**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 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 global impact, World Health Organization (WHO) has recognized it as a global pandemic. After conducting more than yearlong research, several companies manufactured vaccines with different names and as a result, immunization has 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 investigating the perceived data to discover the cause, find the patterns in different countries or demographic areas. So, in this study, we come up with a novel idea for a visualization 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 obtain the predicted results and find the 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 calculated model uncertainties in terms of chromatic aberration and textures in an interactive fashion.

**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 machine learning models for visualization. Some of the popular predictive machine learning algorithms are used for obtaining the forecasting data and corresponding uncertainty and then visualized 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 expansion has occurred all over the world and declared as an international public health crisis by the World Health Organization. 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 259 million people have infected and 5.1 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 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 some vaccines manufactured, and immunization 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 disease along with many other popular machine learning algorithms. 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 even for a single day. 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 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 insights to minimize its impact. Last but not the least the visualization in 2-dimensional space helps the users to perceive dual impacts in single a view. For instance, new cases and new deaths are provided in a single chart with the help of glyph-like texture representations.

**1.2 Basic** **Concepts**To dive into further detail it is very important to introduce thetop-level terms used in the thesis. The following section discusses to raise the basic concept of each of them:

**1.2.1. Uncertainty**

**1.2.2. Texture**

**1.2.3. Chromatic Aberration**

**1.3. Problem statement**   
The primary objective of the research is to present a novel concept of employing chromatic aberration to machine learning model uncertainties by amassing the COVID-19 data. 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 relevant research are conducted in this domain, so we intend to pursue research by working with machine learning models. For example: CNN is used in some research and MLP is used in other research, so we can keep both in our candidate list.

**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 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 three 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 model predictions.

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 data into json file that can be read and sent back to the client on demand.

Finally, 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.

**1.5. 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 four machine learning predictive models ii. Calculate corresponding uncertainties for different countries and visualize uncertainties in terms of Chromatic Aberration (CA) in a graphical presentation surface iii. Conduct user studies to evaluate user perceptions and applicability with commonly used visualizations. In this section, we are going to include some related studies of each component separately conforming to the aspects of the research.

**2.1 Works related to prediction in Machine Learning Models**

On the machine learning forecasting side, Song et. al. [1] compiled monthly data of influenza incidences 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 incidences, using time series analysis to construct an ARIMA model. The same model but different analysis and forecasting approaches was conducted on the coronavirus disease by other researchers [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]. Srivenkatesh applied Naïve Bayes, logistic regression, support vector machines, Random Forest, K Nearest Neighbour for the examination of liver malady. The classifications are assessed with 5 distinctive 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 infections with different machine learning approaches and pick most efficient algorithm [9]. Results of the examination demonstrated that Logistic Regression classifier demonstrated the best outcomes regarding precision with the least execution times.

**2.2 Uncertainty related prior works**

Botchen et al. [29] 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 decision making 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. [30] discusses state-of-the-art approaches such as Quantiﬁcation approach to uncertainty visualization, along with the concept of uncertainty and its sources.

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 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 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. In their experiment, they compare objective uncertainty (computed using the Bayesian framework) with subjective uncertainty (the confidence observers report about their visual perception). To this end, they used a visual task with well-defined statistical properties, discrimination under noise. They report a surprising degree of agreement between objective and subjective uncertainty and discuss possible computational models that could explain this ability of the visual system. 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 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 in regular clinical works. However, still a major problem is the lack of information on the uncertainty of the tissue classification, which is addressed in the paper Lundstrom 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.

A common goal in the communication of uncertainty is uncertainty-aware decision makingwhere the audience should be aware of the risks and rewards of certain decisions, modulate their confidence in their conclusions, and perhaps restrain from deciding when there is high uncertainty perceived. Correl et al. [35] introduced with the idea of allocating smaller ranges of a visual channel to data when uncertainty is high and larger ranges when uncertainty is low. This allocation of visual variables promotes patterns of decision-making that make efficient use of uncertainty information, discouraging comparison of values in unreliable regions of the data, and promoting comparison in regions of high certainty. In traditional bivariate maps, outputs for each combination of value and uncertainty might be represented as a 2D square whereas they approached it as arcs mapping values to smaller and smaller sets of outputs for higher uncertainty. But the main limitation of that research is they have used single color to represent both value and uncertainty in a single cell encoding system and suppresses the values for decision making when uncertainties are high. It also requires imperfect data value quantization.

Being a complex topic, most of the authors try to eliminate the existence of uncertainty from their visualization outcome, so the researcher Hullman conducted a survey and interviewed over 103 visualization authors in [36]. They identified that perceptions, practices, challenges, and attitudes are associated with uncertainty visualization and the majority of them agreed that or at least were sympathetic about the importance of uncertainty communication.

Data analysts also face unique challenges in interpreting the results on applying machine learning and statistical methods to timestamped event sequences to tackle various problems. Through a controlled study, the researcher Guo et. al [37] found that users experience more confidence in making decisions when alternative predictions are displayed alongside uncertainty information, and they consider the alternatives more when deciding between two options with similar top predictions. There are several limitations of this research, for example: they have used darkness to address uncertainty but that is not suitable to determine exact uncertainty values and make accurate decisions. Also, it requires the participants to be domain experts and it also requires data with alternatives.

Since uncertainty is a multi-faceted concept, there are various kinds of uncertainties, and the visualization of such uncertainties are applied in many contexts with different objectives, so there may not be optimal uncertainty visualization technique. The study of Korporaal et al. [38] investigates how data uncertainty visualized in maps might influence the process and outcomes of spatial decision-making, especially when made under time pressure in risky situations. The limitation of the research is that they have not considered the effect of stress along with time constraints. In addition, they have used only one type of texture(dotted) in their visualization experiment. So, the result cannot be generalized with non-texture, non-color based or gradients.

Earthquake models can produce aftershock forecasts but research on uncertainty visualization is often missing from earthquake science. So, Schneider et al [39] conducted research where three different uncertainty visualizations were produced: (1) forecast and uncertainty maps adjacent to one another; (2) the forecast map depicted in a color scheme, with the uncertainty shown by the transparency of the color; and (3) two maps that showed the lower and upper bounds of the forecast distribution at each location. Limitations of the paper includes: they needed to fix either the forecasted aftershock rate or its uncertainty and in the comparative judgment task, geographical features, such as roads and landmarks were omitted from the maps to avoid potential confounding effects on judgments which lowers the ecological validity of the study.

The authors Brodlie et al. [40] have reviewed the state of the art in uncertainty visualization, looking at both the visualization of uncertainty (which considers how to depict uncertainty specified with the data) and the uncertainty of visualization (which considers how much inaccuracy occurs in data processing through the pipeline of Haber and McNabb uncertainty reference model). They note that the visualization research community has enthusiastically taken up the challenge of uncertainty and most of the popular visualization techniques have been extended in some way to handle uncertain data.

When making an inference or comparison with uncertainty, noise, or incomplete data, measurement error and confidence intervals can be as important for judgment as the actual mean values of different groups. The paper [41] investigates drawbacks with the standard encoding and considers a set of alternatives and conducted a series of crowd-sourced experiments that confirms the encoding of mean and error significantly changes and by which viewers make decisions about uncertainty. They use gradient plots with transparency to encode uncertainty and violin plots with width as better alternatives. One area not well-covered by their experimental tasks was decision making and did not collect a great deal of qualitative data such as viewer preferences for different chart types which could be an important consideration for how data are perceived and used, especially for issues of trust and uncertainty.

In daily life, people regularly make decisions based on uncertain data navigating through gadgets or looking at the weather forecast online. The authors Greis et al. [42] published a web-based game on Facebook and compared four representations that communicate different amounts of uncertainty information to the user and compared. The results show that abundance of uncertainty information leads to taking unnecessary risks. Absence of uncertainty information reduces the risk taking and leads to more won turns, but with the lowest money gain. Representations with aggregated detailed uncertainty provide a good trade-off between being understandable by the players and encouraging medium risks with high gains. The paper doesn’t visualize the uncertainties but uses aggregated detailed uncertainty to the representations to offer a good compromise between understandability, encouraging educated risks and achieving credible winning criteria with high gains.

In statistics, people usually quantify uncertainty to help determine the accuracy of estimates, yet this crucial piece of information is rarely included on maps visualizing real data estimates.Lucchesi et al. [43] develop and present three approaches to include uncertainty on maps: (1) the bivariate choropleth map repurposed to visualize uncertainty; (2) the pixelation of counties to include values within an estimate’s margin of error; and (3) the rotation of a glyph, located at a county’s centroid, to represent an estimate’s uncertainty. They have not conducted user studies to determine whether these three methods effectively communicate uncertainty by drawing conclusions and answering questions in visualization. And, although users can see which counties have high uncertainties, they cannot determine the exact quantities of the margins of error by looking at the pixelated map.

Uncertainty is a fact of information; many types of information contain uncertainty, usually of heterogeneous categories. While there have been many calls for research about uncertainty visualization, the understanding of when and why one uncertainty visualization strategy should be used over others remains incomplete. To address the gap MacEachren el al. [44] presents two linked conceptual perspectives focused on uncertainty visualization. First, a typology of uncertainty is used to delineate kinds of uncertainty matched with space, time, and attribute components of data. Second, concepts from visual semiotics are applied to representing different categories of uncertainty. They address representation intuitiveness and relative performance, considering visual variables and iconic representations of uncertainty. The study does not cover finding the best symbolization method by integrating both data and data uncertainty representation into the same sign-vehicles. Also, they have not tested symbol size impact.

Many information fusion applications process and present huge quantities of data to enable an operator to make effective decisions. Reveiro [45] provides a general overview on uncertainty representations techniques and explains why the recognition of uncertainty plays an important role in decision making. In addition, it suggests the techniques developed in information visualization can be applied in information fusion and outlines how information fusion research might proceed further. The major contributions of this paper are (1) to highlight the importance of uncertainty visualization in decision-making, (2) to briefly review relevant modern uncertainty visualization techniques, (3) to propose general theories and results of user experiments for their theoretical analysis, (4) to suggest that techniques developed in information visualization can be applied in information fusion and (5) to outline how information fusion research might proceed further. The limitation of the paper is they only theoretically evaluate the weakness and strengths of the uncertainty visualizations representations.

Visual representations of information are challenged to incorporate a thought of confidence or certainty because the factors that influence the uncertainty of information vary with the type of information. Visualization researchers have no abstract model or framework for describing and constructing visualizations of uncertainty as it relates to intelligence analysis. The paper [46] of Judi Thomson presents a typology describing the aspects of uncertainty related to intelligence analysis, drawing on existing frameworks for uncertainty representation. They do not conduct any uncertainty visualization work but organizes the uncertainties into a logical framework or typology and then explores frameworks for uncertainty that have been developed for representation within the geosciences and scientific visualization community.

Instead of professional data scientists, the authors Boukhelifa et al. [47] engage domain experts with varying skill levels to find pertinent patterns and build a new uncertainty-aware sensemaking model. They describe their various coping strategies to understand, minimise, exploit, or even ignore the uncertainty influenced by accepted domain practices, but appears to depend on the types and sources of uncertainty. Participants of the study have different technical skill levels which may have had an impact on their behaviour and coping strategies. Moreover, the recruitment scheme was in potential bias due to snowball and social network effects.

Evaluating the impact of an uncertainty visualization is complex due to the challenge of defining correct behavior with uncertainty information and difficulties of interpreting uncertainty by people. Hullman et al. [48] present a taxonomy of methods for evaluating uncertainty visualizations and describe the results of a qualitative analysis applying their own framework to 86 publications which represent the state of uncertainty visualization evaluation. The taxonomy differentiates six levels of decisions that comprise an uncertainty visualization evaluation: the behavioral targets of the study, expected effects from an uncertainty visualization, evaluation goals, measures, elicitation techniques, and analysis approaches. They characterize overall trends in evaluation pathsof uncertainty visualization which indicate distinctions between methods for measuring accuracy and decision, as well as different methods for eliciting and assessing subjective confidence. They recommend specific steps that researchers should take when designing uncertainty visualization evaluations to strive for valid and transparent findings.

Understanding how effectively to display uncertain information has become increasingly important because uncertain information can be shown in many formats ranging from simple text to graphical representations. The paper [49] describes two studies in which degraded or blended icons were used to convey uncertainty regarding the identity of a radar contact as hostile or friendly. A classification study first showed that participants could sort, order and rank icons from five sets intended to represent different levels of uncertainty. Contacts and probabilistic estimates of their identities were depicted on a simulated radar screen in one of three ways: with degraded icons and probabilities, with non-degraded icons and probabilities and with degraded icons only. Results showed that participants using displays with only degraded icons performed better, that means the presence of numeric probabilities did not provide a statistically significant advantage in this task. Future research can be conducted to determine the suitability of the display techniques across different and more realistic task situations such as defence applications. The limitation of the paper is uses of icons in combination with numerical probabilities causes decision-makers hesitating and they expect for more assistive information.

Since many visual depictions of probability distributions, such as error bars are difficult for users to accurately interpret, the authors Hullman et al. present a study [50] of alternative representation, Hypothetical Outcome Plots (HOPs). In contrast to the many static representations of distributions, HOPs require relatively little background knowledge to interpret. Results showed that with HOPs, users made more accurate judgments than error bars and violin plots. Authors suspect that viewers of HOPs could make even more accurate probability hypothesis if provided with interactive graphical annotations. The limitations of the paper include: i. they have not tested all abstract, static and special purpose representations of concrete outcomes, ii. They did not raise subjects to explain their conclusions about data and uncertainty and even they know relatively little about the subject pool.

Authors Kay et al. [51] present a novel mobile interface design and visualization of uncertainty for transit predictions on mobile phones based on discrete outcomes. To develop it, they identified domain specific design requirements for visualizing uncertainty in transit prediction through 1) a literature review, 2) a survey of users of a popular real-time transit application, and 3) an iterative design process. In a controlled experiment they found that quantile dotplots reduce the variance of probabilistic estimates by ~1.15 times compared to density plots and facilitate more confident estimation by end-users in the context of real-time transit prediction scenarios. Fernandes et al. [52] noticed that when using uncertainty displays, decision quality may ameliorate over time. In real world, bus riders decide to leave for a bus using a real-time transit prediction application and everyone’s utility function remains personal and changes according to each situation dynamically. But participants of their studies use the same utility functions for all which may make people feel complicit in bad decisions leading to missing bus. Respondents gave mixed opinion about the usefulness of the uncertainty information provided by the app and so future work is necessary to see how widespread such reactions may be in real-world deployments. They both suggested that the presented designs should be evaluated in longitudinal field studies to assess actual acceptability and use.

By developing ways to include uncertainty in traditional information visualizations, we can provide more accurate depictions of critical data sets so that people can make more informed and accurate decisions. Skeels [53] reviewed existing work from several domains on uncertainty and created a classification of uncertainty based on the literature. They empirically evaluated and improved upon their classification by conducting interviews with participants from several domains. Their classification better describes the broad range of uncertainty across domains and provides a structure for more readily understandable uncertainty visualization. One of the most promising aspects of their classification is the concept of ‘layers’ of uncertainty that add complexity to data and is not simple to conceptualize or convey with current techniques. This creates an opportunity for visualization.

Inherent uncertainties from environmental data (e.g., Meteorological stations and doppler radars, etc.) is often omitted from visualization. The authors Whittenbrink et al. [54] showed scientific data collected from different sources, derived uncertainty information, and presented some ideas on designing uncertainty vector glyphs. They have developed a new vector glyph to visualize uncertainty in winds and ocean currents. Their approach is to include uncertainty in direction and magnitude, as well as the mean direction and length, in vector glyph plots. They defined visualization overloading and verity visualization, illustrating how their new glyphs represent the latter. They use both quantitative and qualitative methods to compare their glyphs showing they are superior to traditional ones in terms of uses because of their ease of understanding and information presentation.

**2.3. Chromatic Aberration related prior works**

Again, from a vision 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. Koh et. al. [10] presented a user study to observe the effect on users’ judgment with Lateral Chromatic Aberration (LCA) for Chart Reading in Information Visualization on Display Devices and suggested guidelines for information visualization designers to avoid such issues. LCA occurs when the lens does not focus all lights with different wavelengths to the same convergent point. Although the effect can be observed from natural scenes, they focus on LCA on modern display devices, and they present a series of controlled user experiments to show how people can misjudge information due to LCA. Although humans can compensate for the error especially with monochromatic aberration, the ability to correct errors caused by polychromatic aberration is still limited. There is an open task to investigate different degrees of aberration. A quantitative prediction on the amount of aberration depending on the wavelength and the power of eyeglasses will let us estimate the threshold on which viewers start to misinterpret the chart.

Colour is widely used in information visualisation to deliver different types of information such as extreme values, patterns and attribute values. Colour coding is known to be a particularly effective way to represent extreme values for human viewers due to the nature of pre-attentive vision. Therefore, Hyun Seung Yoo et. al. [11] study undertaken in order to identify appropriate interventions and propose design guidelines for information visualisation, especially in applications where size judgement is critical. The colour size illusion was replicated on an LCD monitor, revealing that yellow images appeared the smallest among a series of red, yellow, green and blue images on a white background.

Lens flare is an effect caused by light passing through a photographic lens in any other way than the one intended by design. In the paper [14] Matthias Hullin et al. present a novel method to interactively compute physically plausible flare renderings for photographic lenses where underlying model covers many components that are important for realism, such as imperfections, chromatic and geometric lens aberrations, and anti-reflective lens coatings. A common problem arises when triangles become smaller than one pixel is rasterization aliasing it can lead to very high intensity, but potentially error-prone rasterization.

Real cameras have an aperture through which light falls on an image plane containing receptors to register an image. For a sharp image, a small aperture is preferable, but then less light would hit these sensors and diffraction becomes an issue. Sungkil Lee et al. [15] nicely present a novel rendering system for defocus blur and lens effects. The efficient solution achieved by approximating the image-capturing process by considering not only aperture but also aspects of the lens interaction itself. They approximate optical aberrations, which is a unique feature for real-time approaches, and sometimes considered as crucial for realism. More precisely, the major contributions of the paper are: i an efficient algorithm for DOF and lens blur effects ii. An interactive and intuitive focus control system iii. A generalized method for expressive DOF rendering. They think combining their approach with single-pass depth peeling can be an interesting avenue for future work and mentioned single-pass decomposition of their depth peeling is slower, but their cache-efficient ray tracing mechanism helps to achieve better quality with a strong speedup.

One of the interesting research projects conducted by Micah K. Johnson et al. [13] shows that inconsistencies in lateral chromatic aberration can be used to detect tampering in visually plausible forgeries. They describe a computational technique for automatically estimating lateral chromatic aberration and show the efficacy of the approach for detecting digital tampering in synthetic and real images. They considered only lateral chromatic aberration for their study where the lateral aberration can be modeled as an expansion/contraction of the color channels with respect to one another. When tampering with an image, these aberrations are often disturbed and fail to be consistent throughout the image.

**2.4. Texture related prior works**

Particle Tracing and Line Integral Convolution (LIC) in Netzel et al. [22] 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 trends properly.

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, Caban et al. [23] 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 Bachthaler et al. [24] have 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 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, 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.

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. They have used a special sampling approach to generate sample distributions that cover the underlying domain densely while significant holes and cluttered areas are avoided. 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.

To improve the use of color in combination with motion where the author Weiskopf [27] 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.

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.

**2.5. Limitations of related works**

As stated in the related works section, a plethora of studies have been conducted in these 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 has not been represented with chromatic aberration. Furthermore, our approach of three dynamic variables visualization in two-dimensional space with texture is also a novel idea.

**3 Data Preparation**

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

**3.1.1 Data Collection and Processing**

Data comes bundled in a csv format from ourworldindata.org. The following table shows the list of fields/properties of each record where many of them are 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**

**A picture containing text, appliance

Description automatically generated**

Table-2: screenshot of sample data

In the above Table-2, we have shown only a snapshot of whole dataset where there are hundreds of thousands of records for Covid data for more than 237 countries and territories. Though there are numerous fields in the data, we only needed few of them as listed in previous section. The dataset is collected as a excel file which includes daily occurances and/or counts of all properties. The total\_\* fields like total\_cases, total\_deaths, etc are cumulative and so every day that is updated with previous day’s counts. Data is ordered by date and name of the country correspondingly. If there is no value in a cell for certain date and country then that cell is kept empty, so that is needed to handle during data preprocessing.

**3.2 Machine Learning**

Though we have not done anything novel in machine learning domain, but it is necessary to briefly introduce the salient components that were used in our research to process the available data and generate the required data.

**3.2.1 Definition**

Machine learning is an approach of artificial intelligence (AI)  to provide ability of automatic learning through the use of data. It doesn’t need any explicit programming to perform the task since the algorithms are designed 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 values for the next duration, for instance, data for the next few weeks using that metric.

Forecasting Algorithms

Training Data

New Data

Predictions

After Training

Figure-1: Predictive modeling workflow

**3.2.3 Time Series Analysis vs Forecasting**

Sometimes ambiguity arises between time series analysis with time series forecasting when working with temporal data. As per 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. The COVID-19 dataset is maintained on a global basis, so it is more trustworthy and with time series forecasting models can be considered as suitable for our research to get the predicted results and hence generate our required uncertainty data to represent chromatic aberration in visualization area.

**3.2.4 Concerns of Forecasting**Time series forecasting is an important area of machine learning. It is important because there are so many prediction problems that involve real life issues like 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, if there are outliers or corrupt data then those need to be addressed. Depending on the frequency, a time series can be of yearly (e.g., annual budget), quarterly (e.g., profit), monthly (e.g., cash flow), weekly (e.g., sales quantity), daily (e.g., weather forecast), hourly (e.g., stock market price), minutes (e.g., calls in a call canter) and even seconds wise (e.g., web traffic). Being the covid pandemic world-wide concerns for whole humanity, we use the daily forecast mechanism to our research. To compare the results side by side we have created prediction for 200 days from every models.

**3.2.5 Example of Forecasting**

Histogram

Description automatically generated

Figure-1: Example of daily covid forecasting

The above figure shows the daily forecasting of number new cases for United States based on previous statistics. So, in the blackish line in left shows the actual occurrences and the reddish part at right shows the predicted number of cases and light pink background surrounding the predicted line represents the ranges of model prediction, that means the model can predict a value between the lower and upper value for a certain day.

**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.

Diagram

Description automatically generated

Figure-2: Basic Architecture of MLP network [ref. 33]

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 the ‘adam’ optimizer for training the network.

The following steps shows the algorithm used to setup our MLP model:

---------------------------------------------------------------------------------------------------------------------

1. Take an instance of ‘Sequential’ Model from Keras deep learning library.
2. Add a Dense layer to the model with stating number of inputs (24), number of nodes (500), number of epochs (100) and batch size (100), rectified linear activation function (relu).
3. Add another Dense layer with number of outputs (1), since we predict a continuous value.
4. Compile the model with mean square error (mse) loss function and ‘adam’ optimizer.
5. Fit the model with training data set for number of epochs (100) and batch size (100).
6. Make an ensemble of models by following the steps 1 to 5.
7. Get prediction ‘yhat’ for each time step (day) from all the models of the ensemble.
8. Calculate the ranges (lower level, mean and upper level) of each prediction.
9. Calculate uncertainty of the model for each day by using the set of yhats using the uncertainty calculating formula explained in 3.7.

---------------------------------------------------------------------------------------------------------------------

Algorithm-1: MLP Model

**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.

Diagram

Description automatically generated

Figure-3: Basic Architecture of CNN network [ref. 34]

We define a CNN with two convolutional layers, one max-pooling layer, one flatten layer, and a dense layer from the input sequences. 1D convolution (Conv1D) layer (e.g., temporal convolution creates a convolution kernel that is convolved with the layer input over a single spatial (or temporal) dimension to produce a tensor of outputs. 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.

To dive into further we need to briefly introduce some of the basic terms used in this model:

**Conv1D:**

A convolution layer transforms the input image in order to extract features from it. This layer creates a convolution kernel that is convolved with the layer input over a single spatial (or temporal) dimension to produce a tensor of outputs

**MaxPooling1D:**

Max pooling is a pooling operation that selects the maximum element from the region of the feature map covered by the filter. Thus, the output after max-pooling layer would be a feature map containing the most prominent features of the previous feature map.

The following steps shows the algorithm used to setup our CNN model:

---------------------------------------------------------------------------------------------------------------------

1. Take an instance of ‘Sequential’ Model from Keras deep learning library.
2. Add a Conv1D layer to the model defining the number of filters (24), kernel size (500), input shape (100), rectified linear activation function (relu).
3. Add another Conv1D layer with same settings but without input shape.
4. Add another MaxPooling1D layer with pool size of 2.
5. Flatten (reshape) the result of previous step into single dimension before interpreted by the next layer.
6. Add a Dense layer with number of outputs (1), since we predict a continuous value.
7. Compile the model with mean square error (mse) loss function and ‘adam’ optimizer.
8. Fit the model with training data set for number of epochs (100) and batch size (100).
9. Create an ensemble of 6 models by following the steps 1 to 8.
10. Get prediction ‘yhat’ for each time step (day) from all the models of the ensemble.
11. Calculate the ranges (lower level, mean and upper level) of each prediction.
12. Calculate uncertainty of the model for each day by using the set of yhats using the uncertainty calculating formula explained in 3.7.

---------------------------------------------------------------------------------------------------------------------

Algorithm-1: CNN Model

**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.

**Diagram

Description automatically generated**

Figure-4: Basic Architecture of LSTM network (ref. 55)

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. A simple grid search of model hyperparameters was performed with the predefined configuration.

The following steps shows the algorithm used to setup our LSTM model:

---------------------------------------------------------------------------------------------------------------------

1. Take an instance of ‘Sequential’ Model from Keras deep learning library.
2. Add an LSTM layer to the model defining the number of nodes (24), input shape (100), rectified linear activation function (relu).
3. Add a Dense layer for 24 input nodes and ‘relu’ activation function.
4. As we predict single value output, add a Dense output layer of 1 node.
5. Compile the model with mean square error (mse) loss function and ‘adam’ optimizer.
6. Fit the model with training dataset for number of epochs (100) and batch size (100).
7. Create an ensemble of 6 models by following the steps 1 to 8.
8. Get prediction ‘yhat’ for each time step (day) from all the models of the ensemble.
9. Calculate the ranges (lower level, mean and upper level) of each prediction.
10. Calculate uncertainty of the model for each day by using the set of yhats using the uncertainty calculating formula explained in 3.7.

---------------------------------------------------------------------------------------------------------------------

Algorithm-1: LSTM Model

**3.6 ARIMA**  
An Autoregressive Integrated Moving Average (ARIMA), is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. A statistical model is autoregressive if it predicts future values based on past values. It is a very popular technique for time series modeling. It describes the correlation between data points and considers the difference of the values. ARIMA models work better with the following assumptions –

* The data series is stationary, which means that the mean and variance should not vary with time. A series can be made stationary by using log transformation or differencing the series.
* The data provided as input must be a univariate series since it uses the past values to predict the future values.

The model has three major components which come from its name – AR (autoregressive term), I (Integrated term) and MA (moving average term). Let us try to briefly explain each of these components –

* AR term refers to predicting the next value using the prior values of dataset. The AR term is defined by the parameter ‘p’ in ARIMA.
* Integrated(I) term represents the number of times the differencing operation is performed on series to make it stationary (i.e., data values are replaced by the difference between the data values and the previous values). Test like ADF can be used to determine whether the series is stationary and help in identifying the d value. Differencing is only needed if the series is non-stationary otherwise, no differencing is needed, and in that case d=0.
* MA term is used to define number of prior/lagged forecast errors used to predict the future values. The parameter ‘q’ in ARIMA represents the MA term.

**3.6.1 Auto ARIMA**Although ARIMA is a very powerful model for forecasting time series data, the data preparation and parameter tuning processes end up being really time consuming. Before implementing ARIMA, it needs to make the series stationary, and determine the values of p and q as stated earlier. Auto ARIMA makes this complicated task simple for us as it eliminates those time-consuming tasks. Below are the steps you should follow for implementing auto ARIMA:

---------------------------------------------------------------------------------------------------------------------

1. Load data: Collect data from the source repository and load into a data table.
2. Preprocess data: As the prerequisite of the model input is to be univariate, drop other columns from the data table and make sure all empty values with NULL so that system does not break during runtime.
3. Fit Auto ARIMA: Fit the model on the univariate series of data
4. Predict values: Make predictions on the validation set by using the prior values.
5. Calculate Series: Calculate series by using the forecasted results in earlier step.
6. Find the lower and upper bound of the series which will be used to calculate the uncertainties of the prediction.

---------------------------------------------------------------------------------------------------------------------Algorithm: ARIMA Model

**3.7 Uncertainty Data Generation**

Uncertainties are calculated from the ranges of predicted values for every time step (day) during the specified 200 days of forecasting period. That means we have a lower bound, mean and upper bound of the predictions for each time step. So the difference between upper and lower limit is the grey area of model prediction. Then find the maximum difference to set out the domain of the difference. Finally, divide each difference by maximum difference and multiply by a scaling factor to keep the maximum result in single digit. Here is given the steps to find the uncertainties using the machine learning models:

1. Read data from filesystem (excel file) to Data-Frame
2. Select Fields for which we need to generate uncertainty data
3. Create Machine Learning model for MLP/CNN/LSTM
4. Split data into training and test set
5. Train model with training set
6. Use model to get predicted or forecasted results
7. Find uncertainties or prediction error from model
8. Continue step 3 to 7 for each field and each model
9. Store uncertainty data as json in filesystem

Algorithm-1: calculate uncertainty using predictive models

**3.7.1 Uncertainty Data Scaling**

We have shown top-level algorithm in the above section to generate uncertainty data. Since the uncertainty values are pretty larger to accommodate in display, so it needed to scale in certain level. The following pseudo code is used to scale the uncertainty data.

1. country\_avg\_error = pred\_errors\_of\_all\_dates/number\_of\_days
2. max\_error = find\_max\_error(all\_country\_avg\_errors)
3. scaling\_factor = 7
4. country\_uncertainty = country\_avg\_error \* scaling\_factor / max\_error;

Algorithm-2: data scaling

**3.7.2 Snapshot of uncertainty data**

Since the pandemic affected all the countries of the world and there are more than 200 countries, so we have trained the models for top 100 countries which were infected severely. Based on that setup, we have sorted the countries by obtained uncertainties in both ascending and descending orders. The following two tables shows the top 10 uncertainty attaining countries and the bottom one shows the lowest 10 uncertainty attaining countries.

**3.7.3 Top 10 uncertainty countries using MLP model**

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **Actual Count** | **Predicted Count** | **Uncertainty** |
| United States | 14,851,118 | 15,652,300 | 7.00 |
| India | 15,693,425 | 7,409,636 | 4.28 |
| Brazil | 7,219,982 | 7,409,636 | 3.64 |
| Kazakhstan | 667,009 | 651,009 | 2.43 |
| France | 2,088,610 | 2,307,005 | 2.15 |
| Peru | 432,034 | 546,901 | 1.28 |
| Germany | 1,700,161 | 1,599,684 | 1.21 |
| Spain | 1,542,012 | 1,510,467 | 1.07 |
| Turkey | 3,645,288 | 3,389,016 | 1.03 |
| Argentina | 2,352,216 | 2,450,255 | 1.02 |

Table-3: Top uncertainty countries in the ordered list

**3.7.4 Lowest 10 uncertainty countries using MLP model**

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **Actual Count** | **Predicted Count** | **Uncertainty** |
| Qatar | 36,256 | 36,796 | 0.013 |
| Albania | 62,292 | 65,515 | 0.016 |
| Estonia | 90,950 | 89,900 | 0.017 |
| Egypt | 118,376 | 124,175 | 0.019 |
| Moldova | 103,270 | 101,832 | 0.019 |
| Australia | 161,819 | 147,134 | 0.021 |
| Algeria | 86,238 | 82,121 | 0.022 |
| Singapore | 178,151 | 175,400 | 0.025 |
| North Macedonia | 57,447 | 57,420 | 0.037 |
| South Korea | 277,584 | 274,766 | 0.037 |

Table-3: Lowest uncertainty countries in the ordered list

From the above two tables, it is clearly noticeable that uncertainty is completely independent on the number of cases (Actual Count). For example: United States has lower number of cases than India but achieved higher uncertainty than India. Again, Kazakhstan and France exhibit same behavior and if we examine other countries then surely, we will get more.

**3.7.5 Uncertainty Comparison among Models**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **MLP** | **CNN** | **LSTM** | **ARIMA** |
| United States | 7.00 | 7.00 | 3.44 | 7 |
| India | 4.28 | 0.61 | 7.00 | 3.52 |
| Brazil | 3.64 | 0.51 | 3.24 | 1.27 |
| Kazakhstan | 2.43 | 0.42 | 0.35 | 0.17 |
| France | 2.15 | 0.31 | 0.81 | 0.56 |
| Peru | 1.28 | 0.23 | 0.28 | 0.22 |
| Germany | 1.21 | 0.19 | 0.50 | 0.51 |
| Spain | 1.07 | 0.19 | 0.67 | 0.33 |
| Turkey | 1.03 | 0.19 | 1.21 | 0.30 |
| Argentina | 1.02 | 0.14 | 1.08 | 0.25 |

From the above comparison table of three different machine learning models, we notice that the uncertainties greatly vary for each country based on the model. There is no country which has identical uncertainty values for all three models. Though the dataset used in each of the models in similar approach, the variation appears due to their internal mechanism of the model algorithms. Since the model superiority examination is not our goal, we are not going to discuss further about it. We use the uncertainty data whatever we obtained from model prediction and uncertainty calculation methods.

**4 Visualization**

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

The followings are major charts which we have implemented in our application and applied chromatic aberration wherever possible.

Bubble chart, Streamgraph,

Describe about Streamgraph here

**4.1 Web Interface**

To visualize different charts, we have developed a 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.

**Chart

Description automatically generated**Figure-5: Initial Web Interface (Left - Bubble chart, right – Color 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. Bubble chart, Parallel Coordinates, Horizontal chart, Impact Chart, Usage Chart are available options in the list.

**Model dropdown:** Names of the predictive models for which we have generated data for finding the uncertainties and presenting as chromatic aberration. MLP, CNN, LSTM and ARIMA are the available options for the list.

**Reset:** Return to the initial state of the drawing for bubble chart. For this chart it has different type of modes listed in the right side of the toolbar.

**Texture Stream:** This is a toggle button to switch the stream graph from color-based filling to texture based filling, that means instead of flat color flow it uses bullet like textures to fill the stream but still they have different colors for their own country region. More detail is shown in section 4.5.

The followings are available operational modes of bubble chart:

**Pan Chart:** Since the bubble chart and stream graphare drawn side by side and they work interactively like filtering the 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.

**Star Fish:** changes the drawing mode to interact with mouse events. In this mode user can click on country bubble to open the corresponding texture stream graph as a wing of star-fish layout. So, when user select 8-10 countries in each side then the resultant chart will look like starfish. We will show further detail about this layout in later sections.

**Drill Models**: In this mode when user selects a country then four stream graphs with aberrated textures are shown in the right panel corresponding to the four predictive models. Detail explanation is shown in later section.

**Bubbles Select:** Select one or more country from the bubble chart and redraw it with the selected countries only. After selection, ‘Go’ button will perform the execution of redrawing task. It helps to compare specific countries because aberrations are not clearly perceivable with all countries.

**Bubbles Remove**: It is opposite feature of bubble select mode. It filters out countries from the bubble chart. In this mode the selected countries are omitted from the chart. After omitting countries on press ‘Go’ button it redraws with the other countries.

**Reshuffle Streams:** Allows to draw main streamgraph with the selected countries of interest from bubble chart. This is handy approach to see the bigger picture and compare streamgraph of one or more countries selectively.

**4.2 Filtering**

We use data for top 100 countries based on the total infection rate. As we see from the Figure-5, 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 a filtering option with different perspectives. In the section below we briefly explain them.

**4.2.1. With Selective Countries**

**Chart, bubble chart

Description automatically generated Chart, bubble chart

Description automatically generated** 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 redraws the bubbles side by side with comparatively bigger sizes.

**4.2.2 By removing countries**

**Chart, shape, bubble chart

Description automatically generated A picture containing honeycomb, outdoor object

Description automatically generated**

Figure-7: Removal of countries of interest

This is the opposite of the earlier one where the user can 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.

**4.3 Legend**

Placed at the top-left corner just below the toolbar and above the bubble chart with 5 consecutive circles. The circles are drawn for representing 5 different level of chromatic aberrations. The circle with 100% uncertainty represents the maximum uncertainty among all the countries drawn in bubble chart. Therefore, to find the amount of uncertainty for lower uncertainty valued countries, it helps users easier understanding.

**4.4 Reshuffling** **Streamgraph**

In Figure-5 we found the stream graph with countries are a bit clumsy to understand, so reshuffling is important to see and compare them side by with lower number of countries.

**Histogram

Description automatically generated with low confidence A picture containing text, comb

Description automatically generated**

Figure-8: Reshuffling main streamgraph

To serve that purpose, in this mode, a user can choose the countries from the bubble chart. 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 grayed out in the same chart (left view). Pressing ‘Go’ button confirms the redraw execution streamgraph with the selected countries as shown in the Figure-8 (right side view).

**4.5 Textures in Web**

**4.5.1 Texture Generation**

**4.5.2 Texture Utilization**

**4.6 Star Fish Inspiration**In this approach, user can draw multiple stream 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 benefit of this chart is it allows one to draw many charts in compact 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.

A picture containing wheel, gear

Description automatically generatedA picture containing wheel, gear

Description automatically generated

Figure-10: Approach to represent multi-country-actual counts(left), predicted counts(right)

**4.7 Parallel Coordinates Chart**

Parallel plots or parallel coordinates plots allows one to compare the features of several individual observations (series) on a set of numeric variables. Each horizontal axis 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. Data is represented in the form of a polyline, and it becomes possible to perceive trends shown by data entries from the visualization.

Diagram

Description automatically generated

Figure-11: Parallel coordinates chart

This plot is helpful in our presentation because we have several variables together to visualize one after another and showing the relationships between them. For example, you can compare number of total cases(total\_cases) with hospitalized patients (hosp\_patients) facilitated by a tooltip showing the country name. Also, it can show the predicted flow (thinner line) along with actual counts (thicker line). The limitation of this chart is frequent occlusions.

**4.8 Impact Chart**

This chart helps to indicate daily uncertainty presentation for every country as a cell. In this way a user can perceive trends for certain day or a set of consecutive days. In other words, the chart provides a useful platform that helps you decide which uncertainty requires your attention. So, if this tool was used by WHO then the administrator could consider which countries are vulnerable tomorrow or the day after tomorrow.

A picture containing text

Description automatically generated

Figure-12: Impact chart

**4.9 Horizontal Chart**

Horizontal 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.

Background pattern

Description automatically generated

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.

Chart, histogram

Description automatically generated

Background pattern

Description automatically generated

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

**4.11 Chromatic Aberration Example**

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 word, chromatic aberration is a color distortion that creates an outline of unwanted color along the edges of objects in a photograph.

**A picture containing plant, tree

Description automatically generated A frisbee flying over a bridge

Description automatically generated with low confidence  
Figure - Example of chromatic aberration (collected)**

**4.11.1 Implementation Mechanism**

**Diagram, engineering drawing

Description automatically generated Shape, circle

Description automatically generated**

(x,y)

Figure- Geometric 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 (, )
* Once for color (0, 0, b) with a shifted location of (, )

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 pictures, we found the aberration is shown as a kind of blurring or fading but here we present one with equal intensity highlighted color though the concept remains the same.

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

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.

**5.3. Future Work**

TBA in the final paper.

From Prof. Mayra/Brooks-  
And note these for future work:

1. Is it possible to have different hues of chromatic aberration? If yes, another possible study can be which CA hue works better.
2. When comparing the CA to other alternatives, you can use eye-tracking to get qualitative data.

… the 2nd one because with Covid we will be doing an online only study, so we don’t be able to use our eye tracking system.

**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. Received March 1, 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.

[3] COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population. Appl. Sci. 2020, 10(11), 3880; 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.  
  
[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.

[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.

[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.

[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. 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.

[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.

[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,

[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.

[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.

[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.

[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.

[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.

[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**.**

[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.

[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.

[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

[32] Jason Brownlee. Deep Learning Models for Univariate Time Series Forecasting. https://machinelearningmastery.com/how-to-develop-deep-learning-models-for-univariate-time-series-forecasting.

# [33] Aayush Agrawal, Building Neural Network from scratch. https://towardsdatascience. com/building-neural-network-from-scratch-9c88535bf8e9

[34] Akinori Hidaka, Takio Kurita. Consecutive Dimensionality Reduction by Canonical   
Correlation Analysis for Visualization of Convolutional Neural Networks. Conference: Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications. December 2017. Pages 160 – 167.

[35] Michael Correll, Dominik Moritz, Jeffrey Heer. Value-Suppressing Uncertainty Palettes.‬‬‬‬‬‬‬‬‬ Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. April 2018. Paper No.: 642 Pages 1–11.

[36] Jessica Hullman. Why Authors Don’t Visualize Uncertainty.‬‬‬‬‬‬‬‬‬ IEEE Transactions on Visualization and Computer Graphic. Jan. 2020, pp. 130-139, vol. 26.

[37] Shunan Guo, Fan Du, Sana Malik, et al. Visualizing Uncertainty and Alternatives in Event Sequence Predictions. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. May 2019. Paper No.: 573. Pages 1–12.

[38] Michelle Korporaal, Ian T. Ruginski, and Sara Irina Fabrikant. Effects of Uncertainty   
Visualization on Map-Based Decision Making Under Time Pressure. Human-Media Interaction, a section of the journal Frontiers in Computer Science. Received: 22 May 2020. doi: 10.3389/fcomp.2020.00032.

[39] Max Schneider, Michelle McDowell et al. Effective uncertainty visualization for aftershock forecast maps. Natural Hazards and Earth System Sciences. Discussion started: 3 September 2021. https://doi.org/10.5194/nhess-2021-237.

[40] Ken Brodlie, Rodolfo Allendes Osorio, and Adriano Lopes. 2012. A review of uncertainty in data visualization. In Expanding the frontiers of visual analytics and visualization. Springer, 81–109. DOI: <http://dx.doi.org/10.1007/978-1-4471-2804-5_6>

[41] Michael Correll and Michael Gleicher. 2014. Error bars considered harmful: Exploring alternate encodings for mean and error. IEEE Transactions on Visualization and Computer Graphics 20, 12 (2014), 2142–2151. DOI: • <http://dx.doi.org/10.1109/TVCG.2014.2346298>

[42] Miriam Greis, Passant El Agroudy, et al. 2016. Decision-Making under Uncertainty:   
 How the Amount of Presented Uncertainty Influences User Behavior. In Proceedings   
 of the 9th Nordic Conference on Human-Computer Interaction. ACM, 52. DOI:   
 http://dx.doi.org/10.1145/2971485.2971535

[43] Lydia R Lucchesi and Christopher K Wikle. 2017. Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. Stat (2017). DOI:http://dx.doi.org/10.1002/sta4.150

[44] Alan M MacEachren, Robert E Roth, James O’Brien, Bonan Li, Derek Swingley, and Mark Gahegan. 2012. Visual semiotics & uncertainty visualization: An empirical study. IEEE Transactions on Visualization and Computer Graphics 18, 12 (2012), 2496–2505. DOI: http://dx.doi.org/10.1109/TVCG.2012.279

[45] Maria Riveiro. 2007. Evaluation of uncertainty visualization techniques for information fusion. In 10th International Conference on Information Fusion. IEEE, 1–8. DOI: http://dx.doi.org/10.1109/ICIF.2007.4408049

[46] Judi Thomson, Elizabeth Hetzler, Alan MacEachren, Mark Gahegan, and Misha Pavel. 2005. A typology for visualizing uncertainty. In Electronic Imaging 2005. International Society for Optics and Photonics, 146–157.

[47] N. Boukhelifa, M.-E. Perrin, S. Huron, and J. Eagan. How data workers cope with uncertainty: A task characterisation study. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pages 3645–3656. ACM, 2017.

[48] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay. In pursuit of error: A survey of uncertainty visualization evaluation. IEEE transactions on visualization and computer graphics, 25(1):903–913, 2019.

[49] R. Finger and A. M. Bisantz. Utilizing graphical formats to convey uncertainty in a decision-making task. Theoretical Issues in Ergonomics Science, 3(1):1–25, 2002.

[50] J. Hullman, P. Resnick, and E. Adar. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. PloS one, 10(11):e0142444, 2015.

[51] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson. When (ish) is my bus?: User-centered visualizations of uncertainty in everyday, mobile predictive systems. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 5092–5103. ACM, 2016.

# [52] M. Fernandes, L.Walls, S. Munson, J. Hullman, and M. Kay. Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, page 144. ACM, 2018.

# [53] M. Skeels, B. Lee, G. Smith, and G. G. Robertson. Revealing uncertainty for information visualization. Information Visualization, 9(1):70– 81, 2010.

# [54] C. M. Wittenbrink, A. T. Pang, and S. K. Lodha. Glyphs for visualizing uncertainty in vector fields. IEEE transactions on Visualization and Computer Graphics, 2(3):266–279, 1996.

[55] Z. Wang et al. Model identiﬁcation of reduced order

using deep learning. International Journal for Numerical Methods in Fluids. Int. J. Numer.