**Visualizing Uncertainty with Chromatic Aberration**

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 an increasing array of research are being conducted by researchers in the field of uncertainty visualization, determine the impact of representation in users’ perception and evaluation of its effectiveness in decision making. Uncertainties are always an integral part of any data and so by nature model predictions also contains significant amount of uncertain information. Since COVID-19 is a respiratory infectious disease caused by novel coronavirus and due to its unprecedented challenges over time and frequent changes of strains, scientists and researchers are investigating the perceived data to discover the cause, find the patterns in different demographic areas and recently looking into the effect of vaccination against different variants. So, in this study, we come up with a novel idea for a visualization to present predictive model uncertainties in terms of Chromatic Aberration (CA). We utilized some common and existing machine learning models to obtain the predictive results and calculate the corresponding model uncertainties for the most impacted countries with respect to number of new-cases, new-deaths, and new-vaccination for different countries. Finally, we visualize the data itself and its associated uncertainties with an artificially and spatially separable channels of red, green, and blue color components. The chromatic aberration representation is visualized in an interactive fashion and experimented with its static and animated version along with some other alternative representations for example: blurriness and noise. The representation is setup in such a way that the people can understand the conveyed uncertainty significantly.

**Chapter 1**

**1 Introduction**

Uncertainty visualization is one of the major areas of research now a days because many people intentionally try to avoid their existence being a complex topic. So how effectively uncertain information can be visualized in front of audience becoming an increasingly important issue in this arena. There are various studies conducted for uncertainty representations for example: textual representation such as caption or tooltip, graphical representation such as glyphs [21, 54], color-based rectangles such VSUP [35], bivariate choropleth maps [43], texture patterns [29] and so on. As far we know, most of the research are conducted by using the inherent uncertainties of the data itself and no uncertainty representation ever used the Chromatic Aberration, we found a gap to introduce the machine learning model uncertainties as chromatic aberration in visual interfaces. To accomplish the purpose, we have categorized the scope of the research in several core components: Firstly, collect real-life data from some authentic source (used WHO delivered COVID data). Secondly, generate uncertainty information in the data (accomplished by feeding collected data to machine learning models and calculated from the forecasting [6] of predictive machine learning algorithms). Thirdly, visualize the uncertainty and data in display devices such as web-interface in human understandable manner. Fourthly, conduct a human-computer interaction experiment to evaluate the effectiveness of the experiment. Fifthly, explain experimental results and numerical analysis and finally, draw a conclusion.

**1.1 Background and Motivation**   
The outbreak of coronavirus COVID-19 first emerged in China in December 2019 and the expansion has occurred all over the world and declared as an international public health crisis by WHO. Since then, the world is quite 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 mortality rate wasn’t in control significantly until immunization started and over 318 million people have been infected and 5.5 million have died in the world over. The infection and death rate are oscillating in different countries due to various reasons such as public insincerity and lack of consciousness about the disease. Moreover, the strain of the virus is changing frequently in different geographical places with more power and variations and a few of the variants like the British variant, the South Africa variant and the Indian variant and finally Omicron became the prime concern for the world community, because sufficient research has not been conducted on them yet. Though a great deal of research is being conducted and wide range of immunization process started holistically but still nobody knows when the world will get rid of this cruel pandemic and return to full 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 diseases such influenza [1], Cholera [5], along with many other popular machine learning algorithms [2, 3, 9]. To our knowledge, most of studies are administered for specific countries. The pandemic started very abruptly and so during the first year, it was difficult to develop any efficient system to forecast anything due to the lack of required data. But after passing more than one year, we have sufficient data to explore, analyze and forecast 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 it and help governments’ regarding contingent policymaking to properly address the consequences of the pandemic and encourage people to be cautious and follow the rules and health guidelines to achieve the maximum benefit by saving valued lives. That’s why 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 a certain period of time and employ those uncertainties in visual representation as chromatic aberration that can help the community administrators for better planning, providing insights to minimize its impact. Last but not the least the visualization of such uncertainties through real life appealing charts helps the users to compare and perceive side by side.

**1.2 Basic** **Concepts**To dig out further detail, it is very important to introduce thetop-level terms used in the dissertation so that the reader can get better insight about the work. The following section discusses the basic concepts of each of them:

**1.2.1 Machine Learning (predictive models)**

Machine learning is an approach of artificial intelligence (AI) to provide ability of automatic learning through the uses of data. It doesn’t need any explicit programming to perform the task since the algorithms are designed to learn new data intuitively.

**1.2.2 Streamgraph**Stream Graphs are ideal for displaying high-volume datasets, to discover shapes, trends, and patterns over time across a wide range of numerical groups side by side. For example, seasonal peaks in the stream shape can suggest a periodic pattern. They work even better when there is an interactive component involved that enables the following of each “flow” or allow filtering the view in some way. In our case, we have data for more than 100 countries, so it greatly helps to get a perception of the country-wise flow for number of cases, death rate, vaccination rate etc.

**1.2.3 D3.js**

D3 is a JavaScript library for manipulating web documents based on data. It creates visualizations by binding the data and graphical elements to the Document Object Model and eventually produce dynamic and interactive data visualization in web browsers with the help of standard web technologies like HTML, CSS, SVG.

**1.2.4 Uncertainty**

Uncertainty is a natural part of life. Everyone feels it in their daily life from big things to small. Being uncertain does not mean someone are lost or a failure. It just means you're alive with higher or lower probability. It is defined lack of sureness or certainty in data. The lack of certainty, a state of limited knowledge where it is impossible to exactly describe the existing state or a future outcome.

 A picture containing nature, smoke, spring, wave

Description automatically generated

Figure: Uncertainty in [real life](https://7summitpathways.com/blog/how-to-accept-uncertainty/) Uncertainty in [data](https://simplicable.com/new/data-uncertainty)

Data uncertainty is the degree to which it is inaccurate, imprecise, or unreliable. It can come from source (e.g.: data provider), data lineage (e.g.: from calculation), noise (e.g.: inaccurate post in social media), abnormalities (e.g.: two sources give different values) and so on. We are considering only the uncertainties calculated from machine learning model predictions and rest of the possible sources of uncertainty are eliminated.

**1.2.5 Texture**

Texture is the disposition or manner of union of the particles of a body or substance, a visual surface characteristic resulting in a certain appearance. Texture has been studied extensively in the field of computer vision, computer graphics, and modeling the low-level human visual system in cognitive psychology. Researchers have used different methods to study the perceptual features inherent in a texture pattern [22, 25]. Textures can be generated in different ways but since our research work is implemented in web, we have used the JavaScript and CSS driven web textures.

Background pattern

Description automatically generated with medium confidence 

Figure: Example of textures: [physical texture](https://fr.dreamstime.com/images-stock-macro-texture-textiles-tissu-image645724)(left), [Implied/web texture](https://github.com/d3/d3-drag) (right)

**1.2.6 Chromatic Aberration**

Chromatic aberration is a color distortion or alteration that is sometimes noticed on high contrast edges of objects in the photographs. It happens when the light of certain wavelengths gets bent. It usually appears in the form of purple, red, blue, cyan, green fringes. It can be seen alongside deep contrast edges and traditionally it means finding colors where they shouldn’t be or found in an unexpected form of color.

**Chart, diagram

Description automatically generated with medium confidence A picture containing plant, tree

Description automatically generated   
Figure - Examples of chromatic aberration (**[**left**](https://iceland-photo-tours.com/articles/photography-tutorials/chromatic-aberration-what-it-is-and-how-to-avoid-it)**,** [**Right**](https://expertphotography.com/remove-chromatic-aberration-photoshop)**)**

In the above figure we see two forms of CA where the left one shows how chromatic aberration occurs in optics as an effect when a lens is not able to properly refract all the wavelengths of colour in the same point. On the other hand, the circle bounded area on right sided picture shows how the quality of the picture gradually distorted.

CA is a problem of an image quality so most of the research about CA are conducted to fix the problem and improve image quality thereby. On the other hand, uncertainty is the problem of data quality and relevant research are conducted mostly regarding reducing it to improve data certainty and some of the research are conducted to visualize uncertainty with traditional approaches such glyph, opacity, and so on. Since our goal is neither to improve image quality nor data quality, we borrowed the term CA for our research to represent uncertainty as a novel approach in the field of visualization.

**1.3. Problem statement**   
The primary objective of the research is to present a novel concept of employing CA to represent uncertainties generated from predictive machine learning algorithms by amassing and feeding the COVID-19 data to the models. We hypothesized that our proposed system would add value in uncertainty visualization.

To implement the system, we needed to consider the following aspects:

1. How to generate the realistic uncertainty data?
2. Which platform or framework to be chosen to implement the visualization?
3. What is the design process of representing uncertainty with CA?
4. How to evaluate CA representation?
5. What is applicability of this representation?

Considering the above aspects, we have chosen to use recent WHO authorized COVID data to feed into three machine learning predictive models and one statistical model to get forecasted results for a certain period [3, 6]. Then calculated uncertainties from the predicted results and those are depicted as CA in D3 driven bubble chart where a single bubble is considered as the basic ingredient to compare with its own static and animated version and alternative options such blurred and noisy aberrations and relevant reference paper VSUP [35]. We conduct a user survey with all the competitor options and conduct numerical analysis to figure out the effectiveness of our novel design of uncertainty representation with CA. The survey is conducted totally online where we developed a questionnaire page to ease the access and participation for the participants.

We have used texture to represent CA and applied on different traditional charts such as streamgraph, bubble chart, impact chart, and usage chart.

**1.4. Approach**

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

Secondly, we had to study different journals about forecasting from temporal data using machine learning models and chose four popular modeling algorithms for our research. Since, finding and comparing the effectives of algorithms’ is out of our scope of work. We randomly chose a reasonable set of the models because we needed to generate the uncertainty data for the countries by using the predictive models and ignoring all inherent uncertainties itself.

Thirdly, having the data generation component in python, we needed to write up APIs to connect and pull the data when drawing the charts. Since the model training and data generation for all countries are long running processes, we precompiled the models to generate the data and stored into json file so that they can be read readily and sent back to the client on demand.

Fourthly, we have chosen D3.js as our front-end library for drawing the charts because it is an efficient platform for visualization prototyping. Since developing the basic drawing algorithms is not our goal, we have fairly taken the benefit of reusing library features but the total data collection, preparation, manipulation, correction and drawing algorithms for various strategic presentations are developed by ourselves.

Fifthly, conducted an experiment to evaluate the study by administering a HCI survey approved by REB of Dalhousie University and with the participation of the members of university community.

Finally, in depth numerical analysis and discussion on the survey response, mention and compare alternative perspectives of reference studies and consolidate the research outcomes to prove the effectiveness of the study.

**1.5. Thesis outline**

The remainder of this thesis is organized as follows. In **chapter 2**, we review the relevant literature on Predictive Machine Learning Models, Texture, Uncertainty, and CA. The whole literature review is subdivided into several sub-sections based on the contents. **Chapter 3** presents data processing, introducing predictive machine learning algorithms and necessary arrangement to setup models, brief description of time series forecasting, snapshots of uncertainty data. **Chapter 4** focuses on user study and numerical analysis for the sake of evaluation. **Chapter 5** shows the example of uses of CA in different charts. Finally, in **Chapter 6**, we discussed and summarized the thesis content, mentioned limitations, and suggest potential directions of future work and associated improvement.