

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

**Applications of artificial intelligence in predicting dengue outbreaks in the face of climate change:  
a case study along coastal India**

*Using machine learning to assess climate-sensitive disease dynamics for public health management*

Krti Tallam<sup>1\*</sup>, Minh Pham Quang<sup>2</sup>

<sup>1</sup> Department of Biology, Stanford University, Stanford, California, United States of America

<sup>2</sup> VNU-HCM High School for the Gifted, Ho Chi Minh City 70000, Vietnam

\* Corresponding author:

Email: [ktallam7@stanford.edu](mailto:ktallam7@stanford.edu) (KT)

21

## 22 **Abstract**

23         The climate crisis will have an increasingly profound effect on the global distribution and burden  
24 of infectious diseases. Climate-sensitive diseases can serve as critical case studies for assessing public  
25 health priorities in the face of epidemics. Preliminary results denote that machine learning-based  
26 predictive modeling measures can be successfully applied to understanding environmental disease  
27 transmission dynamics. Ultimately, machine learning models can be trained to detect climate-sensitive  
28 diseases early, diseases which might represent serious threats to human health, food safety, and  
29 economies. We explore how machine learning can serve as a tool for better understanding climate-  
30 sensitive diseases, taking dengue dynamics along the Godavari River of coastal India as our case study.  
31 We hypothesize that a climate-driven predictive model with controlled calibration can help us  
32 understand several of the most critical relationships and climate characteristics of climate-sensitive  
33 disease dynamics.

34

## 35 **1. Introduction**

### 36 **1.1 The role of climate in ecosystem dynamics.**

37         Climate has always played a critical role in ecosystem dynamics, including in population  
38 dynamics and species interactions (1). However, the mechanisms underlying climate variability and its  
39 consequences remain poorly understood and infrequently field-tested (2). This is particularly true  
40 because climate effects often appear context specific (3). Multiple climate variables can act  
41 synergistically, individual climate variables can affect multiple aspects of ecosystem dynamics, and  
42 impacts can often be nonlinear, challenging to predict in the face of the climate crisis, and presented

43 across a gradient of important conditions (4). Therefore, surprisingly, while climate is known to be one  
44 of the most prominent influencers of ecological processes, its dynamical and directional effects on  
45 ecosystems are often poorly understood and challenging to predict. This has become increasingly true  
46 against the backdrop of the climate crisis and increasing anthropogenic stressors and especially so for  
47 coastal communities (5).

## 48 **1.2 Climate-sensitive diseases.**

49 Climate-sensitive diseases provide a critical case study to model and measure whether climate-  
50 sensitive traits simulated in models can help us more accurately assess the multitude of dynamics  
51 observed in the field. For mosquito-borne diseases in particular, rising average temperatures have often  
52 led to geographic range expansions of disease vectors, decreased incubation periods of some pathogens,  
53 and increased rates of contact of some mosquito species that prey on humans (6). Recent knowledge  
54 provides evidence that many ranges of mosquito-borne diseases will expand dramatically in response to  
55 the climate crisis (7). Models have also helped scientists better understand how average sea levels can  
56 also influence the density of salinity-tolerant climate-sensitive pathogens, namely, mosquitos, along  
57 coastlines (8). Statistical simulations have helped point science to better quantify the influence of  
58 increasingly warmer climates on the hydrological cycle, a phenomenon which has been in part discussed  
59 in the context of the extreme weather events observed increasingly frequently over the last few decades  
60 (9).

## 61 **1.3 Mosquito-borne diseases.**

62 The impacts of climate change on mosquito-borne diseases are intensified by shifting economic,  
63 demographic, and social factors making it even more necessary to better model how environmental  
64 patterns may be shifting in areas that are increasingly prone to climate-sensitive infectious diseases (10).

65 Even going a few decades back, Johansson et al. and Moore et al. discuss evidence of unpredictable  
66 temperature and precipitation influencing the transmission and abundance potential of *Aedes aegypti* in  
67 Puerto Rico (11,6). Suggestive relationships have also been discovered between coastal dengue  
68 outbreaks and sea surface temperatures in places such as New Caledonia and Mexico (12–14). The  
69 transmission of mosquito-borne, viral, climate-sensitive diseases such as dengue occur along a spectrum,  
70 from low levels of year-round endemic transmission (15) to larger-scale period or interannual outbreaks  
71 (16). We hypothesize that the range of dynamics of mosquito-borne disease dynamics are a result of  
72 regional or seasonal differences in climate, where the magnitude or direction of climate effects on  
73 mosquito-borne vectors differ (17–19).

#### 74 **1.4 Prediction and intervention of disease dynamics along coastlines.**

75 Systematically understanding the drivers of climate-sensitive mosquito-borne disease dynamics  
76 will significantly support two core scientific outcomes: the prediction of the climate crisis' effect on  
77 disease dynamics, especially along coastlines, and the intervention and management strategies required  
78 to prevent detrimental outbreaks (20). For instance, predictive models of dengue dynamics that can  
79 capture small-scale mosquito population dynamics along coastal India can be applied to more realistic  
80 projections for outbreaks dynamics across other tropical coastal regions (21,22).

81 Despite the vitality of predictive approaches, validation with vector and disease data remain  
82 limited given the challenges of obtaining and processing such data. Thus, it raises the question of  
83 necessity in that scientists must assess which disease variables a model should capture based on model  
84 parameterization. This must be either scalable or independent from the transmission system researchers  
85 aim to predict. Since disease dynamics cannot be studied in the context of every possible permutation of  
86 climate parameters, transmission setting, or scale, understanding the extent to which models derived  
87 from fundamental science or modeling-based traits will be critical for translational and scalable

88 modeling of future climate-sensitive mosquito-borne disease dynamics (23,24). We hypothesize that a  
89 climate-driven predictive model with controlled calibration can help us understand several of the most  
90 critical relationships and climate characteristics of climate-sensitive disease dynamics.

### 91 **1.5 Dengue outbreaks along coastlines.**

92 In this work, we test the extent to which climate-driven mosquito traits drive disease dynamics  
93 across a coastal region of eastern India, given the climate dynamics of the Bay of Bengal and the  
94 opportunity to characterize additional climatological and ecological factors that may mediate the effects  
95 of climate on dengue disease dynamics. Given the ecology of the primary disease vector of dengue,  
96 *Aedes aegypti*, and the fact that they are anthropophilic, their traits are important for the transmission of  
97 dengue. These traits include their reproduction, survival, development, biting date, and extrinsic  
98 incubation periods. Studies have demonstrated that these traits are heavily temperature-dependent with  
99 an intermediate thermal optimum (25,26). Humidity has been positively associated with mosquito  
100 survival because of the high surface area to volume ratio of mosquitos exposed to desiccation (27).  
101 Standing or stagnating water resulting from rainfall also provides ideal pupal and larval habitats for  
102 mosquitos (28). The relationship is also challenging because heavy rainfall or extreme weather events  
103 off the Bay of Bengal (i.e., cyclones) can also flush away *Ae. aegypti* breeding habitats. Recent studies  
104 have demonstrated that models incorporating temperature-based climate traits can provide insights on  
105 some critical features of dengue disease dynamics (29).

### 106 **1.6 Historical understanding of dengue and climate in India.**

107 Previous studies have attempted to understand the relationship between climate and the presence  
108 of dengue in India by means of statistical and mathematical models. These techniques have included  
109 Poisson regression (30), Naïve Bayes and multivariable regression (31), Bayesian models such as spatial

110 autoregressive (SAR) models or conditional autoregressive (CAR) models (Mudele et al., 2021),  
111 autoregressive integrated moving average (ARIMA) models (32), generalized linear models (33),  
112 Extreme Gradient Boosting and rule-based classification (34), and other types of mathematical modeling  
113 (35). The above methods have focused on multiple parameters influencing dengue and on the discovery  
114 of the climate variables that most influence dengue disease dynamics (36).

115 From our literature review, we observed the establishment of many statistical and mathematical  
116 models, which address stochastic, mechanistic, and deterministic dengue disease models, among others.  
117 However, there is still limited research probing the predictive modeling of the complexities of dengue  
118 disease models with the backdrop of climatically critical variables. One tool that has recently surfaced as  
119 an emerging modeling strategy is machine learning modeling for predictive dengue disease dynamics  
120 (37). While there exists research into the effects of dengue dynamics in relation to climate, these are  
121 often backed by traditional statistical models. Machine learning tools offer a critical opportunity on  
122 capitalize on greater predictive ability. This can expand the literature to include state-of-the-art models  
123 for assessing dengue disease dynamics. In this study, we used advanced machine learning techniques to  
124 assess the impact of hydroclimatic and extreme weather climate parameters on dengue disease dynamics  
125 and outbreak predictions.

## 126 **2. Materials and Methods**

### 127 **2.1. Data development and visualization**

#### 128 **2.1.1 Data cleaning and augmentation**

129 Our data were hand-collected on the ground between 2019-2020, in Andhra Pradesh, India.  
130 Along with historic data collected at those points in time as well, the dataset spans 2016-2020 with  
131 weekly frequency. The months of January to March were not examined in this study due to data quality

132 and collection issues. The following climatic parameters were used in our model: windspeed, pH level,  
133 dissolved oxygen, temperature, total number of extreme weather (measured via cyclonic) events that  
134 took place, rainfall, precipitation. They were chosen based on previous literature and on hypothesized  
135 relationships to dengue fever. The climatic data were collected from the surrounding area of the  
136 Godavari River with primary focus in subsections of the rivers in Andhra Pradesh, India. To evaluate the  
137 accuracy of models in determining the relationship between these variables and dengue disease  
138 dynamics, we used dengue case counts as our predictor benchmarks for both regression and  
139 classification tasks (explained further in Section 2.2).

#### 140 **2.1.2 Model parameters.**

##### 141 **Wind speed.**

142 The model parameter “wind speed” was added based on the provided wind speed data. In the  
143 original dataset, wind speed was represented by weekly occurrence and measurements. After several  
144 model iterations, it became apparent that those weekly changes were unsuitable for the nature of this  
145 long-term dataset. Therefore, when a cyclone was documented, the full month of its occurrence was  
146 labeled with the respective wind speed. For months during the dataset that had multiple cyclones occur,  
147 2-week or 3-week intervals were labeled with the wind speed of the cyclone. For months where no  
148 cyclones occurred, the default value was set to be a normal wind speed of 20 kilometers / hour. The  
149 methodology has potential for further fine-tuning during future model iterations.

##### 150 **pH and dissolved oxygen levels**

151 pH and dissolved oxygen levels were obtained using data from a subsection of River Godavari at  
152 Rajahmundry (16.714456°, 82.331298°). The measurements were originally obtained in monthly  
153 intervals. For the purposes of our model, we had to down-sample the data to weekly intervals. Hence, in

154 the final dataset, the weekly data points are grouped by month. The data was obtained near the outtakes  
155 of water treatment stations through the Water Resource Board of Andhra Pradesh.

### 156 **Dengue counts**

157 Dengue fever is a highly seasonal disease and, for our purposes, is spread through the bite of an  
158 *Aedes aegypti* (*Ae. aegypti* hereafter) mosquito. Being a mosquito-borne disease, dengue demonstrates  
159 high correlation to the mosquito population in the region (38). *Ae. aegypti* populations thrive within  
160 conditions created by monsoons (humid, damp, brackish waters). As such, increased dengue cases, and  
161 more frequent dengue outbreaks, usually occur during the monsoon season.

### 162 **Additional Model Parameters**

163 For model features not mentioned above, they were collected by the lead author and sourced  
164 during the time spent in the region of Andhra Pradesh.

### 165 **2.1.3 Time series analysis**

166 A high degree of seasonality was exhibited across all features. For all climatic variables, we  
167 observed a yearly spike during the months of April through July. The rapid spike in this period for all  
168 variables indicates active environmental activity in the form of increased rainfall and large numbers of  
169 storms and cyclones. These are indicative of typical tropical climates, especially in the monsoon season,  
170 although with climate change have increased severely in the last decade. Moreover, the number of  
171 dengue cases also shows a spike in these months, which suggests there is a correlation between extreme  
172 weather events and dengue outbreak risk in the region. This can be further interpreted as allowing for an  
173 increased exposure to *Ae. aegypti* mosquitoes due to the ideal breeding conditions facilitated by climate  
174 conditions (damp, humid, and stagnant water bodies). Our models show that this often leads to an  
175 overall increased in mosquito population in the region and increased risk of dengue cases [**Figure 1**].



176 By observing the trend lines more closely, there exists a strong correlation between rainfall,  
177 precipitation, water temperature and dengue counts. It is worth noting how dengue counts demonstrate a  
178 lag of 2-4 weeks after a spike in rainfall and precipitation [**Figure 1**].

179 *Fig 1. Our time series analysis of the variables used in our study.*

180 In 2019, we note a particularly abnormal increase in the number of dengue cases, with  
181 approximately 40 percent more cases than the average number of cases in previous years. The outlier  
182 years in dengue cases were also followed by the highest spike in other climatic environmental variables.  
183 This could indicate that there was a heavy dengue outbreak in the region that was likely, at least in  
184 significant part, influenced by climate. Given this, we broke down a correlation matrix to observe the  
185 individual relationships between variables [**Figure 2**].

186 *Fig 2. The correlation matrix of all climatic variables in the study.*

#### 187 **2.1.4 Correlation analysis**

188 We note the strong positive correlation between number of cyclones and wind speed [**Figure 3**].  
189 This is logical given the that cyclonic activities would cause fluctuation in wind speed data. We also  
190 observe the strong positive correlation between precipitation and water temperature, and a strong  
191 positive correlation between total number of cyclones and ocean water pH levels. There is a moderate  
192 positive correlation between wind speed and water temperature. There exists a strong negative  
193 correlation between precipitation and dissolved oxygen levels.

194 *Fig 3. Correlogram of all climatic variables in the study.*

195 For the other features of the model, there was a relatively weak correlation between each  
196 variable, indicating that each of those variables showed greater independence from one another.  
197 However, we use this to our advantage as it can provide clearer insight into what influenced dengue  
198 counts and resulting outbreaks in the region. This is critical for scalability and transferability of our

199 dengue model because it will produce generalized results that are not biased towards a single climatic  
200 feature.

## 201 **2.2 Machine learning models: data augmentation**

### 202 **2.2.1 Experimental Setup and Software**

203 Both classification and regression methods were performed during analysis. In this study, all data  
204 processing and analysis was conducted using Python 3.9 on Google Colab. To produce results, NumPy,  
205 scikit-learn, Keras libraries were employed in the cleaning and model development process.

### 206 **2.2.2 Balancing the dataset**

207 We observed a relatively imbalanced dataset. Hence, we resampled our dataset using a Python  
208 scikit-learn class known as Synthetic Minority Oversampling Technique (SMOTE), which acts as a data  
209 transform object for classification datasets that observe imbalance in their datasets (39). SMOTE  
210 duplicates examples in the minority class and new examples can be synthesized from the existing  
211 examples. We note the newly generated dataset has a near balanced number of samples, with 288  
212 samples in total used for both the training and testing datasets.

### 213 **2.2.3 Data Preprocessing**

214 Given the data collection process, normalization of variables is necessary. As different variables  
215 were measured on different scales such as mm for rainfall and km/h for cyclone speeds, these  
216 differences may influence the weighting of each parameter so that the models are heavily influenced by  
217 variables maximum values, creating bias. Moreover, normalization helps models detect hidden pattern in  
218 data more easily as they helps detect the minor variations in data such as in pH levels more easily. We  
219 normalize all our variables using Min-Max normalization on our dataset to ensure each data point was  
220 scaled to the same range from 0 to 1. Min-max normalization is a common way to normalize data, where  
221 for every feature, the minimum value of that feature gets transformed to 0, the maximum value gets

222 transformed into a 1, and every other value is transformed into a decimal between the scale of 0 and 1.

223 The Min-Max normalization function is defined as equation:

224

$$225 \quad x_{scaled} = (x - x_{min}) / (x_{max} - x_{min})$$

226

227 Where  $x$  is the variable being scaled,  $x_{max}$  is the maximum value  $x$  over all observation of the

228 variable,  $x_{min}$  is the minimum value  $x$  over all observation of the variable and  $x_{scaled}$  is our result.

## 229 **2.2.4 Time-lagged dataset**

230 For both of our classification and regression tasks, a 4-week time-lagged dataset was used. The

231 rationale behind choosing such a timeframe was both from a machine learning and ecology standpoint.

232 We note that after trialing different time spans for the lagged dataset, the 4-week time lagged dataset

233 offered the best performance for the models while it did not cause the models to overfit due to having

234 many parameters the models were inputted. From an ecology standpoint, 4-week time lag would capture

235 the growth cycle of the *Ae. aegypti* mosquito after laying eggs in brackish water. Dengue transmission

236 requires multiple developmental processes to occur in the mosquito and parasite, resulting in time lags

237 that vary with temperature. Time lags also arise between climate and dengue transmission because

238 temperature affects mosquito development rates in maturing stages, as well as during the oviposition

239 cycle and during pathogen extrinsic incubation periods (28).

## 240 **2.3. Machine Learning Models**

### 241 **2.3.1 Regression Methodology**

242 We aimed to produce predictive forecasting curves of dengue counts using long short-term

243 memory networks (LSTM) models. LSTM models are used in deep learning and machine learning as a

244 variety of recurrent neural networks (RNNs) capable of learning long-term dependencies, especially in

245 sequence prediction problems. We used a time-lagged dataset (explained in Section 2.2). The dataset  
246 was built using the 4-week lag time frame, as it turned out to be most adequate in representing  
247 seasonality of dengue transmission, while not still preventing the model from overfitting, which we  
248 discovered from both our own data and other studies (25). We employed the Root Mean Square Error  
249 (RMSE) to calculate the prediction errors of the model. We trained the model on 100 epochs.

250 With initial experimentation, the models indicated a high degree of overfitting. This is seen  
251 through validation scores not decreasing in accordance with a decrease in training score. In fact, it would  
252 increase with more training epochs. This is common in machine learning tasks due to various reason  
253 including noisy data or small datasets.

254 To amend the situation, we added dropout layers to help decrease overfitting. We also designed  
255 multi-layered LSTM model, with 3-layered setup so that each LSTM is independently trained from each  
256 other so that it could generalize more and learn better. Parametric Rectified Linear Unit (PReLU) as an  
257 activation function is used due to its ability to prevent overfitting and because it is commonly used in  
258 similar time series models (40).

### 259 **2.3.2 Model Metrics**

#### 260 **RMSE (Root mean squared error)**

261 RMSE is one of the most used measures for evaluating the quality of predictions. RMSE is equal  
262 to the square root of the average of squared residuals. The RMSE penalizes heavily on large errors as the  
263 residuals are squared before taking the average. In the case of our model, it is particularly useful when  
264 large errors are undesirable as it helps model to produce better predictions (41). High reliability is  
265 particularly desirable as dengue forecasting is used for policy changes, which requires accurate  
266 information to maximize effectiveness.

### 267 **2.3.3 Models**

## 268 **Recurrent Neural Network (RNN) Models**

269 RNN is a type of neural network that handles previous outputs as inputs, factoring in historical  
270 information into the model. As a neural network model, they provide researchers the ability the ability to  
271 extract relationship between complex climatic, hydrological variables and dengue fever risk. The RNN  
272 model allowed for greater predictive ability given the highly seasonal nature of dengue fever, as it is  
273 developed to handle timeseries data (Feng et al, 2021). In this study, we opted to only develop an LSTM  
274 model, an extension of RNN model, as RNN models face vanishing gradient problems. This influences  
275 their ability to work with long-term historical data. That is, the influence of the earlier inputs 'disappears'  
276 when the hidden layer is overwritten with new inputs, this causes the network to 'forget', which hampers  
277 the ability of an RNN to handle higher long-term dependencies.

## 278 **Long-Short Term Memory (LSTM) Models**

279 LSTM is a type of RNN developed to overcome the vanishing gradient problem faced in  
280 traditional RNN models (Hochreiter & Schmidhuber, 1997). LSTM models learn long-term  
281 dependencies and extract and retain information from a much longer timeframe compared to RNN.  
282 LSTM models achieve this with the introduction of memory blocks eliminating the vanishing gradient  
283 problem (Sundermeyer, 2012). In doing so, LSTM models better captured the relationship between  
284 historical trends of dengue cases and the complex climatic conditions for the purposes of our model.

## 285 **LSTM with Attention**

286 LSTM models can also lose important information when passing information across multiple  
287 sequence steps, Moreover, LSTM models primarily work with fixed-length time sequences without  
288 special consideration of the different time of the year, such as pre-monsoon or monsoon season which  
289 influences the importance of different climatic variables. LSTM with Attention is an extension of LSTM  
290 models which introduces attention mechanism which is a method to deal to with non-fixed length

291 sequences of time and allows LSTM models to selectively pay attention to the current input. This  
292 ultimately enabled a stronger understanding of our timeseries data. In our study, LSTM and LSTM with  
293 Attention were also compared to determine the suitability of each model in predicting number of dengue  
294 cases (Shook et al, 2021).

#### 295 **2.3.4 Classification Methodology**

296 Instead of using binary classification to identify which weeks are more prone to dengue case  
297 increases, we decided to use multiclass labels to better represent dengue risks of a specific week or set of  
298 weeks. The categories were: low risk, medium risk, and high risk (with corresponding labels of 0, 1, and  
299 2).

300 We chose to use dengue prediction counts of specific months to categorize our data. Low risk  
301 weeks were weeks with lower dengue counts than the mean dengue counts, medium risk weeks were at  
302 most  $\frac{1}{4}$  standard deviation higher than the mean dengue counts, and high-risk weeks were weeks with  
303 more than  $\frac{1}{4}$  higher standard deviation higher dengue counts than the mean dengue counts.

304 The threshold was defined in the following ways based on prior research on the matter (Pham et  
305 al, 2022). We also considered the outlier weeks where extreme numbers of dengue cases were presented  
306 (i.e., counts that were several times the mean).

#### 307 **2.3.5 Metrics**

##### 308 **Accuracy**

$$309 \quad (1) \quad Accuracy = \frac{\text{total number of accurate prediction}}{\text{total number of predictions}}$$

310

311 Equation (1) demonstrates how well a classifier model will perform in our multiclass  
312 classification problem. Its calculation provides a clear indicator to the performance of a model as it

313 computes the total number of correct predictions divided by the total number of predictions made.

314 Accuracy is the most often used metrics in classification problems.

### 315 **2.3.6 Models**

#### 316 **SVM**

317 Support Vector Machine (SVM) models are methods designed for classification and regression  
318 problems. The algorithm performs predictions by establishing the best hyperplane or hyperplanes that  
319 minimize the loss function over the feature space. An SVM would determine the hyperplane by  
320 calculating the maximal margin of the classes which is the maximum distance of the nearest element to  
321 the hyperplane for each class, with the nearest element of each class to the hyperplane being called  
322 support vector as the affects the orientation of an SVM. However, this can only be effectively done if the  
323 relationship between variables for each feature are linear, for non-linear inputs this problem is resolved  
324 via a kernel. A kernel is a function which allows the model to project the feature space into higher  
325 dimension until each class is linearly separable. The choice for kernel function in this article was radial  
326 basis function based on previous related literature.

#### 327 **KNN**

328 K-Nearest-Neighbor (KNN) is used for a variety of problems in ML. KNN is a stable  
329 deterministic algorithm which produces the same result across trial runs and is computationally efficient,  
330 only needing to calculate the relationship of an observation and “k” of its nearest neighbor. K here is  
331 defined as the number of neighbors being weighted in calculating the cost of each sample point. Here  $k$   
332 = 7 was used to generate the most optimal result (Gauahr et al, 2021). The algorithm works by  
333 calculating the distance functions between each of the nearest neighbors and then classifying them based  
334 on a majority vote based on distance weights.

#### 335 **Random Forest**

336 Random forest models are ensemble models that compute results based on the aggregate of  
337 multiple random decision tree classifiers. The decision trees in random forest are generated from a  
338 random subset of features, compared to normal decision trees which are generated from the entire  
339 dataset. The decision trees in a random forest model are created to have low correlation between one  
340 another, which helps the random forest model perform better than decision tree models (42). Random  
341 forest models are particularly useful in dealing with predicting dengue fever case variability, as they can  
342 deal with numerous independent variables efficiently. Another benefit is that when multiclass  
343 classification is employed, random forest models performed well given that they are designed to handle  
344 such problems without much extension being required of the base model.

### 345 **3. Results**

#### 346 **3.1. Results of regression models**

347 From **Table 1**, we can see that the LSTM captures the trend of cases increasing, indicating the  
348 high spread of dengue cases during in monsoon season. It was interesting to discover that each model  
349 provides similar regression curve with the LSTM with Attention model even providing a worse RMSE  
350 score in the validation dataset compared to the standard LSTM model. A potential explanation can be  
351 attributed to the fact that the time frame used was limited using only 4 years, as such the added  
352 complexity of LSTM with Attention hinders its predictive ability. The dataset is also limited for a  
353 machine learning algorithm. It was also found that the upward and downward trends simulated and  
354 forecasted by the LSTM model provides a reasonable approximation to the reported points, especially  
355 for the identification of high incidence peaks, which supports the results where LSTM can better capture  
356 the trends in dengue cases. Another noteworthy feature is that models can accurately capture the high  
357 jump in dengue cases at 120 weeks after the start of the dataset, the outlier high of dengue cases we see  
358 here can be attributed as a major dengue fever outbreak in the region and the regression results indicate



359 the models can follow such trend accurately, proving their usage in outbreak detection. This study serves  
360 as support that deep learning models such as LSTM can be beneficial in forecasting dengue outbreak,  
361 allowing for authorities to better prevent dengue cases.

Model	Train Score	Test Score
LSTM	0.04	0.12
LSTM With Attention	0.04	0.17

362

363 *Table 1. RMSE scores of all regression models.*

### 364 **3.2. Result of the classification models**

365 Overall, all models show promising classification results [Figure 4]. The accuracy of all models  
366 indicated were above 90 percent, with most misclassifications being in low risk and medium risk weeks  
367 with high reliability. It means that the parameters selected on the time-lagged dataset have contributed  
368 information for the models to analyze and improve from. In the SVM models, there were the most  
369 totaling 3 misclassifications on low dengue risk week prediction, indicating that the model would have  
370 some difficulty differentiating between low dengue risk week and medium risk. An explanation for this  
371 would be due the thresholding value as it medium risk prediction has a smaller range compared to high  
372 risk and low risk value causing models to make wrong predictions more frequently, which indicates that  
373 current models still lack the ability to discern smaller variation in changes between climatic variables  
374 that causes changes between each type of dengue risk classification. Moreover, SVM models do not  
375 support multiclass classification natively which could mean that the extension to handle multiclass  
376 labels which hamper to performance of SVM models.

377 *Fig 4. The regression curve of the LSTM and LSTM with Attention.*

378 Comparatively, both KNN and Random Forest models performed much better in classifying both  
379 low risk and medium risk weeks [**Figure 5**]. All models show capacity to predict high risk dengue week.  
380 This can be attributed to extreme weather events influencing climatic hyperparameters, allowing for  
381 models to better detect high risk weeks. From the current results, it can determine that the models  
382 employed can generate accurate predictions with some models performing at 97 percent prediction on a  
383 time-lagged climatic dataset.

384 *Fig 5. The confusion matrices detailing the prediction result of each model: (a) Confusion Matrix of*  
385 *SVM. (b) Confusion Matrix of KNN. (c) Confusion Matrix of Random Forest.*

## 386 **4. Discussion**

### 387 **4.1 Summary of our results.**

388 The purpose of this study was to dynamically model dengue transmission dynamics in tropical  
389 India by the contributing temperature-dependent entomological parameters of *Aedes aegypti*. Our study  
390 provides a dynamic model of dengue transmission along the Godavari River district of eastern coastal  
391 India. We incorporate seven climate parameters into a predictive transmission model and consider the  
392 sensitivity analysis of dengue cases. The benefits of this model are that they can present how dengue  
393 transmission dynamics are influenced by factors such as extreme weather events (i.e., cyclones). This  
394 model only considered limited climate and dengue parameters to provide a proof-of-concept for more  
395 complex models in the future. Moreover, our models operate without examining less accessible satellite  
396 imagery and GIS data. As these models are operate with only numerical 1-d data points, it provides a  
397 low barriers of entry methodology for integrating ML models and further research in dengue forecasting.

398 To demonstrate the feasibility of the models, we explored dengue modeling through both  
399 regression and classification lenses, to showcase the many potential development further research can  
400 dive deeper in the field. In regression, both LSTM and LSTM with Attention models were explored.  
401 These models produced extremely desirable results in determining the dengue cases in the regions with  
402 RMSE below 0.2. These scores are comparable to similar models explored in different regions (Pham et  
403 al., 2022). In classification, it is determined that SVM and Random Forest models provide the most  
404 accurate result and most suited to identifying and differentiating between low risk and high-risk months  
405 and dealing with multiclass labels. This is due to the nature of these algorithms as they are developed to  
406 natively support the classifying multiclass labels.

407 The results of the study provide interesting insights into dengue fever forecast. Through feature  
408 ranking, it was determined that temperature, windspeed and rainfall are variables of the most important  
409 overall to dengue prediction models. Their value to models can be easily explained through observing  
410 the extreme weather events in India and their effects on *Aedes aegypti* population in the region.  
411 Cyclones often introduces environments where mosquitos' population thrives, therefore increasing the  
412 spread of mosquito-borne diseases. Moreover, these models work well with classifying outliers' dengue  
413 cases week. This means that ML models can handle the task of determining outbreak time periods as  
414 they have shown the ability to both capture and predict radical changes in dengue fever case counts and  
415 accurately classifying high dengue risk week.

#### 416 **4.2 Applications of Machine Learning Models in dengue outbreak forecasting**

417 Climatic variables provide suitable indicators for machine learning models to take advantage of  
418 its predictive ability (43). Given the increasingly unpredictable nature of climate change, traditional  
419 statistical models and master systems that rely heavily upon human calibration and expert knowledge are  
420 increasingly unsuitable and inflexible to changes in environmental systems (44). ML models provide a

421 solution in the face of a rapidly changing environment. As such, these models seek to determine the  
422 indirect relationships between climatic variables and climate-sensitive mosquito-borne disease dynamics  
423 without the need for heavy human monitoring. This minimizes the biases influencing predictive models,  
424 while enabling such models to rapidly adapt to changing ecological systems through continuous updated  
425 data and replenished training. Moreover, ML models can extend well to different conditioning factors  
426 without the need to redesign model architecture (45).

427 ML models can extract information on much more extensive set of inputs as compared to  
428 mathematical models and statistical models. A diverse set of inputs, including hydrological and extreme  
429 weather data, helps models generalize better forecast dengue outbreaks. This is because they can better  
430 simulate ecological changes that influence *Ae. aegypti* populations (25). While more traditional models  
431 operate on more fixed inputs and required extensive efforts to extend their feature space, ML models can  
432 easily work with multiple conditioning factors without much additional extension to the base model,  
433 enabling for better development time and adaptability.

434 ML can provide prediction results on imperfect or missing information, a necessity given that  
435 environmental data is often not extensively available, even more so in rural or under-resourced regions.  
436 This enables better reach and understanding of epidemic diseases as it is often in these regions that these  
437 outbreaks are most widespread and most destructive, given a lack of medical resources and disease  
438 monitoring capabilities. Delayed dengue case reporting or incomplete meteorological data can influence  
439 the effectiveness of ML models; however, such a tradeoff is acceptable in real-world implementation of  
440 the ML models presented in this study, as more traditional models often cannot forecast on missing data.

### 441 **4.3 Dengue and climate change**

442 It is established that we are in a increasing climate crisis and that will influence climate-sensitive  
443 mosquito-borne disease dynamics in different capacities across the world (8). There are many mosquito  
444 vectors that transmit important human parasitic and arboviral diseases, but dengue remains the most  
445 common human arboviral disease. It is reported to have a prevalence of 50 million cases in more than  
446 100 countries, with about 500,000 persons requiring hospitalization each year for dengue hemorrhagic  
447 fever/dengue shock syndrome that has overall case fatality rate of 2.5 percent (46). The mounting  
448 evidence around climate-disease relationships raises many important issues about the potential effects of  
449 global climate changes on the transmission of dengue.

450 Although there is growing statistical evidence indicating that dengue outbreaks are associated  
451 with temperature, rainfall, and humidity, few studies have examined this relationship along coastlines  
452 and with the inclusion of extreme weather events, including both a critical spatial component (47) and a  
453 critical climate event worth examining (37). The authors have identified that other variables such as  
454 river levels are also included along with climate variables (48), however, these studies are out of the  
455 study scope. From our study, it is critical to include a variety of geographic topographies and regions in  
456 dengue transmission dynamics models moving forward, but it is also critical to model which climate  
457 variables are the best predictors of dengue transmission. The research in this study indicates that  
458 cyclones are significant contributors to dengue prediction models and must be included going forward.

#### 459 **4.4 Dengue and Coastlines**

460 Dengue demonstrated an uneven distribution along the Godavari River of Andhra Pradesh and  
461 expressed differently on the spatial gradient from coast to inland. The temporal trends for coastal dengue  
462 showed cyclical trends with 3 to 5-week lags between cyclonic events and a peak in dengue counts. This  
463 is interesting because inland, the inland lag time has been estimated to be an average of up to two  
464 months (25) but coastal lags appear to be much smaller time frames in this case. Our findings

465 demonstrate that coastal dengue transmission dynamics showcase faster transmission rates than dengue  
466 occurring deeper inland. Our results also demonstrate that extreme weather events play a role in the  
467 acceleration of dengue transmission along coastlines. Given the similarities between multiple mosquito-  
468 borne diseases, it is important to consider this trend for other diseases such as Zika and chikungunya.

469 Dengue transmission dynamics along coasts are characterized by non-linear dynamics with  
470 strong seasonality, multi-annual oscillations, and non-stationary temporal variation (49). Seasonal and  
471 multi-annual cycles are fluctuating more unpredictably in the face of climate change, and more irregular  
472 intervals of outbreaks are increasingly observed along coastlines (50).

#### 473 **4.4 Lag times and dengue transmission dynamics**

474 The time lag or delayed effect of the climatic variables on dengue counts in our model can be  
475 reasoned by the factors that indirectly influence dengue counts and transmission. This can happen  
476 through effects on vector and virus life-cycle dynamics. The lag is anticipated to be a shorter duration  
477 for minimum temperatures that are usually related to the mortality of an adult mosquito, and humidity  
478 plays a role in influencing this (29). On the other hand, the lag showed that it would be longer for higher  
479 relative water temperatures. Our model shows that the average temperatures, on the other hand, will take  
480 longer to influence dengue counts because it is involved in all biological cycles of *Ae. aegypti* (3).

#### 481 **4.5 Other factors associated with dengue transmission**

482 There are other factors that must be considered in context of these results, such as socio-  
483 economic and demographic factors associated with dengue transmission, as well as the relative  
484 contribution of these factors based on scale and geographic topography. These include mosquito  
485 management, screen dwellings, the use of insect repellants, protection and prevention programs, bed

486 nets, etc., as well as garden accoutrements, storing water in open containers, etc. These are outside of the  
487 scope of our current project, but models should take such data into account.

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505 **Acknowledgements:** We would like to thank the J. William Fulbright Foreign Scholarship Board for  
506 funding a year's worth of unique in-person, on-the-ground field work and data collection that made this  
507 entire project possible. We would like to thank both the United States Government and the Government  
508 of India for their approval and support for this project. We would also like to give a huge thank you to  
509 the Visakha Society for the Protection and Care of Animals (VSPCA), and more specifically, Pradeep  
510 Nath, for the home institution that was provided to the lead author while the work was being conducted  
511 in India. VSPCA was instrumental in supporting this work and this work would not be possible without  
512 them. We would also like to thank the Gandhi Institute of Technology and Management (GITAM)  
513 University for the incredible support during this program. The lead author was able to serve as a faculty  
514 at GITAM during her time in Andhra Pradesh, and GITAM provided foundational support and guidance  
515 for this project. The authors would like to especially thank Dr. Nalini Bikkina, Dr. Divakar Allavarapu,  
516 and Dr. Leben Johnson. The authors would further like to take a moment to honor Dr. Leben Johnson,  
517 whose life was heartbreakingly taken by COVID-19. We would like to thank the GITAM students  
518 whom the lead author taught, for their initial analyses on these data: Amman Naidu, Abhigna Bathina,  
519 Balaji Appari, and Kumari Gunuru. The authors would also like to thank the countless institutions that  
520 helped them piece this work together: the Andhra Pradesh Forest Department, the Water Resource  
521 Board of Andhra Pradesh, the King George Hospital of Visakhapatnam, the Indian Meteorological  
522 Department, the UNDP-GoI (USAID) project, and Andhra University, specifically Dr. Professor C.  
523 Manjulatha, for her guidance and mentorship. The authors would also like to thank Oxford University,  
524 where the lead author sought initial guidance; and Stanford University, where the lead author had a  
525 home through which to finalize this research. We would also like to thank the VNU-HCM High School  
526 for the Gifted for supporting the second author in his contributions to this work.

527



528 **Author Contributions:** Conceptualization, K.T.; Data Curation, K.T.; Formal Analysis, M.P.Q. and  
529 K.T.; Funding Acquisition, K.T.; Investigation, K.T.; Methodology, K.T. and M.P.Q.; Project  
530 Administration, K.T.; Resources, K.T.; Software, K.T. and M.P.Q.; Supervision, K.T.; Validation,  
531 M.P.Q. and K.T.; Visualization, M.P.Q. and K.T.; Writing – Original Draft Preparation, K.T.; Writing –  
532 Analyses, M.P.Q.; Writing – Review and Editing, K.T; Final Reviews, K.T. and M.P.Q.

533 All authors have read and agreed to the published version of the manuscript.

534

535 **Funding:** This research received funding from the Fulbright Program Research Fellowship on behalf of  
536 the J. William Fulbright Foreign Scholarship Board.

537

538 **Conflicts of Interest:** The authors declare no conflict of interest.

539

540

541

542

543

544

545

546

547 **References**

- 548 1. Naish S, Dale P, Mackenzie JS, McBride J, Mengersen K, Tong S. Climate change and dengue: a  
549 critical and systematic review of quantitative modelling approaches. *BMC Infectious Diseases*. 2014  
550 Mar 26;14(1):167.
- 551 2. Kemp L, Xu C, Depledge J, Ebi KL, Gibbins G, Kohler TA, et al. Climate Endgame: Exploring  
552 catastrophic climate change scenarios. *Proceedings of the National Academy of Sciences*. 2022 Aug  
553 23;119(34):e2108146119.
- 554 3. Dhiman RC, Pahwa S, Dhillon GPS, Dash AP. Climate change and threat of vector-borne diseases  
555 in India: are we prepared? *Parasitol Res*. 2010 Mar;106(4):763–73.
- 556 4. Filho WL, Scheday S, Boenecke J, Gogoi A, Maharaj A, Korovou S. Climate Change, Health and  
557 Mosquito-Borne Diseases: Trends and Implications to the Pacific Region. *International Journal of*  
558 *Environmental Research and Public Health*. 2019 Jan;16(24):5114.
- 559 5. Ebi KL, Nealon J. Dengue in a changing climate. *Environmental Research*. 2016 Nov 1;151:115–23.
- 560 6. Johansson MA, Dominici F, Glass GE. Local and Global Effects of Climate on Dengue  
561 Transmission in Puerto Rico. *PLOS Neglected Tropical Diseases*. 2009 Feb 17;3(2):e382.
- 562 7. Messina JP, Brady OJ, Pigott DM, Golding N, Kraemer MUG, Scott TW, et al. The many projected  
563 futures of dengue. *Nat Rev Microbiol*. 2015 Apr;13(4):230–9.
- 564 8. Ramasamy R, Surendran SN. Possible impact of rising sea levels on vector-borne infectious diseases  
565 [Internet]. 2011 [cited 2022 Dec 15]. Available from: [https://link.springer.com/article/10.1186/1471-](https://link.springer.com/article/10.1186/1471-2334-11-18)  
566 [2334-11-18](https://link.springer.com/article/10.1186/1471-2334-11-18)

- 567 9. IPCC. AR5 Climate Change 2013: The Physical Science Basis — IPCC [Internet]. 2014 [cited 2022  
568 Dec 15]. Available from: <https://www.ipcc.ch/report/ar5/wg1/>
- 569 10. Health (U.S.) IWG on CC and. A Human Health Perspective on Climate Change: A Report  
570 Outlining the Research Needs on the Human Health Effects of Climate Change. National Institute of  
571 Environmental Health Sciences; 2010. 84 p.
- 572 11. Moore CG, Cline BL, Ruiz-Tibén E, Lee D, Romney-Joseph H, Rivera-Correa E. *Aedes aegypti* in  
573 Puerto Rico: environmental determinants of larval abundance and relation to dengue virus  
574 transmission. *Am J Trop Med Hyg.* 1978 Nov 1;27(6):1225–31.
- 575 12. Hurtado-Díaz M, Riojas-Rodríguez H, Rothenberg SJ, Gomez-Dantés H, Cifuentes E. Short  
576 communication: Impact of climate variability on the incidence of dengue in Mexico. *Tropical  
577 Medicine & International Health.* 2007;12(11):1327–37.
- 578 13. Brunkard JM, Cifuentes E, Rothenberg SJ. Assessing the roles of temperature, precipitation, and  
579 ENSO in dengue re-emergence on the Texas-Mexico border region. *Salud Publica Mex.*  
580 2008;50(3):227–34.
- 581 14. Descloux E, Mangeas M, Menkes CE, Lengaigne M, Leroy A, Tehei T, et al. Climate-Based Models  
582 for Understanding and Forecasting Dengue Epidemics. *PLOS Neglected Tropical Diseases.* 2012  
583 Feb 14;6(2):e1470.
- 584 15. Hortion J, Mutuku FM, Eyherabide AL, Vu DM, Boothroyd DB, Grossi-Soyster EN, et al. Acute  
585 Flavivirus and Alphavirus Infections among Children in Two Different Areas of Kenya, 2015. *Am J  
586 Trop Med Hyg.* 2019 Jan;100(1):170–3.

- 587 16. Stewart-Ibarra AM, Lowe R. Climate and Non-Climate Drivers of Dengue Epidemics in Southern  
588 Coastal Ecuador. *The American Journal of Tropical Medicine and Hygiene*. 2013 May 1;88(5):971–  
589 81.
- 590 17. Jury MR. Climate influence on dengue epidemics in Puerto Rico. *International Journal of*  
591 *Environmental Health Research*. 2008 Oct 1;18(5):323–34.
- 592 18. Campbell KM, Haldeman K, Lehnig C, Munayco CV, Halsey ES, Laguna-Torres VA, et al. Weather  
593 Regulates Location, Timing, and Intensity of Dengue Virus Transmission between Humans and  
594 Mosquitoes. *PLoS Neglected Tropical Diseases* [Internet]. 2015 Jul [cited 2022 Dec 15];9(7).  
595 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4519153/>
- 596 19. Aswi A, Cramb SM, Moraga P, Mengersen K. Bayesian spatial and spatio-temporal approaches to  
597 modelling dengue fever: a systematic review. *Epidemiology & Infection*. 2019 ed;147:e33.
- 598 20. Caldwell JM, LaBeaud AD, Lambin EF, Stewart-Ibarra AM, Ndenga BA, Mutuku FM, et al.  
599 Climate predicts geographic and temporal variation in mosquito-borne disease dynamics on two  
600 continents. *Nat Commun*. 2021 Feb 23;12(1):1233.
- 601 21. Gupta N, Srivastava S, Jain A, Chaturvedi UC. Dengue in India. *Indian J Med Res*. 2012  
602 Sep;136(3):373–90.
- 603 22. MV RM, Rajaiah K, Kanungo A. Detection of Shoreline Changes Visakhapatnam Coast, Andhra  
604 Pradesh from Multi-Temporal Satellite Images. 2016 Jan 13;05.

- 605 23. Michael E, Singh BK, Mayala BK, Smith ME, Hampton S, Nabrzyski J. Continental-scale, data-  
606 driven predictive assessment of eliminating the vector-borne disease, lymphatic filariasis, in sub-  
607 Saharan Africa by 2020. *BMC Medicine*. 2017 Sep 27;15(1):176.
- 608 24. Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR. Global expansion and redistribution of Aedes-  
609 borne virus transmission risk with climate change. *PLOS Neglected Tropical Diseases*. 2019 Mar  
610 28;13(3):e0007213.
- 611 25. Mordecai EA, Cohen JM, Evans MV, Gudapati P, Johnson LR, Lippi CA, et al. Detecting the  
612 impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models.  
613 *PLOS Neglected Tropical Diseases*. 2017 Apr 27;11(4):e0005568.
- 614 26. Shocket MS, Ryan SJ, Mordecai EA. Temperature explains broad patterns of Ross River virus  
615 transmission. Lipsitch M, Baldwin IT, editors. *eLife*. 2018 Aug 28;7:e37762.
- 616 27. Costa EAP de A, Santos EM de M, Correia JC, Albuquerque CMR de. Impact of small variations in  
617 temperature and humidity on the reproductive activity and survival of *Aedes aegypti* (Diptera,  
618 Culicidae). *Rev Bras entomol*. 2010;54:488–93.
- 619 28. Ibarra AMS, Ryan SJ, Beltrán E, Mejía R, Silva M, Muñoz Á. Dengue Vector Dynamics (*Aedes*  
620 *aegypti*) Influenced by Climate and Social Factors in Ecuador: Implications for Targeted Control.  
621 *PLOS ONE*. 2013 Nov 12;8(11):e78263.
- 622 29. Huber JH, Childs ML, Caldwell JM, Mordecai EA. Seasonal temperature variation influences  
623 climate suitability for dengue, chikungunya, and Zika transmission. *PLOS Neglected Tropical*  
624 *Diseases*. 2018 May 10;12(5):e0006451.

- 625 30. Mala S, Jat MK. Implications of meteorological and physiographical parameters on dengue fever  
626 occurrences in Delhi. *Sci Total Environ*. 2019 Feb 10;650(Pt 2):2267–83.
- 627 31. Agarwal N, Koti SR, Saran S, Kumar AS. Data mining techniques for predicting dengue outbreak in  
628 geospatial domain using weather parameters for New Delhi, India. *Current Science*.  
629 2018;114(11):2281–91.
- 630 32. Shashvat K, Basu R, Bhondekar A. Application of time series methods for dengue cases in North  
631 India (Chandigarh). *Journal of Public Health*. 2021 Apr 1;29.
- 632 33. Mudele O, Frery AC, Zanandrez LFR, Eiras AE, Gamba P. Modeling dengue vector population with  
633 earth observation data and a generalized linear model. *Acta Trop*. 2021 Mar;215:105809.
- 634 34. Anggraeni W, Sumpeno S, Yuniarno EM, Rachmadi RF, Gumelar AB, Purnomo MH. Prediction of  
635 Dengue Fever Outbreak Based on Climate Factors Using Fuzzy-Logistic Regression. 2020  
636 International Seminar on Intelligent Technology and Its Applications (ISITIA). 2020 Jul;199–204.
- 637 35. Fatmawati, Khan MA. The dynamics of dengue infection through fractal-fractional operator with  
638 real statistical data. *Alexandria Engineering Journal*. 2021 Feb 1;60(1):321–36.
- 639 36. Pol S, Rajderkar S, Dhabekar P, Gokhe S. Effect of climatic factors like rainfall, humidity and  
640 temperature on the dengue cases in the metropolitan city of Maharashtra. *International Journal Of  
641 Community Medicine And Public Health*. 2021 Jan 27;8:672.
- 642 37. Cheng J, Bambrick H, Yakob L, Devine G, Frentiu FD, Williams G, et al. Extreme weather  
643 conditions and dengue outbreak in Guangdong, China: Spatial heterogeneity based on climate  
644 variability. *Environmental Research*. 2021 May 1;196:110900.

- 645 38. Scott TW, Morrison AC. Vector dynamics and transmission of dengue virus: implications for  
646 dengue surveillance and prevention strategies: vector dynamics and dengue prevention. *Curr Top*  
647 *Microbiol Immunol*. 2010;338:115–28.
- 648 39. Xu Z, Shen D, Nie T, Kou Y. A hybrid sampling algorithm combining M-SMOTE and ENN based  
649 on Random forest for medical imbalanced data. *Journal of Biomedical Informatics*. 2020 Jul  
650 1;107:103465.
- 651 40. Brownlee J. A Gentle Introduction to the Rectified Linear Unit (ReLU) [Internet].  
652 *MachineLearningMastery.com*. 2019 [cited 2022 Dec 28]. Available from:  
653 [https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-](https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/)  
654 [networks/](https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/)
- 655 41. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments  
656 against avoiding RMSE in the literature. *Geoscientific Model Development*. 2014 Jun 30;7(3):1247–  
657 50.
- 658 42. TR P. A Comparative Study on Decision Tree and Random Forest Using R Tool. *IJARCCCE*. 2015  
659 Jan 30;196–9.
- 660 43. Medlock JM, Leach SA. Effect of climate change on vector-borne disease risk in the UK. *Lancet*  
661 *Infect Dis*. 2015 Jun;15(6):721–30.
- 662 44. Méndez-Lázaro P, Muller-Karger FE, Otis D, McCarthy MJ, Peña-Orellana M. Assessing climate  
663 variability effects on dengue incidence in San Juan, Puerto Rico. *Int J Environ Res Public Health*.  
664 2014 Sep 11;11(9):9409–28.

- 665 45. Sarker IH. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications  
666 and Research Directions. SN Comput Sci. 2021;2(6):420.
- 667 46. World Health Organization. Dengue guidelines for diagnosis, treatment, prevention and control :  
668 new edition [Internet]. World Health Organization; 2009 [cited 2022 Dec 19]. Report No.:  
669 WHO/HTM/NTD/DEN/2009.1. Available from: <https://apps.who.int/iris/handle/10665/44188>
- 670 47. Patz JA, Martens WJ, Focks DA, Jetten TH. Dengue fever epidemic potential as projected by  
671 general circulation models of global climate change. Environmental Health Perspectives. 1998  
672 Mar;106(3):147–53.
- 673 48. Hashizume M, Dewan AM, Sunahara T, Rahman MZ, Yamamoto T. Hydroclimatological  
674 variability and dengue transmission in Dhaka, Bangladesh: a time-series study. BMC Infectious  
675 Diseases. 2012 Apr 24;12(1):98.
- 676 49. Pham HV, Doan HT, Phan TT, Tran Minh NN. Ecological factors associated with dengue fever in a  
677 central highlands Province, Vietnam. BMC Infectious Diseases. 2011 Jun 16;11(1):172.
- 678 50. Hafsia S, Haramboure M, Wilkinson DA, Baldet T, Yemadje-Menudier L, Vincent M, et al.  
679 Overview of dengue outbreaks in the southwestern Indian Ocean and analysis of factors involved in  
680 the shift toward endemicity in Reunion Island: A systematic review. PLOS Neglected Tropical  
681 Diseases. 2022 Jul 28;16(7):e0010547.

682



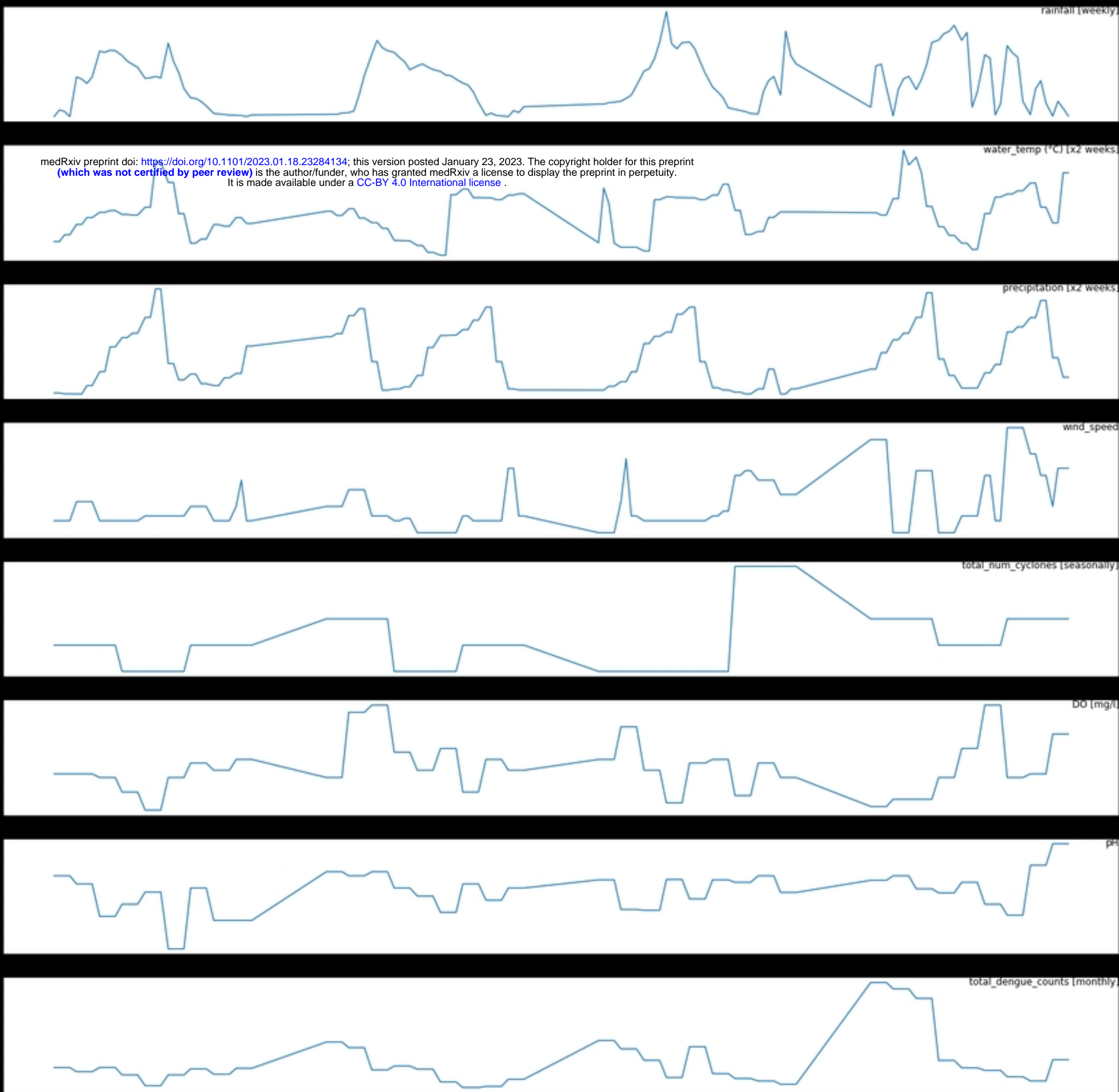


Figure 1



Figure 2

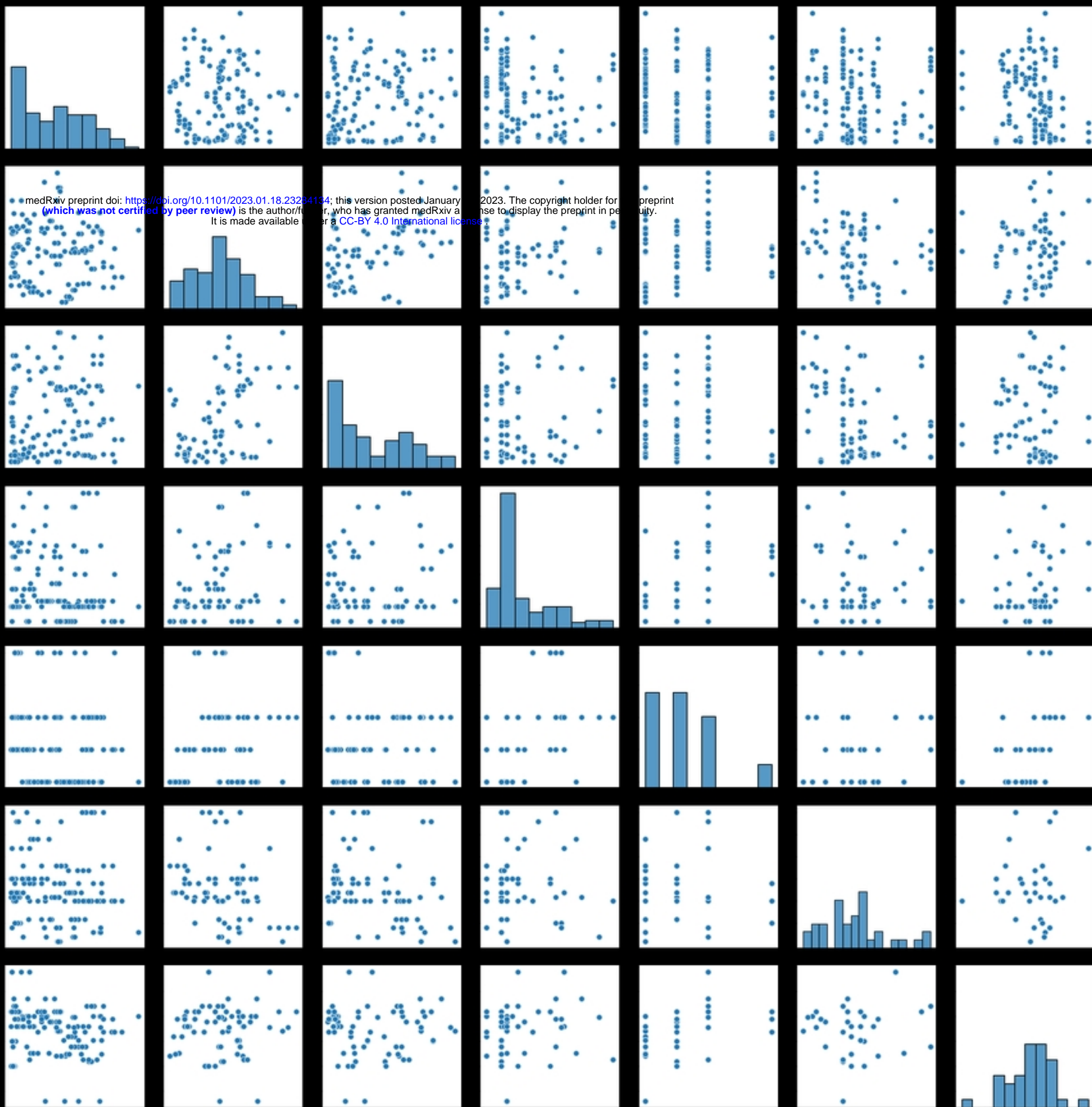


Figure 3

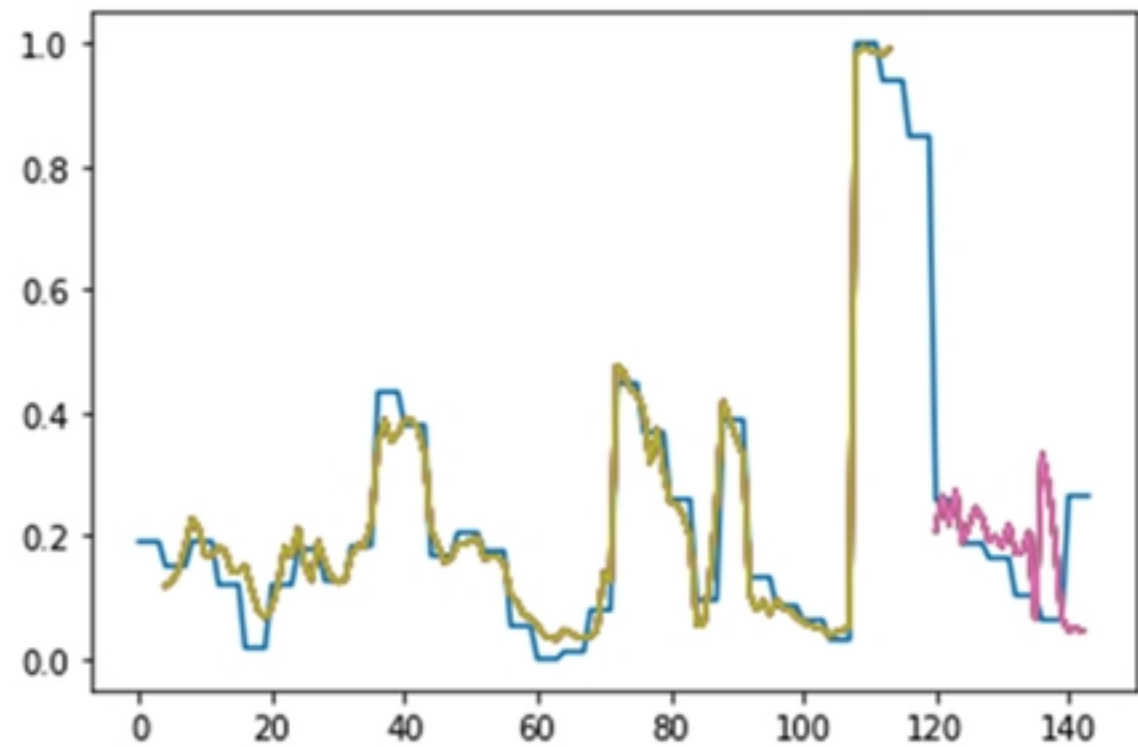
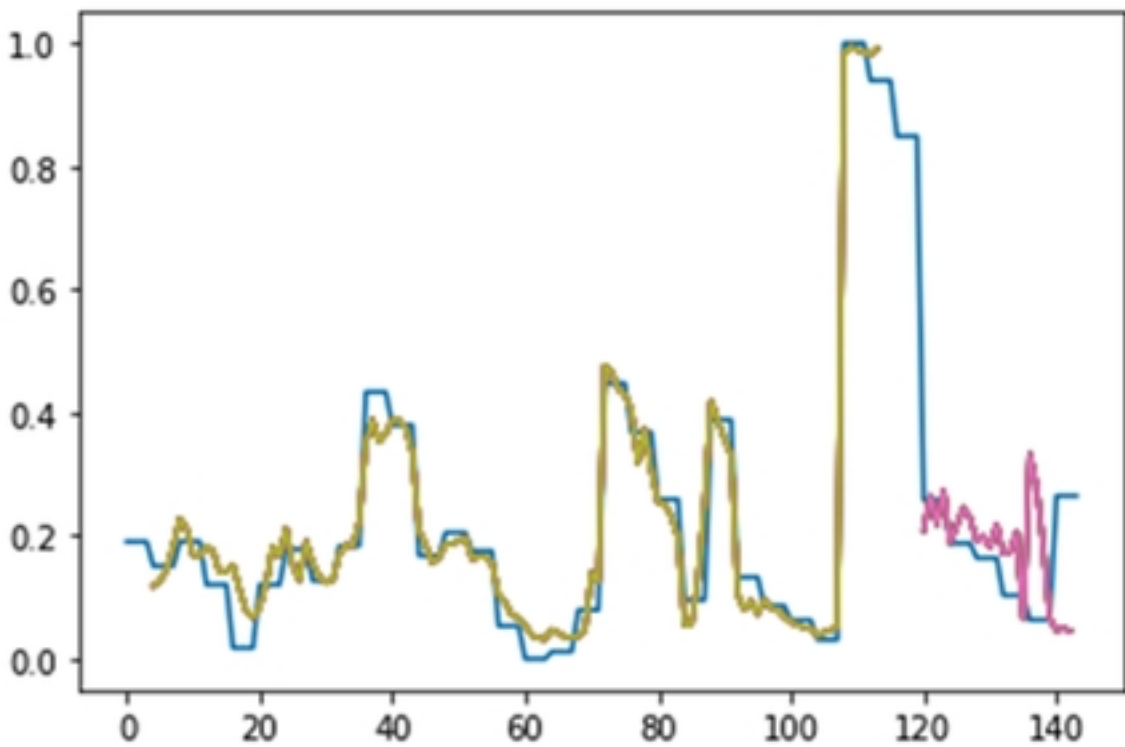
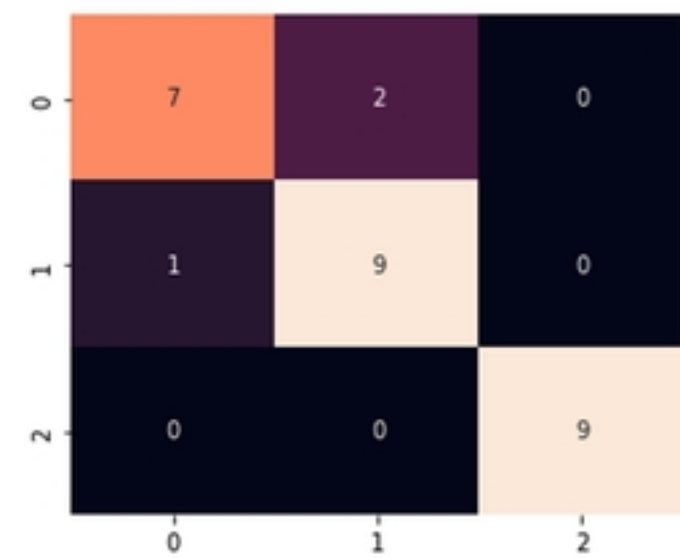
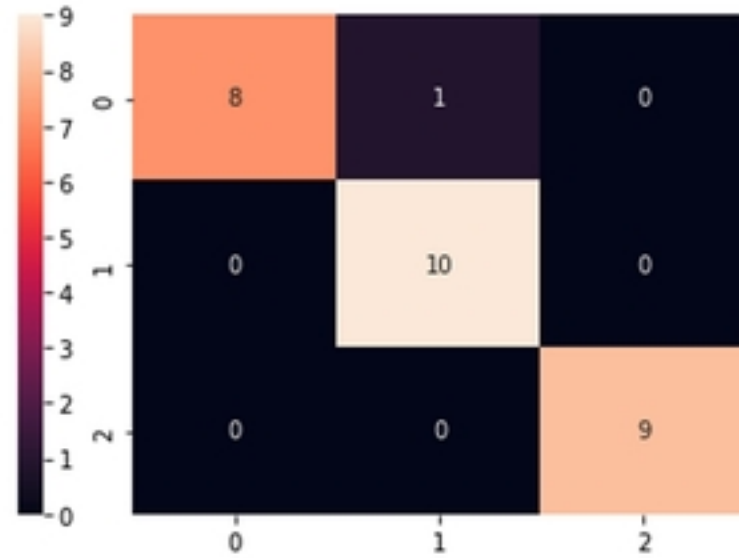


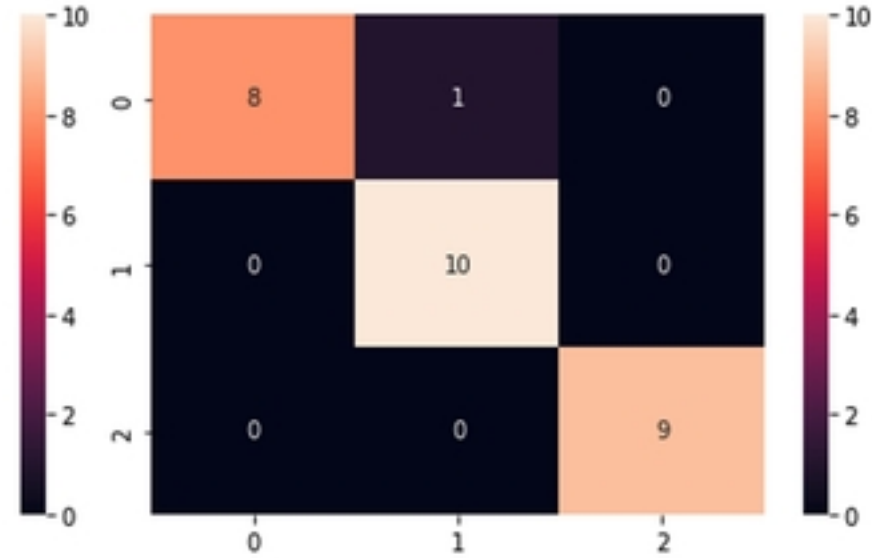
Figure 4



(a)



(b)



(c)

Figure 5