




Article

DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms

Gaurav Gupta ¹, Shakir Khan ^{2,3,*}, Vandana Guleria ⁴, Abrar Almjally ², Bayan Ibrahim Alabdullah ^{5,*}, Tamanna Siddiqui ⁶, Bader M. Albahlal ², Saad Abdullah Alajlan ² and Mashaal AL-subaie ²

¹ Yogananda School of AI, Computers and Data Sciences, Shoolini University, Solan 173229, India

² College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11564, Saudi Arabia

³ Department of Computer Science and Engineering, University Centre for Research and Development, Chandigarh University, Mohali 140413, India

⁴ School of Bioengineering & Food Technology, Shoolini University, Solan 173229, India

⁵ Department of Information System, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11564, Saudi Arabia

⁶ Department of Computer Science, Aligarh Muslim University, Aligarh 202001, India

* Correspondence: sgkhan@imamu.edu.sa (S.K.); bialabdullah@pnu.edu.sa (B.I.A.)

Abstract: The aedes mosquito-borne dengue viruses cause dengue fever, an arboviral disease (DENVs). In 2019, the World Health Organization forecasts a yearly occurrence of infections from 100 million to 400 million, the maximum number of dengue cases ever testified worldwide, prompting WHO to label the virus one of the world's top ten public health risks. Dengue hemorrhagic fever can progress into dengue shock syndrome, which can be fatal. Dengue hemorrhagic fever can also advance into dengue shock syndrome. To provide accessible and timely supportive care and therapy, it is necessary to have indispensable practical instruments that accurately differentiate Dengue and its subcategories in the early stages of illness development. Dengue fever can be predicted in advance, saving one's life by warning them to seek proper diagnosis and treatment. Predicting infectious diseases such as dengue is difficult, and most forecast systems are still in their primary stages. In developing dengue predictive models, data from microarrays and RNA-Seq have been used significantly. Bayesian inferences and support vector machine algorithms are two examples of statistical methods that can mine opinions and analyze sentiment from text. In general, these methods are not very strong semantically, and they only work effectively when the text passage inputs are at the level of the page or the paragraph; they are poor miners of sentiment at the level of the sentence or the phrase. In this research, we propose to construct a machine learning method to forecast dengue fever.

Keywords: dengue fever; machine learning; prediction; classification; SVM; decision tree; random forest; ANN; GNB; opinion mining; sentiment analysis



Citation: Gupta, G.; Khan, S.; Guleria, V.; Almjally, A.; Alabdullah, B.I.; Siddiqui, T.; Albahlal, B.M.; Alajlan, S.A.; AL-subaie, M. DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms. *Diagnostics* **2023**, *13*, 1093. <https://doi.org/10.3390/diagnostics13061093>

Academic Editor: Sameer Antani

Received: 7 February 2023

Revised: 28 February 2023

Accepted: 11 March 2023

Published: 14 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The *Aedes aegypti* mosquito handles transmitting the DENV virus, the causative agent of dengue fever, from person to person. There is currently no vaccination that can protect against all virus serologies. This is because there is no such thing as a vaccine. As a direct consequence, trying to reduce the number of mosquitoes in an area has become the primary focus of the fight against the disease. Researchers are using machine learning (ML) and deep learning (DL) to forecast dengue cases and assist governments in their fight against the disease [1].

Dengue virus is a flavivirus, a genus of flaviviruses, and a family of Flaviviridae [2,3]. Arthropods are the primary vectors for the spread of the dengue virus. It can be broken down into four serotypes, referred to by the names DEN 1, DEN 2, DEN 3, and DEN 4. According to the World Health Organization (WHO), dengue fever poses a considerable

risk to the public health of countries all over the world, especially those nations that are in tropical or subtropical regions (WHO). There are approximately 2.5 billion people who live in dengue-endemic areas [4], with an annual infection rate of 400 million individuals and a mortality rate that can vary from 5 to 20 percent in specific locations [5]. Dengue fever is a disease that may be found all over the world. However, it is particularly prevalent in some regions, such as Europe and the United States of America (USA) [6]. The first recorded case of dengue fever in India occurred in Madras in the year 1780, and the first pandemic of dengue disease in India was confirmed by virological testing in the years 1963–1964 [7].

Dengue fever is a dangerous illness that manifests similarly to the flu and can afflict persons of any age, including infants, children, adolescents, and adults [8]. The *Aedes aegypti* mosquito is to blame for the transmission of the disease to humans, which takes place more often during the wetter months [9]. The World Health Organization (WHO) distinguishes between two levels of severity for dengue fever: moderate and severe [10,11]. Extreme cases are characterized by an abnormal amount of bleeding, impairment of organ function, or significant loss of plasma, while others are considered to be relatively uncomplicated [11]. According to the categorization used in 1997, dengue fever can be broken down into three subtypes: undifferentiated fever, D.F., and DHF [12]. The DHF served as the basis for the creation of grades I–IV. D.F. can develop because of primary diseases as well as secondary diseases, and it is most common in adults and children who have developed into adults. The onset of symptoms is typically accompanied by a high temperature that cycles between phases and can continue anywhere from one week to three weeks [13]. A metallic taste, loss of appetite, diarrhea, nausea, and stomachache are some of the other symptoms. Other symptoms include severe headaches, particularly retrobulbar migraines, fatigue, myalgia, and painful joints. Dengue fever is also referred to as break-bone fever [10,14] because the condition is often accompanied by myalgia and discomfort in the joints.

The worldwide public health system has a continuing need for the early detection of dengue fever, and machine learning algorithms may help medical professionals recognize and prevent infection at an earlier stage. This would save time, money, and the uncomfortable experience of pathology testing [15,16]. Diagnostics would benefit significantly from this. In medical diagnostics, machine learning algorithms have been used to diagnose conditions based on clinical and laboratory signs and produce the outcome, as stated in [17]. This has been carried out so that the outcome will be as accurate as possible. They also note that artificial neural networks (ANN) are one of the most prominent ways of addressing medical diagnostic issues and that support vector machines (SVM) give correct conclusions when evaluating a single ailment. Both concepts are discussed in the article. These two assertions are contained within the same piece of writing altogether.

In another publication [18], the authors use an artificial neural network to assess meteorological data obtained from Singapore's National Environment Agency (SNEA). To make a prediction about dengue illness in Thailand, [19] employed meteorological data collection and implemented feature selection techniques. SVM is another well-known method that can be used to address this matter. Utilizing support vector machines (SVM) on a Singapore meteorological dataset can help with the prediction of dengue fever [20,21]. On a small sample from Brazil, Ref. [22] uses gene expression data and an RBF kernel. Authors of [23] also use the support vector regression (SVR) method. They use data from the Guangdong region to compare several machine learning algorithms (China). Ref. [24] use climatic factors to investigate the incidence of dengue fever in the Philippines. They compare random forest, gradient boosting, general additive modeling, and seasonal autoregressive integrated moving average with exogenous variables [25,26]. These days, deep learning strategies are getting a lot of attention as a potential solution to a variety of challenges, particularly in the discipline of medical imaging. Some of the features that will be included in this system are the ability to suggest differential diagnoses, the composition of preliminary radiology reports, automatic detection, and quantitative characteristics of the lesion in medical imaging. However, this does not imply that the replacement of

radiologists is dangerous; instead, it helps physicians provide more accurate diagnoses to their patients. The subfield of computer vision, known as deep learning, is considered an advanced subfield. The primary aim of computer vision is to carry out a variety of tasks simultaneously, including picture detection and recognition, image analysis, natural language processing, and other similar activities. Over the course of the past few years, interest in computer vision has grown substantially across a variety of academic domains. CNN is used in most computer vision tasks, particularly those involving the classification, recognition, and segmentation of medical pictures. The convolutional neural network (CNN) is a sort of artificial neural network that was developed specifically for processing data related to images and videos. It begins with photographs as input, then extracts and learns features from those images, and then classifies output images depending on the features it has learned from the input images. There have been several different CNN-based model ideas put forward, including AlexNet, SPP-Net, VGGNet, ResNet, GoogleLeNet, and others. Deep convolutional neural network (CNN)-based algorithms have shown promising results in the processing of medical pictures. An introduction to CNN in medical imaging analysis as well as a general discussion of machine learning and deep learning applied to medical pictures are included in this work. The researchers investigate several different machine learning methods, and general additive modeling is just one of them.

The contribution of work:

The purpose of this paper is to pursue an early diagnostic model that helps doctors in the prompt prognosis and diagnosis of dengue disease by using machine learning algorithms. The key steps are as follows:

1. Using techniques from the field of machine learning, such as the KNN classifier, decision tree, random forest, Gaussian naive Bayes, and support vector classifier (SVC), among others.
2. Creating a diagnostic model based on machine learning for fast detection and prognosis of dengue disease to aid medical professionals in making decisions.
3. The K-Fold method is used here for the purpose of result validation.

2. Related Work

The field of computing known as machine learning (ML) enables computers to access information without the requirement for any encoding [27,28]. The study of ML falls under the umbrella of the discipline of computer science. MML has become all-pervasive and essential for resolving intricate problems in any science department, but especially in the field of illness diagnostics [29,30]. Machine learning algorithms and techniques will soon be able to foresee and differentiate between a wide variety of illnesses in the healthcare field [30–32]. This is a direct effect of ongoing technological improvement. Machine learning is often cited as one of the most productive research approaches, mainly when predicting disease occurrence. There are several distinct kinds of ML algorithms, each of which is capable of being applied for the purpose of disease forecasting [33,34]. The findings of an investigation into several different machine learning algorithm approaches are shown in Table 1, along with the research that is pertinent to the topic. According to the findings of the review that was carried out [35,36], several distinct machine learning methods, including SVM, KNN, R.F., D.T., and SVC, are utilized and evaluated for the purpose of dengue prediction.

Table 1. A literature review regarding the dengue fever forecast.

Year	Reference	Machine Learning/Other Techniques	Findings
1995	[37]	<ul style="list-style-type: none"> • CIMSiM (Container-Inhabiting Mosquito Simulation Model) • DENSiM (Dengue Simulation Model) 	<ul style="list-style-type: none"> • Specifics regarding the weather • Bring the model to a close with an example (entomologic, demographic, and epidemiologic)

Table 1. Cont.

Year	Reference	Machine Learning/Other Techniques	Findings
2001	[38]	<ul style="list-style-type: none"> CIMSiM (Container-Inhabiting Mosquito Simulation Model) 	<ul style="list-style-type: none"> Rates of development and survival, global-scale and long-term climate data, monthly mean temperature, precipitation, and cloud cover
2002	[39]	<ul style="list-style-type: none"> Mathematical Approach 	<ul style="list-style-type: none"> Recruitment rate Biting rate Vector mortality The 2-serotype model is considered
2004	[40]	<ul style="list-style-type: none"> ELISA Method, which stands for “Enzyme-Linked ImmunoSorbent Assay.” RNA (RiboNucleic Acid) Extraction 	<ul style="list-style-type: none"> Hemagglutination inhibition Complement fixation The test of neutralization MAC-ELISA (IgM antibody capture enzyme-linked immunosorbent assay) (IgM antibody capture enzyme-linked immunosorbent assay) Indirect immunoglobulin G ELISA
2005	[41]	<ul style="list-style-type: none"> ANN MATLAB Toolbox 	<ul style="list-style-type: none"> Input neurons equal 9, hidden neurons equal 5, and output neurons equal 1. Iterations equal 25, accuracy equals 90%, and error equals 10%. Clinical and epidemiological data: total patients 252: (4 DF and 248 DHF patients). Accuracy in predictions of 90% An inaccuracy in the forecasting of 10%
2007	[42]	<ul style="list-style-type: none"> ELISA Method 	<ul style="list-style-type: none"> (IP-10 and I-TAC, A549, MG-132 and ALLN) New host genes associated with dengue sickness have been identified through objective research on gene expression, and these genes have been validated through available research.
2008	[43]	<ul style="list-style-type: none"> Wavelet, SVM, SVR, GA 	<ul style="list-style-type: none"> Rainfall, humidity, temperature
2008	[44]	<ul style="list-style-type: none"> Decision Tree (C4.5) 	<ul style="list-style-type: none"> Complete blood count, white blood cell/lymphocyte count The temperature of the body, hemoglobin count, or both Neutrophil count The total number of patients is 1200, of which 364 have dengue and 836 do not. The overall error rate is equal to 15.7% (after k-fold validation)
2008	[45]	<ul style="list-style-type: none"> PCR (Polymerase Chain Reaction) 	<ul style="list-style-type: none"> Hematological and biochemical Total patients 104: grade I DHF = 66; grade II DHF = 34; grade III DHF = 4; grade IV DHF = 3
2008	[46]	<ul style="list-style-type: none"> SARIMA (Seasonal Autoregressive Integrated Moving Average) Models 	<ul style="list-style-type: none"> Dataset from January 2000 to September 2007
2008	[47]	<ul style="list-style-type: none"> SPSS mRNA (messenger RNA) Analysis 	<ul style="list-style-type: none"> (NFKB1, NFKB2, TNFR1, IL1B, IL8, and TNFA) (TLR7, TLR4R3, TLR1, TLR2, TLR4R4, and TLR4 co-factor CD14) DHF = 56 children

Table 1. Cont.

Year	Reference	Machine Learning/Other Techniques	Findings
2009	[48]	<ul style="list-style-type: none"> • Rough Set Theory 	<ul style="list-style-type: none"> • Headache • Vomiting • Temperature
2010	[22]	<ul style="list-style-type: none"> • SVM • RBF 	<ul style="list-style-type: none"> • The number of genes is 12, including MYD88, TLR3/7/9, RIG1, IRF3/7, CLEC5A, IFN-γ, and MDA5. • The MYD88 and TLR7 RBF kernel functions, which have a c value of 1.0 and a c value of 10, respectively, are the most efficient genes. • Patients = 28; 15 DF, 13 DHF
2010	[49]	<ul style="list-style-type: none"> • ANN 	<ul style="list-style-type: none"> • Mean temperature. • Relative humidity of the air • total rainfall • Total dataset = 14,209 (dengue confirm cases)
2010	[43]	<ul style="list-style-type: none"> • PDB (Protein Data Bank) codes • cryoEM (Cryo-electron Microscopy) 	<ul style="list-style-type: none"> • E glycoprotein (envelope protein) • M protein (membrane protein)
2010	[4]	<ul style="list-style-type: none"> • SVM • RBF 	<ul style="list-style-type: none"> • Structural proteins consist of three components: capsid (C), membrane (M), and envelope (E). • The number of non-structural proteins equals 7 (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5)
2010	[50]	<ul style="list-style-type: none"> • Rough Set Analysis 	<ul style="list-style-type: none"> • Temperature • Headache • Vomiting
2010	[19]	<ul style="list-style-type: none"> • ANN 	<ul style="list-style-type: none"> • Temperature • Rainfall • Relative humidity • Result: highest accuracy 85.92%
2011	[51]	<ul style="list-style-type: none"> • Regression in a linear fashion (Step-down) • Generalize Boosted regression. • negative binomial Regression • Logistic regression • SVM 	<ul style="list-style-type: none"> • The SVM Model achieved a better result in the logistic regression (in both locations): • Area under curve (AUC) for the SVM models with the use of a cutoff of 0.906 (in Singapore) and 0.960 for the 75th percentile (in Bangkok)
2011	[52]	<ul style="list-style-type: none"> • Spatio-temporal Analysis 	<ul style="list-style-type: none"> • There are four criteria to consider: volume, location, time, and content. • The cluster approach can be used on Twitter to anticipate dengue fever outbreaks locally and in the immediate future.
2012	[53]	<ul style="list-style-type: none"> • CART Method Random Forest Method 	<ul style="list-style-type: none"> • Carry out ten separate experiments, each with tenfold independent validation. • The overall accuracy rate is an average of 84.0% (for D.F.) and 84.6% (for DHF) • AUC = 0.87

3. Materials and Methods

Within the scope of this publication, we constructed a diagnostic and prognostic model for dengue fever. We broke the task down into steps, beginning with the first phase of data collection, then moving on to data preprocessing, and finally employing ML classifiers to evaluate the output according to the accuracy (mean) of disease prediction (see Figure 1).

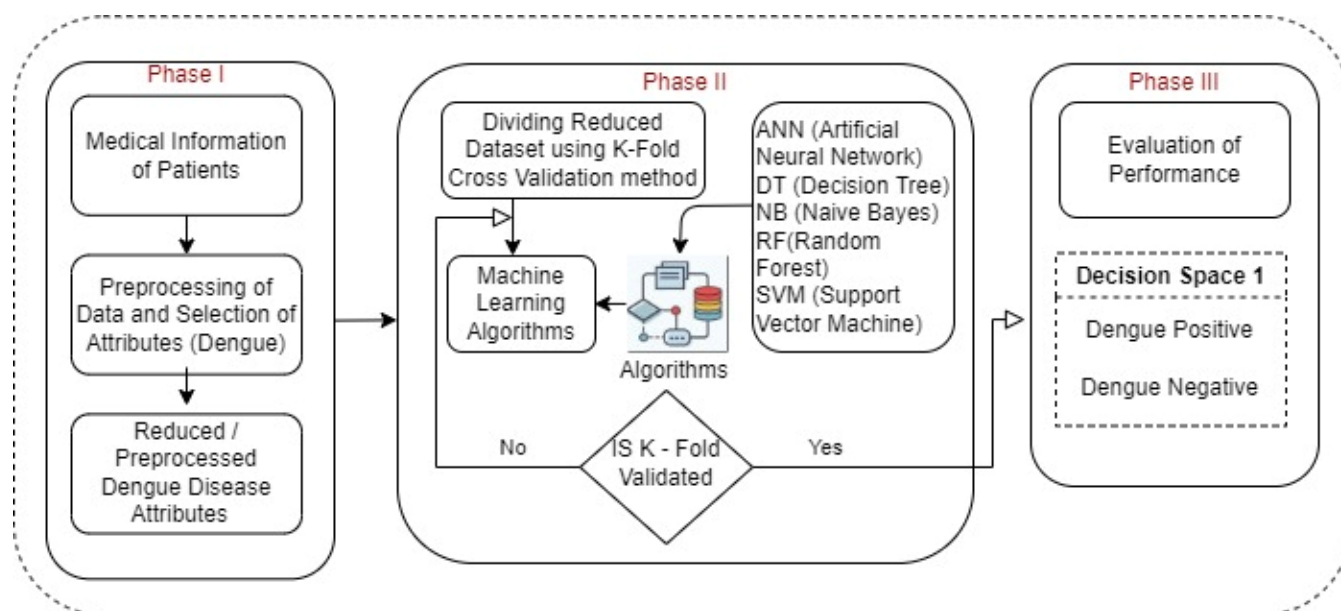


Figure 1. The workflow for the implementation of the proposed diagnostic model for the diagnosis of dengue disease.

3.1. Data Collection

The objective is to accurately predict the total number of dengue cases present in the test set, which will be labeled against each city, year, and week of the year. This study uses data from the DengAI competition (open data of dengue illness competition: DengAI: Predicting Disease Spread (drivendata.org)). The DengAI competition comprises data for two cities, San Juan and Iquitos, extending from three to five years. Every piece of information contributes its own set of forecasts for these cities. The data are separated into two categories: the training and test datasets, as shown in Table 2.

Table 2. Data in a separate dataset.

Dataset	Data
Training	1456
Testing	416
Total	1872

3.2. Data Preprocessing

The machine learning pipeline's most significant component is the step known as "data preprocessing." Data preprocessing converts unprocessed data into processed (meaningful) data. The dataset needs to be cleaned, normalized, and completely free of noise before it can be used for analysis (see Figure 2).

3.3. Features Selection

In building a prediction model, one of the most critical steps is called "feature selection." During this phase, the number of variables (or inputs) is narrowed down to reduce the amount of computing required for the modeling process and, in some cases, to improve

the overall performance of the model. The dataset has missing data for certain of its attributes, so we use the mean method to replace those values. After that, we use the fit and transform method to normalize and standardize the data.

	city	year	weekofyear	week_start_date	ndvi_ne	ndvi_nw	ndvi_se	ndvi_sw	precipitation_amt_mm	reanalysis_air_temp_k
1236	iq	2006	14	2006-04-09	0.241671	0.205400	0.156014	0.283314	41.68	297.774286
1314	iq	2007	41	2007-10-08	0.385186	0.405071	0.310629	0.438129	39.09	299.575714
1076	iq	2003	11	2003-03-12	0.200000	0.190286	0.158943	0.163629	47.97	296.680000
696	sj	2003	38	2003-09-17	-0.093900	0.086425	0.248900	0.224743	38.70	301.142857
75	sj	1991	41	1991-10-08	0.093000	0.093000	0.145083	0.168167	32.10	299.562857

Figure 2. Sample data were taken for the study.

We can see that there are several different features that have extreme values by looking at Figure 3. After investigating the data, it became clear that they are neither outliers nor errors; hence, we are unable to disregard them and will have to take them into consideration. The values of precipitation are taken into consideration here, and given that these are estimates of the amount of rain, it is reasonable to anticipate that the weather can vary significantly depending on the location.

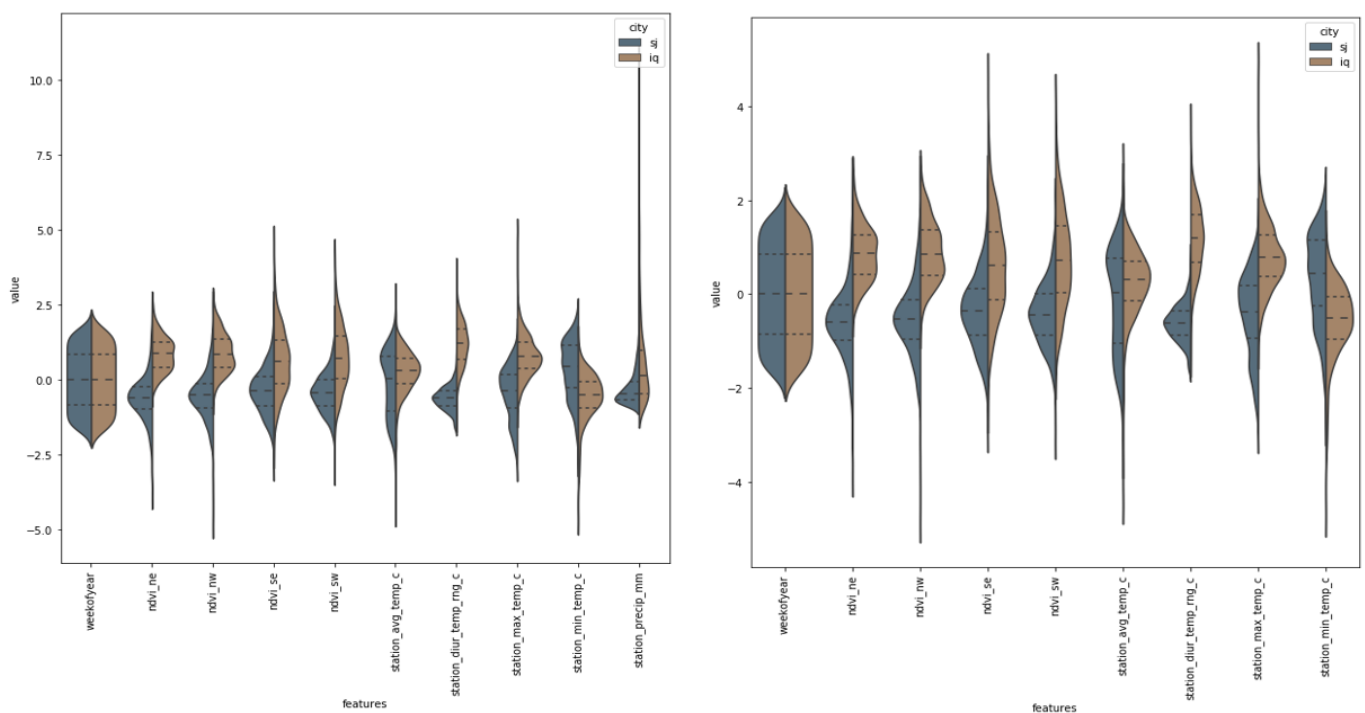


Figure 3. The dengue dataset utilizes a variety of features versus values for its parameters.

The features `reanalysis_avg_temp_k` and `reanalysis_specific_humidity_g_per_kg` appear to be pretty similar in shape; nonetheless, the question that arises here is whether or not they are correlated with one another.

By looking at Figure 4, we can come to the conclusion that certain features are perfectly associated with one another (1), while other features are practically perfectly correlated with one another (0.9). The same information is presented in Table 3.

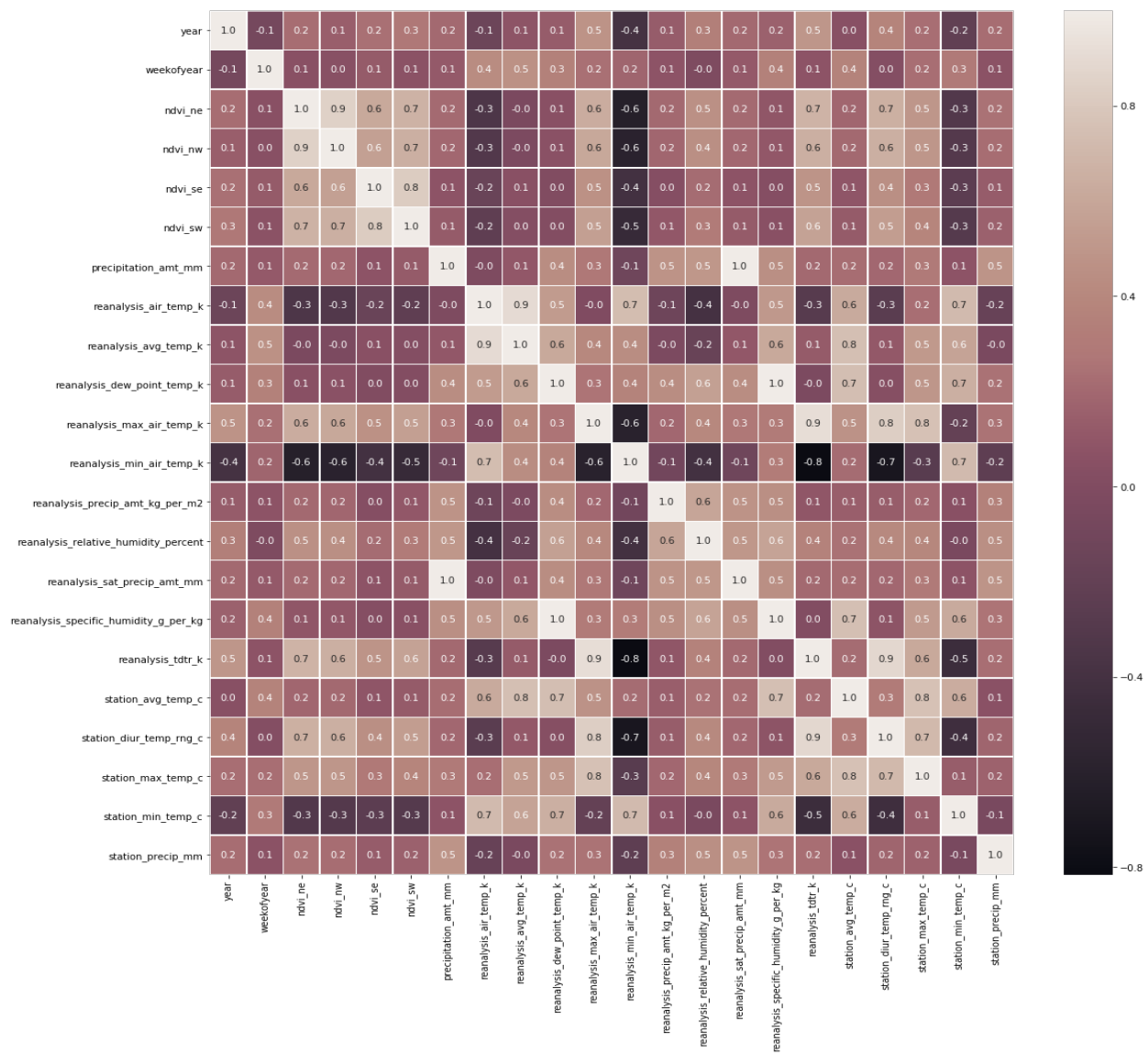


Figure 4. Correlation among features.

Table 3. Statistical characteristics along with their respective correlation values (1 as perfectly correlated and 0.9 as almost perfectly correlated).

Parameter	Correlation (1/0.9)
reanalysis_sat_precip_amt_mm and precipitation_amt_mm	1
reanalysis_specific_humidity_g_per_kg and reanalysis_dew_point_temp_k	1
ndvi_nw	0.9
ndvi_ne	0.9
reanalysis_avg_temp_k	0.9
reanalysis_air_temp_k	0.9
reanalysis_tdtr_k	0.9
reanalysis_max_air_temp_k	0.9
station_diur_temp_rng_c	0.9
reanalysis_tdtr_k	0.9

As we want to detect dengue in this manuscript for the same, if features are far in two cities, then it is suitable for ML classification (reanalysis_tdttr_k); otherwise, if they are near and give mixed information about features, then it is not considered suitable for prediction/classification.

From Figures 5 and 6, and from the dataset, we can create a new data frame, i.e., X_train plus the total_cases column of y_train.

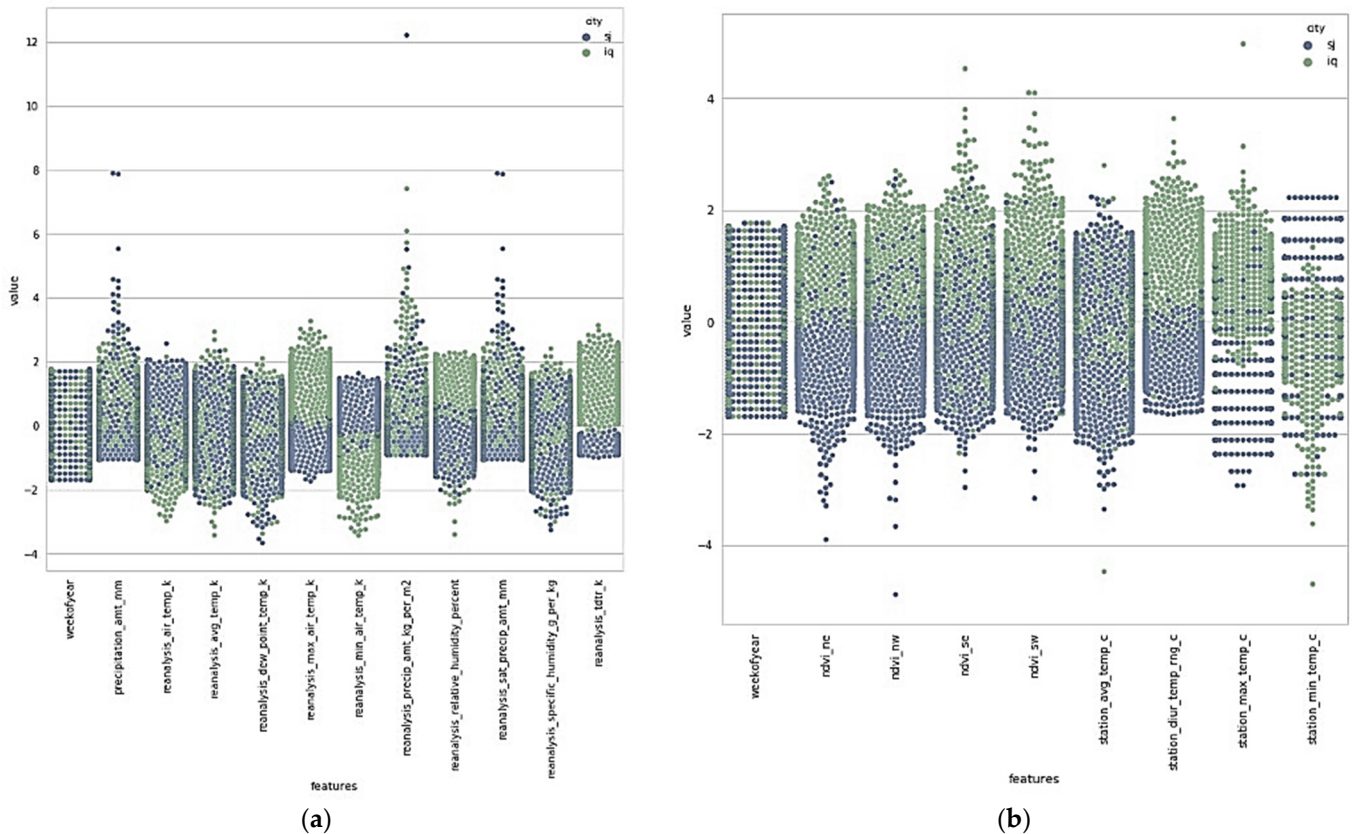


Figure 5. Various parameters were used in the dengue dataset, along with their characteristics. (a) Values along with Features. (b) Various parameters were used in the dengue dataset, along with their characteristics.

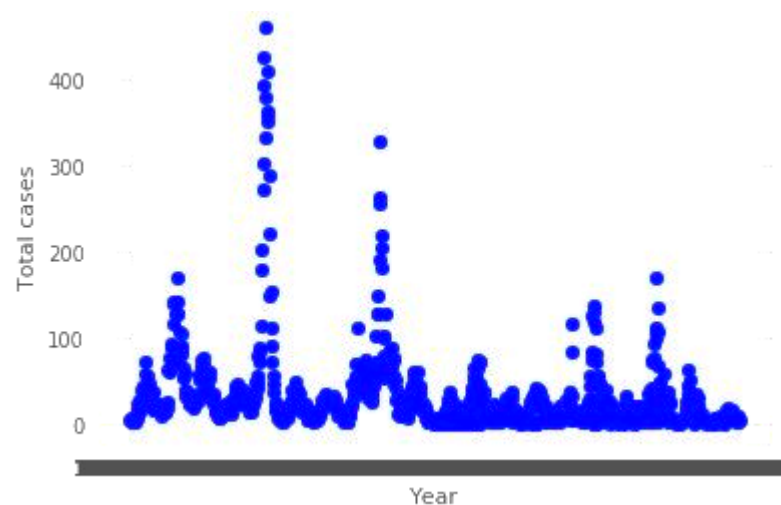
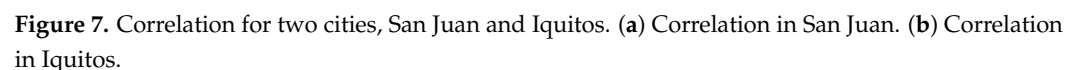


Figure 6. Total cases in the year.

Table 4. Selected features for different cities.

In this feature selection, we are dropping out the two features, i.e., `reanalysis_sat_precip_amt_mm` and `reanalysis_specific_humidity_g_per_kg`. At this time, we are not considering them because they are almost perfectly correlated (0.9), and we want to achieve good accuracy. However, with the present scenario, if we go for machine learning algorithms, i.e., KNN, D.T., R.F., and GNB, the accuracy comes out to be significantly less. This is due to the total number of cases immensely varying from 0 to 400+. The question arises, “How can we improve this accuracy?” The answer to this question is to divide our dataset into two cities, as shown in Table 4.



After finding the correlation between the two cities, we can deduce some information, such as that both cities showed promising results for

reanalysis_specific_humidity_g_per_kg

reanalysis_dew_point_temp_k

reanalysis_min_air_temp_k

The fact that they are perfectly correlated with each other (value 1) is a clear sign that they are. This says that mosquitoes live in areas with high humidity. Since temperature plays a vital role in the spread of mosquitoes, it is correlated both with each other and with the total number of cases. Surprisingly, the weakest part of the year is also highly correlated to San Juan City, and as a result, we will be keeping a close eye on that. In addition, if we plot “a number of years” against “week of the year,” we find that there is an outbreak at the end of the year in both cities. We arrived at this conclusion after outlining the plot between the two variables. The number of reported cases grows, and outbreaks often occur over a few weeks, as illustrated in Table 5 and Figures 8 and 9, respectively.

Table 5. Increase in outbreaks and cases in two cities (San Juan/Iquitos) in weeks.

City	Increase in Cases (Range in Weeks)	Increase in Outbreak (Range in Weeks)
San Juan	35th–45th	35th–45th
Iquitos	45th–50th	45th–50th

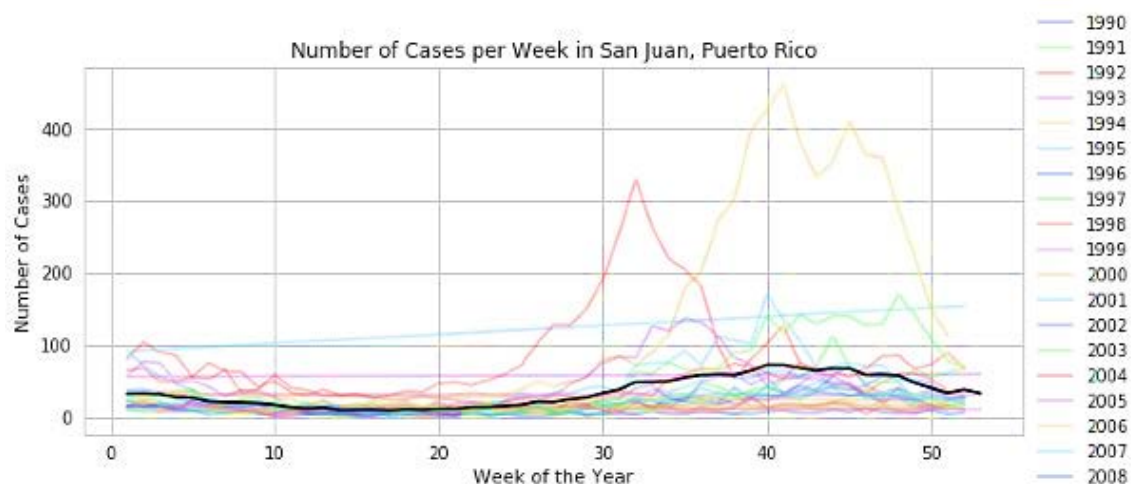


Figure 8. Increase in outbreaks and cases in San Juan city of Puerto Rico in weeks.

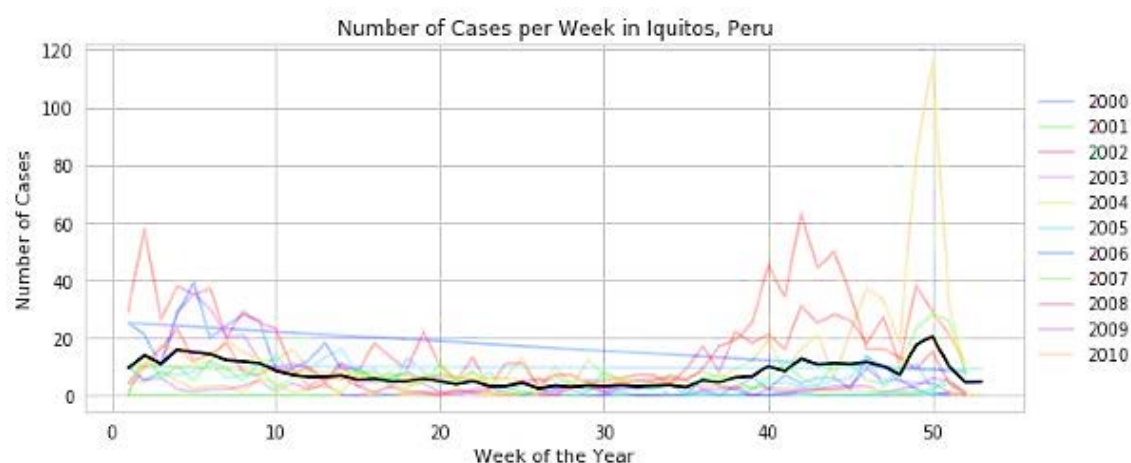


Figure 9. Increase in outbreaks and cases in Iquitos city of Peru in weeks.

4. Results of Different Classifiers

In this instance, we are using a variety of machine learning classifiers, beginning with KNN and moving on to decision trees, random forests, Gaussian neighbor boundaries, and support vector classifiers. In this instance, we are utilizing k-fold cross-validation to partition the data into ten equal portions for the purpose of classification. As a direct consequence of this, the mean value obtained after ten iterations is shown in Table 6.

Table 6. Analysis result of different machine learning classifiers to classify dengue disease by using 10-K-fold cross-validation and mean as a result of 10 iterations.

ML Classifier	K Fold = 10										Mean
	Scoring Accuracies										
	1	2	3	4	5	6	7	8	9	10	
KNN	0.095890	0.034246	0.082191	0.082191	0.082191	0.054794	0.041379	0.027586	0.062068	0.068965	6.315068
Decision Tree	0.075342	0.068493	0.109589	0.068493	0.068493	0.109589	0.055172	0.068965	0.048275	0.048275	7.206896
Random Forest	0.075342	0.082191	0.082191	0.143835	0.047945	0.109589	0.075862	0.096551	0.082758	0.075862	8.721303
Gaussian NB	0.075342	0.082191	0.047945	0.068493	0.047945	0.095890	0.062068	0.048275	0.089655	0.082758	7.005668
Support Vector Classifier	0.068493	0.068493	0.061643	0.068493	0.061643	0.075342	0.062068	0.055172	0.075862	0.089655	6.868682

It can be seen rather plainly from Table 6 that the random forest classifier is the one that turns out to be the best one, with a mean score of 8.72. For a more transparent illustration of the ranking of classifiers (see Figure 10).



Figure 10. Machine learning classifier ranking based on mean achieved after applying 10-K fold.

5. Discussion and Conclusions

As a result of its popularity and widespread application in image segmentation, deep learning has developed into a crucial instrument and is able to achieve ever-higher levels of precision. However, the primary concern is centered on the optimization of deep learning, and optimization encompasses multiple levels. Some of these levels include perfecting the deep network architectures and carrying out ensembled learning; hyperparameter tuning, which is an empirical method; optimizing the loss function in accordance with evaluation metrics; and making use of the appropriate optimizer and activation functions.

The purpose of this research is to develop a diagnostic model for the disease dengue by using machine learning techniques such as KNN, D.T., R.B., SVR, and GNB. The model will be able to make correct predictions regarding the progression of the disease as well as allow for early diagnosis of the disease. As a result of these upcoming initiatives, the focus of prioritization should be placed on cause–effect models for the diagnosis of disease. Not only is it vital to diagnose the sickness, but it is also essential to analyze the elements that have the most considerable influence on the infection. It is essential to do both things in order to be successful. A more profound comprehension of the etiology of the disease, along with the creation of more correct diagnostic models, would be of tremendous assistance in the fight against dengue fever, as well as in the reduction of complications and fatalities caused by the disease. The use of modeling for the purpose of minimizing the impact of data

uncertainty is another vital area. One of the primary challenges that must be surmounted before the quality of previously developed models can be enhanced is the poor standard of epidemiological data about dengue. As a last consideration, the use of independent loops of data analysis works to automate the decision-making process in disease control. Although the D.T., KNN, SVR, and GNB methods all generate better results, the R.F. method requires significantly more time to compute since it generates superior results. Based on the findings, it appears that the R.F. technique is the one to choose. Because of this, it has been determined that, out of all these various machine learning algorithms, the RF-based diagnostic model is the one that is best suited for accurately diagnosing dengue fever at an earlier stage. This conclusion was reached because of the reasons.

The substantial number of optimization factors and schemes that needed to be conducted empirically in order to give our final design requirements were the primary obstacles that needed to be overcome in this effort. Even if we have scaled back the trainable parameters of the network such that they are more compatible with the hardware, there is still the issue of the significant amount of CPU power that must be present to complete the training.

In conclusion, we can say that reason-based models can help with the analysis and interpretation of dengue disease data. This is something that we can assert. Because there is a severe lack of high-quality data in the field of healthcare, machine learning models that can deal with ambiguity can be highly valuable. In conclusion, data decentralization, in conjunction with aggregated learning, may make it possible to cut the costs of computer modeling and may also make it possible to do so without compromising the data's integrity. This may be possible.

Author Contributions: Conceptualization, G.G., S.K., V.G. and B.I.A.; methodology, V.G. and T.S.; software, A.A. and M.A.-s.; validation, resources and data curation, A.A. and S.A.A.; formal analysis, B.I.A. and B.M.A.; writing—original draft preparation, B.M.A.; writing—review and editing, S.A.A.; supervision and funding acquisition, S.K. and B.I.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) for funding and supporting this work through Research Partnership Program no. RP-21-07-06 and Princess Nourah Bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R440), Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) for funding and supporting this work through Research Partnership Program no. RP-21-07-06 and Princess Nourah Bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R440), Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would also like to thank Aleem Ali from Department of Computer Science and Engineering, Chandigarh University for serving as a consultant to critically reviewed the study proposal and participated in technical editing of the manuscript.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Hoyos, W.; Aguilar, J.; Toro, M. Dengue models based on machine learning techniques: A systematic literature review. *Artif. Intell. Med.* **2021**, *119*, 102157. [[CrossRef](#)] [[PubMed](#)]
2. Halstead, S.B. Pathogenesis of Dengue: Challenges to molecular biology. *Science* **1988**, *239*, 476–481. [[CrossRef](#)] [[PubMed](#)]
3. Kurane, I. Dengue hemorrhagic fever with particular emphasis on immunopathogenesis. *Comp. Immunol. Microbiol. Infect. Dis.* **2007**, *30*, 329–340. [[CrossRef](#)] [[PubMed](#)]
4. Guzman, M.G.; Halstead, S.B.; Artsob, H.; Buchy, P.; Farrar, J.; Gubler, D.J.; Hunsperger, E.; Kroeger, A.; Margolis, H.S.; Martínez, E.; et al. Dengue: A continuing global threat. *Nat. Rev. Genet.* **2010**, *8*, S7–S16. [[CrossRef](#)]

5. Linares, E.M.; Pannuti, C.S.; Kubota, L.T.; Thalhammer, S. Immunospot assay based on fluorescent nanoparticles for Dengue fever detection. *Biosens. Bioelectron.* **2013**, *41*, 180–185. [\[CrossRef\]](#)
6. San Martín, J.L.; Brathwaite, O.; Zambrano, B.; Solórzano, J.O.; Bouckenooghe, A.; Dayan, G.H.; Guzmán, M.G. The epidemiology of Dengue in the Americas over the last three decades: A disturbing reality. *Am. J. Trop. Med. Hyg.* **2010**, *82*, 128. [\[CrossRef\]](#)
7. Gupta, N.; Gupta, N.; Srivastava, S.; Jain, A.; Chaturvedi, U.C. Dengue in India. *Indian J. Med. Res.* **2012**, *136*, 373.
8. Thomas, E.A.; John, M.; Bhatia, A. Cutaneous manifestations of dengue viral infection in Punjab (north India). *Int. J. Dermatol.* **2007**, *46*, 715–719. [\[CrossRef\]](#)
9. Arshad, I.; Malik, F.A.; Hussain, A.; Shah, S.A. Dengue fever: Clinico-pathologic correlations and their association with poor outcome. *Prof. Med. J.* **2011**, *18*, 57–63. [\[CrossRef\]](#)
10. Whitehorn, J.; Farrar, J. Dengue. *Br. Med. Bull.* **2010**, *95*, 161–173. [\[CrossRef\]](#)
11. World Health Organization. *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control*; World Health Organization: Geneva, Switzerland, 2009.
12. Ranjit, S.; Kissoon, N. Dengue hemorrhagic fever and shock syndromes*. *Pediatr. Crit. Care Med.* **2011**, *12*, 90–100. [\[CrossRef\]](#)
13. Ahmed, F.U.; Mahmood, C.B.; Sharma, J.D.; Hoque, S.M.; Zaman, R. Dengue and Dengue Haemorrhagic Fever in children during the 2000 outbreak in Chittagong, Bangladesh. 2001. Available online: <https://apps.who.int/iris/handle/10665/163693> (accessed on 9 March 2023).
14. Chen, L.H.; E Wilson, M. Dengue and chikungunya infections in travelers. *Curr. Opin. Infect. Dis.* **2010**, *23*, 438–444. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Jain, A. *Machine Learning Techniques for Medical Diagnosis: A Review*; D.U., Conference Center: New Delhi, India, 2015.
16. Kononenko, I. Machine learning for medical diagnosis: History, state of the art and perspective. *Artif. Intell. Med.* **2001**, *23*, 89–109. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Raval, D.; Bhatt, D.; Kumhar, M.K.; Parikh, V.; Vyas, D. Medical diagnosis system using machine learning. *Int. J. Comput. Sci. Commun.* **2016**, *7*, 177–182.
18. Cetiner, B.G.; Sari, M.; Aburas, H. Recognition of Dengue Disease Patterns Using Artificial Neural Networks. In Proceedings of the 5th International Advanced Technologies Symposium (IATS'09), Karabük, Turkey, 13–15 May 2009.
19. Rachata, N.; Charoenkwan, P.; Yooyativong, T.; Chamnongthai, K.; Lursinsap, C.; Higuchi, K. Automatic prediction system of dengue hemorrhagic-fever outbreak risk by using entropy and artificial neural network. In Proceedings of the 2008 International Symposium on Communications and Information Technologies, Vientiane, Laos, 21–23 October 2008; IEEE: Piscataway, NJ, USA, 2008.
20. Wu, Y.; Lee, G.; Fu, X.; Hu, T. Detect climatic factors contributing to dengue outbreak based on a wavelet, support vector machines, and genetic algorithm. In Proceedings of the World Congress on Engineering 2018, London, UK, 2–4 July 2008.
21. Ahmad, S.; Khan, S.; AlAjmi, M.F.; Dutta, A.K.; Dang, L.M.; Joshi, G.P.; Moon, H. Deep Learning Enabled Disease Diagnosis for Secure Internet of Medical Things. *Comput. Mater. Contin.* **2022**, *73*, 965–979. [\[CrossRef\]](#)
22. Gomes, A.L.V.; Wee, L.J.K.; Khan, A.M.; Gil, L.H.V.G.; Marques, E.; Calzavara-Silva, C.E.; Tan, T.W. Classification of Dengue Fever Patients Based on Gene Expression Data Using Support Vector Machines. *PLOS ONE* **2010**, *5*, e11267. [\[CrossRef\]](#)
23. Guo, P.; Liu, T.; Zhang, Q.; Wang, L.; Xiao, J.; Zhang, Q.; Luo, G.; Li, Z.; He, J.; Zhang, Y.; et al. Developing a dengue forecast model using machine learning: A case study in China. *PLOS Neglected Trop. Dis.* **2017**, *11*, e0005973. [\[CrossRef\]](#)
24. Carvajal, T.M.; Viacrusis, K.M.; Hernandez, L.F.T.; Ho, H.T.; Amalin, D.M.; Watanabe, K. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. *BMC Infect. Dis.* **2018**, *18*, 1–15. [\[CrossRef\]](#)
25. Panwar, A.; Bhatnagar, V.; Khari, M.; Salehi, A.W.; Gupta, G. A Blockchain Framework to Secure Personal Health Record (PHR) in IBM Cloud-Based Data Lake. *Comput. Intell. Neurosci.* **2022**, *2022*, 3045107. [\[CrossRef\]](#)
26. Alrashed, F.A.; Alsubiheen, A.M.; Alshammari, H.; Mazi, S.I.; Al-Saud, S.A.; Alayoubi, S.; Kachanathu, S.J.; Albarrati, A.; Aldaihan, M.M.; Ahmad, T.; et al. Stress, Anxiety, and Depression in Pre-Clinical Medical Students: Prevalence and Association with Sleep Disorders. *Sustainability* **2022**, *14*, 11320. [\[CrossRef\]](#)
27. Kalansuriya, C.S.; Aponso, A.C.; Basukoski, A. Machine Learning-Based Approaches for Location Based Dengue Prediction. In *Fourth International Congress on Information and Communication Technology*; Springer: Berlin/Heidelberg, Germany, 2020.
28. Haq, A.U.; Li, J.P.; Ahmad, S.; Khan, S.; Alshara, M.A.; Alotaibi, R.M. Diagnostic Approach for Accurate Diagnosis of COVID-19 Employing Deep Learning and Transfer Learning Techniques through Chest X-ray Images Clinical Data in E-Healthcare. *Sensors* **2021**, *21*, 8219. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Gupta, G.; Salehi, A.W.; Sharma, B.; Kumar, N.; Sonia, G.; Vaidya, P. COVID-19: Automated Detection and Monitoring of Patients Worldwide Using Machine Learning, in *Modeling, Control and Drug Development for COVID-19 Outbreak Prevention*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 731–761.
30. Thomas, S.J.; L'Azou, M.; Barrett, A.D.; Jackson, N.A. Fast-track Zika vaccine development—Is it possible? *New Engl. J. Med.* **2016**, *375*, 1212–1216. [\[CrossRef\]](#)
31. Gupta, G.; Gupta, N.; Gupta, A.; Vaidya, P.; Singh, G.K.; Jaiswal, V. Prediction of Alzheimer associated proteins (PAAP): A perspective to understand Alzheimer disease for therapeutic design. *Int. J. Bioinform. Res. Appl.* **2021**, *17*, 363–374. [\[CrossRef\]](#)
32. Thakur, B.; Kumar, N.; Gupta, G. Machine learning techniques with ANOVA for the prediction of breast cancer. *Int. J. Adv. Technol. Eng. Explor.* **2022**, *9*, 232.

33. Yousef, R.; Gupta, G.; Yousef, N.; Khari, M. A holistic overview of deep learning approach in medical imaging. *Multimedia Syst.* **2022**, *28*, 881–914. [\[CrossRef\]](#)
34. Gill, H.S.; Murugesan, G.; Khehra, B.S.; Sajja, G.S.; Gupta, G.; Bhatt, A. Fruit recognition from images using deep learning applications. *Multimedia Tools Appl.* **2022**, *81*, 33269–33290. [\[CrossRef\]](#)
35. Macrae, A.; de Colella, C.S.; Sebastian, E. CS229 Project: Classification of Dengue Fever Outcomes from Early Transcriptional Patterns. Available online: <https://www.semanticscholar.org/paper/CS-229-Project-%3A-Classification-of-Dengue-fever-Macrae-Colella/b8e1ba4d0d2555fef4466be15207ff276b61d49f> (accessed on 9 March 2023).
36. Iqbal, N.; Islam, M. Machine learning for Dengue outbreak prediction: An outlook. *Int. J. Adv. Res. Comput. Sci.* **2017**, *8*, 93–102.
37. Focks, D.A.; Daniels, E.; Haile, D.G.; Keesling, J.E. A Simulation Model of the Epidemiology of Urban Dengue Fever: Literature Analysis, Model Development, Preliminary Validation, and Samples of Simulation Results. *Am. J. Trop. Med. Hyg.* **1995**, *53*, 489–506. [\[CrossRef\]](#)
38. Hopp, M.J.; Foley, J.A. Global-Scale Relationships between Climate and the Dengue Fever Vector, *Aedes Aegypti*. *Clim. Chang.* **2001**, *48*, 441–463. [\[CrossRef\]](#)
39. Hartley, L.; Donnelly, C.; Garnett, G. The seasonal pattern of Dengue in endemic areas: Mathematical models of mechanisms. *Trans. R. Soc. Trop. Med. Hyg.* **2002**, *96*, 387–397. [\[CrossRef\]](#)
40. De Paula, S.O.; Fonseca, B.A.L.D. Dengue: A review of the laboratory tests a clinician must know to achieve a correct diagnosis. *Braz. J. Infect. Dis.* **2004**, *8*, 390–398. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Ibrahim, F.; Taib, M.N.; Abas, W.A.B.W.; Guan, C.C.; Sulaiman, S. A novel dengue fever (D.F.) and dengue hemorrhagic fever (DHF) analysis using artificial neural network (ANN). *Comput. Methods Programs Biomed.* **2005**, *79*, 273–281. [\[CrossRef\]](#)
42. Fink, J.; Gu, F.; Ling, L.; Tolfvenstam, T.; Olfat, F.; Chin, K.C.; Aw, P.; George, J.; Kuznetsov, V.A.; Schreiber, M.; et al. Host Gene Expression Profiling of Dengue Virus Infection in Cell Lines and Patients. *PLOS Neglected Trop. Dis.* **2007**, *1*, e86. [\[CrossRef\]](#) [\[PubMed\]](#)
43. A Gadkari, R.; Srinivasan, N. Prediction of protein-protein interactions in dengue virus coat proteins guided by low resolution cryoEM structures. *BMC Struct. Biol.* **2010**, *10*, 17. [\[CrossRef\]](#)
44. Tanner, L.; Schreiber, M.; Low, J.G.H.; Ong, A.; Tolfvenstam, T.; Lai, Y.L.; Ng, L.C.; Leo, Y.S.; Puong, L.T.; Vasudevan, S.; et al. Decision Tree Algorithms Predict the Diagnosis and Outcome of Dengue Fever in the Early Phase of Illness. *PLOS Neglected Trop. Dis.* **2008**, *2*, e196. [\[CrossRef\]](#)
45. Butt, N.; Abbassi, A.; Munir, S.M.; Ahmad, S.M.; Sheikh, Q.H. Haematological and biochemical indicators for the early diagnosis of dengue viral infection. *J. Coll. Physicians Surg. Pak.* **2008**, *18*, 282–285.
46. Choudhury, Z.M.; Banu, S.; Islam, A.M. Forecasting Dengue Incidence in Dhaka, Bangladesh: A Time Series Analysis. 2008. Available online: <https://apps.who.int/iris/handle/10665/170465> (accessed on 9 March 2023).
47. De Kruif, M.D.; Setiati, T.E.; Mairuhu, A.T.A.; Koraka, P.; Aberson, H.A.; Spek, C.A.; Osterhaus, A.D.M.E.; Reitsma, P.H.; Brandjes, D.P.M.; Soemantri, A.; et al. Differential Gene Expression Changes in Children with Severe Dengue Virus Infections. *PLOS Neglected Trop. Dis.* **2008**, *2*, e215. [\[CrossRef\]](#)
48. Rissino, S.; Lambert-Torres, G. Rough Set Theory—Fundamental Concepts, Principals, Data Extraction, and Applications. In *Data Mining and Knowledge Discovery in Real Life Applications*; IntechOpen: Cambridge, UK, 2009.
49. Aburas, H.M.; Cetiner, B.G.; Sari, M. Dengue confirmed-cases prediction: A neural network model. *Expert Syst. Appl.* **2010**, *37*, 4256–4260. [\[CrossRef\]](#)
50. Madhu, G.; Reddy, G.; Kiranmai, C. Hypothetical description for intelligent data mining. *Int. J. Comput. Sci. Eng.* **2010**, *2*, 2349–2352.
51. Althouse, B.M.; Ng, Y.Y.; Cummings, D.A.T. Prediction of Dengue Incidence Using Search Query Surveillance. *PLOS Neglected Trop. Dis.* **2011**, *5*, e1258. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Gomide, J.; Veloso, A.; Meira, W.; Almeida, V.; Benevenuto, F.; Ferraz, F.; Teixeira, M. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In *Proceedings of the 3rd International Web Science Conference*, Koblenz, Germany, 15–17 June 2011; p. 3.
53. Brasier, A.; Scott, T.W.; Morrison, A.C.; Kochel, T.J.; Spratt, H.M.; Bazan, I.; Forshey, B.M.; Garcia, J.; Victor, S.S.; Rocha, C.; et al. A Three-Component Biomarker Panel for Prediction of Dengue Hemorrhagic Fever. *Am. J. Trop. Med. Hyg.* **2012**, *86*, 341–348. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.