

AI For Predicting Disease Outbreaks: Developing AI Systems Capable of Predicting Disease Outbreaks By Analyzing Global Health Data and Trends

Vivek Yadav*

Citation: Yadav V. AI For Predicting Disease Outbreaks: Developing AI Systems Capable of Predicting Disease Outbreaks By Analyzing Global Health Data and Trends. *J Artif Intell Mach Learn & Data Sci* 2023, 1(1), 946-953. DOI: doi.org/10.51219/JAIMLD/vivek-yadav/226

Received: 02 January, 2023; **Accepted:** 18 January, 2023; **Published:** 20 January, 2023

*Corresponding author: Vivek Yadav, USA, E-mail: Yadav.Vivek@myyahoo.com

Copyright: © 2023 Yadav V., This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

This study investigates disease outbreak prediction using artificial intelligence (AI) based on machine learning. Using random forests, decision trees, and gradient models to analyze massive data sets, the study shows the potential for early risk detection and intervention. The results show the effectiveness of the ensemble approaches, reaching 100% forecast accuracy. The study emphasizes that improving the efficiency, transparency, and ethical use of models in public health systems requires optimizing and integrating multiple data sources.

Keywords: Artificial Intelligence, Machine Learning, Disease Outbreak Prediction, Random Forest, Gradient Boosting, Decision Tree, Public Health, Epidemic Preparedness, Data Analysis, Model Evaluation

1. Introduction

“Artificial Intelligence (AI)” based on machine learning impacts disease outbreaks by determining utilizing capable algorithms on huge datasets. The AI model can predict the likelihood of infection spread based on verifiable information and patterns in current information with negligible mistakes. This study examines the precision of a few exceptions such as random forest, decision tree, and gradient boosting models in the context of infection outbreaks. The performance of these epidemic models should be compared in this study using metrics such as accuracy, confusion matrix, and ROC curves to improve epidemic preparedness and response strategies.

2. Aims and Objectives

2.1. Aim

The main purpose of this research is to enhance the prediction of disease outbreaks using several machine-learning algorithms.

2.2. Objectives

- To Analyze the feature distribution and correlation of the

dataset using exploratory data analysis (EDA).

- To develop and evaluate several machines learning models, including Decision Tree, Random Forest, and Gradient Boosting, to predict disease outbreaks.
- To Compare the accuracy and performance of these models using confusion matrices and ROC curves.
- To Identify the best-performing model based on accuracy and AUC scores.
- To determine the optimal model based on accuracy ratings and AUC.
- Provide an analysis and recommendations for improving epidemic preparedness and response plans based on the model results.

2.3. Background

Nowadays in the era of AI, such complex diseases can be properly analyzed as well as easily predicted before any outbreak disease [1]. Modern approaches to predict diseases as a function of health risk assessment are the main process that has been done over the past few years with the help of AI. It transits from purely

statistical models as well as past data analysis with the help of AI techniques. AI not only helps in improving the healthcare sector but also its concerns about social welfare, and the environment which is affected are properly recognized by the AI in terms of recognizing the most outbreaking features.

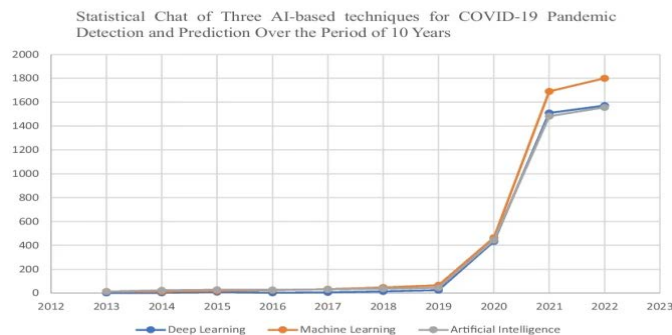


Figure 1: AI-based techniques for disease detection.

These models utilize machine learning algorithms such as neural systems and decision trees, in this way having the capability of learning from information and improving precision in their forecasts as time advances. This integration of technology and public health not only empowers the arrangement of proactive measures but also makes a difference within the coordination of effort and policymaking as it looks to decrease the impacts of communicable illnesses around the world.

3. Literature Review

3.1. Historical approaches to disease prediction

Disease prediction has not always been in the fashion revealing its historical background focused mainly on epidemiological models and statistical analysis. Some of this includes compartmentalized models for instance Susceptible, Infected, and Recovered (SIR) models that have helped in understanding the spread of disease in a population². Such models track the progression of illnesses due to certain features such as infection risks, population density, and usage of prevention methods including immunity boosters like vaccines or the implementation of isolation procedures. Yet it often depends on such assumptions and may therefore not be very accurate in certain conditions such as when business environments are dynamic and complex. Statistical methods also use data for analysis and coming up with results that show the possibility of an epidemic. In this case, such methodologies as time series analysis and spatial statistics have been used to identify statistically significant increases in disease rates or clustering of diseases by geography. Though useful for diagnostic studies, these methods are less successful in prognosis because of their dependency on past trends and stationary assumptions. Some of the drawbacks that accompany traditional methods include the fact that they build on the assumption that there is perfect information hence in resource-limited settings or during emerging epidemics the information may not be complete or accurate. Also, the time taken to propagate these approaches in offering results is highly ineffective in a constantly changing epidemic or the appearance of a new disease³. Even though they have their shortcomings, historical methods are instrumental in developing the fundamentals of modelling disease behavior and identifying plans of action. These give a benchmark against which newer and more vibrant approaches such as the use of artificial intelligence and machine learning can be compared.

3.2. Applications of AI in disease outbreak prediction

AI can predict diseases by using advanced computation and

integration of big and complex data and has positively impacted the prediction of disease outbreaks. One main application area of AI in this field is the reinforcement of the current epidemiological data with the help of machine learning models. AI models can generate insights from large volumes of data such as patients' records, environmental characteristics, global mobility trends, and people's interactions on social networks. Thus, AI is capable of identifying preliminary signs of emerging diseases by analyzing similarities and differences between these datasets⁴. Machine learning techniques allow AI to pay attention to the textual data obtained from news articles, social media postings, epidemic announcements, and the like, in real-time. This capability enables the quick identification of the disease signals such as the symptoms or instances that are not usually picked by the normal surveillance systems. Moreover, high-level Thinking Machines are capable of adjusting the predictions made by them in real time based on new data that is fed to them hence making them more accurate and perceptive. For example, the appropriate disease transmission path mathematical models and analysis of control measures' efficiency are provided by machine learning algorithms to help officials. The evaluation of heterogeneously structured data streams is also possible and strengthens the setting up of early warning systems. Such systems can be used to warn relevant authorities of possible epidemics before they worsen; this will allow for timely actions like the administration of vaccines or the isolation of people who have contracted the virus. Also, AI adds to augmenting health security on the worldwide level because it can also be used for prognostications of the future epidemics of new pathogens and for modifying the existing models that can be adapted to a novel pathogen.

3.3. Case studies and success stories

Google Flu Trends: In the later year of 2008, Google developed a system called flu trends aimed at tracking flu activity in almost real-time using analyses of the frequency of the related search terms. Google even managed to predict certain flu outbreaks up to two weeks before the official surveillance data due to the analysis of the search terms associated with flu symptoms, proving that non-conventional data sources can be indeed particularly useful for early warning systems⁵.

Blue Dot's COVID-19 Prediction: A Canadian startup for artificial intelligence called Blue Dot also became a center of attention for the accurate calculation of the coronavirus emergence. Through intelligent linguistics and data mining, Blue Dot offered vital travel data, newspapers, and other significance to the client's governments and healthcare units about the potential reach of the outbreak days before the official discourse⁶.

Predicting Dengue Outbreaks in Brazil: To provide a solution to the problem, researchers from the University of São Paulo proposed an AI model to predict dengue fever cases. Specifically, demographic conditions, and climate information, as well as the historical presence of the diseases and mosquitoes, allowed the model to forecast disease outbreaks several months in advance. This early warning helped public health workers to plan for targeted action to control the diseases and thus the disease incidence is brought down.

Primmed-Mail: Even though not AI-based, Primmed-Mail employs digital surveillance to capture new infectious diseases by specialists for identification around the world. that it provides

a human-to-machine interface for sharing and processing reports on outbreaks of diseases, as an example of the efficient synergy of people and computer solutions in increasing the chances of early identification and eradication of threats.

The aforementioned cases show how AI can enhance conventional practices in surveillance; the algorithms make predictions more quickly and with higher accuracy, allowing for early action to halt the spread of diseases⁷. They stress the innovative roles of AI in international health emergencies and management and bring about improvements for further development of disease predicting and contagion prevention methodology.

4. Challenges and Ethical Considerations

The adoption of AI for disease outbreak prediction presents several challenges and raises important ethical considerations that need careful attention: The adoption of AI for disease outbreak prediction presents several challenges and raises important ethical considerations that need careful attention:

- **Data Quality and Bias:** It looks like more heavily relying on data quality and representativeness of the models. Assumptions present in data-sample bias for example-make the model's predictions less accurate from an organizational equity perspective and therefore unhelpful to equality of outcomes in health care. Diversification of the data collected is essential to eliminate these biases.
- **Interpretability and Transparency:** Machine learning that is used in AI is largely predictive and tends to be opaque which means that it is difficult to discern a given set of inputs to get to a given output. With regards to the topic, lack of transparency can worsen the relations between different health stakeholders including the health sector employees, the government, and other individuals in society. It is critical to make AI systems explainable and provide the rationale for action and/or decision-making.
- **Privacy and Data Security:** The main issue regarding the employment of the data is privacy since using personal health data in the models. The privacy of individuals especially their personal detailed information is something that needs to be safeguarded against fraud and other violations. This indicates that while the sharing of patients' data with healthcare research is commendable to promote the provision of best practices and innovation, there is a need to balance it with state and individual patients' right to privacy of their information⁸.
- **Integration into Public Health Systems:** However, folding AI-aided predictions into existing public health environments provides practical and organizational concerns. Such AI outputs may be unfamiliar to many healthcare professionals and, thus, they may need education/training on how to interpret these outputs and how these could be integrated into their decision-making⁹. Moreover, there is a need to advance the principles of applied AI technologies' stability and universality in various healthcare domains.
- **Literature Gap:** The literature reveals a gap in the integration of AI predictions with existing public health systems, particularly concerning the need for training healthcare professionals and ensuring the stability and universality of AI applications across diverse healthcare environments. Additionally, addressing data quality and bias remains a significant challenge.

5. Methodology

5.1. Data Collection

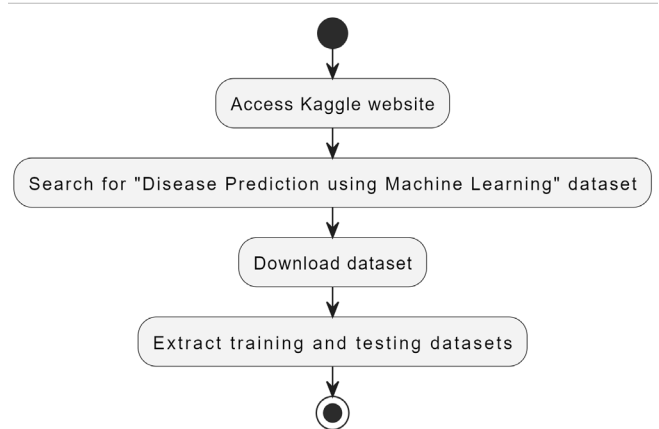


Figure 2: Data collection flow chart.

Data for this research is sourced from Kaggle, a reputable platform for datasets, specifically from the “Disease Prediction using Machine Learning” dataset available at this link: <https://www.kaggle.com/datasets/kaushil268/disease-prediction-using-machine-learning>. This dataset contains various types of rows and two CSV files one is train and test¹⁰.

5.2. Data Preprocessing

Accuracy

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

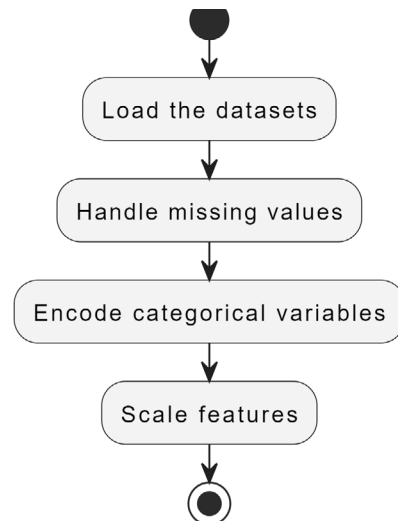


Figure 3: Data preprocessing.

Data preprocessing is one of the most important steps for machine learning, namely the preparation of the learning dataset¹¹. Therefore, the model training process involves the conversion of categorical variables to numerical form by applying one hot encoding or label encoding technique to the disease names and symptoms. Further, this process normalizes the range of independent variables using feature scaling to make models such as Gradient Boosting work efficiently. These preprocessing steps are important as they provide clear data sets that must be used for optimal machine learning model design.

5.3. Decision Tree

Entropy

$$H(X) = - \sum_{i=1}^n p_i \log_2(p_i)$$

Where;

$H(X)$ is the entropy.
 i is the probability of class i

Information Gain

$$IG(T, X) = H(T) - \sum_{i=1}^n \frac{|T_i|}{|T|} H(T_i)$$

$IG(T, X)$ is the information gained for the feature X

$H(T)$ is the entropy of the entire dataset.

$H(T_i)$ is the entropy of the subset T_i after splitting by feature X

5.4. Exploratory Data Analysis (EDA)

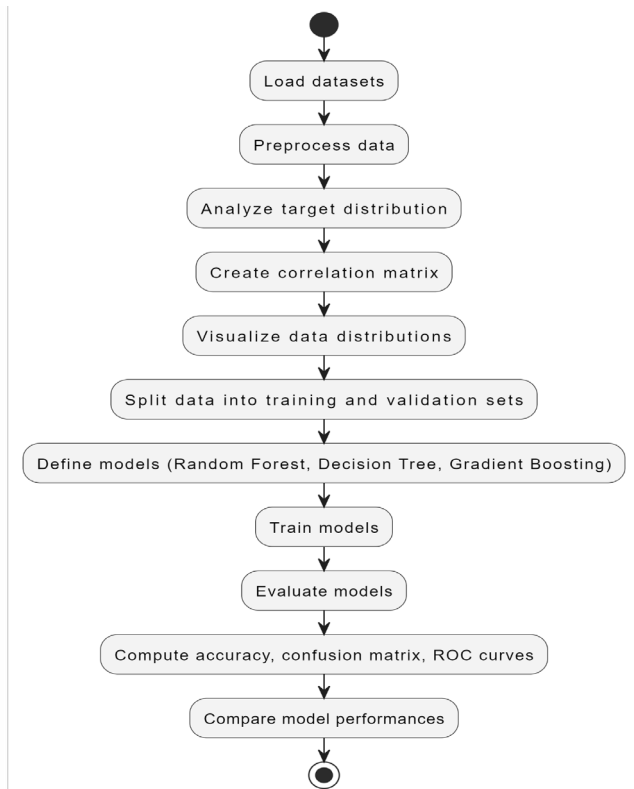


Figure 4: Flowchart of disease outbreak prediction methodology.

The Analysis of the Characteristics of the Dataset is also called Exploratory Data Analysis to get an idea of the data set’s distribution. This involves making a density plot of the target variable, which assists in the determination of class balance and bias¹². Together with correlation analysis cross-tabulations and quantile analysis are used to find relationships between features, which is useful to perform feature selection and feature engineering. Charts such as histograms, bar plots, and heat plots are applied to display the distribution and interdependence of data, which give the viewer a tangible initial sense of the data being analyzed. By integrating EDA, it becomes easier to identify patterns and other peculiarities as a lead-up to the modelling phase.

5.5. Model training and evaluation

Gradient Boosting

$$F_m(x) = F_{m-1}(x) + \eta h_m(x)$$

$F_m(x)$ is the m -th model

η is the learning rate

$h_m(x)$ is the base learner.

Model preparation and assessment include building numerous machine-learning models and surveying their execution. In this study, Random Forest, Decision Tree, and Gradient Boosting models are prepared utilizing the preprocessed data. A set of parameters is changed for each show to optimize execution [13]. Metrics counting precision, confusion matrix, and ROC curves are utilized to assess the models within the approval set after preparation. These measurements give a comprehensive diagram of each model’s prescient capacity and offer assistance select the optimal demonstration. To guarantee that the models are exact and reliable in predicting blasts, they are tested.

6. Result and Discussion

51:

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...	scurring	skin_peel
0	1	1	1	0	0	0	0	0	0	0	...	0	0
1	0	1	1	0	0	0	0	0	0	0	...	0	0
2	1	0	1	0	0	0	0	0	0	0	...	0	0
3	1	1	0	0	0	0	0	0	0	0	...	0	0
4	1	1	1	0	0	0	0	0	0	0	...	0	0
5	0	1	1	0	0	0	0	0	0	0	...	0	0
6	1	0	1	0	0	0	0	0	0	0	...	0	0
7	1	1	0	0	0	0	0	0	0	0	...	0	0
8	1	1	1	0	0	0	0	0	0	0	...	0	0
9	1	1	1	0	0	0	0	0	0	0	...	0	0

10 rows × 134 columns

Figure 5: First few rows of train dataset.

The above figure shows the first rows of the training dataset representing the various symptoms associated with the diagnosis of “fungal infection”. The lines of the table, which are quiet clinical cases, and the cells, which show a twofold framework where 1 demonstrates the presence of indications and demonstrates their absence, reflect the side effects, which incorporate itching, skin rash, and nodal skin emissions¹⁴. The first observation provides data corresponding to 1 for skin rash, 1 for itching, and 1 for nodal skin eruptions. This dataset is very useful for training machine learning models that forecast diseases based on observable symptom patterns.

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...	blackheads	scurring
0	1	1	1	0	0	0	0	0	0	0	...	0	0
1	0	0	0	1	1	1	0	0	0	0	...	1	0
2	0	0	0	0	0	0	0	1	1	1	...	1	0
3	1	0	0	0	0	0	0	0	0	0	...	0	0
4	1	1	0	0	0	0	0	1	0	0	...	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0
6	0	0	0	0	0	0	0	0	0	0	...	0	0
7	0	0	0	0	0	0	0	0	0	0	...	0	0
8	0	0	0	0	0	0	0	0	0	0	...	0	0
9	0	0	0	0	0	0	0	0	0	0	...	0	0

Figure 6: First Few rows of the test dataset.

This figure represents the first rows of the test dataset that contain seven symptoms. The first row marks such features as itching, skin rash, nodal skin eruptions as well a diagnosis of “Fungal infection”. Likewise, other rows depict groups of symptoms causing such diagnoses as “Allergy”, “GERD”, and “Chronic cholestasis”. This dataset is used to assess the efficacy of the trained Artificial Intelligence models on a particular platform.

A Horizontal bar chart showing the distribution of different diseases in the data set is illustrated in the above figure. Bar lengths for each disease represent the actual count of instances in a given dataset¹⁵. This type of visualization helps identify the distribution of classes-If it is balanced or if there are one or more diseases that dominate the dataset. The creation of balanced datasets is preferred for training, as it reduces the risk of the model favoring diseases with higher numbers.

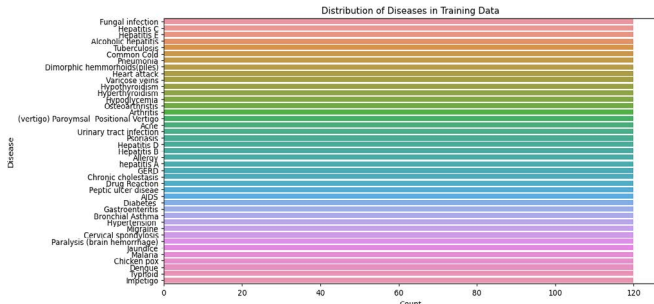


Figure 7: Distribution of diseases in training data.

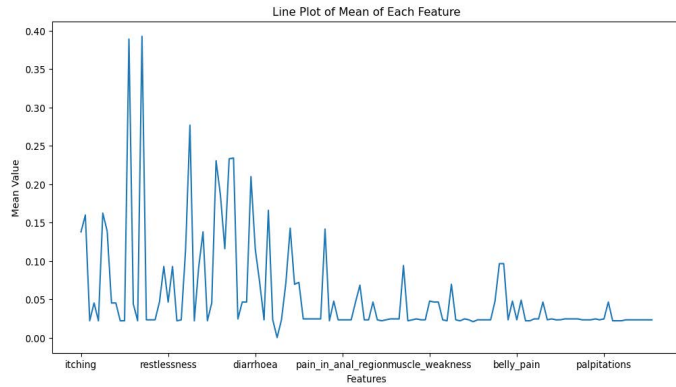


Figure 8: Line plot of mean of each feature.

This figure shows a line of the mean for each symptom over the entire data set. Symptoms are listed on the x-axis and averages are shown on the y-axis. This number facilitates the calculation of the average frequency of each symptom¹⁶. The peaks of the line indicate the symptoms that occur most often in the cases and can be more important signs of the disease.

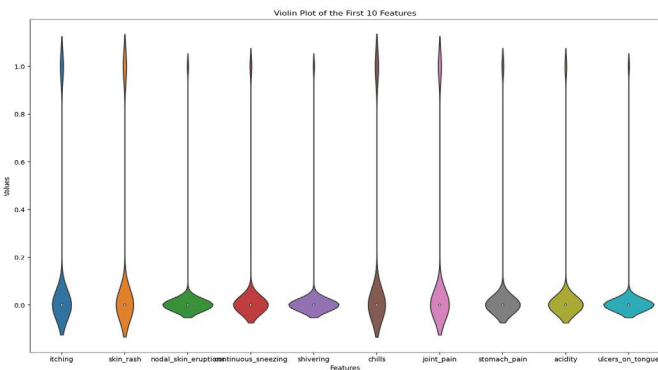


Figure 9: Violin plot of the first 10 features.

The above figure shows violin plots for the first ten symptoms in the dataset. Each vial chart shows the distribution of data points for a particular symptom. The width of the curve for different y values indicates the density of the data points¹⁷.

The uniform distribution of the itch symptom is shown in the image, highlighting areas of high and low frequency. The “Kernel Density Estimation (KDE)” plot of the first symptom of the dataset is shown in the image above¹⁸. Unlike histograms, KDE plots help show the underlying distribution of a function without being limited to individual cells.

In the figure above, the histogram shows the frequency distribution of the “it c hing” symptom. The x-axis shows alternative values (0 or 1), while the y-axis shows the number of occurrences of each value¹⁹. This histogram shows the significant imbalance in the data set for this feature, as most patients do not show any symptoms of “itching”.

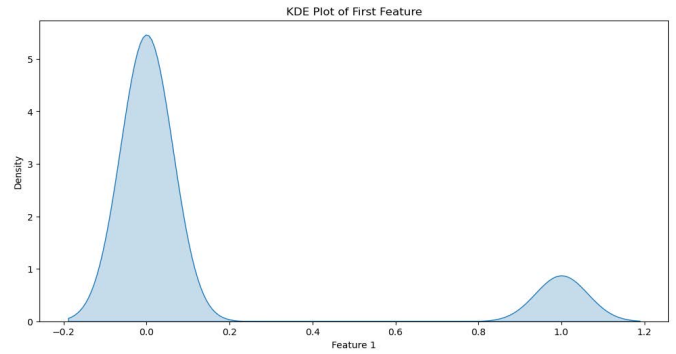


Figure 10: KDE Plot of First Feature.

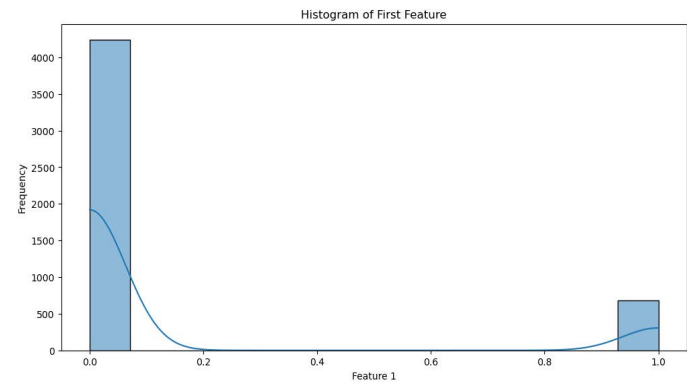


Figure 11: Histogram of first feature.

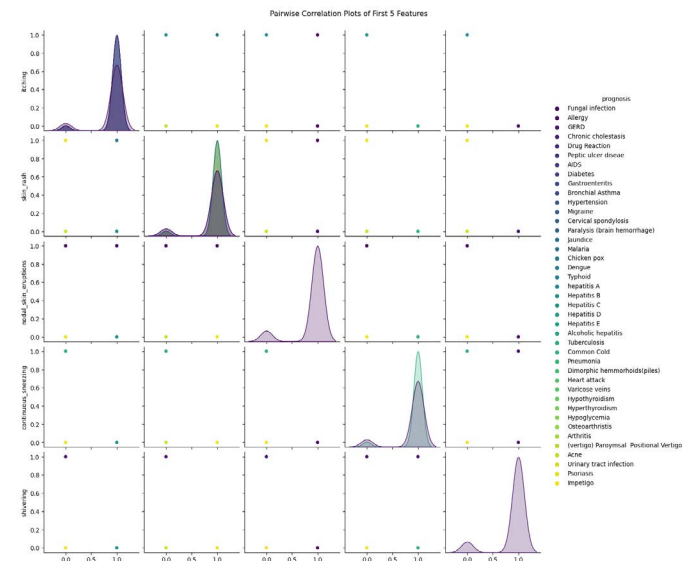


Figure 12: Pairwise correlation plots of first 5 features.

The above figure shows pairwise correlation plots of the first five symptoms. These spreads show the associations between each set of symptoms with different colors indicating different diseases. Diagonal plots show the distribution of each different symptom²⁰. Using these graphs to see potential relationships between significant symptoms can help determine how predictive features are when added to a machine-learning model.

The accuracy results for three models “Random Forest, Decision Tree, and Gradient Boosting”. The Random Forest and Gradient Boosting models both achieve perfect accuracy with scores of 1.0000, indicating their exceptional performance in predicting disease outbreaks. The Decision Tree model shows a significantly lower accuracy of 0.1128, highlighting its inadequacy for this task²¹. This comparison underscores the superiority of ensemble methods in handling complex prediction problems.

```

j]: # Training and evaluating the models
accuracy_scores = {}
conf_matrices = {}
roc_curves = {}

for model_name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_val)
    accuracy = accuracy_score(y_val, y_pred)
    accuracy_scores[model_name] = accuracy
    print(f"{model_name} Accuracy: {accuracy:.4f}")

# Confusion Matrix
conf_matrix = confusion_matrix(y_val, y_pred)
conf_matrices[model_name] = conf_matrix

# ROC Curve and AUC
y_val_bin = label_binarize(y_val, classes=y.unique())
y_pred_prob = model.predict_proba(X_val)
fpr, tpr, _ = roc_curve(y_val_bin.ravel(), y_pred_prob.ravel())
roc_auc = roc_auc_score(y_val_bin, y_pred_prob, multi_class='ovr')
roc_curves[model_name] = (fpr, tpr, roc_auc)

Random Forest Accuracy: 1.0000
Decision Tree Accuracy: 0.1128
Gradient Boosting Accuracy: 1.0000
    
```

Figure 13: Model Accuracy.

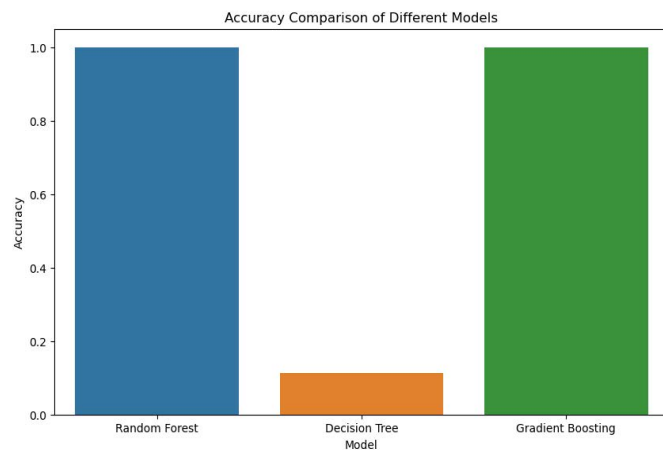


Figure 14: Accuracy comparison of different models.

It presents the results of the three models in terms of accuracy as represented by the bar chart below. Under the x-axis, there are the models indicated, while the y-axis represents the accuracy ratio. From the chart, it is evident that some models are better than others whereas the Random Forest and Gradient Boosting models are better than the Decision Tree model. This comparison is very useful in arriving at the right model to use in predicting disease outbreaks as changes in colors allow for easily distinguishable boundaries between the models²².

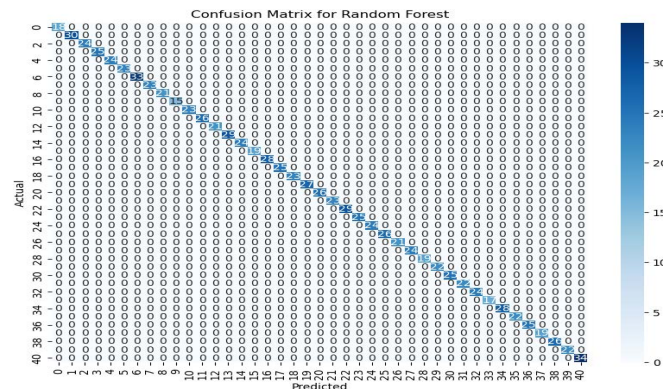


Figure 15: Confusion matrix for random forest.

The Confusion Matrix of the Random Forest model depicts the efficiency of the model that distinguishes between actual

and predicted classes. The diagonal form shows the number of instances actually in the class for classification while the off-diagonal elements show the number of wrong classifications made by the inherent classifiers²³. The elements of this matrix reflect a perfect classification of all the data points, indicated by the off-diagonal element of zero, and this is well supported by the accuracy of one for the Random Forest model 1.0000.

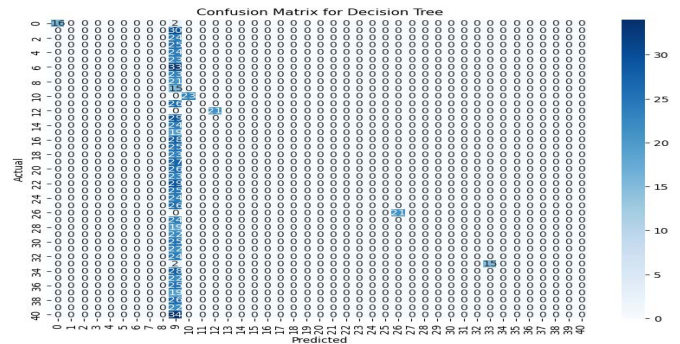


Figure 16: Confusion matrix for decision tree.

From the confusion matrix, it becomes evident that the Decision Tree model has numerous off-diagonal figures, thus portraying a poor ability to classify the assets correctly. The probability remains ambiguous for separating the features and the model fails to predict the right class where the accuracy is only 0.1128. This matrix demonstrates that the proposed model is unable to adequately accommodate the features of this dataset²⁴.

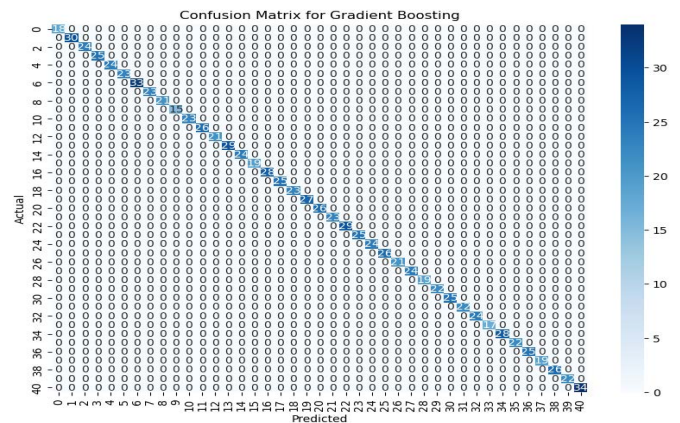


Figure 17: Confusion matrix for gradient boosting.

The confusion matrix for the Gradient Boosting model shows a near-perfect classification with most instances correctly predicted along the diagonal. The sparse off-diagonal elements can manage the suggestions in very few misclassifications, aligning with the Gradient Boosting model's high accuracy of 1.00²⁵.

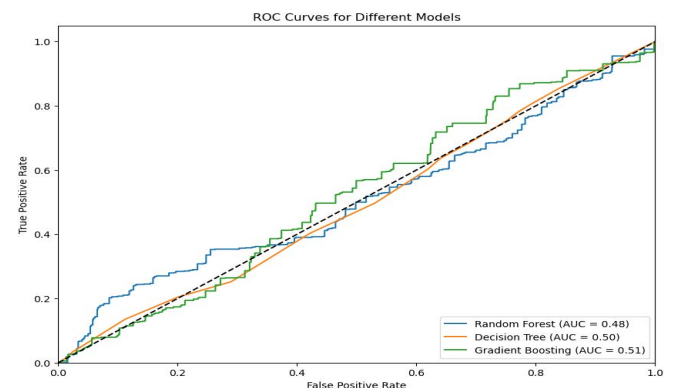


Figure 18: ROC curves for different models.

This specific ROC curve displays a multi-line graph between the “True Positive rate” and “False positive rate”²⁶. It illustrates that the AUC of “Random Forest” is 0.48, the AUC of “Decision tree” demonstrating is 0.50 and lastly, the AUC for Gradient boosting states is 0.51. Thus, it is clearly showing that the true positive rate is highest in Gradient boosting modeling.

	prognosis	predicted
0	Fungal infection	Fungal infection
1	Allergy	Allergy
2	GERD	GERD
3	Chronic cholestasis	Chronic cholestasis
4	Drug Reaction	Drug Reaction

Figure 19: First few rows of the test set with predictions.

The few rows of the test developed alongside the predictions assembled by the best-performing measures, Random Forest as well as Gradient Boosting, document the authentic disease (prognosis) or the predicted disease. The predictions align

Table 1: Summary of model performance.

Metric	Random Forest	Decision Tree	Gradient Boosting
Accuracy	1.0000	0.1128	1.0000
AUC	0.48	0.50	0.51
Key Observations	Perfect accuracy, High misclassification, low misclassification	Near-perfect accuracy, rate, low accuracy	low misclassification
Confusion Matrix	Diagonal dominance, no off-diagonal elements	Scattered, many off-diagonal elements	Diagonal dominance, few off-diagonal elements
ROC Curve Characteristics	Moderate true/false positive rate	Balanced true/false positive rate	Moderate true/false positive rate
Test Set Predictions	Accurate predictions, high generalization	Inaccurate predictions, low generalization	Accurate predictions, high generalization

8. Conclusion

The study shows the enormous capacity when it comes to the use of machine learning models in disease outbreak prediction based on prominent symptoms. Along with that, the implementation of random forest and the gradient booster have been performed in the report. Various kinds of models have been implemented throughout the training and testing process in Python. On the other hand, the performance of the decision tree model is poorer when evaluated for an accuracy of 0. The GB classifier, for its part, achieved a score of 0.1128 and is found to be less effective when addressing more complicated data sets. This is the case of the random forest where the AUC of the ROC is 0.48, while the decision tree had a worse AUC of 0.50, and the gradient boost had a moderate AUC of 0.51 even though the corresponding ensemble models had very high accuracy. The analysis signifies the advantages of the methods of greater complexity in the problems of complex quantitative forecasting with the selection of a solid groundwork for further investigations in AI-aided outbreak forecasting.

9. Future Recommendations

The future study of this research may consider how the tunings and parameters of the Random Forest and Gradient

Boosting models, increase the AUC scores. It may also be worth experimenting with techniques like hyperparameter optimization, k-fold cross-validation, and modern ensemble learning techniques to enhance the results. To enhance early disease signal detection, it is proposed to expand the existing data-combining process with the help of data from social platforms, patient records, and environmental monitoring devices²⁹. It may also have to create models that should incorporate real-time data to improve the system’s efficiency to make timely interventions and ensure adequate preparedness in the events of epidemics. Further work must be done towards engineering the models and methods used in AI to make them more transparent and interpretable to the various stakeholders in healthcare. Also, privacy and data security issues should be solved by implementing strong data protection measures. ethical use of patient data is significant in this context. The application of these recommendations could result in improved accuracy, credibility, and ethical means of applying AI systems for prognostication of disease outbreaks³⁰.

7. Discussion

The “Random Forest” and “Gradient Boosting” models achieve perfect accuracy as evidenced by their confusion matrices and accuracy scores of 10000. The decision tree model is significantly weaker with an accuracy of 0.1128, indicating its inadequacy for this task in the areas that carry the high accuracy means in such onsen\mile modules. Despite the high accuracy of the ensemble models, the ROC curves show moderate AUC values, suggesting a possible need for further optimization to improve sensitivity and specificity²⁸. The first lines of the test series with predictions confirm the strong generalization ability of the Random Forest model. Overall, the results highlight the superiority of ensemble methods in solving complex forecasting problems and suggest that although accuracy is high, the balance of true and false positives should be improved to ensure comprehensive outbreak forecasting.

10. References

1. Shah H, Shah S, Tanwar S, Gupta R, Kumar N. Fusion of AI techniques to tackle COVID-19 pandemic: models, incidence rates, and future trends. Multimedia Systems 2022; 1-34.

2. Jamshidi MB, Roshani S, Talla J, et al. A review of the potential of artificial intelligence approaches to forecasting COVID-19 spreading. *AI* 2022;3: 493-511.
3. Kaur I, Behl T, Aleya L, et al. Artificial intelligence as a fundamental tool in management of infectious diseases and its current implementation in COVID-19 pandemic. *Environmental Science Pollution Research* 2021;28: 40515-40532.
4. Agbehadji IE, Awuzie BO, Ngowi AB, Millham RC. Review of big data analytics, artificial intelligence, and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing. *Int J environmental research public health* 2020;17: 5330.
5. Singh P, Kaur R. An integrated fog and Artificial Intelligence smart health framework to predict and prevent COVID-19. *Global transitions* 2020;2: 283-292.
6. Aldahiri A, Alrashed B, Hussain W. Trends in using IoT with machine learning in health prediction systems. *Forecasting* 2021;3: 181-206.
7. Khan FN, Khanam AA, Ramlal A, Ahmad S. A review on predictive systems and data models for covid-19. *Computational intelligence methods in COVID-19: Surveillance, prevention, prediction and diagnosis* 2020; 123-164.
8. Arora G, Joshi J, Mandal RS, Shrivastava N, Virmani R, Sethi T. Artificial intelligence in surveillance, diagnosis, drug discovery and vaccine development against COVID-19. *Pathogens* 2021;10: 1048.
9. Jamshidi M, Roshani S, Daneshfar F, et al. Hybrid deep learning techniques for predicting complex phenomena: A review on COVID-19. *AI* 2022;3: 416-433.
10. Kaur S, Singla J, Nkenyereye L, et al. Medical diagnostic systems using artificial intelligence (ai) algorithms: Principles and perspectives. *IEEE Access* 2020;8: 228049-228069.
11. Khakharia A, Shah V, Jain S, et al. Outbreak prediction of COVID-19 for dense and populated countries using machine learning. *Ann Data Science* 2021;8: 1-19.
12. Kiru MU, Belaton B, Xingying C, Aminu M, Abubakar BS. A proposed framework for pandemic control using Artificial Intelligence: A case study of covid-19 global pandemic. *J Computer Science and Its Application* 2021;28.
13. Heidari A, Jafari Navimipour N, Unal M, Toumaj S. Machine learning applications for COVID-19 outbreak management. *Neural Computing and Applications* 2022;34: 15313-15348.
14. Piccialli F, Di Cola VS, Giampaolo F, Cuomo S. The role of artificial intelligence in fighting the COVID-19 pandemic. *Infor Systems Front* 2021;23: 1467-1497.
15. Sinha A, Rathi M. COVID-19 prediction using AI analytics for South Korea. *Applied Intelligence* 2021;51: 8579-8597.
16. Swayamsiddha S, Prashant K, Shaw D, Mohanty C. The prospective of artificial intelligence in COVID-19 pandemic. *Health and Technology* 2021;11: 1-10.
17. Surianarayanan C, Chelliah PR. Leveraging artificial intelligence (AI) capabilities for COVID-19 containment. *New generation computing*, 2021;39: 717-741.
18. Musulin J, Baressi Šegota S, Štifanić D, et al. Application of artificial intelligence-based regression methods in the problem of COVID-19 spread prediction: a systematic review. *Int J environmental research public health* 2021;18: 4287.
19. Tambo E, Ai L, Zhou X, et al. Surveillance-response systems: the key to elimination of tropical diseases. *Infectious diseases of poverty* 2014;3: 1-11.
20. Humayun M, Alsayat A. Prediction model for coronavirus pandemic using deep learning. *Computer Systems Science Engineering* 2022;40: 947-961.
21. Demertzis K, Taketzis D, Tsiotas D, Magafas L, Iliadis L, Kikiras P. Pandemic analytics by advanced machine learning for improved decision making of COVID-19 crisis. *Processes* 2021;9: 1267.
22. Martin-Moreno JM, Alegre-Martinez A, Martin-Gorgojo V, Alfonso-Sanchez JL, Torres F, Pallares-Carratala V. Predictive models for forecasting public health scenarios: practical experiences applied during the first wave of the COVID-19 pandemic. *Int J environmental research public health* 2022;19: 5546.
23. Devaraj J, Elavarasan RM, Pugazhendhi R, et al. Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant? *Results in Physics* 2021;21: 103817.
24. Corsi A, de Souza FF, Pagani RN, Kovaleski JL. Big data analytics as a tool for fighting pandemics: A systematic review of literature. *J ambient intelligence and humanized computing* 2021;12: 9163-9180.
25. Shariaty F, Shariati S, Navvabi S, Oshnari MN, Novikov B. Application of Artificial Intelligence for Rapid Prevention of Epidemic Diseases (COVID-19) *Int J Health Policy Planning* 2022;1: 18-27.
26. Alam F, Almaghthawi A, Katib I, Albeshri A, Mehmood R. IResponse: An AI and IoT-enabled framework for autonomous COVID-19 pandemic management. *Sustainability* 2021;13: 3797.
27. Taj RM, El Mouden ZA, Jakimi A, Hajar M. Towards using recurrent neural networks for predicting influenza-like illness: Case study of covid-19 in Morocco. *Int J Advanced Trends in Computer Science Engineering* 2020;9.
28. Chakraborty S, Paul H, Ghatak S, et al. An AI-based medical chatbot model for infectious disease prediction. *IEEE Access* 2022;10: 128469-128483.
29. Imran M, Zaman U, Imran, Imtiaz J, Fayaz M, Gwak J. Comprehensive survey of IoT, machine learning, and blockchain for health care applications: A topical assessment for pandemic preparedness, challenges, and solutions. *Electronics* 2021;10: 2501.
30. Imran M, Zaman U, Imran Imtiaz J, Fayaz M Gwak J. Comprehensive survey of iot, machine learning, and blockchain for health care applications: A topical assessment for pandemic preparedness, challenges, and solutions. *Electronics* 2021;10: 2501.