Deep Learning Approach for Automated Fact-Checking in Key Climate Change Topics

Bushraa Yousuf
School of Science and Engineering
Habib University
bm07609@st.habib.edu.pk

Hijab Fatima
School of Science and Engineering
Habib University
hf07068@st.habib.edu.pk

Zainab Haider
School of Science and Engineering
Habib University
zh07104@st.habib.edu.pk

Abdul Samad
School of Science and Engineering
Habib University
abdul.samad@sse.habib.edu.pk

Sandesh Kumar
School of Science and Engineering
Habib University
sandesh.kumar@sse.habib.edu.pk

***Abstract*—Climate change misinformation poses a significant threat to informed public discourse, policy-making, and climate advocacy. To address this challenge, we propose an automated fact-checking system specifically designed to verify climate change claims using scientific abstracts. Our approach focuses on four essential subdomains: Global Greening, CO2 Emissions, Extreme Weather, and Global Warming. By leveraging a retrieval system (FAISS), large language models (LLMs) and a dataset of scientific abstracts and claims, our system aims to enhance the accuracy of detecting misleading climate information. We fine-tune existing LLMs (LLaMA 3.2, SciBERT, and DeBERTa) on our curated dataset and evaluate their performance using accuracy as the evaluation metric. Our results show that DeBERTa achieves the highest accuracy of 89.29%, followed by LLaMA 3.2 3B, SciBERT, and LLaMA 3.2 1B. The use of data augmentation significantly improves model performance, highlighting the effectiveness of our approach. This study contributes to the development of reliable climate literacy solutions, ensuring that climate-related decisions are driven by accurate and reliable information. Future work involves expanding the dataset, exploring additional deep learning models, and integrating our system with online platforms to combat climate change misinformation.**

I. Introduction

The recent spread of climate change misinformation has become a serious threat to informed public discourse, policy-making, and climate advocacy. Incorrect information about climate change can easily mislead the general public, hindering efforts to address environmental challenges and pushing public policies in misguided directions. As awareness of the issues regarding the climate crisis grows, so does the volume of false information, especially across social media and online platforms. Checking these assertions is a very time-consuming process, leading to the demand for automatic and reliable systems to judge and verify whether such claims are true.

The fact-checking domain which is mediated by automated systems with the help of deep learning has received much research attention. While there is a notable advancement in the larger fact-checking practices, climate change-related fact-checking applications is still relatively under-explored.

Few investigations on this subject applied various machine learning structures to either verify claims or predict the onset of climate-related phenomena. Some of these studies use NLP techniques in extracting information from climate-related publications and public databases, while others apply architectures from deep learning to assess the validity of the claims. The current research initiative is founded upon these principles, aiming to create an automated fact-checking system specifically designed to tackle the distinct obstacles associated with verifying misinformation related to climate change. In contrast to prior studies that address climate-related issues in a generalized manner, our methodology concentrates on four essential subdomains: **Global Greening, CO2 Emissions, Extreme Weather, and Global Warming.** This specialized approach facilitates enhanced accuracy in fact-checking by systematically extracting and scrutinizing claims from credible scientific abstracts. Utilizing a dataset of scientific abstracts and claims obtained from credible sources and the CLIMATE-FEVER dataset, our system aims to enhance the accuracy of detecting misleading climate information. This would satisfy the growing need for climate literacy solutions thereby ensuring that all climate-related decisions are driven by accurate and reliable information.

This study is driven by the central research question: *Given a claim about climate change, assess its factual accuracy and determine whether it is true or false.* To address this, the research focuses on fine-tuning existing large language models using a dataset of climate-related claims and scientific abstracts, enabling accurate classification of claims. The subsequent sections of this report provide a comprehensive literature review followed by the the methodology used, experimentation with different deep learning models, results obtained, and analysis of the findings, culminating in a discussion of the system's potential limitations and future directions for research.

II. Literature Review

In recent years, the increase of climate change misinformation has become a growing concern. Various studies have aimed to develop automated fact-checking systems using machine learning and deep learning models to address this issue. The relevant research carried out in this domain will be reviewed in this section, synthesizing key points of each research.

During our literature review, we found that significant research has been conducted on fact-checking in general however; efforts specifically targeting climate change claims were limited. The oldest study [5] we found was from 2017, presenting a general-purpose fact-checking framework that can be applied to any domain. This paper is important for building the foundation of automated fact-checking. The model employs a combination of Long Short-Term Memory (LSTM) networks and Support Vector Machines (SVMs) to fact-check claims using web-based sources. The system retrieves snippets from search engines and compares them with claims using text embeddings. The framework performed well on rumor detection tasks, achieving an accuracy of 80% in detecting false claims. However, it was limited by the need for external information, which is not always available.

In a related vein, another study [4] presents a deep learning approach focused on predicting climate-induced disasters using historical climate data and literature. The authors developed a model that forecasts Climate-Induced Disasters (CID) by processing climate-related literature in the form of climate change indices and flood disaster data. By using neural networks with different architectures, they achieved impressive accuracy of around 96% in predicting flood disasters. Although centered on disaster prediction rather than direct fact-checking, this research underscores the significance of integrating authoritative scientific information into machine learning models, which aligns with the need for reliable climate science sources in building effective fact-checking systems.

Building on this foundation, in 2024, researchers introduced the FACT-GPT [2] framework to address fact-checking challenges related to COVID-19 misinformation. Using a dataset of 1,225 debunked claims from Google Fact Check Tools and PolitiFact, they generated 3,675 synthetic tweets using models like GPT-4, GPT-3.5-Turbo, and Llama-2-70b-chat-hf. The framework fine-tuned these models on the synthetic dataset, with GPT-4-generated data delivering the best results. FACT-GPT achieved high precision and recall in identifying relevant claims, with a notable F1 score improvement in entailment and neutral classifications.

After reviewing the general landscape of fact-checking research, we found a study [8] from 2021 specifically addressed the challenge of combating misinformation related to climate change using the CLIMATE-FEVER dataset, which contains 1,535 climate-related claims. The researchers developed a fact-checking pipeline based on RoBERTa, which was fine-tuned using the CLIMATE-FEVER and FEVER datasets. The key innovation in this paper was the integration of Unsupervised Data Augmentation (UDA), which allowed the model to leverage 4,127 additional unlabeled claims. By augmenting the training data with this semi-supervised approach, the model achieved a state-of-the-art F1 score of 0.7182.

In 2022 [7], researchers further advanced this field by fine-tuning a GPT-2 model using a climate-specific corpus of 360,233 abstracts from leading climate scientists, creating the climateGPT-2 model. This model was then used for fact-checking tasks on the CLIMATE-FEVER dataset. The fine-tuned model showed significant improvements, increasing the F1 score for fact-checking from 0.67 to 0.72. These results demonstrate the effectiveness of fine-tuning GPT-2 on domain-specific data for improving performance on climate change-related fact-checking.

The study [6] from 2024 further contributes to the field by introducing Climinator, an AI-driven system designed for fact-checking climate change claims, which will serve as the main research we follow in our project. Climinator employs large language models (LLMs) and integrates an innovative Mediator-Advocate framework, where different LLMs (Advocates) evaluate claims from various perspectives. Disagreements between advocates trigger iterative debate processes until a consensus is reached. The study evaluated Climinator on 414 climate change claims from sources like Climate Feedback and Skeptical Science, leveraging the GPT4 model and RAG system to achieve an accuracy of 97.06%. This research aligns closely with our proposed approach, which will also leverage both claims and literature to build a robust fact-checking system.

From the literature reviewed above, it is evident that significant work has been done on fact-checking in general, but specific efforts targeting climate change claims remain limited. Our project aims to address this gap by performing fact-checking on climate change claims using knowledge derived from scientific abstracts. While many of the existing studies target climate change as a broad subject, covering various topics, our approach will focus on four specific subtopics to achieve more fine-tuned results. This focused methodology will enhance the precision and reliability of our fact-checking efforts in the context of climate change misinformation.

III. Methodology

In the methodology section, we provide a comprehensive explanation of the dataset curation process, including its detailed acquisition steps and criteria. This is followed by an overview of the deep learning models selected for the study, highlighting their architecture and relevance to the task.

A. Dataset

Our corpus consists of scientific literature abstracts and claims. We have collected a total of 721 abstracts taken from the scientific literature for the four main topics that we are targeting. We accessed the scientific journals and articles from the links available on first 20 pages of Google Scholar. We have 770 claims regarding climate change with main focus on our targeted topics that have been sourced from the online dataset called Climate Fever [3] which has real-world climate claims for verification.

1) *Acquisition*:

We utilized the Beautiful Soup library to scrape links to relevant scientific literature from Google Scholar, categorized by topic. For each topic, we collected approximately 200 links. While we were able to automate the extraction of some abstracts, the remaining ones required manual review, where we visited the links individually to retrieve the abstracts. Additionally, we excluded the links leading to books, as they typically lacked abstracts and were not suitable for efficient use in training and testing our models. We relied solely on Google Scholar because it is a trusted source for scientific literature and offers access to a wide range of freely available papers and journals. This ensured the credibility of our dataset while maximizing access to open-source materials.

During our literature review, we identified the CLIMATE-FEVER dataset [3], which contains 1,535 climate-related claims. Since our model focuses on only four specific topics, we initially filtered the relevant claims using the Pandas library. Following this, we performed a manual review to further refine the data and ensure that only the most relevant claims were selected for our analysis which left us with 770 claims. The dataset originally categorized claims into four labels—supports (0), refuted (1), not enough information (2), and disputed (3). For simplicity, we reduced these categories to three by combining the *refuted* and *disputed* labels into a single *false* category. In our revised labeling system, claims were labeled as 0 if supports, 1 if refutes, and 2 if there was not enough information, aligning with the classification tasks for our model.

2) *Characterization*:
Our scientific abstracts data is distributed across the four topics as illustrated in the table I:

Topic	*Number of Abstracts*
Global Greening	188
CO2 Emissions	189
Extreme Weather	177
Global Warming	167
Total	721

TABLE I
ABSTRACTS CORPUS BREAKDOWN

Distribution of claims according to the categories is illustrated in the table II:

Category	*Label*	*Number of Claims*
Supports	0	356
Refutes	1	288
Not enough information	2	221
Total	-	**770**

TABLE II
CLAIMS CORPUS BREAKDOWN

Our dataset comprises of the four topics for the following reasons:

a) In the initial phase of the project, we conducted research on trending climate-related topics using Google Trends along with personal reviews of social media platforms such as Instagram and Twitter. We observed that *global warming* and *CO2 emissions* consistently ranked among the most discussed topics. Based on this popularity, we prioritized these topics for inclusion in our dataset.
b) *Global greening* emerged as a climate change topic with relatively little controversy within the scientific community, as its increase is widely recognized. However, public awareness on this subject remains limited. We included global greening to incorporate well-established factual information, aiming to increase public understanding of this lesser-known issue.
c) We selected *extreme weather* as the fourth topic due to the prevalence of misinformation surrounding it, which can hinder effective policy-making. Misinformation about extreme weather events often misleads the public and policymakers, potentially allowing for poor decision-making. By including this topic, we aim to promote informed public discourse and support better decision-making processes.

B. Deep Learning Models

In this project, we utilized the following open-source Deep Learning Models for Climate Change Claim Verification: LLaMA 3.2 (1B and 3B), SciBERT, and DeBERTa. Each model was chosen for its unique strengths in language understanding and processing, particularly for scientific and factual claims.

1) *LLaMa 3.2:*
LLaMA (Large Language Model Meta AI) is a family of foundational models designed for efficient and high-performance natural language understanding and generation. Widely recognized as one of the best open-source LLMs, LLaMA combines state-of-the-art performance with accessibility, making it an excellent choice for research and practical applications. In this study, we employed two versions of the LLaMA 3.2 model: 1B and 3B, representing the number of parameters. These models are decoder-only transformer architectures trained on diverse datasets, optimized for tasks like text classification, completion, and question answering.

- *LLaMA 3.2 1B:* The 1B model, with one billion parameters, strikes a balance between computational efficiency and performance. Its smaller size makes it suitable for scenarios requiring faster inference. For this study, we fine-tuned the model by adding a classification head tailored to our task, enabling it to classify climate change claims into the categories 'supports,' 'refutes,' or 'not enough info'.
- *LLaMA 3.2 3B:* The 3B model, with three billion parameters, provides greater capacity to capture nuanced linguistic and contextual information, enhancing classification accuracy. Similar to the 1B

model, a classification head was added to adapt it to the claim verification task. While offering improved performance, the 3B model requires higher computational resources due to its increased parameter size.

2) *SciBERT*
SciBERT is a domain-specific variant of BERT (Bidirectional Encoder Representations from Transformers) designed specifically for scientific text. Trained on 1.14 million scientific papers from the Semantic Scholar corpus [1], SciBERT is adept at understanding technical and scientific language. It outputs embeddings in a 768-dimensional space and is well-suited for tasks requiring comprehension of scholarly and factual information.
In this study, SciBERT was utilized to classify climate change claims by leveraging its specialization in scientific text, particularly for claims supported by abstracts from scientific literature.
For fine-tuning, a classification head was added to SciBERT to predict labels corresponding to the claims. The input to the model consisted of claims and abstracts tokenized together which allowed the model to leverage context from both inputs, improving its decision-making ability for classification.

3) *DeBERTa*
DeBERTa (Decoding-enhanced BERT with disentangled attention) is a transformer-based model known for its unique disentangled attention mechanism and enhanced position encodings. These features allow it to model semantic relationships in text more effectively than traditional BERT-based models. DeBERTa has demonstrated state-of-the-art performance across multiple natural language understanding benchmarks.
In this project, we employed the open-source DeBERTa-v3-base model, fine-tuning it for the specific task of classifying climate change claims into 'supports,' 'refutes,' or 'not enough info.'
The fine-tuning process involved adding a custom classification head to the model, enabling it to process both claims and their corresponding abstracts simultaneously. To handle claims and abstracts effectively, both inputs were tokenized separately with a maximum sequence length of 256 tokens each. The concatenation of their hidden state outputs allowed the model to leverage information from both sources for improved classification performance.

IV. Experiments

1) *Model Workflow:*
The model works by first passing the claim and abstract corpus to a retrieval model (FAISS), which extracts the top k relevant abstracts for each claim. These top k abstracts are then combined with the claim and fed into the deep learning models (LLaMA, SciBERT, DeBERTa) for classification. The final output classifies the claim as supports, refutes, or not enough info. A detailed workflow of this process is illustrated in figure 1.

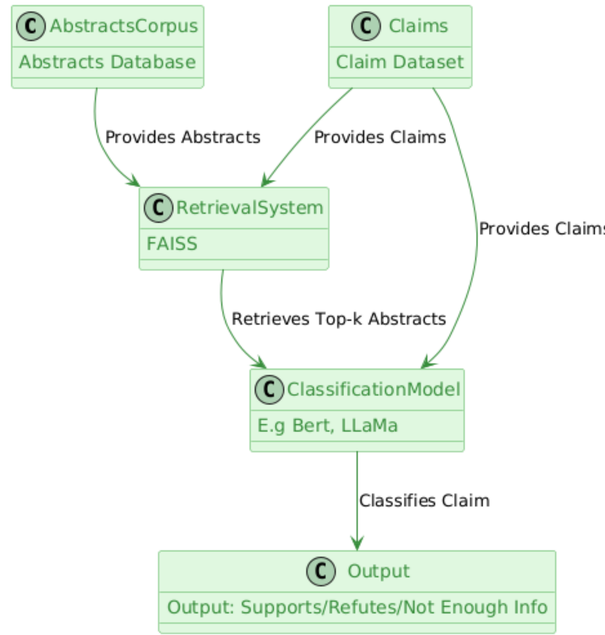

Fig. 1. Model Workflow

2) *Retrieval*
The retrieval process leverages the *FAISS* (Facebook AI Similarity Search) library to identify the top k abstracts most relevant to the claim. These relevant abstracts are essential for providing context to the claim, enabling the deep learning models to make more accurate predictions. Different values of k (number of top abstracts to retrieve) were tested: k = 1, k = 2, and k = 3. Based on the evaluation results, the best performance was achieved with k = 1, so this value was selected for the subsequent experiments. This means that for each claim, only the most relevant abstract is retrieved and passed to the deep learning models for classification.

3) *Data Augmentation*
To enhance the dataset and improve model performance, we employed data augmentation using the *Gemini API*, specifically focusing on *Synonym Replacement (SR)*. In the SR method, synonyms were substituted for selected words in the claims while preserving their contextual meaning. This technique introduced variation in the textual data, helping the models generalize better and mitigate overfitting.
The augmented claims were added to the original dataset to expand the training data. The breakdown of the augemneted claims dataset is shown in table III. This process was particularly useful in handling the relatively small size of the dataset, ensuring that the models were exposed to a broader range of claim representations while maintaining semantic consistency.

Category	*Label*	*Number of Claims*
Supports	0	686
Refutes	1	439
Not enough information	2	412
Total	-	**1537**

TABLE III
CLAIMS CORPUS BREAKDOWN

4) *Computational Resources*
Model training was conducted on *Kaggle Kernels*, which offers 29GB RAM and two NVIDIA Tesla T4 GPUs with 15GB memory each. The environment provided the necessary computational power for training large-scale models of LLaMa 3.2 as well.

5) *Hyperparameters for Deep Learning Models*
Two distinct sets of hyperparameters were employed for the Deep Learning models: one for LLaMA 3.2 and another for SciBERT and DeBERTa. This distinction was necessary due to the larger size and computational demands of the LLaMA model, which required adjustments to the training epochs and other parameters to optimize performance within the available computational resources.
The hyperparameters for LLaMA 3.2 are detailed in Table IV, while those for SciBERT and DeBERTa are presented in Table V.

Hyperparameter	*Value*
Epochs	20
Batch Size	4
Learning Rate	2e-5
Optimizer	AdamW
Loss Function	Cross-Entropy Loss
Weight Decay	0.01
Train-test Split	0.2

TABLE IV
HYPERPARAMETERS FOR LLAMA 3.2 1B AND 3B

Hyperparameter	*Value*
Epochs	100
Batch Size	4
Learning Rate	1e-5
Optimizer	Adam
Loss Function	Sparse Categorical Cross entropy
Train-test Split	0.2

TABLE V
HYPERPARAMETERS FOR SCIBERT AND DEBERTA

6) *Evaluation Metric*
The evaluation criterion which was used for assessing model performance was accuracy, calculated as follows:

$$\text{Accuracy} = \frac{\text{Total Number of Correct Predictions}}{\text{Total Number of Samples}}$$

V. RESULTS

Our experiments involved training both LLaMa 3.2 and BERT-based models on the claims dataset along with top-k abstracts to classify claims into the three categories: Supports, Refutes, and Not Enough Information. The evaluation metric used was accuracy. The results of our experiments are summarized below.

1) *Initial Experiments*
The initial experiments were conducted on the original claims dataset. Using the LLaMa 3.2 1B model resulted in an accuracy of 59.01% and similarly, when SciBERT and DeBERTa models were tested on the same dataset, their performance was also suboptimal, with accuracies ranging between 55% and 60%. The trends in validation accuracy over epochs and the confusion matrix for the LLaMa 3.2 1B model are presented in Figures 2 and 3. These preliminary results indicated the necessity for improvements, leading to the implementation of data augmentation, as described in section V, to enhance model performance in subsequent experiments.

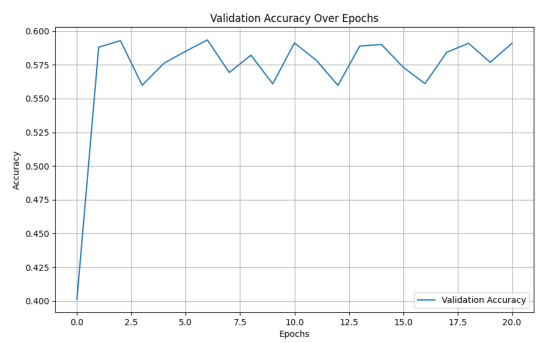

Fig. 2. Plot of Validation Accuracy vs Epochs for LLaMa 3.2 1B on original claims dataset

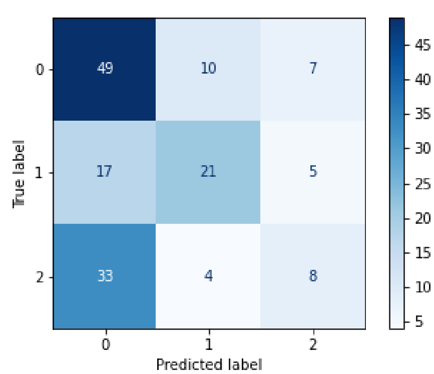

Fig. 3. Confusion Matrix for LLaMa 3.2 1B on original claims dataset

2) *Deep Learning Model Results*

After applying data augmentation to the original claims dataset, we used the augmented dataset for further experimentation with deep learning models. The models tested include LLaMa 3.2 1B, LLaMa 3.2 3B, SciBERT, and DeBERTa. The results presented below show the performance of each model using the augmented dataset, demonstrating significant improvements in accuracy compared to the initial experiments.

a) *LLaMa 3.2 1B*
The LLaMa 3.2 1B model, when trained on the augmented dataset, showed a significant performance boost, with accuracy increasing to 80.19%, up from 59.01% on the original dataset. The plot of validation accuracy versus epochs for this model (Figure 4) illustrates a steady improvement in model learning. The confusion matrix for this model (Figure 5) also shows a better classification performance across the three claim categories.

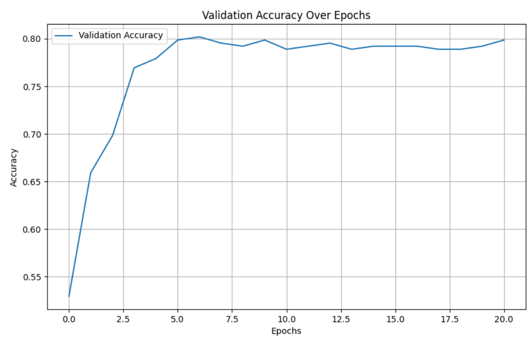

Fig. 4. Plot of Validation Accuracy vs Epochs for LLaMa 3.2 1B on augmented claims dataset

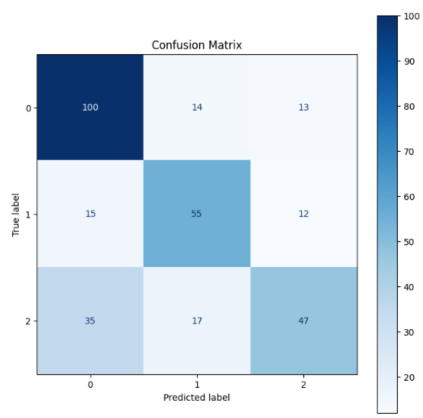

Fig. 5. Confusion Matrix for LLaMa 3.2 1B on augmented claims dataset

b) *LLaMa 3.2 3B*
The larger LLaMa 3.2 3B model achieved the second-highest accuracy of 83.44% with the augmented dataset. As seen in the accuracy vs. epochs plot (Figure 6), the model demonstrated stable learning, showing a higher performance compared to the 1B variant. The confusion matrix in Figure 7 highlights a further refinement in the model's predictions across the categories.

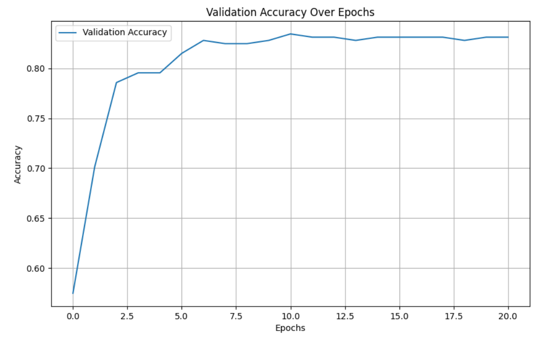

Fig. 6. Plot of Validation Accuracy vs Epochs for LLaMa 3.2 3B on augmented claims dataset

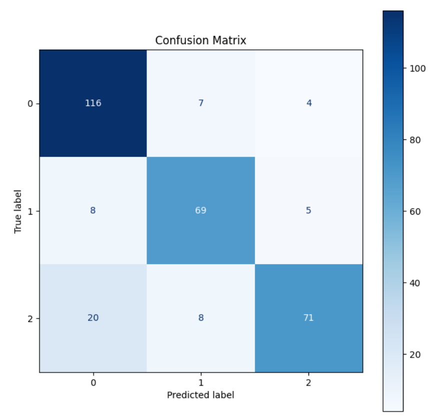

Fig. 7. Confusion Matrix for LLaMa 3.2 3B on augmented claims dataset

c) *SciBERT*
SciBERT, a transformer-based model tailored for scientific text, reached an accuracy of 81.17% with the augmented dataset. The model performed well, with the accuracy vs. epochs plot (Figure 8) showing steady improvement. The confusion matrix in Figure 9 demonstrates that SciBERT was able to distinguish between the claim categories with reasonable accuracy, though it lagged behind LLaMa 3.2 3B.

d) *DeBERTa*
DeBERTa achieved the highest accuracy of 89.29% when trained on the augmented dataset. This performance was reflected in the accuracy vs. epochs

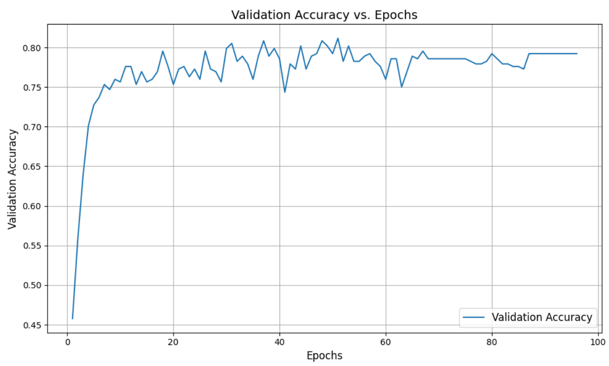

Fig. 8. Plot of Validation Accuracy vs Epochs for SciBERT on augmented claims dataset

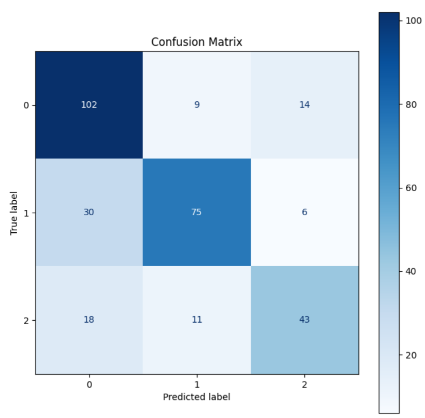

Fig. 9. Confusion Matrix for SciBERT on augmented claims dataset

plot 10, which shows a steady trend in validation accuracy throughout the training. The confusion matrix in 11 reveals that DeBERTa had excellent classification performance across all three categories, particularly when distinguishing between Supports and Refutes.

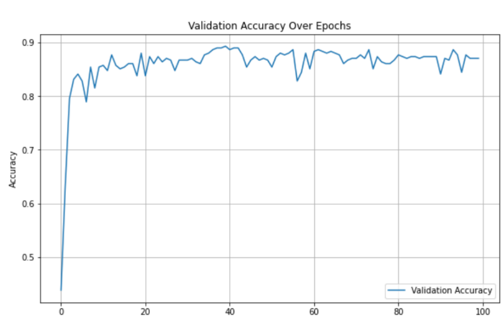

Fig. 10. Plot of Validation Accuracy vs Epochs for DeBERTa on augmented claims dataset

3) *Analysis of Results*

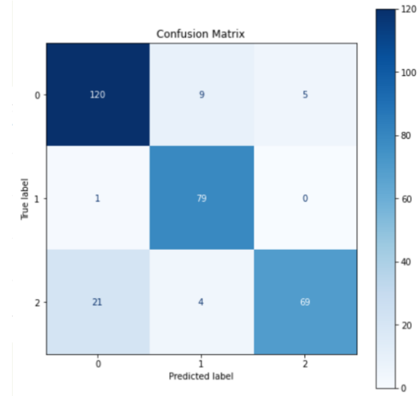

Fig. 11. Confusion Matrix for DeBERTa on augmented claims dataset

In this section, we analyze the results of the models after applying augmentation to the dataset, focusing on model performance, the impact of data augmentation, and the training trends observed during experimentation.

a) *Model Performance*
The DeBERTa model achieved the highest accuracy at 89.29%, outperforming all other models, followed by LLaMa 3.2 3B with an accuracy of 83.44%. SciBERT and LLaMa 3.2 1B also showed significant improvement with the augmented dataset, reaching 81.17% and 80.19% accuracy, respectively. These results are summarised in table VI

Model	*Accuracy*	*Dataset Type*
LLaMa 3.2 1B	0.5901	Original Claims Dataset
LLaMa 3.2 1B	0.8019	Augmented Claims Dataset
LLaMa 3.2 3B	0.8344	Augmented Claims Dataset
SciBERT	0.8117	Augmented Claims Dataset
DeBERTa	0.8929	Augmented Claims Dataset

TABLE VI
MODEL PERFORMANCE ON ORIGINAL AND AUGMENTED CLAIMS DATASET

b) *Effect of Data Augmentation*
The use of the augmented dataset led to significant performance improvements across all models, with accuracy increasing from the 55%-60% range to 80%-90%, highlighting the effectiveness of the data augmentation process in enhancing model performance.

c) *Training Trends*
The training plots of validation accuracy versus epochs and the confusion matrices for each model illustrate the improvement in learning after data augmentation. For instance, the LLaMa 3.2 1B model shows a steady increase in validation accuracy, as seen in Figure 4, while the confusion matrix in Figure 5 highlights better class-wise pre-

dictions. The same trend is seen for LLaMa 3.2 3B, SciBERT, and DeBERTa, with each model showing improvements in accuracy and class balance after data augmentation.

VI. Limitations and Future Work

While this study aims to fact-check claims about climate change, certain limitations restrict the scope and performance of the model. Below, we discuss these limitations and propose future directions to address them.

A. Limited Scope of Climate Issues

Climate change encompasses a vast array of issues; however, this work focuses on four of the most common topics. Due to resource and time constraints, we were unable to include emerging climate concerns such as ocean acidification, biodiversity loss, and microplastic pollution.

Future Work: Expanding the dataset to include a broader spectrum of climate-related topics will improve the model's relevance and applicability. Collaborating with domain experts to curate an extensive list of critical issues could ensure comprehensive coverage.

B. Dataset Constraints

Since no publicly available dataset of scientific abstracts specific to climate change was found, we curated a dataset from Google Scholar, limiting the collection to the first 20 search pages. This restriction likely introduced limited diversity in the data. Additionally, relying solely on Google Scholar which excluded other valuable academic repositories.

Future Work: Future efforts can involve incorporating datasets from academic repositories like JSTOR to enhance the variety and depth of the dataset. Data from other platforms like IPCC (Intergovernmental Panel on Climate Change) reports could facilitate the creation of a richer, more representative dataset.

C. Data Augmentation Limitations

To address the scarcity of labeled data, we used synonym-based data augmentation to expand the claims dataset. While this method increased the dataset size, it provided limited semantic variation. As a result, the model's ability to generalize to new claims remained constrained.

Future Work: Advanced data augmentation techniques, such as leveraging large language models (LLMs), can generate more diverse claims while preserving the underlying meaning. Models like GPT or Llama can be fine-tuned on the curated dataset to synthesize new claims and counterclaims, enriching the training data and enhancing model robustness.

By addressing these limitations, future iterations of this project can improve the accuracy, robustness, and generalizability of the model, making it a more reliable tool for fact-checking claims related to climate change.

References

[1] Iz Beltagy, Kyle Lo, and Arman Cohan. Scibert: A pretrained language model for scientific text, 2019.

[2] Eun Cheol Choi and Emilio Ferrara. Fact-gpt: Fact-checking augmentation via claim matching with llms, 2024.

[3] Timon Diggelmann, Jordan Boyd-Graber, Jannis Bulian, Massimiliano Ciaramita, and Markus Leippold. Climate-fever: A dataset for verification of real-world climate claims. *arXiv preprint arXiv:2012.00614*, 2020.

[4] M. Haggag, A.S. Siam, W. El-Dakhakhni, et al. A deep learning model for predicting climate-induced disasters. *Natural Hazards*, 107:1009–1034, 2021.

[5] Georgi Karadzhov, Preslav Nakov, Lluis Marquez, Alberto Barron-Cedeno, and Ivan Koychev. Fully automated fact checking using external sources, 2017.

[6] Markus Leippold, Saeid Ashraf Vaghefi, Dominik Stammbach, Veruska Muccione, Julia Bingler, Jingwei Ni, Chiara Colesanti-Senni, Tobias Wekhof, Tobias Schimanski, Glen Gostlow, Tingyu Yu, Juerg Luterbacher, and Christian Huggel. Automated fact-checking of climate change claims with large language models, 2024.

[7] Saeid Vaghefi, Veruska Muccione, Christian Huggel, Hamed Khashehchi, and Markus Leippold. Deep climate change: A dataset and adaptive domain pre-trained language models for climate change related tasks. In *NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning*, 2022.

[8] Gengyu Wang, Lawrence Chillrud, and Kathleen McKeown. Evidence based automatic fact-checking for climate change misinformation. In *ICWSM Workshops*, 2021.