Chromatic Aberrations for Model Uncertainties and Textures for Tertiary Property Visualization in 2-D Surface with COVID-19 Pandemic Data

Md Rashidul Islam
Faculty of Computer Science, Dalhousie University
6050 University Ave, Halifax, Nova Scotia, Canada, B3H 4R2
md313724@dal.ca

Abstract:

In recent years ~~tons of~~ an increasing array of research are being conducted by researchers in the field of visualization and in time series forecasting by employing various machine learning models. Since COVID-19 is a respiratory infectious disease caused by novel coronavirus (also known as SARS-CoV-2) and due to its unprecedented challenges over time and ~~impacts to the whole human beings~~ global impact., World Health Organization (WHO) has recognized it as a global pandemic. After ~~administering~~ conducting more than a yearlong research, several companies manufactured vaccines with different names and as a result, immunization has ~~been~~ started in many countries and that significantly helps to reduce the spread and severity of the infections. From the very beginning scientists and researchers are ~~having discreet~~ investigating ~~on~~ the perceived data to discover the cause, find the patterns in different countries or demographic areas. So, in this study, we ~~come up with~~ introduce. a novel idea ~~to develop~~ a visualization ~~tool~~ to present predictive model uncertainties and visualize textures to represent a third property in 2D space. We utilized some common and existing machine learning models to ~~get~~ obtain. the predicted results and find the model uncertainties for the most impacted countries with respect to number of new cases, total cases, death rate and recovery rate for different countries. Finally, we visualize the calculated model uncertainties in terms of chromatic aberration and textures ~~in web interface with user~~ in an. interactive ~~way~~ fashion.

for

LIST OF ABBREVIATIONS USED

D3	-	Data Driven Documents
JSON	-	JavaScript Object Notation
API	-	Application Programming Interface
ANN	-	Artificial Neural Network
RNN	-	Recurrent neural networks
MLP	-	Multilayer Perceptron
CNN	-	Convolutional Neural Network
LSTM	-	Long Short-Term Memory
MAE	-	Mean Absolute Error
RMSE	-	Root Mean Square Error
WHO	-	World Health Organization

1 Introduction

The scope of the research includes ~~involvement of~~ machine learning models ~~to~~ for visualization ~~tools in web interface~~. Some of the popular predictive machine learning algorithms are used for ~~getting~~ obtaining the forecasting data and corresponding uncertainty and then visualize d ~~in web by using the handy~~ with the JavaScript library D3.

1.1 Background and Motivation

The outbreak of coronavirus COVID-19 first emerged in China in December 2019 and the ~~surprising~~ expansion has ~~been~~ occurred all over the world and declared as an international public health crisis ~~in a couple of weeks~~ by the World Health Organization. Since then, the world is quite ~~stagnant~~ affected in almost all respects. Various preventive health measures are imposed, and different short-term restrictions are applied to the habitants in different countries at different times. But the ~~death procession~~ mortality rate isn't yet controlled significantly and over 217 million people ha~~s~~ve been infected and 4.1 million ~~has been~~ have died ~~in~~ the world over. The infection and death rate are oscillating in different countries due to various reasons ~~like~~ such as public insincerity and lack of consci~~ou~~sness encious about the disease. Moreover, the strain of the virus is changing frequently in different geographical place~~s~~ with more power and variations~~,~~ and a few of the variants like the British variant, the South Africa variant, and the Indian variant became the prime ~~headache~~ concern for the world communit~~ies~~y, because sufficient research has not been conducted on them yet. Though ~~hundreds of thousands of~~ a great deal of research are being ~~proceeding but~~ conducted nobody knows when the ~~whole~~ world ~~human being~~ will get rid of this cruel pandemic and ~~come back in~~ return to normal life again.

In recent days, many studies have been conducted to forecast the trend of the spread of the COVID-19 pandemic using various statistical models as well as machine learning models. The autoregressive integrated moving average (ARIMA) model has been widely used in the previous studies to analyze and predict the spread of the disease along with many other popular machine learning algorithms. ~~In~~ To our knowledge, most of studies are administered for specific countries. ~~Since~~ The pandemic started very abruptly and so during the first year ~~period~~, it was ~~quite impossible~~ difficult to develop any efficient system to forecast anything ~~about~~ even for a single day. But after passing more than one year ~~of painful time~~, we have sufficient data to explore, analyze and forecast ~~with maximum possible efficacy~~ with the help of modern machine learning algorithms. The ability to identify the expansion rate at which the disease is spreading is very important to confront ~~against~~ it and help ~~the~~ governments' regarding contingent policymaking to properly address the consequences of the pandemic and ~~make bound the~~ encourage people to follow the rules and health guidelines to achieve the maximum benefit by saving valued lives. The principal objective behind the current research is: to perceive the extent of country-wise uncertainty by discovering property driven predicted results of COVID-19 for certain period of time and employ those uncertainties in visual representation as chromatic aberration that can help the community administrators' for better planning, providing ~~guidelines~~ insights to minimize its impact. Last but not the least, the ~~novelty of the third property~~ visualization in 2-dimensional space helps the users to ~~get~~ perceive dual impacts in a single ~~go~~ view. For instance, new cases and new deaths are provided in a single chart with the help of glyph-like texture representations.

1.2. Problem statement

The primary objective of the research is to present a novel concept of ~~presenting~~ employing chromatic aberration ~~from~~ to machine learning models' uncertainties ~~in graphical display device~~ by amassing the COVID-19 data. ~~On top of that~~ We also present how a tertiary property can be displayed over a two-dimensional viewing system by using the color blended texture elements and convey dual valued message from a single content.

As stated earlier many working with relevant research are conducted in this domain, so we intend to pursue ~~a~~ research by ~~taking some~~ best performer machine learning models ~~which were not used in a single study before~~. For example: ARIMA is used in some research and SVM (Support Vector Machine) is used in ~~some~~ other research, ~~so~~ we ~~can~~ keep both in our candidate list. We will select three models for our final execution of the experiment but before that we must ~~do~~ perform a preliminary scrutiny from ~~that big~~ a large set of candidates ~~list~~.

1.3. ~~Contributions~~ Approach.

At the ~~very~~ first step we ~~headed to look for~~ sought a suitable dataset in terms of completeness for the whole duration of covid. By analyzing numerous data repositories, we came to know that OWID is the most ~~competent~~ comprehensive one among all others.

Secondly, we had to study different journals about forecasting from temporal data using machine learning models and ~~henceforth chosen~~ chose three popular modeling algorithms for our research. Since, finding and comparing the effectives of algorithms' is out of our scope of work chose a reasonable set of

~~so~~ we randomly ~~chosen the~~ models because we need ed to generate the uncertainty data for the countries by using ~~the~~ thier predictions.

it basic is

Thirdly, we have chosen D3.js as our front-end library for drawing the charts ~~in web platform~~ because ~~that is much lighter in weight and lot of common stuffs are bundled inside and easily reusable~~. Since developing the drawing algorithms ~~are~~ not our goal, we have fairly taken the benefit of reusing library features.

an efficient platform for visualization prototyping

Finally, having the data generation component in python, we needed to write up APIs to connect and pull the data ~~and employ in~~ drawing the charts ~~in visualization displays~~. Since the model training and data generation for all countries are long running processes, ~~so made it standalone to generate the data as json and stored in file~~ so that that can be read and sent back to the client on demand.

when

it pre computed the data into a json file.

1.4. Thesis outline

TBA in the final paper.

2. Prior Works:

This study involves three major components i. generate time series forecasted data from COVID-19 data using some of the predictive models and calculate corresponding uncertainties for different countries ii. Visualizing Chromatic Aberration to represent the perceived uncertainties ~~in~~ a graphical presentation surface iii. Incorporate data of three variables and orient in a two-dimensional geometric setting. In this section, we are going to include some closely related studies conforming to ~~one or other side~~ aspects of the research ~~context~~.

2.1 the Prediction in Machine Learning Models

~~On~~ machine learning forecasting side, Song et. al. [1] compiled monthly data of influenza incidence s from all provinces in mainland China from January 2004 to December 2011, comprehensively evaluated and classified these data, and then randomly selected 4 provinces with higher, median and lower incidence s, using time series analysis to construct an ARIMA model. ~~With~~ The same model but different analysis and forecasting approach es was conducted on the coronavirus disease by ~~some~~ other researcher s ~~in Pakistan~~ [2]. Recent studies of [3, 4] use Facebook's Prophet Forecasting Model and ARIMA Forecasting Model to compare their performance and accuracy on the dataset containing the confirmed cases, deaths, and recovered numbers, obtained from the Kaggle website. The forecast models are then compared to the last 2 weeks of the actual data to measure their performance against each other. The result shows that Prophet generally outperforms ARIMA. Several neural network predictive models are used to evaluate their performance against more common machine learning models in a Dengue forecasting project [7]. ~~Muktevi~~ Srivenkatesh applied Naïve Bayes, logistic regression, support vector machines, Random Forest, K Nearest Neighbour for the examination of liver malady. The classification s ~~exhibitions~~ are assessed with 5 distinctive ~~by and large~~ execution measurements, i.e., precision, kappa, Mean absolute error (MAE), Root mean square error (RMSE), and F measures. The objective of this query work is to foresee liver infection s with different machine learning approaches, and pick most efficient algorithm [9]. Results of the examination demonstrated that a Logistic Regression classifier demonstrated the best outcomes regarding precision ~~and~~ with the least execution time s.

2.2 Uncertainty in Visualization (reference #?)

~~R.P~~. Botchen et al. focuses on uncertainty that occurs during data acquisition and demonstrates the usefulness of the methods for the example of real-world fluid flow data measured with the particle image velocimetry (PIV) technique. They present two novel texture-based techniques to visualize uncertainty in time-dependent 2D flow fields where in the first method, texture advection is employed to show flow direction by streaklines and convey uncertainty by blurring these streaklines and in a second method isotropic diffusion implemented by Gaussian filtering to continuous change of the density of flow representation.

Error in data is inherent so it cannot be ignored in visualization. Improper or eliminated presentations in visualizations can mislead ~~in~~ decision making ~~to~~ for data analysts. The goal of uncertainty visualization is to minimize the errors in judgment and represent the information as accurately as possible. This survey Aasim Kamal et al. (#?) discusses state-of-the-art approaches (such as?) to uncertainty visualization, along with the concept of uncertainty and its sources.

2.2. Chromatic Aberration in Visualization

Again, from a vision ~~on visualization~~ perspective, chromatic aberration leads to various forms of color imperfections in the image. When tampering with an image, these aberrations are often disturbed and fail to be consistent across the image. ~~K~~. Koh et. al. presented a user study to observe the effect on users' judgment with Lateral Chromatic Aberration for Chart Reading in Information Visualization on Display Devices [10] and suggested guidelines for information visualization designers to avoid such issues. Hyun Seung Yoo et. al. explained Colour illusion on liquid crystal displays and design guidelines for bioinformatics tools to enhance the usability and design of LCD monitors [11]. Lens flare rendering in real-time applications [14] are explained by ~~M.~~ Hullin et. al. whereas ~~S.~~ Lee ~~nicely~~ presented ~~the~~ blur effects and focus control to retain a realistic look of the display elements [15]. One of the interesting research projects conducted by ~~M. K.~~ Johnson shows that inconsistencies in lateral chromatic aberration can be used to detect tampering in visually plausible forgeries [13].

Bonneau et al. [16] explores uncertainty in the visualization domain by comparing different results, such as a weather forecast generated with different parameters and to detect similarities

(Should this be in 2.2??)

or differences in the results a comparative visualization technique is employed. To compare certain regions in more detail, e.g., borders, they suggested to consider larger comparison areas than individual pixels and it is crucial that data sets which should be compared are visualized next to each other to get a direct comparison for a certain area.

Objective uncertainty of a visual system is evaluated by ~~Simon~~ Barthelme et al. [17] where they discuss the natural perceptual systems involvement with systematic uncertainty because sensory information is imperfect and insufficient to uniquely designate the environment. In their experiment, observers were presented with pairs of images of oriented objects embedded in high levels of noise and had to report the orientation of the image of their choice. Even though the two images contained the same extent of noise, one particular noise structure made an image orientation more obvious than the other. Eventually, observers reliably chose the more obvious of the two images, thereby providing evidence of a capacity to accurately evaluate objective uncertainty.

A statement on the position of uncertainty visualization today is explained in ~~Henning~~ Griethe et al. [18] that defines the basic concept of uncertainty and discusses sources and necessary measures. Visualization is an indispensable approach to the exploration and communication of large data sets of different domains where data sets may contain an unavoidable amount of uncertainty that needs to be included in the visualization process to enable the correct cognition of hidden facts and figures. In addition, it explains how existing approaches could be systematically presented to the acquisition and display of uncertainty can be transferred to new fields, e.g., the visualization of uncertainty in structures.

Uncertainty visualization is a research area that integrates visualization with the study of uncertainty. Among many uncertainties representation of participant-based empirical techniques, there is little evidence in Deitrick et al. [19] to suggest that uncertainty visualization influences in results or decisions. Through a human-subjects experiment, this research evaluates uncertainty visualization methods and indicates that it may affect decisions, but the degree of influence is affected by how the uncertainty is expressed.

State-of-the-art visualization techniques have been successfully engaged in diagnostic medical imaging and Direct Volume Rendering (DVR) sectors and attained maturity ~~by serving faithfully at~~ regular clinical works. However, still a major problem is the lack of information

in

on the uncertainty of the tissue classification, which is addressed in the paper Lundstr¨om et al. [20] by proposing animation methods to convey uncertainty in the rendering. The rendering is animated by sampling the probability domain over time that allows direct user interaction with the classification and it outperforms traditional rendering in terms of assessment accuracy.

Most of the visualization research has ignored the presentation of uncertainty from data because of the inherent difficulty in defining, characterizing and controlling the uncertainty in the visualization process. The paper Pang et al. [21] introduced a wide variety of new uncertainty visualization methods like adding glyphs, adding geometry, modifying attributes, modifying geometry, animation and applied to many applications. The results of the research show that there are a wide variety of possible means to map uncertainty into a scene. The methods presented in the paper represent significant steps toward achieving the goals of uncertainty visualization.

(in 2.2?)

2.3 Textures in Visualization

(introduce topic first)

(ref #?)

Particle Tracing and Line Integral Convolution (LIC) in Rudolf Netzel et al. are parallelly and independently used on every pixel of the texture to reduce the computational cost. On top of that a Gaussian low-pass filter with sparse input noise is used for phase shifting along the streamlines. But there is no indication of how high pass filter and/or variable input noise impacts on the result and performance in terms computation and rendering. Streamline computations were replaced by texture advection that works well for both steady and unsteady flow and provides extremely quick results. But the disadvantage of this setup is coupling exponential filter that cannot handle ~~the~~ trends properly.

(ref #?)

Existing techniques are not capable of accurately aligning and tracking dynamic time-varying data because of the segmentation problem, key feature identification or absence of overlap in consecutive timestep. So, ~~Jesus J~~. Caban et al. introduces a texture-based feature tracking technique capable of tracking multiple features over time by analyzing local textural properties and finding correspondent properties from synthetic and real-world time varying volumetric data. The main limitation specified in the paper is the cumulative error issue that is caused from the "drifting problem" which exists when small errors are introduced to the texture-based multi-dimensional feature vector over time.

The authors ~~Sven~~ Bachthaler et al. [24] have ~~come up with~~ introduced a new technique of utilising the overlay of two different LIC (line integral convolution) textures to combine the visualization of the tangential and orthogonal vector fields. They have applied a weaving of high-frequency spatial textures of different colors and avoided avoid a direct color blending for compositing. Different filter kernels and filter methods are compared and discussed in terms of visualization quality and speed to obtain a consistent and temporally coherent animation. A perception study was ?? ~~can be~~ carried out to measure the discrimination and perceived speed of moving patterns under realistic settings. Also, there is an open question to study the implication of global motion perception and the effectiveness of conveying flow structures since they have focused on low-level local motion perception only. The approach of the study is restricted to 2D manifolds and cannot be extended to higher dimensions.

To avoid color blurring and inconsistencies in popular Line Integral Convolution (LIC) scheme and mitigate the expensive computation or memory cost, eliminating surface parameterization, ~~Jin~~ Huang et al. [25] have introduced a novel image-space surface flow visualization approach that preserves the coherence during user interactions. They have employed a precomputed sequence of triangle textures on coordinates of each vertex to ensure noise textures under different viewpoints remain coherent. Although the approach works fine for most models, popping artifacts can be still visible for some complicated models. For example: when the viewpoint is very far away from or very close to the surface.

~~Andrea~~ Kratz et al. [26] have presented a method for the generation of anisotropic sample distributions in the planar and the two-manifold domains. They also presented interactive rendering of anisotropic Voronoi cells. The have used a special sampling approach to generate sample distributions that cover the underlying domain densely while significant holes and cluttered areas are avoided. ~~Have used~~ They use quadratic textures as GPU data structures, which results in some redundant storage that consumes higher memory than it should be required. The most time-consuming step during initial sampling and relaxation in the two-manifold domain is the back-projection. Influence of adding noise to the cell boundaries are not tested in their experiment but have plan to do in future.

(ref #?)

To improve the use of color in combination with motion where the author ~~D.~~ Weiskopf has distinguished between the detection of patterns in motion (seeing the existence) and the actual perception of motion (recognizing speed and direction). It discussed on how calibration is needed to represent data by the perceived speeds of colored patterns and demonstrated how the guidelines of design of animated graphics and the calibration approach can be used. Although they defined and explained the guidelines, they were not able to make a well-established computational model. Finally, they have mentioned several of possible future works, firstly - user studies could be conducted to test the proposed guidelines for various application scenarios, secondly - evaluate the calibration process in more detail by statistically significant user tests, thirdly - address specific combinations of chromatic motion and further perceptual features like texture.

~~C.G.~~ Healey et al. [28] presents a new method for using texture to visualize multidimensional data elements arranged on an underlying three-dimensional height field. Perceptual texture elements are built by controlling three separate texture dimensions: height, density, and regularity. They conducted a set of controlled experiments to measure the effectiveness of these dimensions, and to identify any visual interference that may occur when all three are displayed simultaneously at the same spatial location. Ad-hoc mapping often introduces visual artifacts that actively interfere with a user's ability to perform their visual analysis tasks. Additionally, it is found that taller, shorter, denser, and sparser pexels can be easily identified, but that certain background texture patterns must be avoided to ensure accurate performance.

2.4 Limitations of related works

As stated in the related works section, a plethora of study ies ~~has~~ have been conducted. been ~~done and still going on in its~~ in these ~~own~~ domains, for example: predicting modeling and augmentation of algorithms, time series analyses and comparisons on different diseases and/or on other temporal data, real time predictions from models, measuring chromatic aberration from image distortion, effect of color and light on display devices, uncertainty visualization and decision making, texture analyses and assessments, perceptual textures to represent multi-dimensional dataset, and etc. In our knowledge predictive uncertainty ~~was never used in representing~~ has not been represented with chromatic aberrations and (presenting three dynamic variables in two-dimensional space with texture is also completely new idea.)

(not sure what you mean by this)

3 Data Preparation

Data preparation is one of the most important factors in the research. In the following sub-sections, we explain ~~about received~~ the raw data and it's processing to achieve the ~~tuned and precise~~ data for the visualization module.

3.1.1 Data Collection and Processing

Data comes ~~as~~ bundled in a csv format in the distributions. [reference the data source here] The following table shows the list of fields/properties of each record where many of them ~~are useless for~~ not relevant to our research. For example: date, location, new_cases, total_cases are some of the useful attributes bolded in the following table.

iso_code	continent	**location**
date	**total_cases**	**new_cases**
new_cases_smoothed	**total_deaths**	**new_deaths**
new_deaths_smoothed	total_cases_per_million	new_cases_per_million
new_cases_smoothed_per_million	population_density	new_deaths_per_million
new_deaths_smoothed_per_million	stringency_index	**population**
new_cases_smoothed_per_million	median_age	aged_65_older
aged_70_older	gdp_per_capita	extreme_poverty
cardiovasc_death_rate	diabetes_prevalence	female_smokers
male_smokers	handwashing_facilities	hospital_beds_per_thousand
life_expectancy	human_development_index	

Table-1: COVID Data property list

3.1.2 Sample Data

iso_code	location	date	total_cases	new_cases	total_deaths	new_deaths	new_tests	total_tests	new_vaccinations	population
USA	United States	31/1/21	26249342	112152	449340	1862	943985	309077233	1545397	331002647
USA	United States	1/2/21	26384317	134975	451420	2080	1032022	310109255	1099103	331002647
USA	United States	2/2/21	26499620	115303	454846	3426	1632097	311741352	558458	331002647
USA	United States	3/2/21	26621311	121691	458728	3882	1869029	313610381	1097394	331002647
USA	United States	4/2/21	26745317	124006	462473	3745	1928145	315538526	1325456	331002647
USA	United States	5/2/21	26879739	134422	466138	3665	1829259	317367785	1615502	331002647
USA	United States	6/2/21	26983915	104176	468821	2683	1332964	318700749	2218752	331002647
USA	United States	7/2/21	27073661	89746	470257	1436	828602	319529351	2172973	331002647
USA	United States	8/2/21	27164099	90438	471877	1620	1042164	320571515	1206680	331002647
USA	United States	9/2/21	27259364	95265	474927	3050	1691635	322263150	788573	331002647
USA	United States	10/2/21	27354614	95250	478231	3304	1638478	323901628	1563780	331002647
USA	United States	11/2/21	27460378	105764	481447	3216	1414964	325316592	1620300	331002647
USA	United States	12/2/21	27560048	99670	484374	2927	1411948	326728540	2020288	331002647
USA	United States	13/2/21	27647267	87219	486570	2196	1101360	327829900	2231326	331002647
USA	United States	14/2/21	27712402	65135	487741	1171	663849	328493749	2242472	331002647
USA	United States	17/2/21	27899318	70139	492854	2397	1470698	332013338	1061463	331002647
USA	United States	18/2/21	27969229	69911	495370	2516	1414506	333427844	1455940	331002647
USA	United States	19/2/21	28048511	79282	497994	2624	1353482	334781326	1847276	331002647
USA	United States	20/2/21	28120207	71696	499829	1835	1058252	335839578	1704457	331002647
USA	United States	21/2/21	28177359	57152	501065	1236	683633	336523211	1801134	331002647
USA	United States	22/2/21	28233518	56159	502384	1319	1038714	337561925	1086840	331002647
USA	United States	23/2/21	28305788	72270	504661	2277	1643226	339205151	854609	331002647
USA	United States	24/2/21	28380537	74749	507843	3182	1687287	340892438	1432864	331002647
USA	United States	25/2/21	28458041	77504	510279	2436	1611102	342503540	1809170	331002647
USA	United States	26/2/21	28535390	77349	512357	2078	1475665	343979205	2179947	331002647
USA	United States	27/2/21	28600016	64626	513878	1521	1074248	345053453	2352116	331002647
USA	United States	28/2/21	28651438	51422	514970	1092	633047	345686500	2429823	331002647
USA	United States	1/3/21	28709536	58098	516487	1517	993543	346680043	1663984	331002647
USA	United States	2/3/21	28766634	57098	518430	1943	1611124	348291167	1731614	331002647
USA	United States	3/3/21	28833825	67191	520911	2481	1736631	350027798	1908873	331002647
USA	United States	4/3/21	28901885	68060	522833	1922	1551765	351579563	2032374	331002647
USA	United States	5/3/21	28968304	66419	524636	1803	1430138	353009701	2435246	331002647
USA	United States	6/3/21	29026558	58254	526146	1510	1066184	354075885	2904229	331002647
USA	United States	7/3/21	29067631	41073	526848	702	608815	354684700	2439427	331002647
USA	United States	8/3/21	29112548	44917	527585	737	976094	355660794	1738102	331002647
USA	United States	9/3/21	29170215	57667	529391	1806	1540048	357200842	1602746	331002647

Table-2: screenshot of sample data (need to briefly explain and refer to Table 2)

3.2 Machine Learning

the

Though we have not done anything ~~new~~ in machine learning domain, but it is ~~fair~~ to briefly introduce the salient components that were used in our research to generate the required data.

novel necessary

3.2.1 Definition

a branch the

Machine learning is ~~an application~~ of artificial intelligence (AI) ~~to play with computer algorithms and develop systems~~ to provide ability of automatic learning through ~~experience and by~~ the use of data. It doesn't need any explicit programming to perform the task since the algorithms are designed ~~with such intellectual mastery~~ to learn new data intuitively.

3.2.2 Predictive/Forecasting Models

A Time series forecasting model comprises a sequence of data points captured, using time as the input parameter. It uses the historical data to develop a numerical metric and predicts for the next ~~certa~~in duration, for instance, data for next ~~couple of~~ weeks using that metric. values

the. few

(incomplete?)

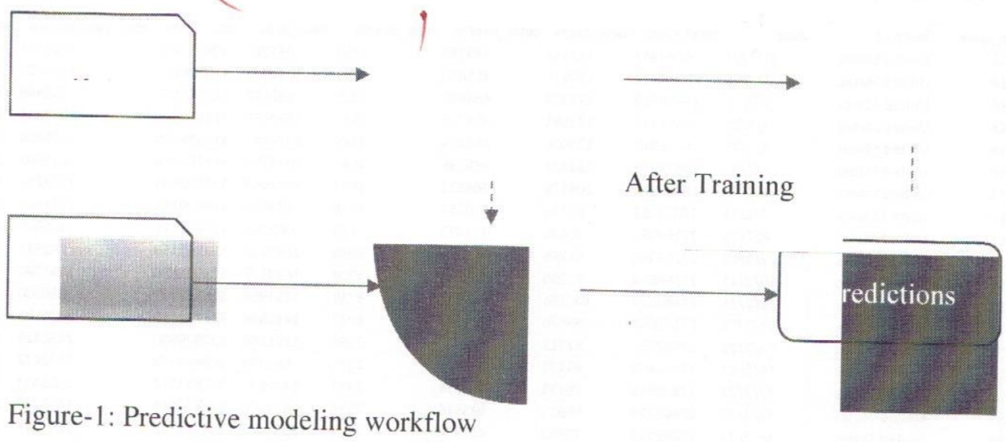


Figure-1: Predictive modeling workflow

3.2.3 Time Series Analysis vs Forecasting

Sometimes ambiguity ~~raises~~ arises between time series analysis with time series forecasting when working with temporal data~~, because in general these two major goals~~. As per ~~Galit~~ Shmueli el al. [31] in time series analysis, a time series is modeled to determine its components in terms of seasonal patterns, trends, and relation to external factors. In contrast, time series forecasting uses the information in a time series (perhaps with additional information) to forecast future values of that series. ~~Since~~ The COVID-19 dataset is ~~preserved and~~ maintained ~~worldly~~ on a global basis. so it is more trustworthy and with ~~this~~ the time series forecasting models it can be considered as suitable for our research to get the predicted results and hence generate our required uncertainty data.

3.2.4 Concerns of Forecasting

Time series forecasting is an important area of machine learning ~~that is often neglected~~. It is important because there are so many prediction problems that involve real life issues involving a time component. In forecasting it is very important to understand the goal of the problem and the nature of the available data. For instance, the volume of data, time horizons (short, medium or long term), frequency of update etc. plays an important role in forecasting. Sometimes time series data requires cleaning, scaling and even transformation, for example: if there are gaps/missing data ~~then need to fill-up with interpolation~~, if there are outliers or corrupt data then those need to ~~clean up.~~ be addressed.

3.3 MLP

A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN). It is a neural network connecting multiple layers in a directed graph, which means that the signal passes through the nodes only in one direction. It can be used for time series forecasting by taking multiple observations at prior time steps, called lag observations, and using them as input features and predicting one or more time steps from those observations. The training dataset is therefore a list of samples, where each sample has some number of observations from days prior to the time being forecasted, and the forecast is the next days in the sequence.

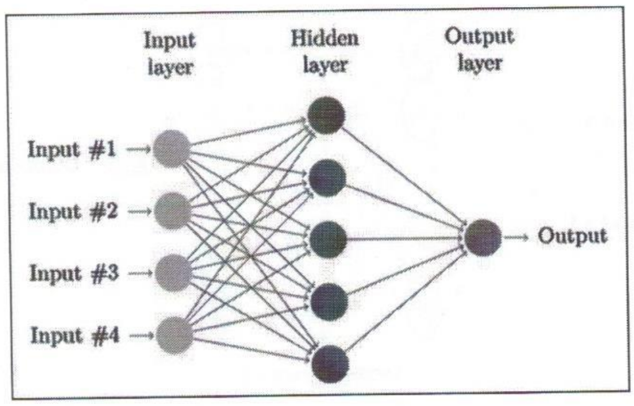


Figure-2: Basic Architecture of MLP network(collected)

(lighter font color :)

We use the rectified linear activation function on the hidden layer as it performs well and a linear activation function on the output layer because we are predicting a continuous value. We use root squared error as loss function and 'adam' optimizer for training the network.

the

3.4 CNN

Convolutional Neural Networks are a type of deep neural network developed for computer vision; for instance, two-dimensional image data, although they can be used for one-dimensional data such as sequences of text and time series forecasting. When operating on one-dimensional data, the CNN reads across a sequence of lag observations and learns to extract features that are relevant for making a prediction.

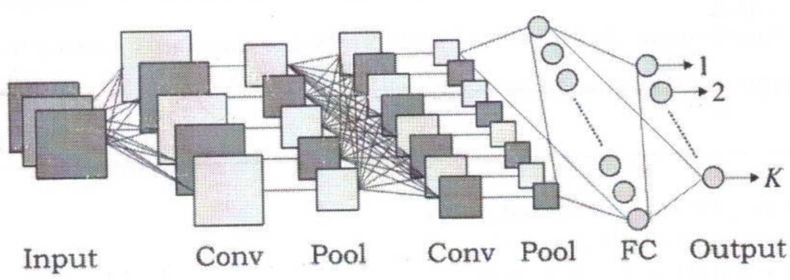


Figure-3: Basic Architecture of CNN network(collected)

We define a CNN with two convolutional layers, one max-pooling layer, one flatten layer, and a dense layer from the input sequences. They have a configurable number of filters, kernel size, pool size and rectified linear activation function is used as loss function. The number of filters determines the number of parallel fields on which the weighted inputs are read and projected. A max pooling layer is used after convolutional layers to distill the weighted input features into those that are most salient, reducing the input size by 1/2. The pooled inputs are flattened to generate a long vector before being interpreted and used to make the prediction.

3.5 LSTM

The LSTM neural network is a member of RNN and it can be used for univariate time series forecasting. It uses an output of the network from a prior step as an input in attempt to automatically learn across sequence data. The LSTM has an internal memory allowing it to accumulate internal state as it reads across the steps of a given input sequence.

(make sure you directly refer to all figures)

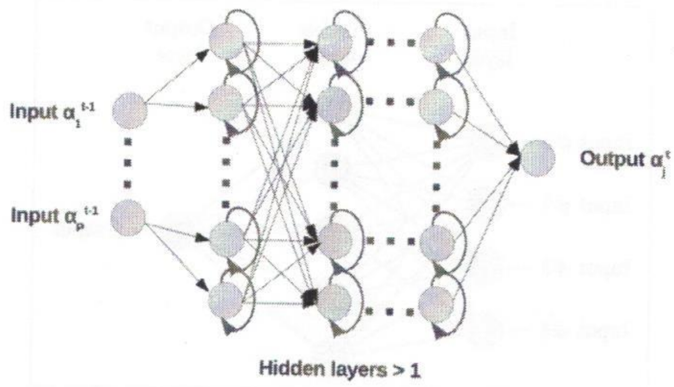


Figure-4: Basic Architecture of LSTM network(collected)

For this model we define a LSTM layer from inputs and subsequently two dense layers. Like other models, rectified linear activation function is used in LSTM layer and in one of dense layer. The LSTM does perform better if the data is stationary and ~~being our~~ data ~~static from the excel files~~, we can expect about achieving better prediction outcome. given that our data is static.

3.7 Uncertainty Tables

Uncertainties are calculated from the actual and predicted values. The current calculation is ~~very~~ straight forward where we find the difference between actual and predicted values. Then we find the maximum difference to set out the domain of the difference. Finally~~,~~ we divide each difference by the maximum difference and multiply by a scaling factor to keep the maximum result in single digits. ~~Here is given the pseudo code:~~ Algorithm 1 provides the pseudocode.

```
diff = actual_value - predicted_value
max_diff = find_max(list_of_diff)
uncertainty = diff * 7 / max_diff
```

algorithm. 1

3.7.1 Top 10 uncertainty countries

Country	Actual	Predicted	Uncertainty
France	312,023	377,734	7.00

(need to briefly discuss these next 2 tables).

Argentina	518,305	464,893	5.69
Brazil	1,303,729	1,253,563	5.34
Turkey	286,366	325,556	4.17
Germany	219,476	257,156	4.01
United States	645,536	681,911	3.87
Italy	141475	168,383	2.87
Iran	276,931	301,707	2.64
Nepal	169823	194,396	2.62
India	6,469,614	6,493,921	2.59

Table-3: Top uncertainty countries in the ordered list

3.7.2 Lowest 10 uncertainty countries

Country	Actual	Predicted	Uncertainty
Myanmar	392	385	0.00
South Korea	12,198	12,208	0.00
Czechia	23,274	23,304	0.00
China	278	234	0.00
Ireland	4,580	4,532	0.01
Ghana	937	857	0.01
Sri Lanka	50,525	50,439	0.01
Nigeria	820	912	0.01
Libya	4,976	5,099	0.01
Montenegro	1,530	1,657	0.01

Table-3: Bottom uncertainty countries in the ordered list the

4 Visualization

As stated earlier we generate required data from the standalone python program with the predictive models and save it in JSON format in file system. Then for front-end we pull it through API call and feed in client-side scripts for drawing charts since chart presentation is the key part of our research.

an.

4.1 Web Interface

To visualize different charts, we have developed a ~~very straight forward~~ web-interface with several html input controls in the top toolbar and all charts are presented in the main container placed just below the toolbar.

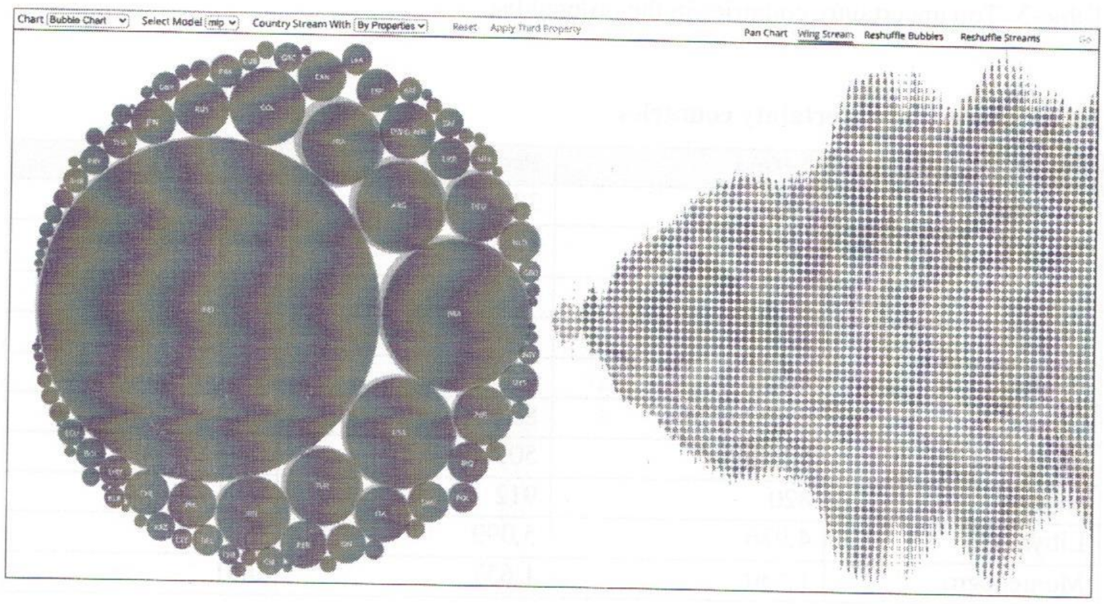


Figure-5: Initial Web Interface (Left - Bubble chart, right – Streamgraph)

In the following section, we briefly explain the basic functionalities of the input fields in toolbar.

Chart dropdown: List of chart names, on selection it will automatically draw the corresponding chart in the main container.

Select Model: Names of the predictive models for which we have generated data for charting.

Country stream with: We have an option to draw stream graph by using the actual property data or with the predicted data.

(explain).

Return.

Reset: ~~Get back~~ to the initial state of the settings.

Pan Chart: Since the bubble chart and stream graph are drawn side by and they work interactively like filtering streamgraph with the selection from bubble chart, so sometimes it is necessary to zoom-in/out of the charts and consequently panning the charts in its own space is also advantageous. the

Wing Stream: changes the drawing mode to interact with mouse events. We explain later how wing stream works.

the

Bubbles Select: ~~Allows to~~ Select ~~country~~ country from the chart and redraw with the selected ones only.

Bubbles Remove: ~~Allows~~ to Filter-out countries from the bubble chart. In this mode the selected countries omitted from the chart. After omitting countries on press 'Go' button it redraws with the other countries.

are

Reshuffle Streams: Allows to draw main streamgraph with the selected countries of interest from bubble chart.

Apply Third Property: Toggles to impose third property in right side stream graph.

4.2 Filtering

We use data for top 100 countries based on the ~~number of~~ total infection rate. As we see from the above figure, it is difficult to read the label of the country and difficult to identify the extent of aberration for the smaller circles having lower uncertainties. That's why we implement filtering options with different perspectives. In the ~~below~~ section we briefly explain them.

a below

4.2.1. With Selective Countries

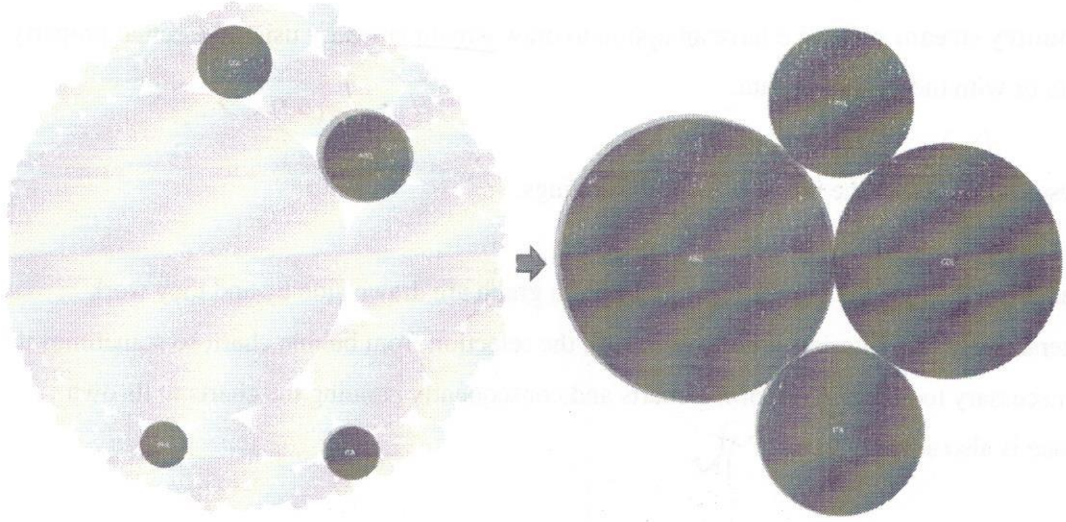


Figure-6: With selected countries of interest

In this mode, it allows users to select the countries of interest on first click and toggles on the next one. So, when all preferred countries are selected the 'Go' button ~~allows to~~ redraws the bubbles side by side with comparatively bigger sizes.

4.2.2 By removing countries

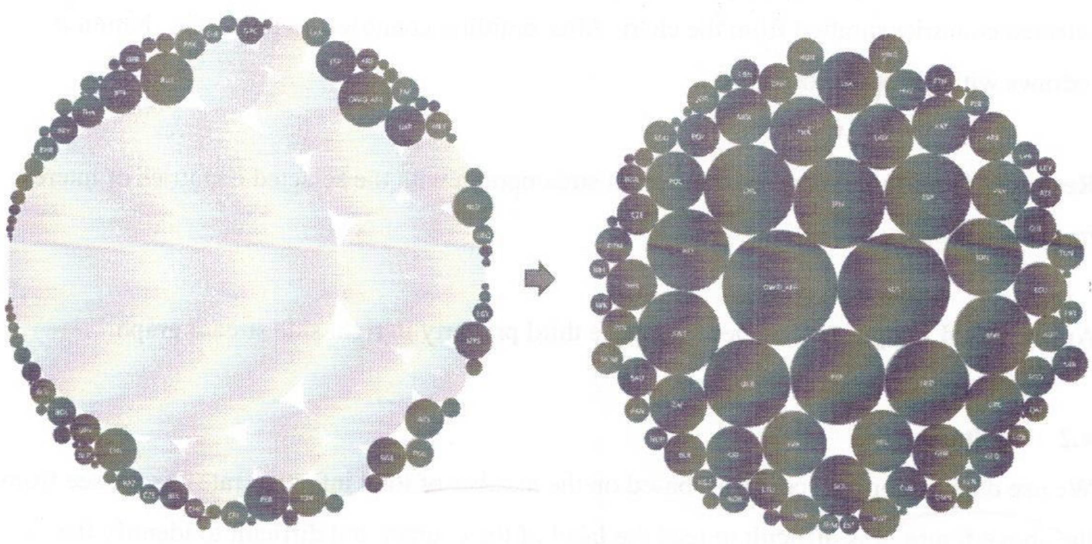


Figure-7: Removal of countries of interest

This is the opposite of the earlier one where the user can ~~intuitively~~ select the countries to remove from the chart, for instance, removing bigger ones help to find the status of the countries having a smaller size.

4.4 Main Streamgraph

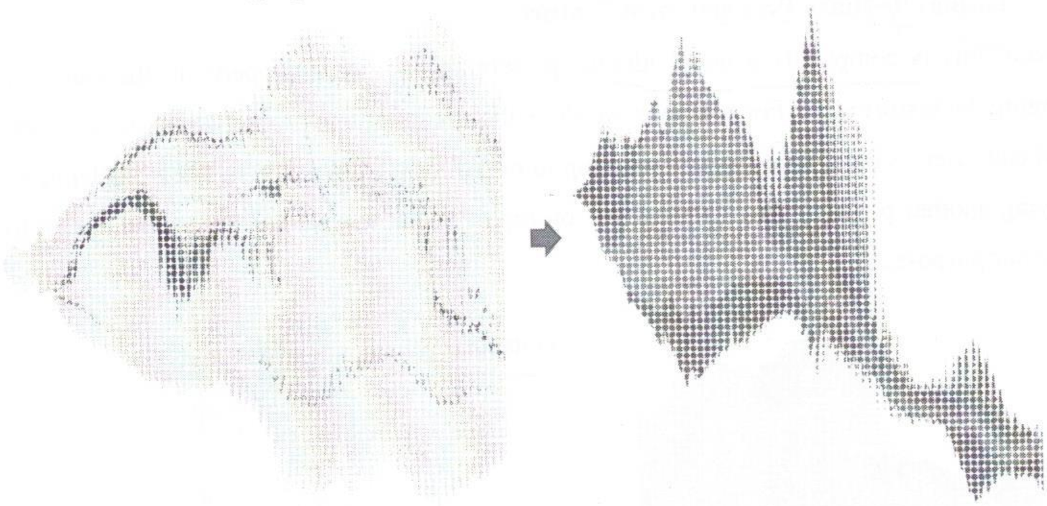


Figure-8: Reshuffling main streamgraph

In this mode, a user can choose the countries from the bubble chart. ~~on~~ On select the countries, the corresponding ones will be highlighted in the streamgraph to represent the selection and the rest of the country-streams will be blurred in the same. To confirm the execution streamgraph redraws with the selected countries as shown in the above figure.

??

(I don't think this is completely novel. It is like using glyphs on "small multiples").

4.5 Display Tertiary Property with Texture

As said, this is completely a novel idea of presenting additional property in the same by changing the texture style. For example: we show the stream of one property like 'New Cases' of all countries vs date. Since we are working in two-dimensional space so it's a challenge of showing another property like 'Death Rate' on top of it. So, texture can be a good option to serve our purpose.

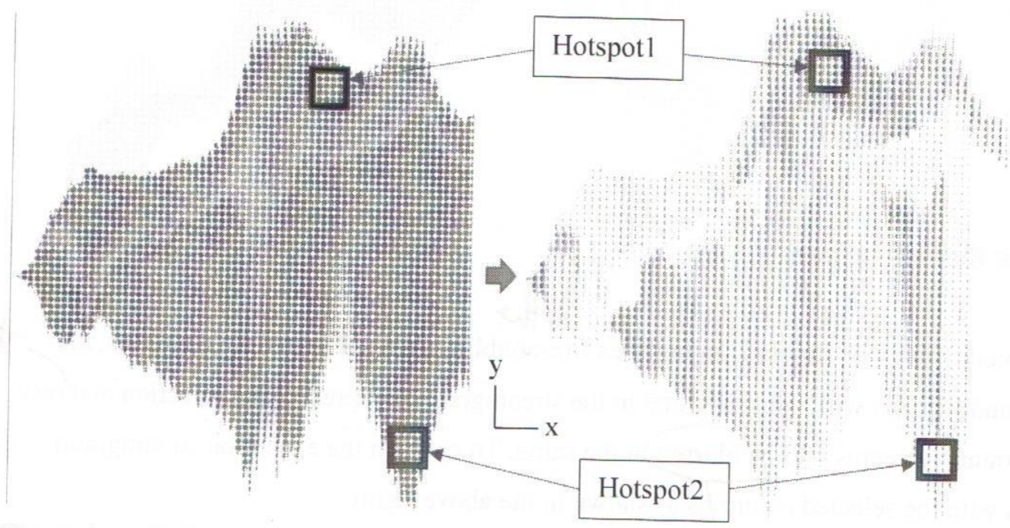


Figure-9(A): Realization of third property effect

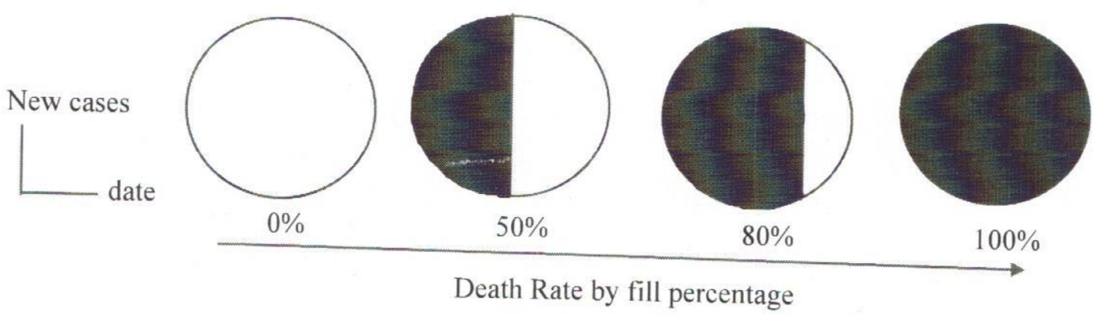


Figure-9(B): Three properties representation

Now the question is how does it work~~s~~? To answer this question, we have marked two hotspots in the above Figure-9(A) where we see in hotspot1 the smaller circles are much closer, but for hotspot2 this is changing a lot. This is happening because for hotspot1 the death rate is much higher than the hotspot2.

In Figure-9(B) it gives indication of how three properties work in the 2D space where 'Date' and 'New Cases' represents traditional (x, y) axes respectively and 'Death Rate' is represented by the percentage of fill of the circle.

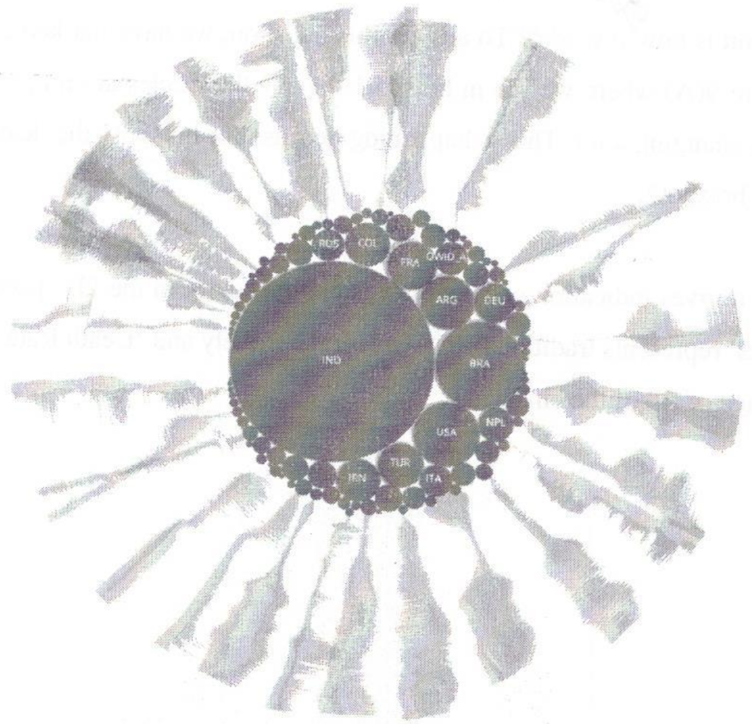
benefit.

one.

4.6 Star Fish Inspiration

In this approach, user can draw multiple streams graphs by dynamically calculating the position of the country cell and its corresponding start point in the cell center and angle to place it without (or possibly minimum) overlapping the other countries' streams. If we call each individual stream as a wing, then the ~~beauty~~ of this chart is it allows to draw ~~so~~ many charts in ~~smaller space~~ in compact ~~but precise~~ way. Another interesting feature is that we accommodate multiple properties in each stream, for example: mlp, cnn and lstm predictions are used in bottom chart whereas total_cases, new_cases, new_deaths, icu_patients, hosp_patients, new_tests are used for top chart.

(line spacing changed)



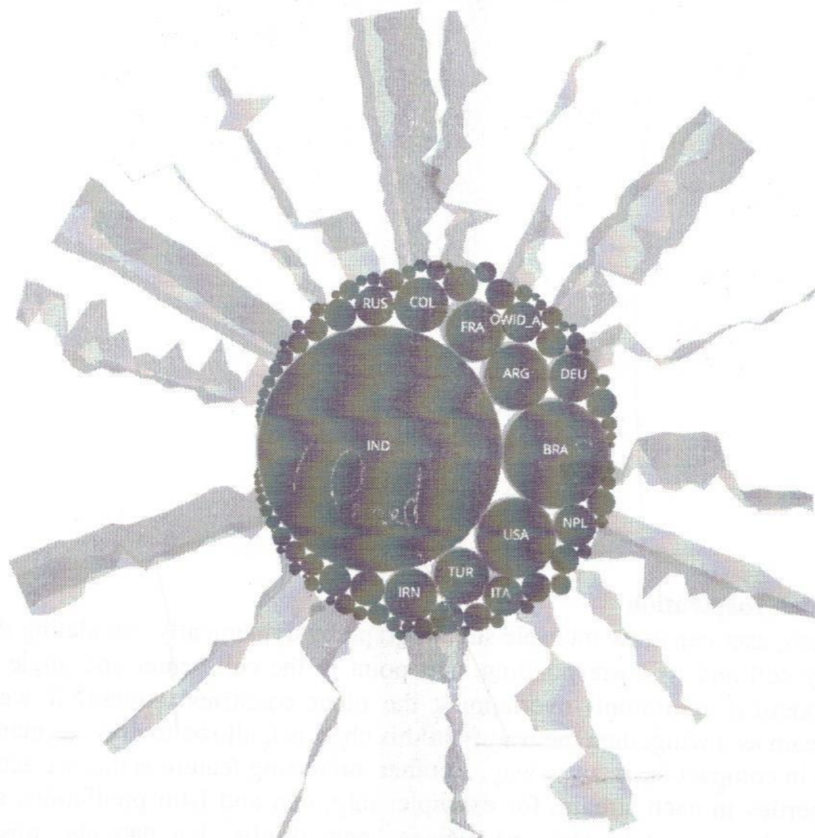


Figure-10: Approach to represent multi-country- actual counts(top), predicted counts(bottom)

(are they not horizontal in figure 11?)

one

4.7 Parallel Coordinates Chart

Parallel plot or parallel coordinates plot allows to compare the feature of several individual observations (series) on a set of numeric variables. Each ~~vertical bar~~ represents a variable and often has its own scale and the units can even be different, that is the strength of this special kind of plots. The main advantage offered by parallel coordinate is the representation of high dimensional data as a 2-dimensional visualization. ~~As~~ data is represented in the form of a poly line, and. it becomes ~~easy~~ possible. to perceive ~~the~~ trend shown by data entries from the visualization.

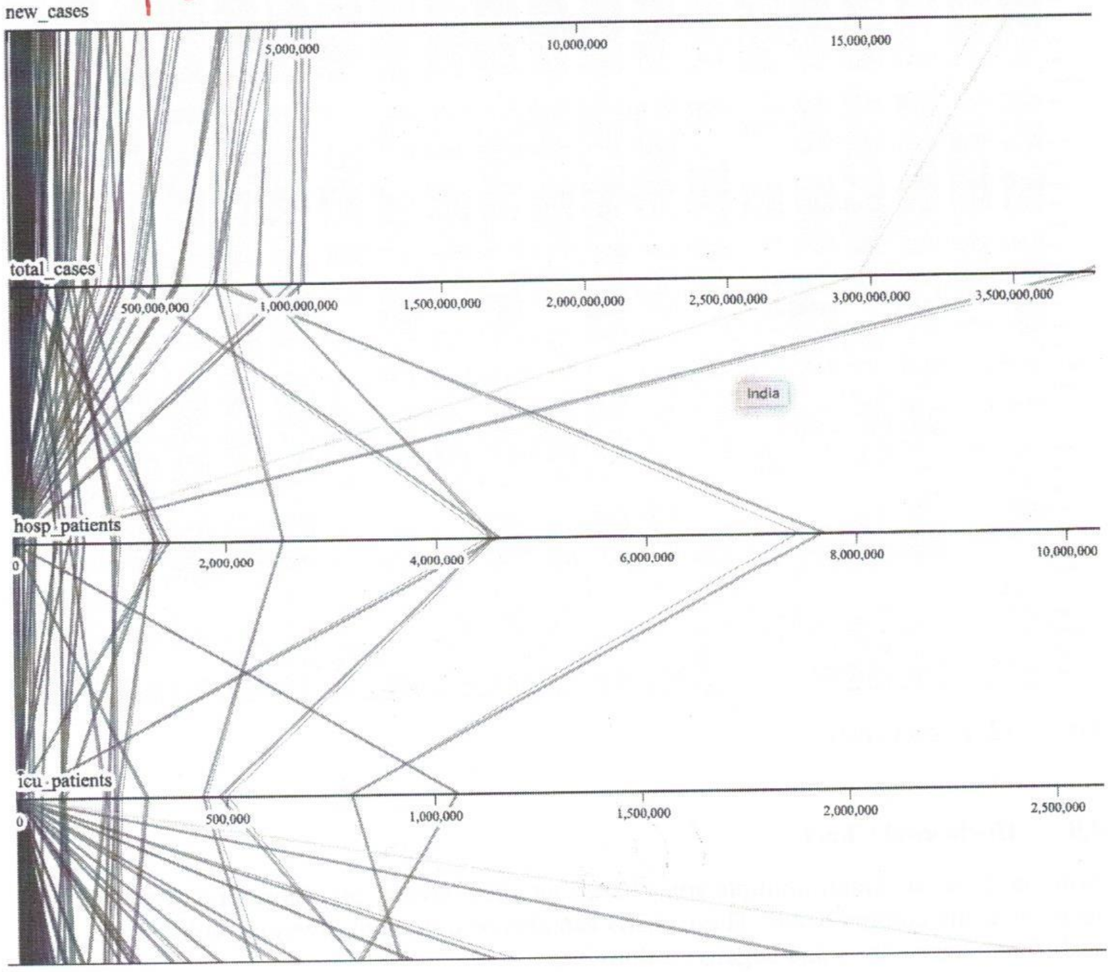


Figure-11: Parallel coordinates chart

showing.

you can

This plot is helpful in our presentation because we have several variables together to visualize one after another ~~and nicely seeing~~ the relationships between them. For example, ~~if you had to~~ compare number of total cases(total_cases) with hospitalized patients (hosp_patients) ~~then it gives a clear insight for the countries since it shows a tooltip with~~ country name. Also, it is ~~interestingly showing~~ the predicted flow (thinner line) along with actual counts (thicker line). The ~~bad side~~ of this chart is ~~too much overlapping~~.

limitation. able to show. facilitated by a tooltip showing the.

frequent occlusions.

preserve trends.

was.

4.8 Impact Chart

This chart helps to indicate daily uncertainty ~~presentation~~ for every country as a cell. In this way user can ~~easily get indication or trend~~ for certain day or a set of consecutive days. In other words, the *chart* provides a useful platform that helps you decide which uncertainty ~~need~~ your attention. So, if ~~we consider~~ this tool ~~is~~ used by WHO then the administrator ~~can easily be concerned about~~ which countries are vulnerable ~~for~~ tomorrow or the day after tomorrow.

a

the. requires.

Could consider,

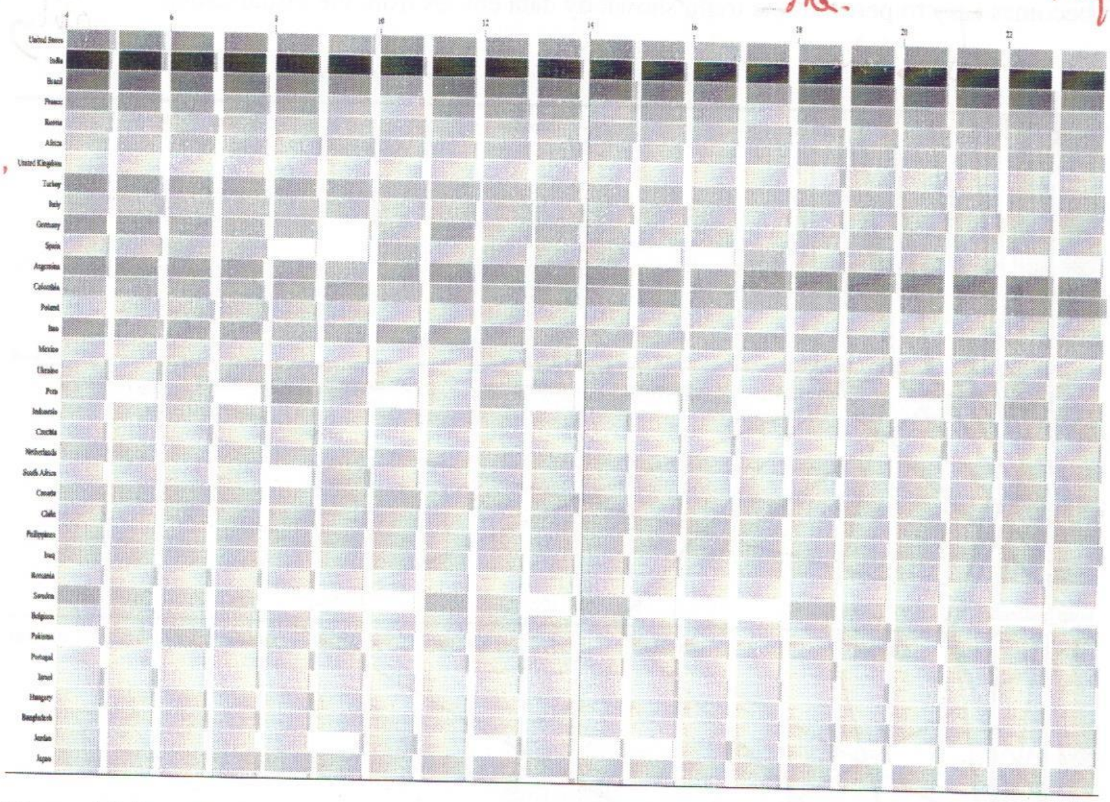


Figure-12: Impact chart

4.9 Horizontal Chart

Horizon charts are small-multiple area charts that allow greater precision for a given vertical space by using colored bands. These charts can also be used with diverging color scales to differentiate positive and negative values.

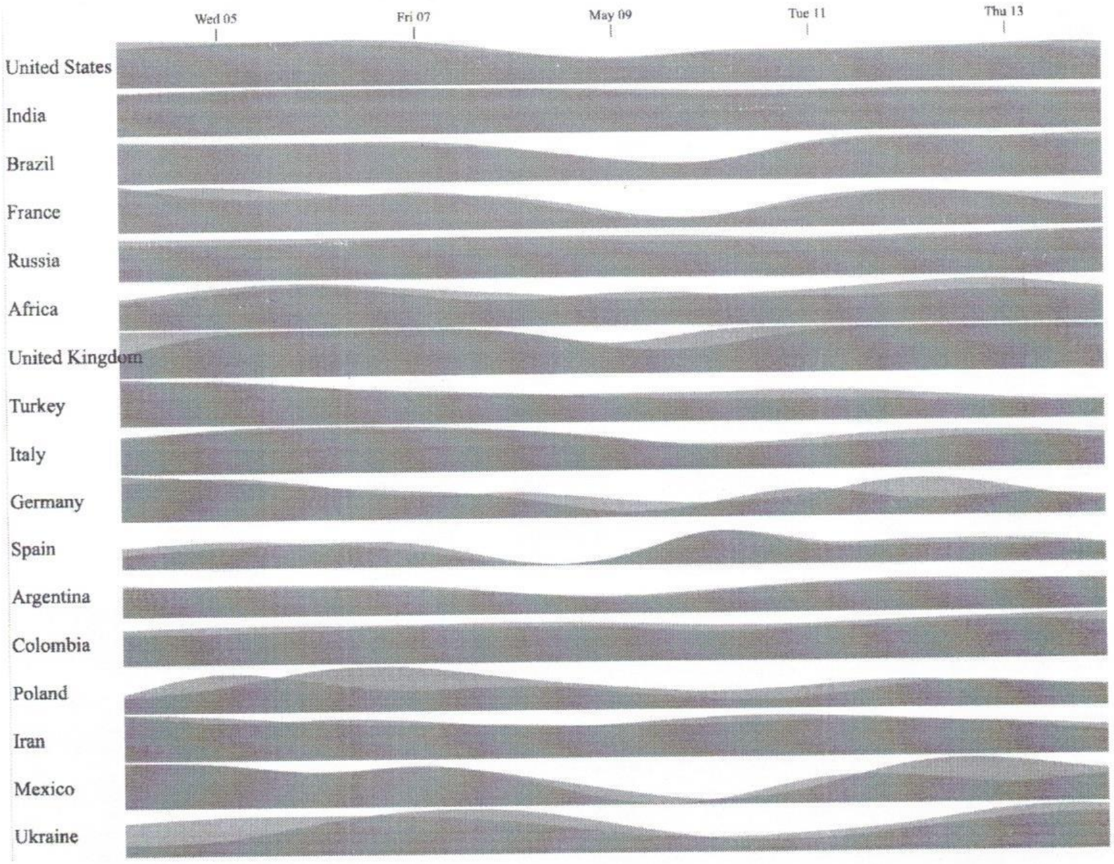


Figure-13: Horizontal chart

4.10 Usage Chart [Work in progress]

This chart is more much like impact chart because their construction style is mostly similar to each other, though the axes are used in reverse order. Need to brainstorm on it about how this chart can be used in our research in meaningful way.

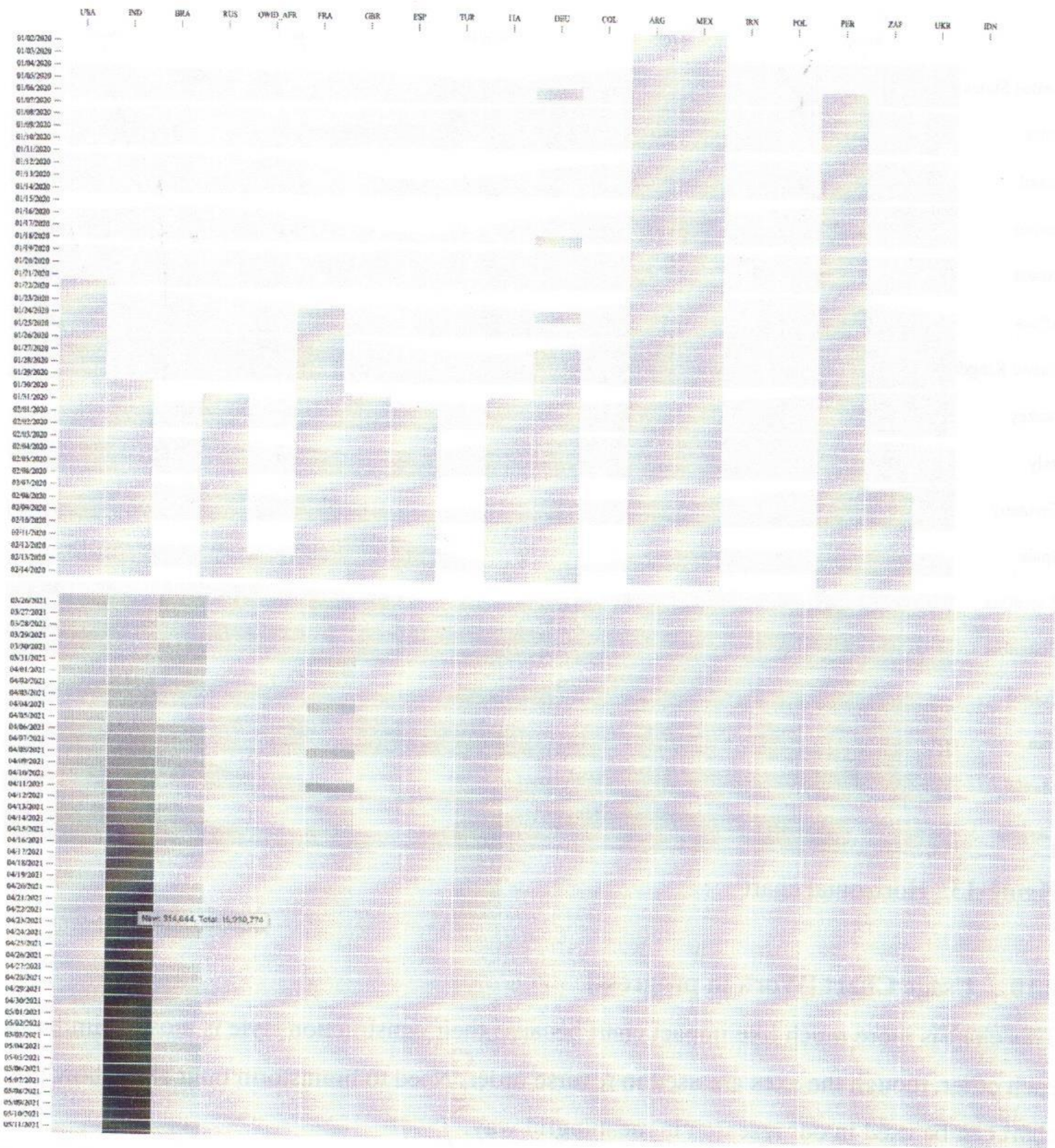


Figure-14: Charts of Daily counts (yet to impose uncertainty)

4.11 Chromatic Aberration

Chromatic aberration is term in optics where it refers to distortion and spherochromatism that happens due to the failure of a lens to focus all colors to the same point. In other words, chromatic aberration is a color distortion that creates an outline of unwanted color along the edges of objects in a photograph.

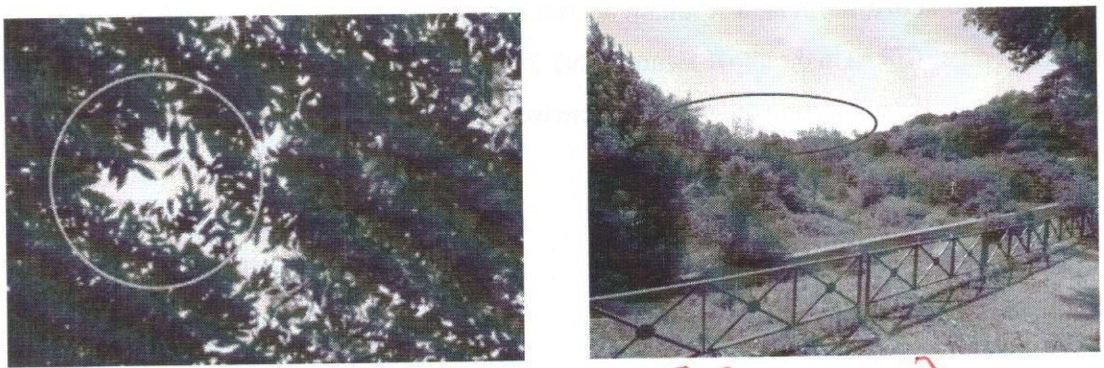


Figure - Example of chromatic aberration (collected) [references?]

4.11.1 Implementation Mechanism

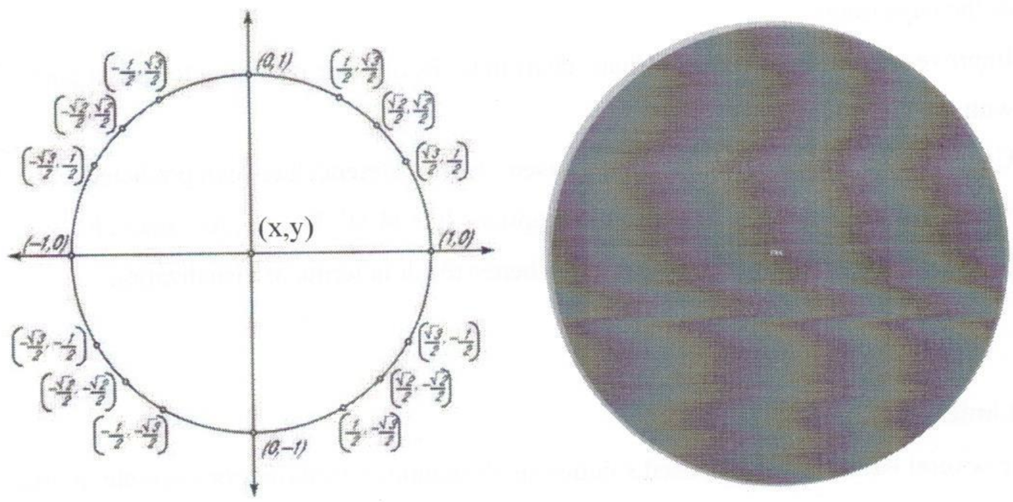


Figure- Geometic concept(left), Implementation with a circle(right)

To draw a circle representing aberration, we draw 3 circles internally, let's call them 3 chromatic circles. The following technique is applied on each of the chromatic circles -

- Once for color (r, 0, 0) with a shifted location of (x, y + r)
- Once for color (0, g, 0) with a shifted location of $(x + r * \frac{+\sqrt{3}}{2}, y + r * \frac{-1}{2})$
- Once for color (0, 0, b) with a shifted location of $(x + r * \frac{-\sqrt{3}}{2}, y + r * \frac{+1}{2})$

Where 'r' is the radial offset of each of the 3 circles from the center of the original circle located at (x, y).

By using the above formula, a resultant aberration is presented with the uncertainty for the country France (FRA) in the above figure(right). Though in real picture we found the aberration is shown as a kind of blurring or fading but here we present one with equal intensity highlighted color though concept remains the same.

the.

5 Discussion

TBA in the final paper.

5.1. Known issues to work out

The following issues are yet pending to implement

- Finish 'Usage Chart' by finding a way to inject aberration information on each slice of the day/country.
- Improve aberration of 'Impact Chart' chart in terms of color matching for count part with the uncertainty part.
- Uncertainties are currently calculated based on the difference between predicted values and actual values. Try with more options like MAE, RMSE, Accuracy, F1 Score, etc. and check which option gives better result in terms of visualization.

5.2. Limitation

There are several issues in our proposed solution of chromatic aberration. For example: in real aberration in picture the blurring happens very slowly from inner edge to outer edge but in our case, it just gives us a range of uncertainty for the prediction, so the whole edges are with bright color.

In texture presentation we have generated texture patterns with linear gradient so the color intensity in left of the bullet point higher than the right side. So, it is an open problem to improve and ensure the intensity of the color for the visible part of the circular textures.

(not sure what this means).

However, our simplified implementation allows us to reduce the abberation to a single parameter, which facilitates chromatic abberation tuning with regards to the amount of represented uncertainty.

5.3. Future Work

TBA in the final paper.

Your references need to have: authors, title, Journal or conference, page #'s, volume (issue #) (for journals), year, location (for conferences)

References:

[1] Song, Xin; Xiao, Jun PhD; Deng, Jiang PhD; Kang, et al. Time series analysis of influenza incidence in Chinese provinces from 2004 to 2011. doi: 10.1097/MD.0000000000003929, Received March 1, 2016, Received in revised form May 10, 2016, Accepted May 20, 2016, Medicine: June 2016 - Volume 95 - Issue 26 - p e3929

[2] Muhammad Ali, Dost Muhammad Khan, et al. Forecasting COVID-19 in Pakistan, received: August 17, 2020; Accepted: November 10, 2020; Published: November 30, 2020, https://doi.org/10.1371/journal.pone.0242762

[3] COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population. Appl. Sci. 2020, 10(11), 3880; https://doi.org/10.3390/app10113880 Received: 5 May 2020 / Revised: 23 May 2020 / Accepted: 29 May 2020 / Published: 3 June 2020

[4] Christophorus Beneditto, Aditya Satrio et al. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET, https://doi.org/10.1016/j.procs.2021.01.036

[5] Leo J, Luhanga E, Michael K. Machine Learning Model for Imbalanced Cholera Dataset in Tanzania. The Scientific World Journal. 2019 Jul; 2019: p. 1–12.

[6] Emrah Gecili, Assem Ziady, Rhonda D. Szczesniak. Forecasting COVID-19 confirmed cases, deaths and recoveries: Revisiting established time series modeling through novel applications for the USA and Italy. Received: June 30, 2020; Accepted: December 5, 2020; Published: January 7, 2021. https://doi.org/10.1371/journal.pone.0244173.

[7] Sathler C, Luciano J. Predictive modeling of dengue fever epidemics: A Neural Network Approach. 2017. Data Science for Drug Discovery, Health and Translational Medicine. December 10, 2017. I590.

[8] Miranda GHB, Baetens JM, Bossuyt N, Bruno OM, Baets BD. Real-time prediction of influenza outbreaks in Belgium. Epidemics. 2019 Sep; 28: p. 100341. https://doi.org/10.1016/j.procs.2021.01.036

[9] Muktevi Srivenkatesh, Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver Disease. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-2, December 2019. DOI: 10.35940/ijitee.L3619.129219

[10] K. Koh, B. Kim & J. Seo. 2014. Effect of lateral chromatic aberration for chart reading in information visualization on display devices. Advanced Visual Interfaces. Como, Italy, 289-292.

[11] H. S. Yoo. 2007. Color illusions on liquid crystal displays and design guidelines for information visualization. Master of Science, Virginia Tech.

(The http reference is not usually required unless you are referencing a webpage or web resource itself).

[12] T. Boult & W. Wolberg. 1992. Correcting chromatic aberrations using image warping. CVPR, Champaign, IL, 684–87.

[13] M. K. Johnson & H. Farid. 2006. Exposing digital forgeries through chromatic aberration. Multimedia and security, Geneva, Switzerland, 48-55.

[14] M. Hullin, E. Eisemann H.P. Seidel & S. Lee. 2011. Physically-based real-time lens flare rendering. ACM SIGGRAPH, Vancouver, 108:1–108:9.

[15] S. Lee, E. Eisemann & H.P. Seidel. 2010. Real-time lens blur effects and focus control. ACM SIGGRAPH, Los Angeles, 1-7.

[16] Bonneau et al. Overview and State-of-the-Art of Uncertainty Visualization, The University of Grenoble, France e-mail: Georges-Pierre.Bonneau@ujf-grenoble. fr. DOI:10.1007/978-1-4471-6497-5_1. ISBN: 978-1-4471-6496-8

[17] Simon Barthelme, Pascal Mamassian. Evaluation of Objective Uncertainty in the Visual System. Received June 8, 2009; Accepted August 12, 2009; Published September 11, 2009. doi:10.1371/journal.pcbi.1000504

[18] Henning Griethe et al. The Visualization of Uncertain Data: Methods and Problems. Computer Graphics, 18051 Rostock, Germany. January 2006, 2,988 reads, 36 publications, 262 citations. https://www.researchgate.net/publication/216756553.

[19] Deitrick, S., Edsall, R.: The influence of uncertainty visualization on decision making: An empirical evaluation. In: Progress in Spatial Data Handling, pp. 719–738. Springer Berlin Heidelberg (2006).

[20] Lundstr¨om, C., Ljung, P., Persson, A., Ynnerman, A.: Uncertainty visualization in medical volume rendering using probabilistic animation. IEEE Transactions on Visualization and Computer Graphics 13(6), 1648–1655 (2007).

[21] Pang, A., Wittenbrink, C., Lodha., S.: Approaches to uncertainty visualization. The Visual Computer 13(8), 370–390 (1997).

[22] Rudolf Netzel and Daniel Weiskopf, Texture Based Flow VisualizaTion. November 2013, Computing in Science and Engineering 15(6):96-102, DOI:10.1109/MCSE.2013.131

[23] Jesus J. Caban, Alark Joshi, and Penny Rheingans. Texture-based feature tracking for effective time-varying data visualization, IEEE Transactions on Visualization and Computer Graphics (Volume: 13, Issue: 6, Nov.-Dec. 2007). Page(s): 1472 – 1479. DOI: 10.1109/TVCG.2007.70599.

[24] Sven Bachthaler, Daniel Weiskopf. Animation of Orthogonal Texture Patterns for Vector Field Visualization. IEEE Transactions on Visualization and Computer Graphics (Volume: 14, Issue: 4, July-Aug. 2008), Page(s): 741 – 755. DOI: 10.1109/TVCG.2008.36

[25] Jin Huang, Zherong Pan, Guoning Chen, Wei Chen, Hujun Bao. Image-Space Texture-Based Output-Coherent Surface Flow Visualization. IEEE Transactions on Visualization and Computer Graphics (Volume: 19, Issue: 9, Sept. 2013). Page(s): 1476 – 1487. DOI: 10.1109/TVCG.2013.62

[26] Andrea Kratz, Daniel Baum, and Ingrid Hotz. Anisotropic Sampling of Planar and Two-Manifold Domains for Texture Generation and Glyph Distribution. IEEE Transactions on Visualization and Computer Graphics (Volume: 19, Issue: 11, Nov. 2013). Page(s): 1782 – 1794. DOI: 10.1109/TVCG.2013.83

[27] D. Weiskopf. On the role of color in the perception of motion in animated visualizations. Conference: 10-15 Oct. 2004, Austin, TX, USA. IEEE Visualization 2004. ISBN: 0-7803-8788-0. DOI: 10.1109/VISUAL.2004.73

[28] C.G. Healey; J.T. Enns. Building perceptual textures to visualize multidimensional datasets. 18-23 Oct. Research Triangle Park, NC, USA. 1998. Proceedings Visualization '98 (Cat. No.98CB36276). ISBN: 0-8186-9176-X. DOI: 10.1109/VISUAL.1998.745292

[29] R.P. Botchen; D. Weiskopf; T. Ertl. Texture-based visualization of uncertainty in flow fields. VIS 05. IEEE Visualization, Minneapolis, MN, USA. 23-28 Oct. 2005. ISBN: 0-7803-9462-3. DOI: 10.1109/VISUAL.2005.1532853

[30] Aasim Kamal · Parashar Dhakal, et al. Recent advances and challenges in uncertainty visualization: a survey. May 2021, Journal of Visualization 24(5):1-30. DOI:10.1007/s12650-021-00755-1

[31] Galit Shmueli, Kenneth C. Lichtendahl Jr. Practical Time Series Forecasting with R: A Hands-On Guide [2nd Edition] (Practical Analytics) Paperback – July 19, 2016. Page 18-19. ISBN-13 978-0997847918