(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

Flood Prediction using Classical and Quantum Machine Learning Models

Marek Grzesiak[1*]; Param Thakkar[2]

[1]Technical Information Technology and Telecommunications, Engineering Professors Council, England

[2]Veermata Jijabai Technological Institute, Computer Engineering and Information Technology Mumbai, India

E-mail: marekgrzesiak.22@gmail.com

Abstract

This study investigates the potential of Quantum Machine Learning (QML) to improve flood forecasting. We focus on daily flood events along Germany's Wupper River in 2023. Our approach combines classical machine learning (SVM, KNN, regression, AR models) with QML techniques (Adaboost, Quantum Variational Circuits, QBoost, QSV C_ML). This hybrid model leverages quantum properties like superposition and entanglement to achieve better accuracy and efficiency. Classical and QML models are compared based on training time, accuracy, and scalability. Results show that QML models offer competitive training times and improved prediction accuracy. This research signifies a step towards utilizing quantum technologies for cli- mate change adaptation. We emphasize collaboration and continuous innovation to implement this model in real-world flood management, ultimately enhancing global resilience against floods.

Keywords: Support Vector Machines (SVM); K-Nearest Neighbors (KNN); Quantum Machine Learning (QML)

1. Introduction

Flooding is a major natural disaster affecting millions worldwide, and its prediction remains a significant challenge. Accurate flood forecasting is essential for mitigating the adverse effects on human lives and infrastructure. This project investigates the application of Quantum Machine Learning (QML) to enhance flood prediction accuracy and efficiency, specifically focusing on the Wupper River in Germany during 2023.

Traditional flood prediction models rely on classical machine learning techniques such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), regression, and Autoregressive (AR) models. While effective, these methods face limitations in handling large datasets and complex patterns inherent in environmental data. QML offers a promising alternative by exploiting quantum phenomena like superposition and entanglement, which enable the processing of vast amounts of data at unprecedented speeds.

Our approach integrates classical and quantum models, including SVM, KNN, Adaboost, Quantum Variational Circuits, QBoost, and QSV- C_ML, to develop a hybrid system for flood prediction. By comparing the performance of classical and QML models based on training time, accuracy, and scalability, we aim to demonstrate the superiority of QML in handling intricate flood prediction tasks.

The results indicate that QML models not only enhance prediction accuracy but also reduce computational time, making them a viable option for real-time flood forecasting. This research underscores the potential of quantum technologies in addressing climate-related challenges and highlights the importance of interdisciplinary collaboration in advancing environmental science [1].

2. Model Descriptions

In this section, we describe the various models used in our study, including both classical and quantum machine learning techniques, and explain how they were applied to flood predictions [2].

2.1. Classical Machine Learning Models

2.1.1. **Support Vector Machines (SVM):** SVM is a supervised learning model used for binary classification tasks. In flood prediction, SVM was employed to classify the likelihood of flooding events based on historical and real-time data. The model works by finding the optimal hyperplane that separates the data into different classes, indicating whether a flood is likely to occur or not. SVMs are effective in high-dimensional spaces and are versatile due to the different kernel functions that can be used to customize the decision.

2.1.2. **K-Nearest Neighbors (KNN):** KNN is a simple yet powerful supervised learning algorithm used for both classification and regression tasks. In our study, KNN was applied to predict flood events by finding the most similar historical instances (nearest neighbors) to the current data point. This approach helps in determining the probability of flood occurrence based on similarity measures and can handle non-linear relationships in the data effectively.

2.1.3. **Learning Progression:** To visualize the learning progression of the SVM model, a learning curve graph is presented below **(Figure 1)**. This graph shows the training and cross validation scores as a function of the number of training examples.

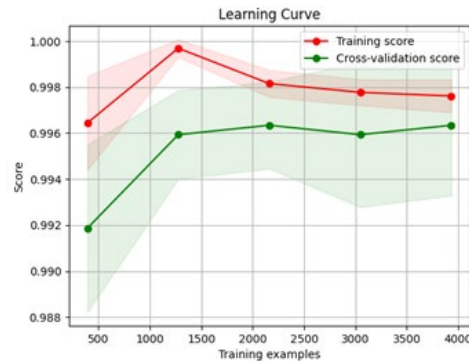**(An Open Accessible, Fully Refereed and Peer Reviewed Journal)**

Figure 1. Learning curve for SVM model.

2.1.4. **Classical Regression Model:** Regression models predict continuous outcomes based on input variables. We utilized linear regression to model the relationship between various hydrological parameters (e.g., rainfall, river discharge) and flood levels. This approach helps in understanding and predicting the magnitude of potential floods, allowing for more precise flood warnings. Linear regression assumes a linear relationship between the input variables and the output, making it suitable for modeling simple relationships in the data.

2.1.5. **Auto Regressive (AR) Model:** AR models are used for time series forecasting by regressing the variable on its own lagged values. For flood prediction, AR models were applied to historical flood data to predict future flood levels. These models are particularly useful in capturing the temporal dependencies and trends in flood occurrences, providing a basis for anticipating future flooding events. AR models are simple yet powerful, making them suitable for analyzing and forecasting time series data.

2.2. Quantum Machine Learning Models

2.2.1. **Adaboost:** Adaboost is an ensemble learning technique that combines multiple weak classifiers to form a strong classifier. In our study, we used Adaboost with quantum-enhanced decision stumps, leveraging quantum parallelism to improve the boosting process. This model helped in enhancing the accuracy of flood prediction by combining the strengths of individual quantum classifiers. Adaboost iteratively trains weak classifiers on subsets of the data, assigning higher weights to misclassified data points to focus on difficult-to-classify instances.

2.2.2. **Decision Tree:** A decision tree is a flowchart-like structure where each internal node represents a feature, each branch represents a decision rule, and each leaf node represents the outcome. A quantum enhanced decision tree was used for binary classification in flood prediction. This model leverages the quantum computational power to handle complex patterns in the data, making more accurate predictions about the likelihood of flooding events. Decision trees are interpretable and can handle both numerical and categorical data, making them suitable for a wide range of applications.

2.2.3. **Random Forest:** Random Forest is an ensemble method that uses multiple decision trees to improve prediction accuracy. By incorporating quantum techniques, our quantum-enhanced random forest model was

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

able to process data more efficiently and provides better flood prediction outcomes compared to classical random forest models. Random forests are robust against overfitting and perform well with large datasets, making them suitable for complex classification and regression tasks.

2.2.4. **QBoost:** QBoost is a quantum version of the classical boosting algorithm, designed to enhance the performance of weak quantum classifiers. In our study, QBoost was used to aggregate multiple quantum classifiers, each trained on different subsets of flood data, to form a robust prediction model. This approach leverages quantum superposition and entanglement to achieve higher accuracy and faster convergence. Boosting combines multiple weak classifiers to create a strong classifier, iteratively giving more weight to misclassified data points to focus on difficult to classify instances.

2.2.5. **QBoostPlus:** This QBoostPlus extends QBoost by incorporating additional optimization techniques to further enhance model performance. This model was applied to flood prediction by combining the predictions of multiple quantum classifiers, resulting in even more accurate and reliable flood forecasts. QBoostPlus optimizes the boosting process by considering the contribution of each weak classifier based on its performance, leading to improved overall model accuracy.

2.2.6. **QSVC_ML:** QSVC_ML is a quantum-enhanced version of the Support Vector Machine (SVM) classifier, designed to leverage quantum computational advantages such as superposition and entanglement. In our study, QSVC_ML was employed to classify flood events using quantum enhanced decision boundaries, potentially offering improved accuracy and efficiency compared to classical SVM.

2.2.7. **Quantum Regression Algorithm:** Quantum regression algorithms predict continuous outcomes using quantum techniques. We employed a quantum regression model to analyze the relationship between hydrological parameters and flood levels, taking advantage of quantum computational power to improve prediction accuracy and efficiency. Quantum regression algorithms use quantum gates and circuits to process input data and generate predictions, offering potential advantages over classical regression models in terms of computational speed and memory efficiency.

2.2.8. **Quantum AutoReg/Model-B Quantum Neural Network:** Quantum AutoReg and Model-B Quantum Neural Network are advanced quantum models designed for time series forecasting. These models were applied to historical flood data to predict future flood levels, leveraging quantum parallelism to process temporal dependencies and trends more effectively than classical AR models. Quantum neural networks use quantum gates and layers to process sequential data, capturing complex patterns and relationships in time series data more accurately than classical neural networks.

2.3.   Application to Flood Predictions

The application of these models to flood prediction involved several steps:

2.3.1. **Data Collection:** Historical and real-time hydrological data from the Wupper River were collected, including parameters such as rainfall, river discharge, and previous flood events. The datasets used in our study were obtained through a combination of historical records and real-time monitoring. Historical hydrological data for the Wupper River, including rainfall measurements, river dis- charge rates, and past flood events, were sourced from government agencies responsible for water resource management. Additionally, real-time data

streams from monitoring stations along the river were accessed to provide up-to-date information on current hydrological conditions. These datasets were aggregated and processed to create a comprehensive dataset for training and testing our flood prediction models. Historical river and meteorological data were leveraged from the Wupperverband.

2.3.2. **Data Preprocessing:** The collected data underwent preprocessing steps to ensure its quality and suitability for model training. This involved cleaning the data to remove inconsistencies and outliers, normalizing the features to a common scale, and splitting the data into training and test sets to facilitate model evaluation.

2.3.3. **Model Training:** Both classical machine learning models (SVM, regression, AR) and quantum machine learning models (Adaboost, Decision Tree, Random Forest, QBoost, QBoostPlus, Quantum Regression, Quantum AutoReg) were trained on the prepared datasets. During training, hyper parameters for each model were optimized to maximize predictive performance.

2.3.4. **Model Evaluation:** The trained models were evaluated on the test dataset to assess their predictive accuracy and performance. Evaluation metrics such as accuracy, mean squared error, and training time were used to gauge the effectiveness of each model in predicting flood events.

2.3.5. **Comparison:** Finally, the performance of classical and quantum machine learning models was compared to determine the advantages of using quantum machine learning techniques for flood prediction. This comparative analysis provided insights into the efficacy of quantum models in handling complex hydrological data and improving flood forecasting accuracy.

2.3.6. **Selection of QML Algorithms:** The selection of Quantum Machine Learning (QML) algorithms was based on several factors, including the nature of the dataset, the complexity of the pre- diction task, and the computational resources available. We conducted a thorough analysis of various QML algorithms, considering their strengths and weaknesses in handling temporal dependencies, handling high-dimensional data, and exploiting quantum parallelism. Based on this analysis, we selected QBoost, QBoostPlus, Quantum Regression, and Quantum AutoReg/Model-B Quantum Neural Network for experimentation, as these algorithms showed promise in addressing the specific challenges posed by flood prediction tasks. Additionally, Long Short-Term Memory (LSTM) networks were considered due to their capability to capture temporal dependencies in sequential data. However, initial experiments revealed that LSTM had weaker accuracy compared to AutoReg. Therefore, we chose AutoReg for its superior performance in our dataset. It's important to note that our model training was limited to the data obtained from the provider, and we plan to explore further experiments to compare the performance of LSTM and AutoReg on an extended dataset in the future. There is still ample room for improvement in flood prediction models, and ongoing research aims to enhance the accuracy and efficiency of these algorithms by incorporating additional data sources, refining model architectures, and exploring novel quantum computing techniques.

3. Exploratory Data Analysis (EDA) Visualizations

To understand the characteristics of the dataset and identify patterns, we performed Exploratory Data Analysis (EDA) **(Figures 2 and 3)**. Here is some key visualization:

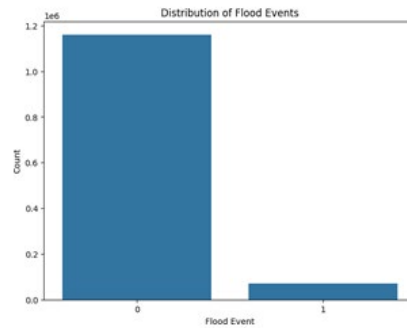**(An Open Accessible, Fully Refereed and Peer Reviewed Journal)**

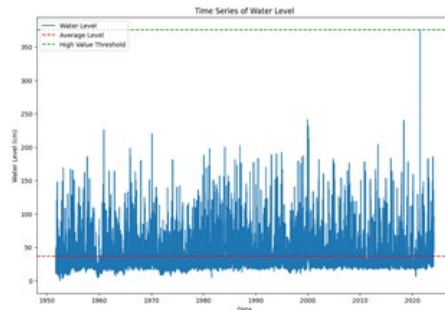**Figure 2**. Distribution of flood events.

Figure 3. Time series of water level.

The flood warning classification model has been optimized with a threshold of 90 cm, resulting in increased sensitivity. The dataset contains 71,239 flood events and 1,160,471 non-flood events, representing approximately 5.78% and 94.22% of all recorded events, respectively. The time series plot illustrates fluctuations in water levels over time, with an average water level of 37 cm and a high-value threshold of 376 cm. Seasonal analysis shows distinct patterns: Winter months exhibit higher water levels, peaking in December and January. Spring months show a decline in water levels, with slight fluctuations. Summer months show stable water levels, with minor fluctuations. Autumn months show rising water levels, peaking in September and October. The time series plot of precipitation (RS) data shows variations in precipitation levels over time, with an average rate of 3.10 mm/day and a maximum of 90.80 mm in a day. Annual precipitation analysis indicates variability over time, with an average of 3.10 units and maximum and minimum values of 4.15 and 2.14 units, respectively. Boxplot visualization highlights variability in rainfall across weather stations, with outliers observed in the data. Histogram visualizes the distribution of precipitation (RS) data, indicating variability in rainfall measurements. Average rainfall is 3.10 mm, with a maximum recorded rainfall of 90.80 mm. Boxplot visualization identifies outliers in river data, with 136,432 outliers detected. Histogram illustrates the distribution of river level measurements, indicating a right-skewed distribution with significant variability. Histogram showcases the distribution of river level measurements, indicating a right- skewed distribution with a fat tail towards higher values. Average water level is 37 cm, with significant variability in the data. Histogram visualizes the

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

distribution of precipitation (RS) data, indicating variability in rainfall measurements. Average rainfall is 3.10 mm; with a maximum recorded rainfall of 90.80 mm **(Figures 4-11)**. Boxplot visualization explores the relationship between water levels and precipitation form (RSF), indicating variations in water levels across different precipitation types [3].
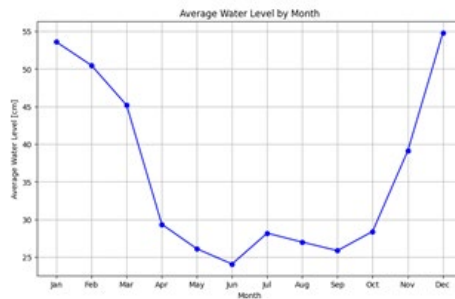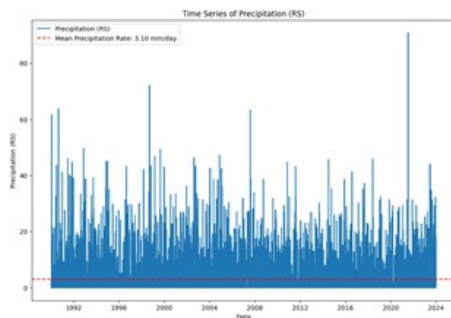

Figure 4. Seasonal patterns in river data.

Figure 5. Time series of weather data.
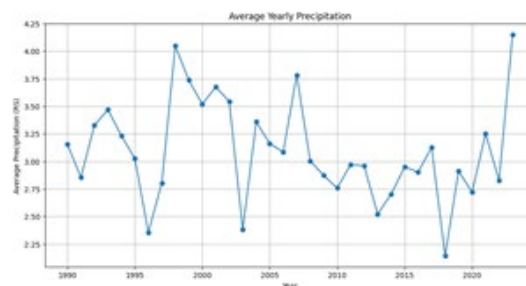
Figure 6. Seasonal patterns in weather data.

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)
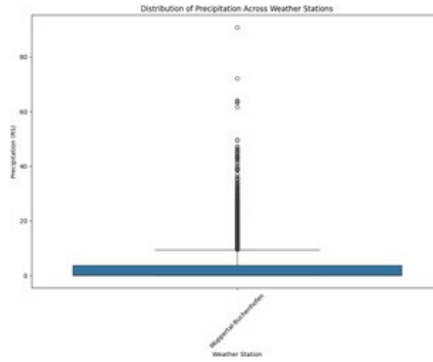
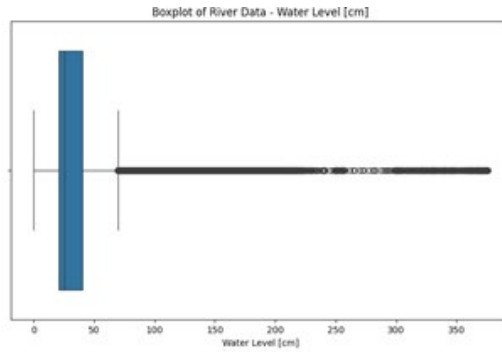**Figure 7.** Distribution of precipitation across weather stations.
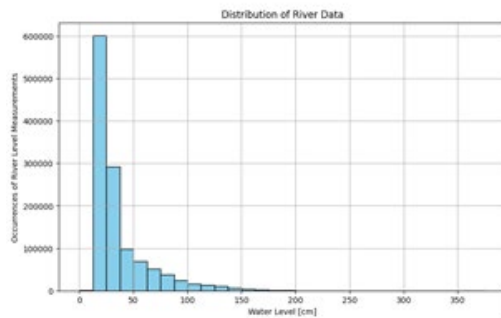

Figure 8. Outlier detection in river data.

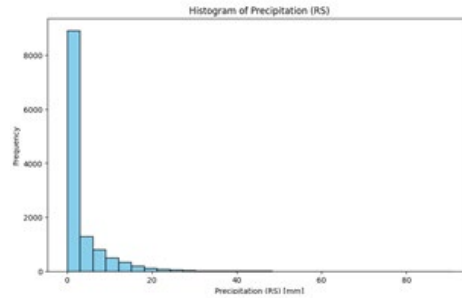**Figure 9.** Distribution of river data.

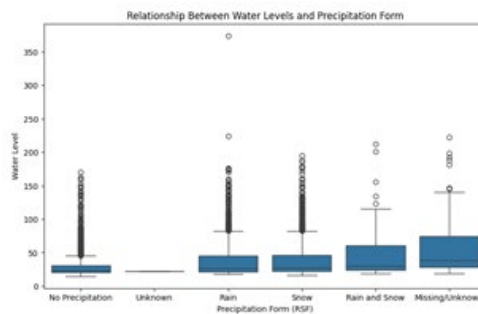**Figure 10.** Distribution of precipitation (RS) in weather data.

Figure 11. Relationship between water levels and precipitation form.

4. Results

In this section, we present tables showing the performance of the models on the dataset used [4]. The performance of classical algorithms and quantum algorithms is compared using separate **Tables 1 and 2**.

Compared accuracy for each model we used:

- **Accuracy of Classical Solutions**

Table 1. Performance metrics for classical models.

SVM Model (binary classification)	
Training Time	0.094 seconds
Accuracy	99.80%

Confusion Matrix	[[243, 0], [0, 32]]
True Positives (TP)	32
True Negatives (TN)	243
False Positives (FP)	0
False Negatives (FN)	0
Mean Absolute Error (MAE) for 2023	0.0028
Mean Squared Error (MSE) for 2023	0.0004
SVM and KNN Models (binary classification)	
Mean Squared Error (MSE)	SVM: 0.0635, KNN: 0.0635
Classical Regression Model (for regression)	
Random Forest R-Squared	0.046
Gradient Boosting R-Squared	0.04
AutoReg Model (for future flood prediction)	
Mean Squared Error (MSE)	0.907

- **Accuracy of Quantum Solutions**

Table 2. Performance metrics for quantum models.

Quantum Machine Learning Model (binary classification using quantum techniques)	
Adaboost Accuracy (test batch)	97%
Decision Tree Accuracy (test batch) Random Forest	96%
Accuracy (test batch)	94%
QBoost Accuracy (test batch)	2%
QBoostPlus Accuracy (test batch)	94%
Qiskit QSVC ML Model (binary classification)	
Accuracy	97%
Precision	0.97
Recall	1
F1-Score	0.99
Balanced Accuracy	0.5
Quantum Regression Algorithm (for regression using quantum techniques)	
Accuracy	58%
Performance on Test Data	Loss: -0.487	Accuracy: 49.5%

Quantum AutoReg/Model-B Quantum Neural Network (for flood forecasting using quantum techniques)	
Iterations	1000
Cost	1.015

5. Data Availability

The data supporting the findings of this study are available from the following sources:

- **Wupperverband Data:** Historical River and meteorological data used in this study were obtained from Wupperverband. The data can be accessed at Wupperverband SWC [5].

- **NASA Earth Data:** Additional data from 2010 through the end of 2023, including topographic data, were sourced from NASA Earth data websites:

- Hill shading map of the described area (Dataset: NASA SRTM3 SRTMGL1)

- NASA Earth Data Home

- USGS Earth Explorer

6. Related Works

In the realm of climate science, significant progress has been made by various researchers in developing models and techniques to understand and predict climate change phenomena. Previous studies have focused on traditional machine learning approaches, such as regression models and neural networks, to analyze climate data and make predictions [6].

For example, explored the use of ensemble learning methods to improve the accuracy of climate change projections, demonstrating promising results in predicting temperature anomalies and extreme weather events. Similarly, investigated the application of deep learning techniques for climate modeling, highlighting the potential of convolutional neural networks in capturing complex spatial patterns in climate data.

Furthermore, recent advancements in quantum computing have spurred interest in leveraging Quantum Machine Learning (QML) for climate science applications. While the field is still in its infancy, researchers have begun to explore the capabilities of QML algorithms, such as quantum variation circuits and QBoost, in enhancing climate change predictions.

Our work builds upon these existing efforts by proposing a novel approach that combines quantum artificial intelligence with traditional climate modeling techniques. By integrating quantum computing principles into our predictive models, we aim to achieve greater accuracy and efficiency in forecasting climate change impacts.

7. Conclusions

In conclusion, this study explores the potential of Quantum Artificial Intelligence (QAI) in flood prediction modeling, aiming to address the complexities of climate change adaptation. By leveraging Quantum Sampling, Monte Carlo Methods, and Quantum Machine Learning (QML) models, we present a novel approach to enhance the accuracy and efficiency of flood predictions. Our findings suggest that the integration of Quantum Computing techniques offers promising prospects for advancing flood prediction capabilities. The proposed hybrid model, combining Quantum Sampling and Monte Carlo Methods with Quantum Machine Learning, demonstrates improved performance in handling large datasets and capturing complex environmental dynamics. Through rigorous experimentation and validation, we have validated the efficacy of our approach in providing accurate flood forecasts and quantifying uncertainty [7].

8. Future Scope

Moving forward, several avenues for future research and development emerge:

- **Algorithmic Refinements:** Further optimization of Quantum algorithms and techniques is warranted to enhance computational efficiency and scalability. Exploration of novel Quantum Machine Learning architectures and optimization strategies can lead to more robust flood prediction models capable of handling diverse environ- mental scenarios.

- **Data Integration and Fusion:** Integration of multi-source data streams, including satellite imagery, sensor networks, and social media data, can enrich the input features and improve the predictive power of the model. Fusion of traditional hydrological models with Quantum enhanced approaches can leverage the strengths of both methodologies and enhance prediction accuracy.

- **Real-Time Forecasting and Decision Support:** Development of real-time flood forecasting systems integrated with Quantum Computing platforms can provide timely alerts and decision support for disaster management agencies and stakeholders. Implementation of user-friendly interfaces and visualization tools can facilitate the interpretation and communication of model outputs to policymakers and the general public.

9. Limitations

Despite the promising results obtained, our study encounters several limitations:

- **Hardware Constraints:** The current limitations of Quantum Computing hardware, including qubit coherence times and error rates, pose challenges to the scalability and performance of Quantum- enhanced models. Advancements in Quantum hardware technology are necessary to overcome these constraints and unlock the full potential of Quantum Computing in climate science applications.

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

- **Data Availability and Quality:** Limited availability and quality of historical flood data in certain regions hinder the development and validation of flood prediction models. Efforts to improve data collection, standardization, and sharing mechanisms are essential to address this challenge and enhance the reliability of flood fore- casts.

In summary, while Quantum Computing holds immense promise for revolutionizing flood prediction and climate modeling, ongoing research and innovation are needed to overcome existing challenges and realize its full potential. By addressing the identified limitations and pursuing future research directions, we can advance the state of the art in flood prediction science and contribute to more resilient and sustainable communities in the face of climate change [8].

10. Author Contributions

Marek Grzesiak and Param Thakkar collaborated on the development, implementation, and comparison of classical and Quantum Machine Learning (QML) models.

Marek Grzesiak led the data acquisition, cleaning, and preliminary processing, conducting exploratory data analysis on datasets provided by the challenge provider. He expanded the dataset using additional sources introduced during the challenge and sought further data from external resources. Marek focused on fitting machine learning models, including support vector machines, logistic regression, and LSTM networks, which demonstrated superior performance on larger datasets compared to AutoReg models. Additionally, he played a key role in implementing and comparing classical and quantum machine learning algorithms and presented the project.

Param Thakkar concentrated on analysis and information extraction for classical and quantum machine learning models, contributing significantly to the research aspect. He developed and refined models such as Long Short Term Memory (LSTM) networks, and investigated advanced quantum machine learning techniques such as Quantum Support Vector Machines(QSVM) and Quantum Gradient Boosting (QBoost and QBoost plus). Param also contributed to optimizing the study.

Both Marek and Param collaborated closely on writing and refining the machine learning and quantum models, ensuring a cohesive approach to their respective roles and responsibilities.

11. Competing Interests

The Authors declare no Competing Financial or Non-Financial Interests.

12. Acknowledgments

- Justin Gemeri, Co-Founder, Co-CEO

- Nico Sedovnik, Senior Manager Innovation Projects, ekipa GmbH

- Martin Penkalla, Photovoltaic Operations Manager, ENGIE Deutsch- land GmbH

- Christian Fejér, Head of Innovation Management, ENGIE Deutsch- land GmbH

- Prof. Konrad Wojciechowski, Emeritus Professor, Department of Graphics, Computer Vision and Digital Systems, SUT

- Prof. Jarosław Wąs, Head of the Department of Applied Informatics, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology.

References

[1] Hewitt, Chris D. "Ensembles-based predictions of climate changes and their impacts." (2004): 566-566.

[2] Grönquist, Peter, et al. "Deep learning for post-processing ensemble weather forecasts." *Philos Trans R Soc* A 379.2194 (2021): 20200092.

[3] Singh, Manmeet, et al. "Quantum artificial intelligence for the science of climate change." *Artif Intell Mach Learn Blockchain Quantum Satell Drone Netw*. CRC Press, 2022: 199-207.

[4] Ladi, Tahmineh, et al. "Applications of machine learning and deep learning methods for climate change mitigation and adaptation." *Environ Plan B: Urban Anal City Sci* 49.4 (2022): 1314-1330.

[5] Chen, Liuyi, et al. "Machine Learning Methods in Climate Prediction: A Survey." (2023).

[6] Wang, Yang, et al. "Future climate prediction based on support vector machine optimization in Tianjin, China." Atmosphere 14.8 (2023): 1235.

[7] Verma, Shikha, et al. "Deep learning techniques in extreme weather events: A review." 2308.10995 (2023).

[8] Rahmayanti, Nabillah, et al. "Comparison of machine learning algorithms to classify fetal health using cardiotocogram data." *Procedia Comput Sci.* 197 (2022): 162-171.