STEM Fellowship High School Big Data and AI Challenge 2024-2025

The effects of climate change on education

Shariq[1], Timothy[1], Henry[1], Misgana[1], and Dinesh[1]

[1]Merivale High School

January 11, 2025

Abstract

The increasing frequency and intensity of climate-induced disasters pose a significant threat to education systems. Educational access is critical to reducing poverty and improving quality of life, yet inequalities are exacerbated by climate change. Adolescents, at a critical stage of development, are particularly vulnerable to the long-term effects of interrupted education. This study examines how extreme weather events amplify global educational disparities among adolescents. The primary objective is to explore this relationship across diverse socioeconomic contexts and varying levels of risk and vulnerability to climate change. Data was sourced from databases including UNICEF's *Child Climate Risk Index* and socioeconomic indicators from Our World in Data. Country-specific data facilitated the inclusion of case studies from regions disproportionately affected by climate change. A mixed-methods approach was employed, combining statistical analysis to identify patterns of inequality with machine learning models for predictive comparisons. Findings reveal that extreme weather events significantly disrupt educational access and opportunities, with adolescents in low-income regions, rural areas, and marginalized communities being the most affected. Key challenges include the destruction of school infrastructure and forced displacement. These barriers exacerbate existing inequalities, reducing the likelihood of affected adolescents overcoming cycles of poverty and vulnerability.[1] The study underscores the urgent need for climate-resilient educational policies prioritizing vulnerable populations. Integrating disaster preparedness into school systems, establishing alternative learning platforms, and addressing systemic inequities are critical to mitigating the educational impacts of climate change. These insights aim to support global efforts towards SDG 4 to reduce educational disparities and ensure equitable opportunities for adolescents worldwide.

Keywords
Extreme weather events, Climate-induced displacement, Climate Risk and Vulnerability, Educational inequalities, Machine Learning, Predictive Modeling

1 Introduction

Climate change has become an urgent global issue, with its effects increasingly felt through extreme weather events. These events, including floods, cyclones, and wildfires, are more frequent and severe due to climate change, causing devastating social and economic consequences. Extreme weather disrupts communities, destroys infrastructure, and endangers lives. In 2024 alone, statistics from Our World in Data reveal that climate-driven disasters affected over 148 million people and resulted in 11,500 deaths. [2]Our study focuses on how these climate-induced disasters exacerbate existing educational inequalities among adolescents. The guiding question for our research is: How does differing risk and vulnerability regarding climate change exacerbate educational inequalities among adolescents?

Why, specifically, did we choose to examine the impacts of climate change on adolescents? This focus stems from recognizing the unique challenges adolescents face, particularly in vulnerable regions As young people living in Canada, we have benefitted from consistent access to high-quality education and robust mitigation strategies to address the challenges posed by climate change. However, this privilege is not universal. Around the world, climate-driven events disproportionately affect vulnerable populations, particularly in low- and middle-income countries that lack adequate resources for disaster preparedness and recovery.

To better understand the future implications, we used artificial intelligence to predict the impact of natural disasters on school enrollment rates. Our model analyzed historical data on climate events, demographic vulnerabilities, and education statistics to forecast enrollment trends in affected regions. This predictive tool enabled us to identify areas at heightened risk of severe educational disruptions, emphasizing the urgency of targeted interventions. By leveraging AI, our research provides valuable insights into mitigating the long-term consequences of climate change on educational equity, empowering policymakers to act proactively.

2 Materials & Methods

2.1 Datasets

This study draws on data from a variety of reliable sources, including Our World in Data and the World Bank, to analyze global trends in education, impacts of natural disasters, and economic conditions. The first dataset, sourced from Our World in Data, focuses on education enrollment rates, offering a detailed historical account of participation across primary, secondary, and tertiary levels. It spans several decades, from the 1960s to the present, and includes country-specific data categorized by gender and education stage. This resource highlights long-term trends and disparities in access to education. Another dataset from Our World in Data tracks the annual number of people affected by natural disasters, covering a broad time frame from the early 20th century to the present. It categorizes disasters by type, such as floods, droughts, and earthquakes, and provides country-level insights into the effect of natural disasters on people. Lastly, the World Bank dataset on GDP per capita provides essential context for measuring economic development. It includes annual data for countries worldwide, measured in current US dollars, and covers an extensive period from 1960 to the present. Adjusted figures, accounting for inflation and purchasing power, further enhance the dataset's ability to capture economic trends and disparities. By integrating data on education, disaster exposure, and economic performance, this study explores the complex intersections between social and environmental factors.

2.2 Technologies Used

The study used several technologies for data processing and analysis. Microsoft Excel was used to view data, for manual processing of the data, for producing graphs, and for generating predictions. Python was used for data processing and analysis, and to create the machine learning model. Python packages used include Pandas, Numpy, MatplotLib, and Scikit-Learn. Git was used for version control and a GitHub repository was created for the project.

Once the user is captivated, we direct them to another page and ask them if they're a student or not. Students are directed to a page with an AI[3], where adolescents can tell the AI the grade they're in and the last thing they learn and based on this it will help guide the student through future topics. This will help them catch up or continue school if their education has been interrupted due to external factors. We also implemented an algorithm that can help teachers and schools reduce their carbon footprint[4], by helping them better use their energy. There are around 4 million schools[5] in the world. The collective contribution of the greenhouse emissions of schools exclusively in the U.S., which accounts for 2.8% of schools is more than 72 million metric tons[6]. Most of these carbon emissions are exclusively caused by electricity usage, thus we offer schools to upload their electricity bill as shown in Figure 2 which will then be processed. Once processed, there will be a compilation of stats along with AI-generated suggestions to make their schools more efficient based on the electricity bill. Our algorithm will also provide a suggestion for a greener alternative and how it will positively impact the environment and themselves through cost savings.

2.3 Data Processing methods

The data had to be processed before it was ready for analysis and for training the machine learning model. This included linear interpolation to fill in missing values, as well as filtering out data points that were unusable. Some examples of filtered out data points include non-countries (regions, continents, categories), data points not within the chosen range of years (1980-2023) and any duplicates. These steps were automated using Python to manipulate csv files.

2.4 Data Analysis Methods

The data was analyzed for correlations, specifically the Pearson correlation coefficient. This was done with the builtin .corr() function in Pandas. This was done at a global scale for the correlation between GDP per capita and enrollment and between people affected by natural disasters and enrollment. The correlation between people affected by natural disasters and enrollment was also calculated for two countries specifically. **Lesotho** is a small, landlocked country in Southern Africa that has a high risk to natural hazards including floods, drought, frost, strong winds, and heavy snowfall[7]. **Canada** is also subject to a wide range of natural hazards because of the extent and diversity of its geography, such as floods, wildland fires, and winter storms[8]. However, it is less affected by these events due to better disaster preparedness and infrastructure.

2.5 Machine Learning

A neural network was trained to predict future enrollment rates given data on past enrollment rates, people affected by natural disasters, and GDP. This was done using random forest regression, which involves partitioning the data into a "training" section and "testing" section, with the model being trained on the training section, and having its predictions compared against the values from the testing section.

However, there were issues with the neural network and for predictions involving the GDP data. It was getting too computationally intensive, the model had a large amount of mean absolute error (MAE) and because the relationship was mostly linear anyways, it was decided to instead use the FORECAST excel function to create predictions using linear regression. The function is calculated as described below.[9]

$$y = a + bx$$

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$y = a + bx$$

2.6 Visualization

It was also decided to build an interactive map (Figure 1) to complement the analysis, providing a way to visualize the data used in this study. The map is located on the project website, and is a choropleth world map with the colour of each country signifying its value in the selected data set and selected year. The data set (total people affected by natural disasters, enrollment rates, GDP) is selected with a button in the bottom right and the year (1980 - 2023) is selected with a scrollable timeline at the bottom of the page. There is also a search bar in the top left to find specific locations.

Google Maps was used to create the map, and the map was overlaid with coloured polygons for each country. The coordinates for each polygon was sourced from Azure Maps.

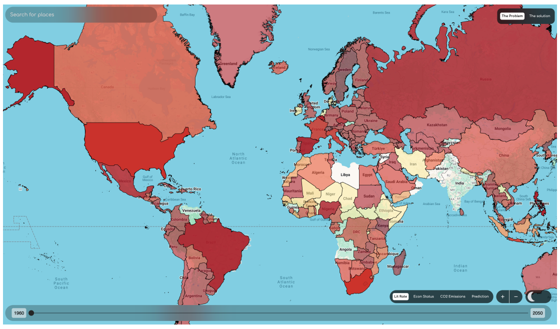

Figure 1: Our interactive map showcasing how climate change is impacting education

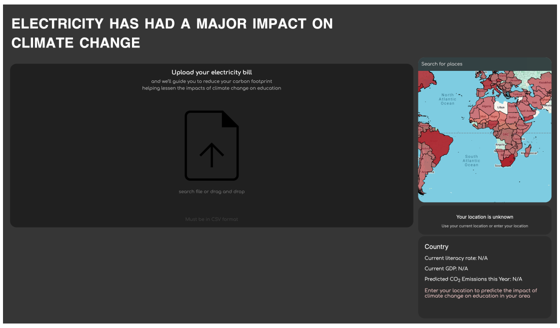

Figure 2: The section of the website where schools can upload their power bill

2.7 Proposed Solution

It was also decided to propose a possible solution to the impacts of climate change on education. On the project website, there is a separate page (Figure 2) that contains AI-based technology to help students, teachers, and schools. For students, an AI assistant provides lesson material to students who are unable to attend school due to external circumstances, only requiring their grade and the last lesson they learned. For teachers and schools, the AI requires their electricity bill and then provides suggestions on energy efficiency.

3 Results

3.1 Dataset

There are some issues to be noted about the dataset. Some of the data are inaccurate and have missing values. Steps were taken to mitigate this effect as much as possible by interpolating the dataset, but this still leaves room for error. This likely greatly influenced the correlation between variable and the outcome of the model.[10]

3.2 Correlation from the dataset

The data in this analysis proved to have weak correlations for all the variables tested against literacy rates. CO_2 emissions had the highest correlation with 21% followed by the annual GDP which showed a 12.08% correlation. The number of Disasters was next which had a 6.49% correlation, followed by the number of deaths due to natural disasters with -1.86%, after which came the total number of people affected by natural disasters with -1.1186% correlation and lastly economic growth from year to year in percentage -0.6868%.

3.3 Randomforest Regression

The version of the machine learning model proved to be highly accurate. When using economic growth data, and CO_2 emissions to predict literacy rates, the model had an accuracy score of around 96%. When CO_2 emissions data was replaced with the number of deaths caused by natural disasters the accuracy fell drastically, to 88%. Testing again with the total number of people affected by natural disasters and economic growth data had an accuracy of 86%. The version of the machine learning model proved to be highly accurate. When using with economic growth data, and CO emissions to predict literacy rates, the model had an accuracy score of around 96%. When CO_2 emissions data was replaced with the number of deaths caused by natural disasters the accuracy fell drastically, to 88%. Testing again with the total number of people affected by natural disasters and economic growth data had an accuracy of 86%.

GDP data and the number of natural disasters seemed to be a decent indicator of literacy rates. It had adequate correlation and the model had a high accuracy of 96%.

```
\begin{figure}
    \centering
    \includegraphics[width=0.4\textwidth]{test.png}
    \caption{Hello!}
\end{figure}
```

4 Discussion

4.1 Creating Literacy Rate Model

4.1.1 Filling missing values

Filling missing values in our dataset is important since it is essential to have the least gaps when inputting data into a machine learning model. With gaps in the dataset ML models

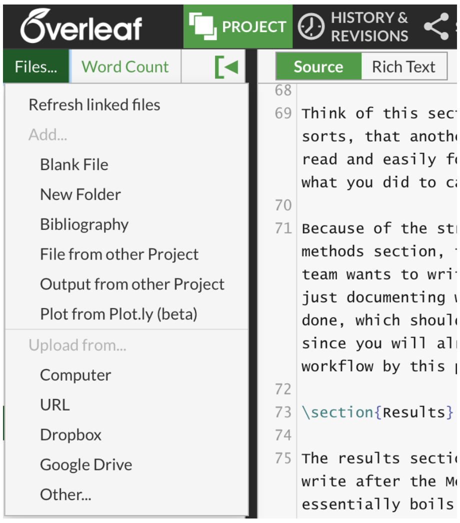

Figure 3: Notice how LaTeX automatically numbers this figure.

will often assume the values in that position based on the rest of the dataset, this introduces bias because the statistics of other countries will have an influence on the missing data for other countries[6]. This means that if there is more data missing for certain countries the data will be filled with the influence of other countries creating bias. To avoid this type of bias we manually ran the interpolation

4.2 Data Sources

It's important we first talk about our data sources. All of our data is from reputable sources, but often there are inherent shortcomings in the data.

And here is the 'meat' of the paper, so to speak. This is where you interpret your results, pointing out interesting trends within your data and how they relate to your initial hypothesis. This is also the place to justify your methodology if you're so inclined (i.e. Why did you specifically use a certain statistical test over another? Why this tool over that tool?). Lastly, you're going to want to discuss potential sources of error. Make sure to make explicit reference to figures/tables when discussing your data; it can be helpful to walk the reader through your own personal interpretation of each figure in order. Although we recommend looking at past winning papers over at the STEM Fellowship Journal's website anyway, referring to those papers might prove most helpful when it comes to writing your discussion.

5 Conclusions

In conclusion, our research underscores the critical intersection of climate change and educational inequality, emphasizing the disproportionate impact of climate-induced disasters on adolescents in vulnerable regions. By analyzing historical data and employing advanced artificial intelligence techniques, we have demonstrated the cascading effects of extreme weather events on school enrollment rates, particularly in low- and middle-income countries. These findings highlight the urgent need for targeted interventions to mitigate the adverse impacts of climate change on educational equity.

A key component of our study was the development of an innovative neural network model, which utilized artificial intelligence to predict the effects of climate change on educational outcomes. This model analyzed complex datasets, including climate event patterns, demographic vulnerabilities, and education statistics, to identify trends and forecast areas at heightened risk. The neural network's ability to process and interpret large amounts of data with high accuracy allowed us to uncover critical insights into the relationships between climate factors and educational disruptions.Recognizing the complexity of these challenges, we went a step further by developing an application designed to incorporate a wider range of factors into climate change prediction and analysis. This app integrates data on socioeconomic conditions, and climate change vulnerability, providing a more holistic view of the vulnerabilities and risks faced by various regions. By enabling users to visualize and analyze these multidimensional interactions, the app serves as a valuable tool for policymakers, educators, and researchers. It empowers stakeholders to make informed decisions and design proactive strategies to address the root causes of educational disparities exacerbated by climate change.

Our findings call for a collaborative, global approach to ensure that adolescents worldwide can access uninterrupted education, even amidst the growing threats posed by a changing climate[11]. By leveraging advanced technologies such as artificial intelligence and fostering cross-disciplinary partnerships, we can build resilient educational systems that adapt to the challenges of the future. The development of our neural network model and accompanying app marks a significant step forward in this effort, providing actionable insights and fostering greater awareness of the interconnectedness between climate resilience and educational equity.

Acknowledgements

Anyone to thank/credit for helping your team along the way? This is the place to do it!

References

[1] Minsun Kim Chairat Chongvattanakij. Lives displaced and disrupted: how climate change threatens the right to education. *UNESCO*, 2023.

[2] H Ritchie and P Rosado. Natural disasters, how many people die from disasters, and how are these impacts changing over time, 2022.

[3] Jure Zupan. Introduction to artificial neural network (ann) methods: what they are and how to use them. *Acta Chimica Slovenica*, 41(3):327, 1994.

[4] Nikolay V Lobus, Maria A Knyazeva, Anna F Popova, and Maxim S Kulikovskiy. Carbon footprint reduction and climate change mitigation: A review of the approaches, technologies, and implementation challenges. *C*, 9(4):120, 2023.

[5] Chris Padilla. How many schools are there in the world? – save our schools march. *Expert Beacon*, 2023.

[6] Parikhit Sinha, William A Schew, Aniket Sawant, Kyle J Kolwaite, and Sarah A Strode. Greenhouse gas emissions from us institutions of higher education. *Journal of the Air & Waste Management Association*, 60(5):568–573, 2010.

[7] UNDRR. Strengthening climate change and disaster statistics to improve lesotho's resilience. *UNDRR*, 2023.

[8] Gouvernment of Canada. Natural hazards.

[9] Sanjida Ahmed. Forecast function in excel (with other forecasting functions. 2024.

[10] IBM. Organizing data sets. 2021.

[11] Emily P Backes, Richard J Bonnie, Engineering National Academies of Sciences, Medicine, et al. Adolescent development. In *The Promise of Adolescence: Realizing Opportunity for All Youth*. National Academies Press (US), 2019.