Predicting Dengue Fever cases:

San Juan, Puerto Rico

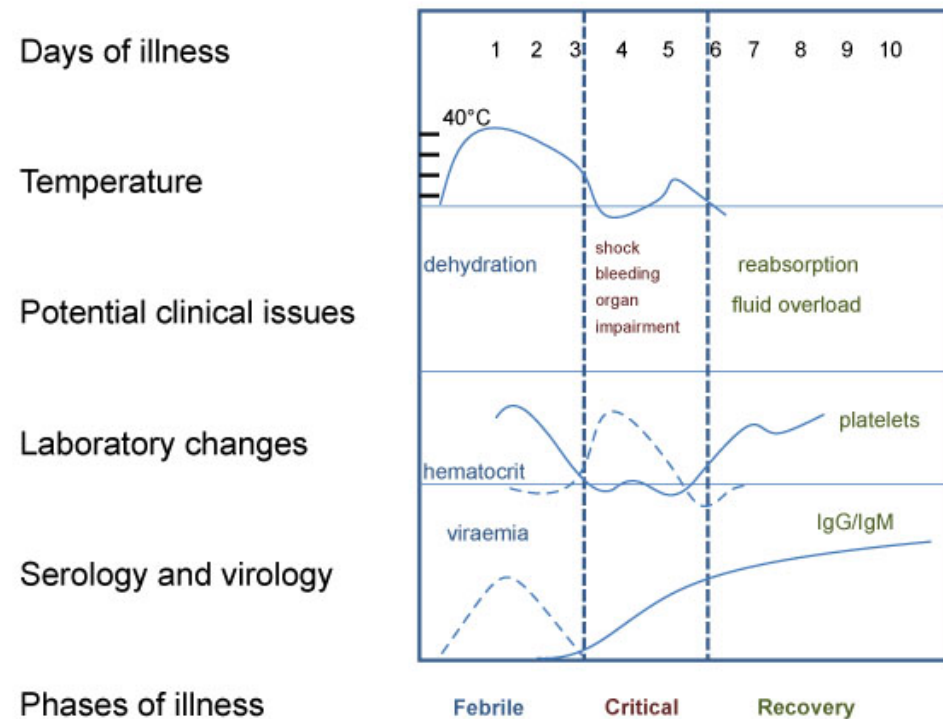## Iquitos, Peru

Giovanni Rosati
September 26, 2019

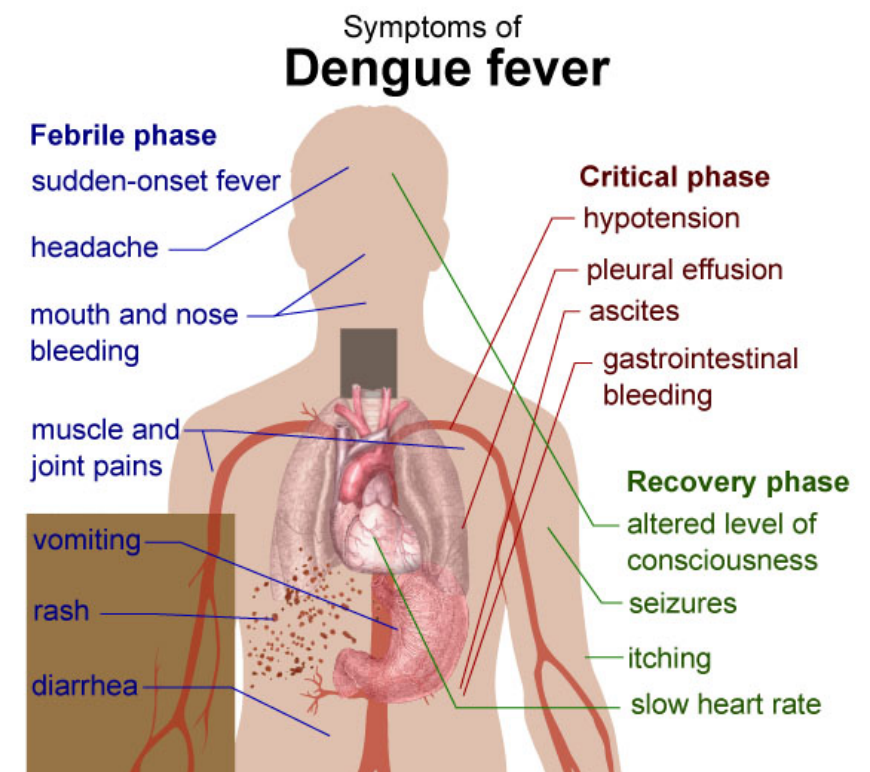# Dengue Fever

❖ The Problem:

- • Dengue Fever infects up to 400 million humans every year.

- • Infection is spread by female mosquitos

- • Severe cases can result in death.

❖ The Goal:

- • Create a machine learning model that can accurately predict the number of cases.

By Dr Graham Beardsderivative work: UAwiki - This file was derived from: Course of Dengue illness.png:, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=27425148
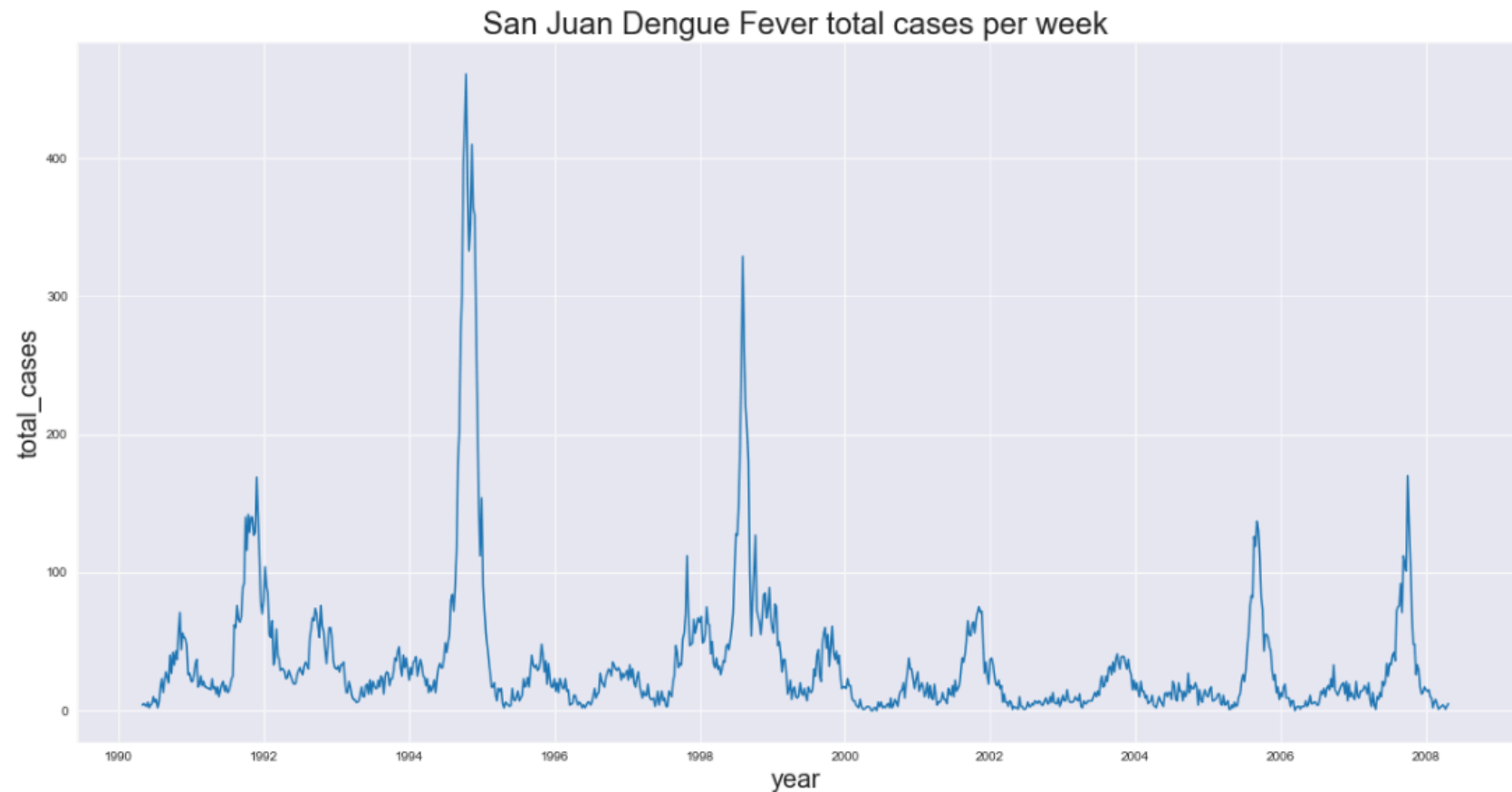

By Mikael Häggström - All used images are in public domain; references are found in Wikipedia article at Wikipedia:Dengue fever., Public Domain, https://commons.wikimedia.org/w/index.php?curid=12712988
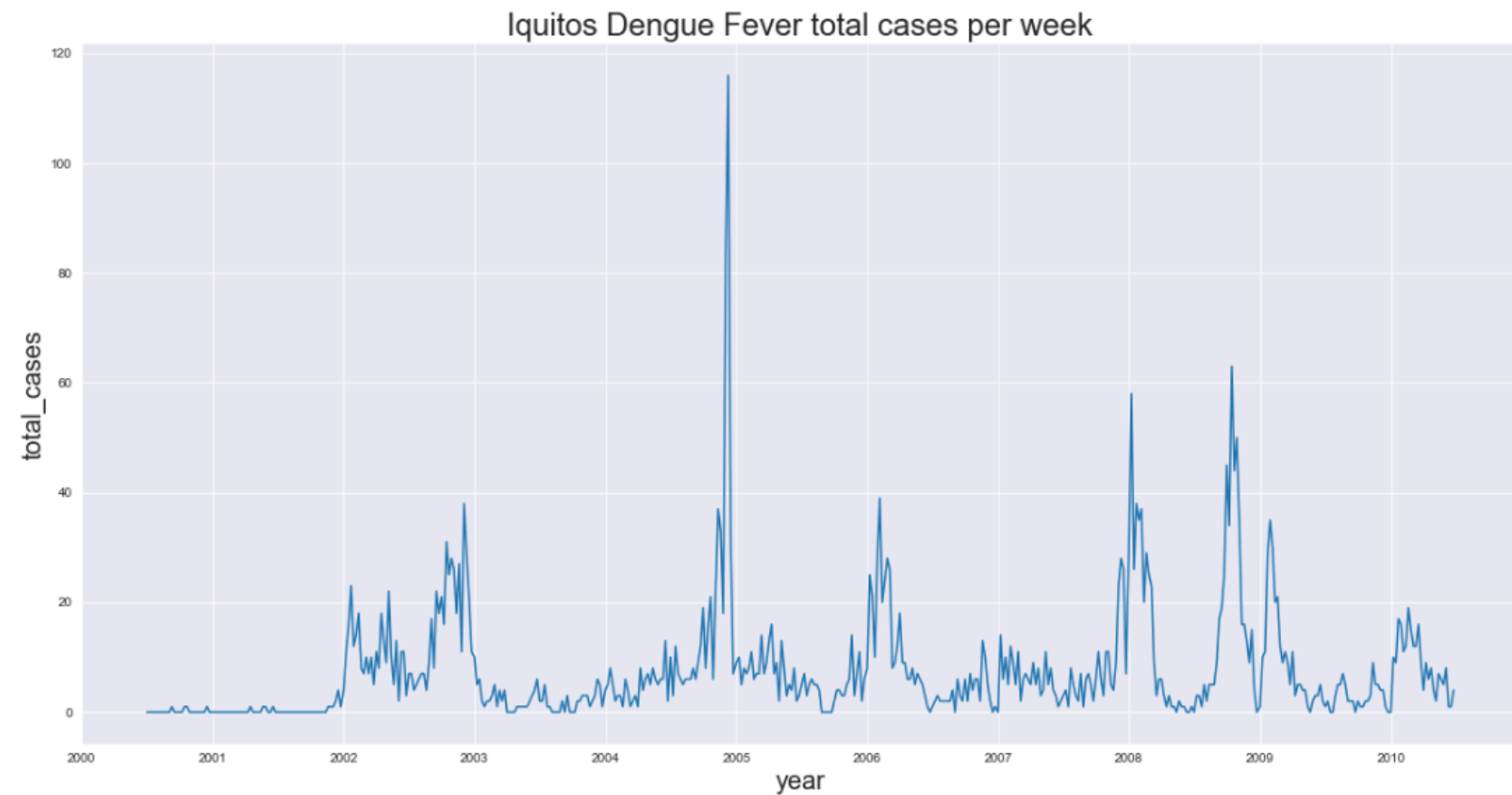
Dengue Fever Outbreaks

San Juan Dengue Fever total cases per week

❖ San Juan, Puerto Rico

• Data from mid-1990 through early-2008

❖ Iquitos, Peru

• Data from mid-2000 through early 2011.

Iquitos Dengue Fever total cases per week

Weather Data

Looking for missing values in
6,937
Dark Sky records

- ❖ Sources:

  - ❖ U.S. National Oceanic and Atmospheric Administration

  - ❖ Dark Sky API

- ❖ Issues:
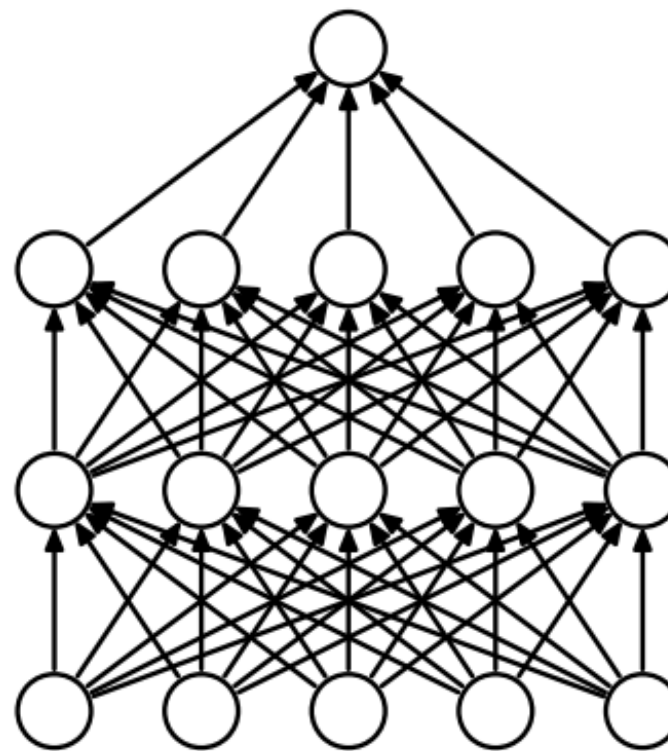
  - ❖ Missing information for some dates.

  - ❖ Missing values for many variables.

  - ❖ Dark Sky data is daily - NOAA data is weekly.

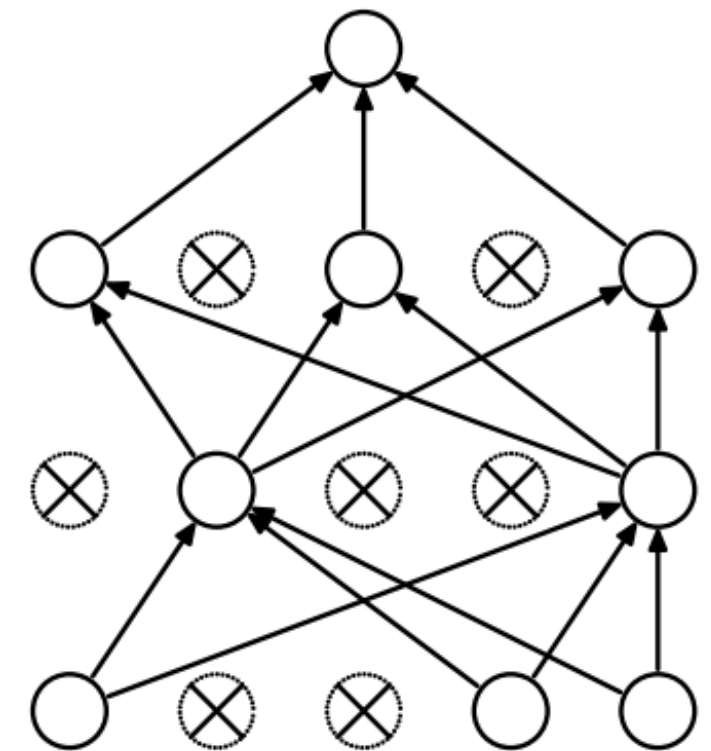|	
date	16
latitude	0
longitude	0
timezone	0
offset	0
apparentTemperatureHigh	18
apparentTemperatureHighTime	18
apparentTemperatureLow	17
apparentTemperatureLowTime	17
apparentTemperatureMax	18
apparentTemperatureMaxTime	18
apparentTemperatureMin	18
apparentTemperatureMinTime	18
cloudCover	175
dewPoint	18
humidity	18
icon	16
moonPhase	16
precipIntensity	3695
precipIntensityMax	3695
precipIntensityMaxTime	5041
precipProbability	3695
precipType	1561
pressure	18
summary	16
sunriseTime	16
sunsetTime	16
temperatureHigh	18
temperatureHighTime	18
temperatureLow	17
temperatureLowTime	17
temperatureMax	18
temperatureMaxTime	18
temperatureMin	18
temperatureMinTime	18
time	16
uvIndex	16
uvIndexTime	16
visibility	3918
windBearing	18
windGust	4854
windGustTime	4854

Technology - Ketas & Tensorflow

❖ Sequential Model w/4 layers

- LSTM (Long Short Term Memory)

- Dropout (50% chance of removing a node)

- Dense (50 nodes)

- Dense (prediction output)

(a) Standard Neural Net   (b) After applying dropout.

Image from: https://medium.com/@bingobee01/a-review-of-dropout-as-applied-to-rnns-72e79ecd5b7b

Tuning Hyperparameters

Dataset name: iq_final
Lookback window periods: 4
Training proportion: 0.79
Dropout proportion: 0.5
last epoch loss: 0.03095
last epoch val_loss: 0.03681
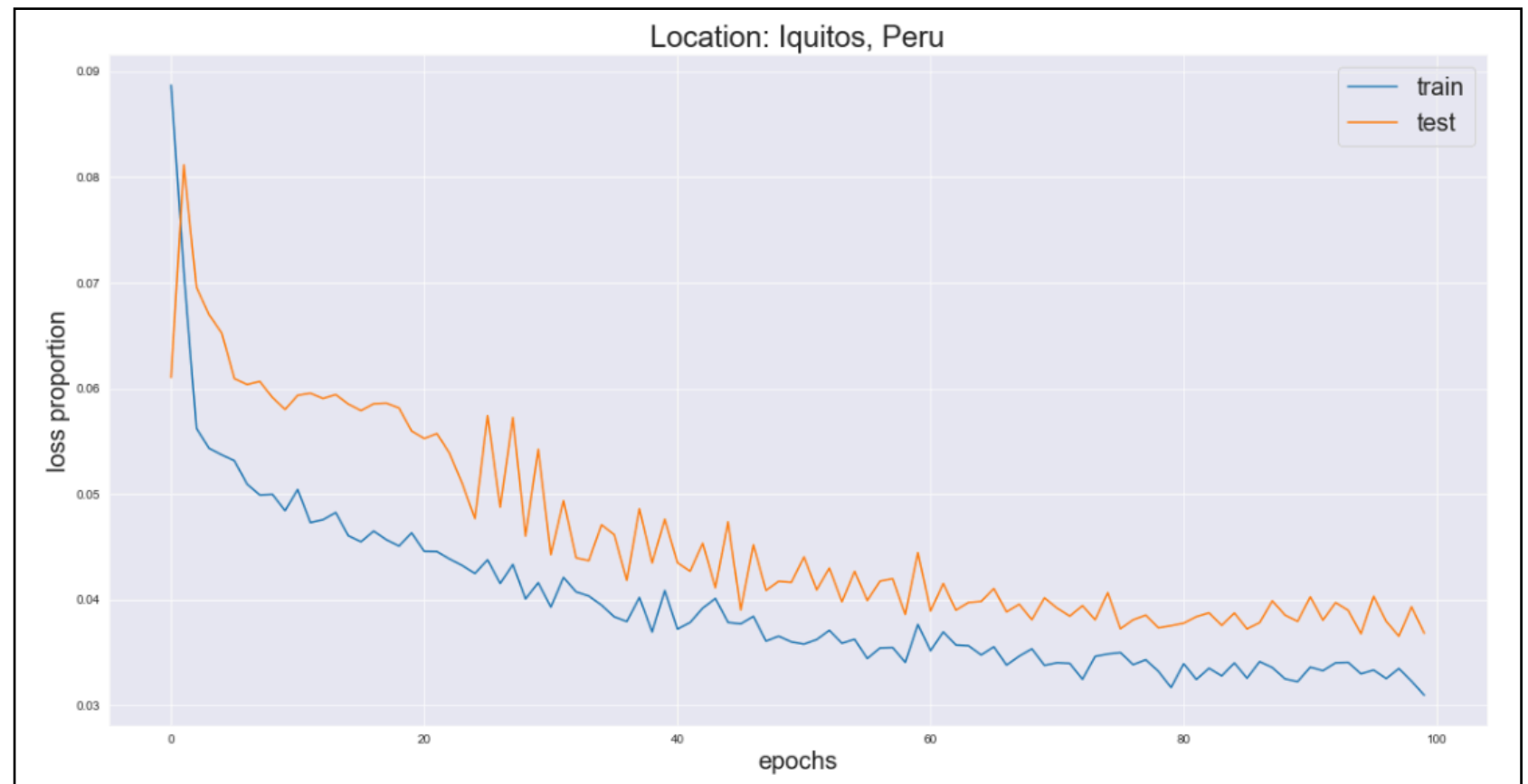
Dataset name: iq_final
Lookback window periods: 1
Training proportion: 0.79
Dropout proportion: 0.4
last epoch loss: 0.03221
last epoch val_loss: 0.03264
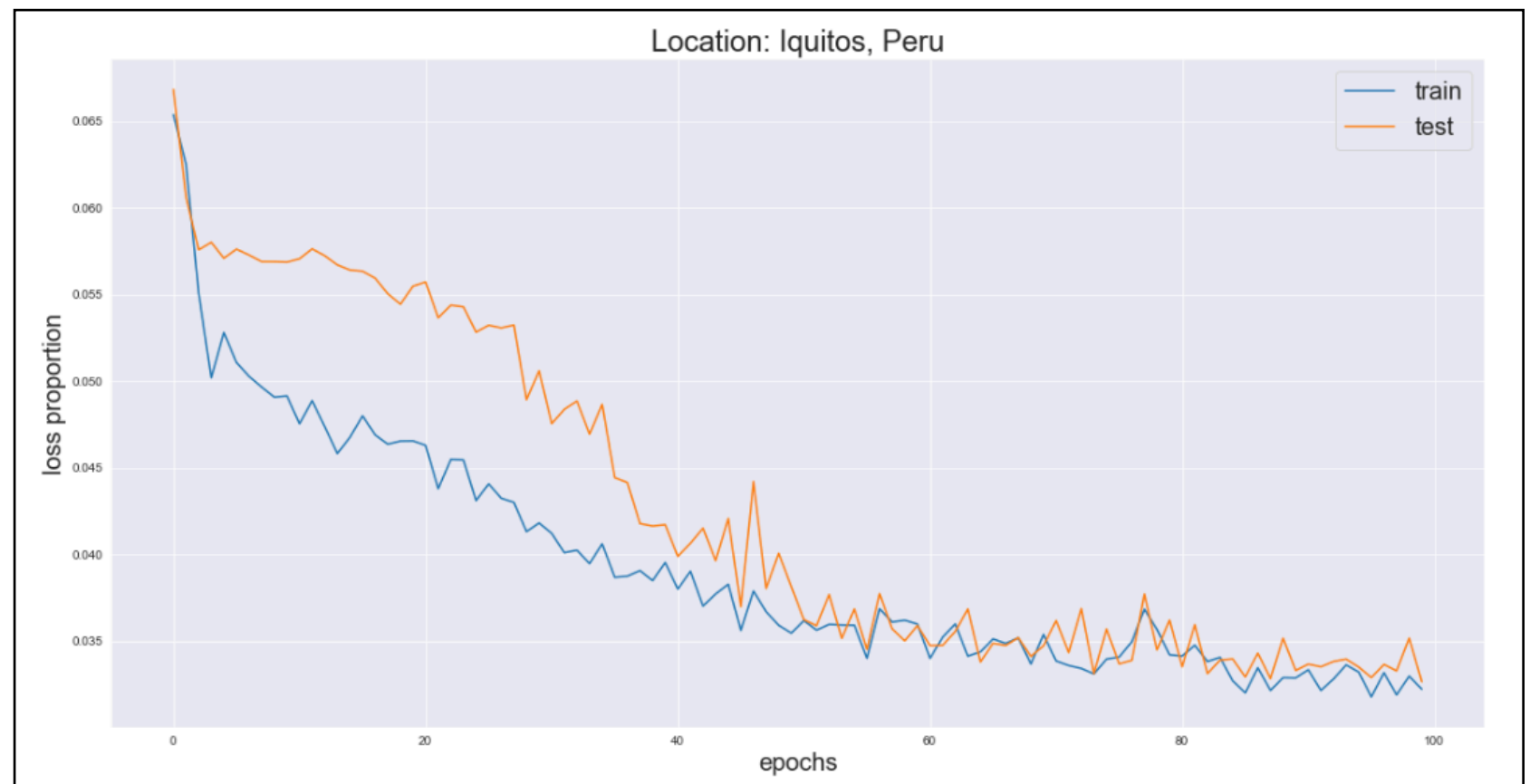
Different Models for Iquitos & San Juan

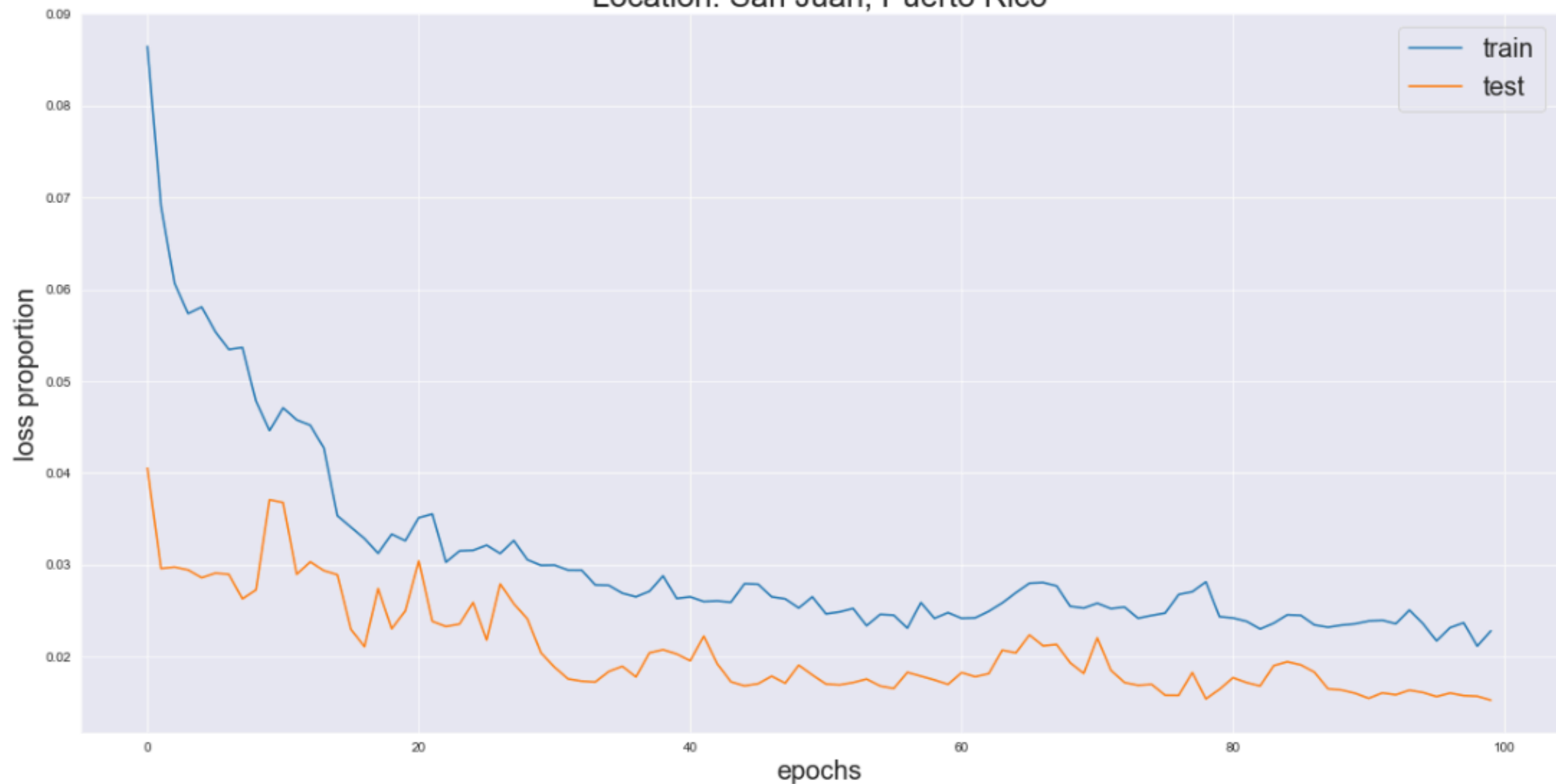Location: San Juan, Puerto Rico
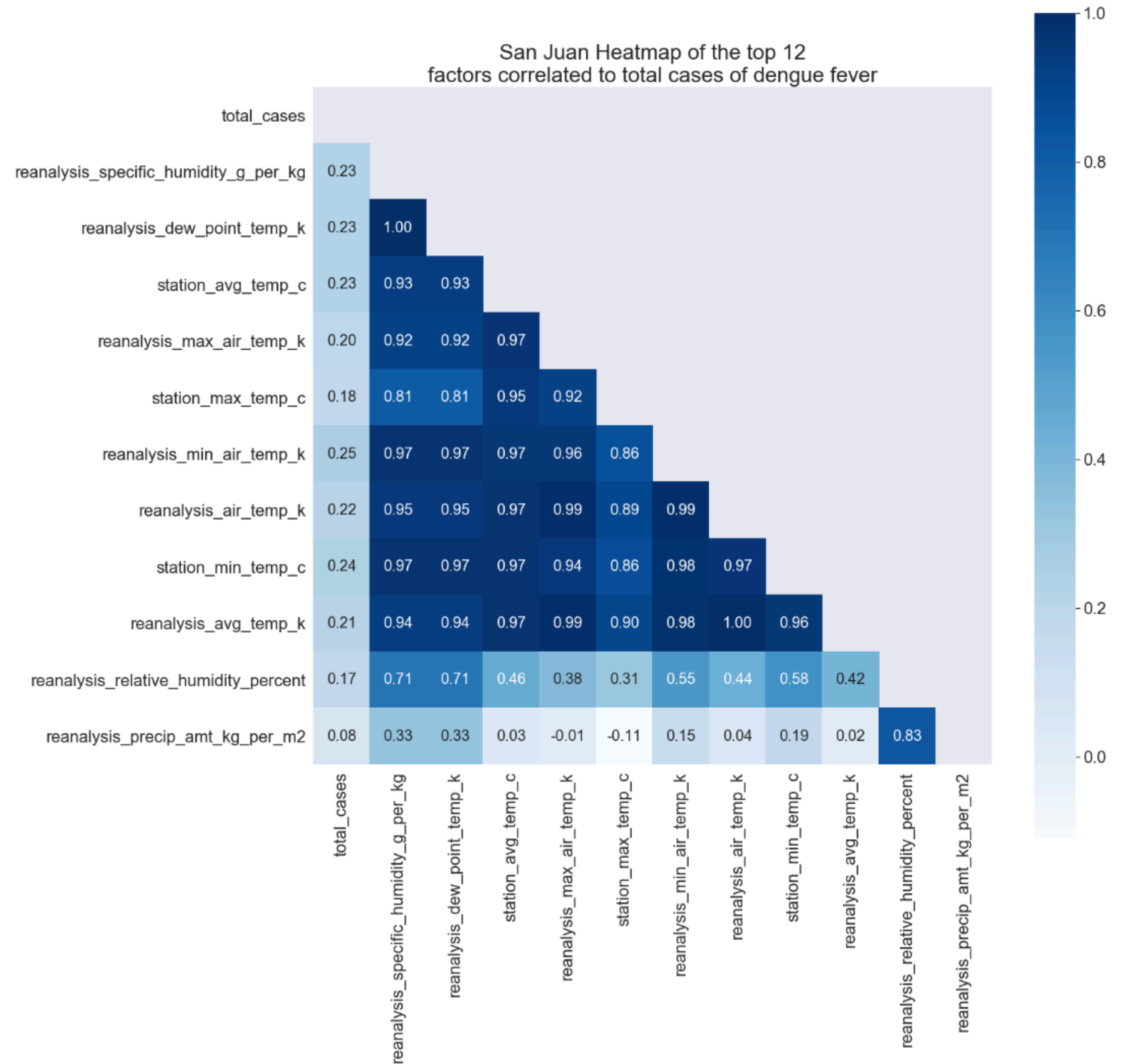
Dataset name: sj_final
Lookback window periods: 4
Training proportion: 0.87
Dropout proportion: 0.5
last epoch loss: 0.02283
last epoch val_loss: 0.01528

Further Work

- Test different ways of creating weekly data from the Dark Sky data (eg., minimum or maximum instead of the mean).

- Test differencing the data (Subtracting each weeks values from the prior weeks values).

- Test other machine learning models designed for time series data.

- Test removing variables with multicollinearity.

San Juan Heatmap of the top 12
factors correlated to total cases of dengue fever

	total_cases	reanalysis_specific_humidity_g_per_kg	reanalysis_dew_point_temp_k	station_avg_temp_c	reanalysis_max_air_temp_k	station_max_temp_c	reanalysis_min_air_temp_k	reanalysis_air_temp_k	station_min_temp_c	reanalysis_avg_temp_k	reanalysis_relative_humidity_percent	reanalysis_precip_amt_kg_per_m2
total_cases												
reanalysis_specific_humidity_g_per_kg	0.23											
reanalysis_dew_point_temp_k	0.23	1.00										
station_avg_temp_c	0.23	0.93	0.93									
reanalysis_max_air_temp_k	0.20	0.92	0.92	0.97								
station_max_temp_c	0.18	0.81	0.81	0.95	0.92							
reanalysis_min_air_temp_k	0.25	0.97	0.97	0.97	0.96	0.86						
reanalysis_air_temp_k	0.22	0.95	0.95	0.97	0.99	0.89	0.99					
station_min_temp_c	0.24	0.97	0.97	0.97	0.94	0.86	0.98	0.97				
reanalysis_avg_temp_k	0.21	0.94	0.94	0.97	0.99	0.90	0.98	1.00	0.96			
reanalysis_relative_humidity_percent	0.17	0.71	0.71	0.46	0.38	0.31	0.55	0.44	0.58	0.42		
reanalysis_precip_amt_kg_per_m2	0.08	0.33	0.33	0.03	-0.01	-0.11	0.15	0.04	0.19	0.02	0.83	

The Future

Back to the Future: Using Historical Dengue Data to Predict the Next Epidemic

JUNE 5, 2015 AT 1:10 PM ET BY DYLAN GEORGE

"Accurate dengue predictions would help public health workers, like Johansson, and people around the world take steps to reduce the impact of these epidemics. But predicting dengue is a hefty task that calls for the consolidation of different data sets on disease incidence, weather, and the environment."*

*https://obamawhitehouse.archives.gov/blog/2015/06/05/back-future-using-historical-dengue-data-predict-next-epidemic