8 body parameters. All the collected samples there are 4484 infected (Adenovirus) and 950 healthy (non-Adenovirus). Based on the collected dataset train the Machine Learning algorithms so that the Machine Learning algorithm can work as an alternative to diagnose and predict Adenovirus and non-Adenovirus accurately. **Findings:** For the best result of the Machine Learning classifiers, this study used Decision Tree, K-Nearest Neighbors, Support Vector Machines, Logistic Regression, Naive Bayes, Random Forest, and Gradient Boosting Classifier. The decision tree gives the best results when compared to other Machine Learning algorithms. This study demonstrates that the Decision Tree classifier performed the most effectively with an accuracy of 95% when making comparisons between Adenovirus and non-Adenovirus. **Novelty:** The major uniqueness of this proposed work is the recognizing of the Adenovirus from the human body so that general people can be health conscious and take precautions to prevent the Adenovirus infection.

Keywords: Adenovirus; Child Healthcare; Machine Learning Algorithm; Fuzzy Dilation Membership function; Intelligent Diagnosis and Treatment

1 Introduction

The unique features of this model are that when individual inputs his/her body parameters, the model will accurately identify if the individual is infected with Adenovirus or not. As a result, the number of appointments for physical check-up will be reduced in the hospital. In rural areas where people can't effort an appointment with the doctor, they can use this ML model for Adenovirus check-ups. The World Health Organization (WHO) said that 20 nations throughout the world, including those in

South East Asia, had found over 300 suspected instances of kids suffering from severe hepatitis. The World Health Organization (WHO) has confirmed one fatality. To further stop or reduce the epidemic, everyone should use the proper adenovirus-specific infection control techniques. Using this research anyone can detect Adenovirus on their own because it is based on a Machine Learning algorithm and also remotely accessible from any place. DNA viruses known as adenoviruses commonly cause minor infections of the upper or lower respiratory tract, gastrointestinal system, or conjunctiva. Hepatitis, hemorrhagic colitis, hemorrhagic cystitis, pancreatitis, nephritis, and meningoencephalitis are uncommon symptoms of adenovirus infections. Because they lack humoral immunity, young children are more likely to contract adenoviruses infections. In closed or congested environments, epidemics of adenoviruses infection may affect healthy children or adults (particularly military recruits). In patients with compromised immunity, the disease is more severe and spread is more likely.

Following are the key contributions of this proposed work:

- A detailed explanation of the Adenovirus detection process by using several ML algorithms for clinical support.
- The distribution of the data set’s cases has been balanced and carefully handled during the data pre-processing procedure, which has produced accurate classification outcomes for the prediction of the Adenovirus.
- In this proposed work all, the steps are taken starting from collecting the data, pre-processing the data, multiple ML models, and ML model selection, until the result analysis is clearly mentioned.
- The part of ML model selection used multiple classifiers, the cross-validation method, and hyperparameter tuning to improve the productivity of the ML model.
- The effectiveness of all the machine learning models implemented for this particular project was evaluated, and the best ML model was chosen after a discussion of the project’s potential future.

Generally, when someone filling physical discomfort the first option is for that person to be admitted to a hospital and checked for physical condition. Based on the complexity there are some regular tests and a doctor confirms the disease. So, for this case doctor check the patient report and confirm the Adenovirus. There are three conditions:

- **Condition-1:** If the patient is Adenovirus negative then the doctor discharges the patient and the patient is good to go.
- **Condition-2:** If the patient is suspected of being Adenovirus positive then the doctor observes the patient for 24/7 and tracks the physical activity of patient. If the patient recovered then discharged otherwise start the treatment.
- **Condition-3:** If the patient is confirmed Adenovirus positive then the doctor starts the treatment evidently based on the patient’s physical condition. Also, admit to ICU if required.

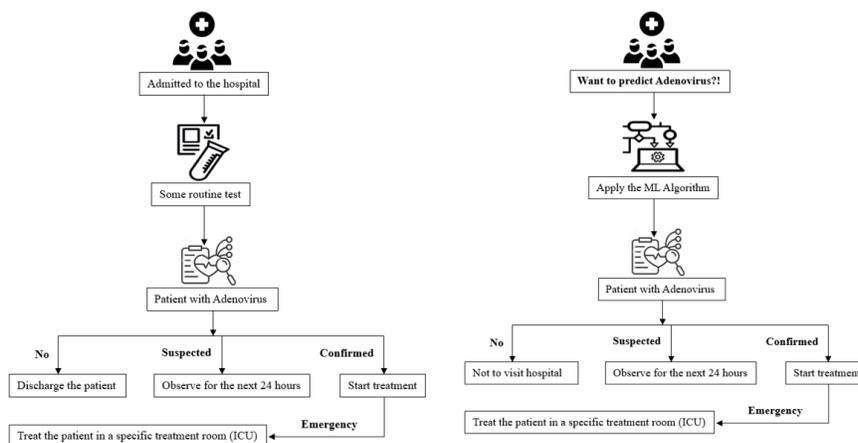


Fig 1. Traditional diagnosis and Treatment (left), and Intelligent diagnosis and Treatment (right)

This proposed study will help the patient to predict the Adenovirus by just putting some physical sample in the Machine Learning algorithm as shown in Figure 1 in contrast with traditional diagnosis depicted. The patient admits to the hospital only when the Machine Learning algorithm predicts the Adenovirus otherwise it’s not required to go to the hospital. So, the main outcome of this proposed study is that the patient is not taking extra pasture to predict the Adenovirus and admit the hospital only when the patient is confirmed Adenovirus positive. This study is based upon the recent disease Adenovirus classification

of several physical activities based upon different Machine Learning algorithms. Those Machine Learning algorithms will differentiate between Adenovirus and non-Adenovirus based on the people's physical health condition inputs. Because the whole Adenovirus testing process is based on Machine Learning algorithms that's why anyone can easily test the Adenovirus by providing some body parameters also if they haven't any idea about the Adenovirus as well as the Machine Learning technology. The following literatures are studied for finding the research gaps.

The child's oral intake was decreased, and he or she experienced loose stools, a stuffy nose, and mild atopic dermatitis in the past. This is the first instance of a baby who had both an adenovirus and COVID-19 infection. 10 out of 257 individuals tested positive for confection with COVID-19 and adenovirus, according to a study by Zhu et al., however, the majority of these patients were between the ages of 15 and 65. 16 Notwithstanding the few adult cases that have been reported, 16–18 the majority of the literature discusses confections with other viruses; adenovirus confections are not mentioned in any of the patients⁽¹⁾. Adenovirus is a member of the family Adenoviridae. Adenovirus is a viral disease that causes mild to severe infection in the human body. It mainly affects the human respiratory system. The symptoms are commonly the same as cold or flu. There are almost 50 types of adenoviruses that can infect the human body. Anybody can be infected by adenovirus but children below 5 are more prone to this infection because children are used to putting the dirty hand in their mouth often lead to disease. Adenovirus can be easily spread through Close contact, The air, Surfaces & Objects, Poop, and Water. If someone is infected through adenovirus and if it is mild it can be treated at home and no need to go to a healthcare provider but if it causes a severe problem then we need to consult a healthcare provider⁽²⁾.

The state's health department is extremely concerned about the fast rise of viral cases in West Bengal. The health department's concern has grown as a result of the deaths of three infants in a 24-hour period in the month of February. The victims are all infants less than 18 months. Due to this incidence, children's wards may be found in both public and private hospitals. According to reports, the B.C. Roy hospital's lack of a pediatric intensive care unit (PICU) bed was to blame for the nine-month-old baby's death⁽³⁾.

Adenovirus was detected in about 33 percent of the 500 samples examined between the third week of January and February. The patient needs to be ventilated as well when the disease worsens. There is no special prescription for this adenovirus, but we can take care of ourselves by keeping ourselves and our environment clean. All age groups are susceptible to the mild cold and flu-like symptoms caused by adenovirus, but children are particularly at risk. Even though adenovirus cases have been on the rise over the previous two months, the conversion of gynecological wards into children's wards has just begun⁽⁴⁾. A co-infection of COVID and adenovirus is possible. A study on the first instance of a baby who had both an adenovirus and COVID-19 infection were released in the BMJ in June 2020. The infant, a 4-month-old boy, displayed symptoms that were typical with an adenovirus infection, but due to the child's exposure at home, the team decided to test for COVID-19. The study concluded that "this youngster may not have undertaken COVID-19 testing due to the positive adenovirus infection" if the child had not had in-home exposures⁽⁵⁾.

The Pandemic of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) raises the question of which viruses could pose a global challenge. The traits that make adenoviruses a risk, according to this Opinion, include efficient intra- and interspecies transmission, thermos-table particles, chronic infections in diverse hosts, and the ability to readily recombine and escape immune systems⁽⁶⁾. The first is lytic infection, which occurs when an adenovirus enters human epithelial cells and continues through an entire replication cycle, which results in cytolysis, cytokine production, and induction of host inflammatory response⁽⁷⁾. Adenoviruses that infect humans adenoviruses are categorized as belonging to the genus Adenoviruses and are classified into seven species (A to G), with approximately 100 types classified by serology Encyclopedia of Virology - 4th Edition⁽⁸⁾.

Artificial Intelligence and Machine Learning are playing important roles in medical diagnosis for predicting and detecting complex diseases. Deep learning, a recent advancement in machine learning, has acquired enormous capability of learning with large volumes of dataset. This technological progress enables doctors to provide quick and precise prognosis, diagnosis, and treatment⁽⁹⁾. The advancement of these approaches have increased the functionality of doctors to the point where the simulation results can be thought of as an AI physician in parallel. As there haven't been any prospective, randomized treatment trials, the management of adenoviruses infections is debatable. Although the preferred medication for treating severe adenoviruses infections is cidofovir, not all individuals need to be treated. Although routinely used in the American military and very effective in lowering the risk of respiratory adenoviruses infection, live oral vaccinations are not yet accessible to civilians⁽¹⁰⁾.

According to WHO first, we have to identify additional Adenovirus cases both in currently affected countries and elsewhere. If we don't recognize the consequence of Adenovirus at an early stage this may cause serious issues like COVID-90⁽¹¹⁾. The research gaps is that there is insufficient relevant AI/ML model to predict the Adenovirus cases and this research gaps are the motivation behind this study.

2 Methodology

In this research, used several Machine Learning classifiers under Artificial intelligence to predict the Adenovirus and non-Adenovirus other methods of diagnosis. The Adenovirus is an active member of the over 50 different types of viruses that make up the Adenoviridae family. According to the recent news in Kolkata itself till February 2023, 115 patients are admitted to AMRI Hospitals with respiratory issues and most of the patients are suffering from Adenovirus.

Figure 2 explains the methodology for deploying our proposed work like collecting human body samples, data preprocessing, feature engineering, and building a Machine Learning model to predict the final Adenovirus.

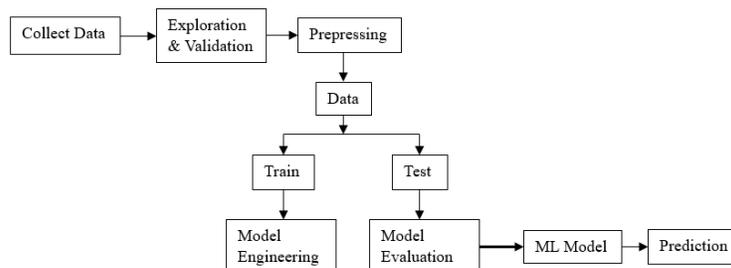


Fig 2. Work Flow Diagram

2.1 Data Collection

Among a total of 5434 physical samples with 8 body characteristics, we discovered 4484 samples that were contaminated with Adenovirus and 950 samples that were immune (non-Adenovirus). The primary novel aspect of the proposed effort is the development of an Adenovirus diagnostic tool based on certain physiological characteristics. Collect the patient report from various hospitals in order to achieve a more efficient Machine Learning model. Currently, using 5434 data to train the model and, in the future, we incorporate more data to train the model in order to achieve an efficient model. In the process of assembling the adenovirus dataset, we looked at a number of datasets, including COVID-19, and the COVID-19 dataset significantly supported in the completion of this study.

The dataset demonstrates that the patients who are not yet infected with the virus are classified as having an Adenovirus, but those who are already infected with the virus are classified as having a non-Adenovirus. This research uses the dataset which is mostly populated with adenovirus samples (approx. 4/5) and the remaining sample is of non-adenovirus in order to achieve better outcomes.

2.2 Data Pre-Processing

To get the best result from a Machine Learning model it's very important to train the model with processed or good-quality data. In the complete process of data pre-Processing, breaking the process into multiple sub-processes like Data cleaning, Data reduction, Data scaling, Data transformation, and Data partitioning.

- **Data cleaning:** By removing outliers and imputed missing values, data cleaning tries to improve the overall quality of the dataset. Using the Pandas library, we find the duplicate values and drop the duplicate values. Filter the unwanted outliers that help analyze the dataset more efficiently and help improve the Machine Learning model performance.
- **Data reduction:** Data reduction can be carried out in two ways: row-wise data reduction and column-wise data reduction. In this dataset, there are some irrelevant records like 'Heart Disease' and 'Diabetes'. These records are irrelevant to predict the Adenovirus that's why we use a column-wise data reduction process to drop 'Heart Disease' and 'Diabetes' from our dataset.
- **Data scaling:** Support vector machines (SVM) or k-nearest neighbors (KNN) algorithms are generally used for Data scaling. This research used both methods for scaling the dataset. In order to ensure every characteristic equally contributes to the outcome used data scaling in the Machine Learning model. Used the scikit-learn library to achieve the data scaling goal.
- **Data transformation:** Every algorithm used in machine learning is based on mathematics. The dataset contains only Categorical data so, each column must be converted to a numerical format. The dataset is converted to a numerical

format by replacing true equal to '1' and false equal to '0'. NumPy, Matplotlib, Pandas, Scikit-Learn, and Tensor-Flow are the library used for data transformation but we used NumPy, Matplotlib, Pandas, and Scikit-Learn in our proposed work. The Figure 3 showing Clean Data and Data transformation.

```
[ ] adenoviruses.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5434 entries, 0 to 5433
Data columns (total 9 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Breathing Problem     5434 non-null   object
 1   Pink Eye              5434 non-null   object
 2   Pneumonia             5434 non-null   object
 3   Fever                 5434 non-null   object
 4   Acute Gastroenteritis 5434 non-null   object
 5   Dry Cough             5434 non-null   object
 6   Sore throat          5434 non-null   object
 7   Bladder Infection     5434 non-null   object
 8   Adenoviruses         5434 non-null   object
dtypes: object(9)
memory usage: 382.2+ KB

[ ] from sklearn.preprocessing import LabelEncoder
e=LabelEncoder()

[ ] adenoviruses['Breathing Problem']=e.fit_transform(adenoviruses['Breathing Problem'])
adenoviruses['Pink Eye']=e.fit_transform(adenoviruses['Pink Eye'])
adenoviruses['Pneumonia ']=e.fit_transform(adenoviruses['Pneumonia '])
adenoviruses['Fever']=e.fit_transform(adenoviruses['Fever'])
adenoviruses['Acute Gastroenteritis ']=e.fit_transform(adenoviruses['Acute Gastroenteritis '])
adenoviruses['Dry Cough']=e.fit_transform(adenoviruses['Dry Cough'])
adenoviruses['Sore throat']=e.fit_transform(adenoviruses['Sore throat'])
adenoviruses['Bladder Infection']=e.fit_transform(adenoviruses['Bladder Infection'])
adenoviruses['Adenoviruses']=e.fit_transform(adenoviruses['Adenoviruses'])
```

Fig 3. Clean Data (left) and Data transformation (right)

- **Data partitioning:** When the dataset is ready to process for the training then we split our dataset into two parts one is called the training dataset and another is called the test dataset. The dataset is split with an 80% and 20% ratio for the training and testing respectively. To improve the model performance used most of the data(80%) for training purposes and the rest(20%) of the data was used to check the model performance. Under the scikit-learn library, here used the train_test_split() function for splitting the dataset.
- **Classification Process:** In this research, several ML algorithms are used including Decision Tree, K-Nearest Neighbors, Support Vector Machines, Logistic Regression, Naive Bayes, Random Forest, and Gradient Boosting Classifier for classification and checked the accuracy of each model individually. In the ideal Machine Learning algorithm, we instruct the system to carry out this analysis and choose the best model architecture on its own. In this research used hyper-parameter tuning to find the appropriate model architecture by adjusting parameters that describe the model architecture, or hyper-parameters. Using hyper-parameters get the actual structure of the model, whilst the model parameters indicate how to modify the model parameters into the expected outcome. So, the primary goal of hyper-parameter tuning is to identify the parameters that improve the effectiveness of our model.

3 Results and Discussion

- **Decision Tree:** Classification and regression, under supervised learning Decision Tree is the most efficient algorithm. Already name suggests 'tree' means, Decision Tree is like the Tree model. The main concept behind Decision Tree is that it splits the entire dataset into sub-datasets by some values and properties and this process continues recursively until the model returns the best results. Previously mentioned Classification trees and Regression trees are mainly two types of Decision Tree. Under Classification trees 'fit' or 'unfit' is the only output given by any variable on the other hand Regression trees the output is continuous.

Classification trees are used for this proposed work. Classification: $G = \sum(pk * (1 - pk))$ Here, pk is the share of inputs from the same class that are present in a given group. Using the Decision Tree algorithm accuracy rate of the model is 94.94% and it returns the highest accuracy with compare to any other model used as illustrated in Figure 4.

- **K-Nearest Neighbors (KNN):** Classification, under supervised learning K-Nearest Neighbor (KNN), is the most efficient algorithm. For the Classification, K-Nearest Neighbor (KNN) model returns a pretty good result. The KNN method makes the assumption that the new case and the existing cases are equivalent, and it places the new instance in the classification that is most similar to the available classifications. The KNN algorithm saves all the information that is available and categorizes new inputs according to their similarities. The main advantage is When fresh data is present, it may be quickly and accurately categorized by using the previous data. Using the K-Nearest Neighbor (KNN) algorithm accuracy rate of the model is 94.48% and it returns the second-highest accuracy with compare to other models as illustrated in Figure 4.
- **Support Vector Machines (SVM):** Classification and regression, under supervised learning Support Vector Machines(SVM), is the most efficient algorithm. Although it's used for Classification and regression-related problems Mostly, it's utilized to solve Classification challenges in Machine Learning. The main working principle of SVM is to

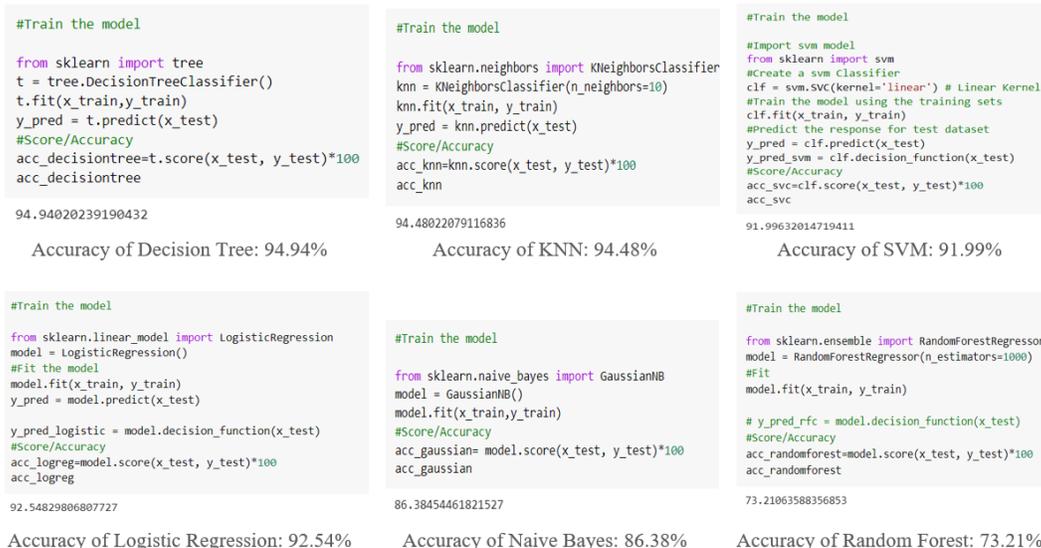


Fig 4. Accuracy of Decision Tree, KNN, SVM, Logistic Regression, Naive Bayes and Random Forest Algorithms

establish the optimal decision threshold or edge that can divide n-dimensional circumstances into classifications so that subsequent data points may be quickly assigned to the appropriate category. The term "hyper-plane" refers to this optimal decision boundary. Using the Support Vector Machines (SVM) algorithm accuracy rate of the model is 92% and it also returns a pretty good result as illustrated in Figure 4.

- Logistic Regression:** Classification, under supervised learning Logistic Regression, is the most efficient algorithm. For the Classification, the Logistic Regression model returns a pretty good result. A binary classification procedure called logistic regression uses a collection of independent factors to forecast the likelihood of a binary event (i.e 0 or 1). For example, Logistic Regression is used in such conditions whether a patient is affected with Adenovirus or not based on some physical body parameters.

Equation for logistic regression:

$$y = \frac{e^{(b_0 + b_1 x)}}{1 + e^{(b_0 + b_1 x)}} \tag{1}$$

Here, y is the anticipated result, b_0 is bias, and b_1 is the coefficient of (x) for the single input value.

Using the Logistic Regression algorithm accuracy rate of the model is 92.54% and it also returns a pretty good result as illustrated in Figure 4.

- Naive Bayes:** The Nave-Bayes algorithm is a supervised learning method for classification issues that are developed on the Bayes theorem. It is primarily utilized in training datasets with high dimensional datasets. The equation for Nave Bayes' algorithm:

$$P(A | B) = P(B | A) \frac{P(A)}{P(B)} \tag{2}$$

Here, $P(A|B)$ is the posterior probability, and $P(A)$ is called the prior The accuracy rate of the Logistic Regression algorithm is 86.38%as illustrated in Figure 4.

- Random Forest:** Classification and regression, under supervised learning Random Forest, is the most efficient algorithm. Every decision tree is supplied a portion of the sample once it has been partitioned. So, every decision tree generates a prediction performance during the training cycle. When a new data item is encountered, the random forest classification

algorithm evaluates the outcome based on the overwhelming results. The Random Forest model accuracy depends upon the number of trees present in the Random Forest algorithm. For the testing dataset, Random Forest is ineffective, and the model does not produce satisfactory results. The accuracy rate of the Random Forest algorithm is 73.21% as illustrated in Figure 4.

- **AUC-ROC Curve:** The AUC-ROC curve provides a visual representation of how effectively the Machine Learning classifier works. AUC-ROC curve is commonly used for binary classification concerns.
 - **Sensitivity:** Sensitivity is a measure of the percentage of instances that were actually positive but were misclassified as positive (or true positive). Another percentage of genuine positive instances will be forecasted mistakenly as negative (called false negative).

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

- **Specificity:** The percentage of genuine negatives that were projected as negatives is known as specificity (true negative). False positives are genuine negatives that were anticipated to be positive but turned out to be true negatives.

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \tag{4}$$

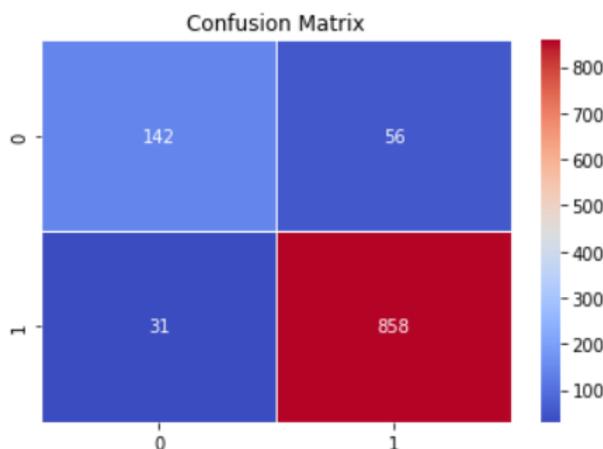


Fig 5. Confusion Matrix

- **ROC Curve :** The AUC ROC curve essentially serves as a gauge for how well a machine learning model performs, which is estimated from confusion matrix as illustrated in Figure 5. The ROC curve is summarized by AUC, which assesses a binary classifier’s ability to differentiate between classes. A higher X-axis value on a ROC curve denotes a greater proportion of false positives compared to true negatives. In contrast, a higher Y-axis value denotes a greater proportion of True positives than False negatives as shown in Figure 6. The capacity to balance False positives and False negatives will thus influence the threshold selection.

The ROC curve is compared and observed with fuzzy dilation membership function⁽¹²⁾; the observation is that SVM and logistic algorithm yield somewhat better accuracy as per the theory of fuzzy dilation membership function⁽¹³⁾ that is the square root of the fuzzy membership function as depicted in Figure 7. After analyzing the complexity of Adenovirus cases from the early stages to the emergency stage we observe the main symptom of Adenovirus. Based on the Adenovirus symptom we proposed this ML model. This ML model will predict the Adenovirus cases with an accuracy of 95% approximately as shown in Figure 8.

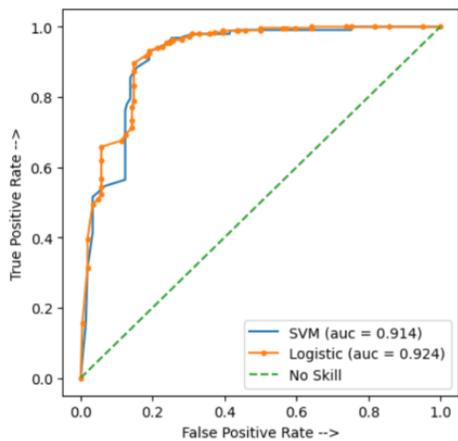


Fig 6. ROC Curve

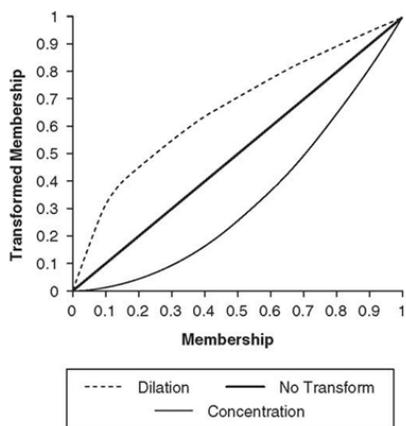


Fig 7. Fuzzy dilation membership function ⁽¹²⁾

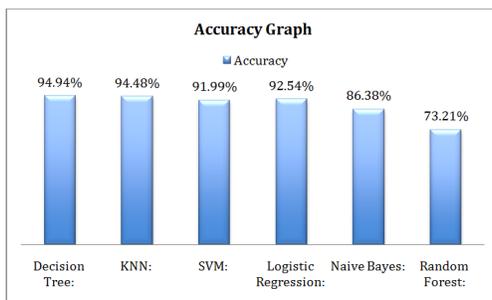


Fig 8. Graphical Representation of ML Model Accuracy

After applying all the models to the dataset, we found that Decision Tree gives the most efficient outcome. So, we use Decision Tree for our proposed work. Human adenovirus (HAdV) is a major cause of acute respiratory infections (ARIs) in children⁽¹⁴⁾. The principal diagnostic tools and the immune response in HAdV infections are described and whether markers based on the response of the host may help early recognition of H adenoviruses and avoid inappropriate antimicrobial prescriptions in acute airway infections is evaluated⁽¹⁵⁾.

```

✓ [52] result = model.predict([[0,0,1,0,0,1,0,0]])
0s /usr/local/lib/python3.9/dist-packages/sklearn/base.py:42:
    warnings.warn(

✓ [53] if result==1:
0s     print("The Patient is Adenovirus Positive(+ve)")
     else:
     print("The Patient is Adenovirus Negative(-ve)")

The Patient is Adenovirus Negative(-ve)

```

Fig 9. Final Adenovirus positive(+ve) prediction using the Machine Learning model

```

✓ [50] result = model.predict([[1,1,1,1,1,1,1,0]])
0s /usr/local/lib/python3.9/dist-packages/sklearn/base.py:42:
    warnings.warn(

✓ [51] if result==1:
0s     print("The Patient is Adenovirus Positive(+ve)")
     else:
     print("The Patient is Adenovirus Negative(-ve)")

The Patient is Adenovirus Positive(+ve)

```

Fig 10. Final Adenovirus negative(-ve) prediction using the Machine Learning model

4 Major Findings

This proposed study is based on a machine learning algorithm that has been trained to accurately forecast Adenoviruses. Figures 9 and 10 depict two real-time Adenovirus case studies, as well as how accurately the Machine Learning algorithm predicts Adenovirus and Non-Adenovirus. The K-Nearest Neighbors (KNN) algorithm and the Decision Tree algorithm both performed well throughout the testing phase, both with a 94% success rate⁽¹⁶⁾. The results of the machine learning anticipated outcome were repeatedly cross-verified with medical professionals; Clinico–pathologic correlation within the specific disease entity caused by adenoviruses is reviewed to better understand this common viral infection in pediatric population⁽¹⁷⁾. So, the alternative of diagnosing performs very well, as well as correct for predicting Adenovirus. This machine-learning-based alternative to diagnosing will reduce the hospital readmission rate.

5 Conclusion

This proposed work is assembled using a number of Machine Learning techniques. Choosing a specific machine learning method for this project is really difficult. A number of actions are done, including as cross-validation and hyper-parameter tweaking, to increase the Machine Learning algorithm's efficiency and ensure that it produces the maximum output. The dataset and the algorithm are both much improved via cross-validation. The over-fitting issue is resolved via cross-validation, allowing an independent sample input to the Machine Learning model to generalize the independent sample and predict the right outcome. To regulate a Machine Learning algorithm's behavior, hyper-parameter tuning is performed before the training assignment is performed. Regarding training time, they may have a significant influence on model training. To achieve

maximum accuracy, we trained the Machine Learning model with 80% of the dataset and tested the Machine Learning model's accuracy using the remaining 20% of the dataset. Used a decision tree algorithm for this proposed work with an accuracy of 94.94%. Another three algorithms K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Logistic Regression show outstanding results with accuracy of 94.48%, 92%, and 92.54%, respectively. Naive Bayes and Random Forest algorithms are used for testing purposes, and the model accuracy is 86.38% and 73.21%, respectively.

Currently, this model is only able to predict Adenovirus or non-Adenovirus but, in the future, we will enhance the model using a Deep Learning algorithm and as a result, the model will be able to predict various diseases besides Adenovirus. This ML model is highly recommended to the health ministry and scientists for further research. This ML model is available at open source medium free of cost so that each and every individual can avail of the benefits from this model.

Acknowledgments

The authors thank the management, JIS College of Engineering, and JIS GROUP for providing all of the necessary R&D resources and facilities.

References

- 1) Danley K, Kent P. 4-month-old boy coinfecting with COVID-19 and adenovirus. *BMJ Case Reports*. 2020;13(6):e236264. Available from: <https://doi.org/10.1136/bcr-2020-236264>.
- 2) Adenovirus: Symptoms, Causes & Treatment. 2023. Available from: <https://my.clevelandclinic.org/health/diseases/23022-adenovirus>.
- 3) Why adenovirus has doctors in West Bengal concerned?. 2023. Available from: <https://www.indiatoday.in/india-today-insight/story/why-adenovirus-has-doctors-in-west-bengal-concerned->.
- 4) Bengal: 5 Children Died In Kolkata Due To Respiratory Infections Amid Adenovirus Scare. 2023. Available from: <https://news.abplive.com/news/india/west-bengal-5-children-died-in-kolkata-due-to-respiratory-infections-amid-adenovirus-scare-1585199>.
- 5) What is adenovirus?. 2023. Available from: <https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/acute-severe-hepatitis-in-kids-covid-adenovirus-co-infection-detected-in-20-cases-over-170-infected-so-far-across-11-countries-know-the-signs-and-symptoms/photostory/91122206.cms>.
- 6) Kremer EJ. What is the risk of a deadly adenovirus pandemic? *PLOS Pathogens*. 2021;17(9):e1009814. Available from: <https://doi.org/10.1371/journal.ppat.1009814>.
- 7) Adenovirus: Background, Pathophysiology, Epidemiology. Accessed. 2021. Available from: <https://emedicine.medscape.com/article/211738-overview>.
- 8) Encyclopedia of Virology - 4th Edition. 2023. Available from: <https://www.elsevier.com/books/encyclopedia-of-virology/bamford/978-0-12-814515-9>.
- 9) Das S, Sanyal MK. Machine intelligent diagnostic system (MIDS): an instance of medical diagnosis of tuberculosis. *Neural Computing and Applications*. 2020;32(19):15585–15595. Available from: <https://doi.org/10.1007/s00521-020-04894-8>.
- 10) Iii JPL, Kajon AE. Adenovirus: Epidemiology, Global Spread of Novel Serotypes, and Advances in Treatment and Prevention. *Semin Respir Crit Care Med*. 2016;37(04):586–602. Available from: <https://doi.org/10.1055/s-0036-1584923>.
- 11) Coronavirus. 2023. Available from: [https://www.who.int/health-topics/coronavirus\(accessedMay05](https://www.who.int/health-topics/coronavirus(accessedMay05).
- 12) Fuzzy dilation membership function. 2023. Available from: <https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/fuzzy-membership.htm>.
- 13) Fuzzy Sets and Fuzzy Logic. 2023. Available from: [http://www.computing.surrey.ac.uk/ai/PROFILE/fuzzy#Prop\(accessed](http://www.computing.surrey.ac.uk/ai/PROFILE/fuzzy#Prop(accessed).
- 14) Wen S, Lin Z, Zhang Y, Lv F, Li H, Zhang X, et al. The Epidemiology, Molecular, and Clinical of Human Adenoviruses in Children Hospitalized With Acute Respiratory Infections. *Frontiers in Microbiology*. 2021;12. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.629971>.
- 15) Biserni GB, Scarpini S, Dondi A, Biagi C, Pierantoni L, Masetti R, et al. Potential Diagnostic and Prognostic Biomarkers for Adenovirus Respiratory Infection in Children and Young Adults. *Viruses*. 2021;13(9):1885. Available from: <https://doi.org/10.3390/v13091885>.
- 16) Machine_Learning/Adenoviruses at main , AmartaKundu/Machine_Learning. 2023. Available from: [https://github.com/AmartaKundu/Machine_Learning\(accessedMay05](https://github.com/AmartaKundu/Machine_Learning(accessedMay05).
- 17) Shieh WJ. Human adenovirus infections in pediatric population - An update on clinico-pathologic correlation. *Biomedical Journal*. 2022;45(1):38–49. Available from: <https://doi.org/10.1016/j.bj.2021.08.009>.