NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCES
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

BSc THESIS

Machine Learning Snowfall Retrieval Algorithms for Satellite Precipitation Estimates

Ioannis Th. Dravilas

Supervisors: **Manolis Koubarakis,** Professor
Stavros Dafis, Research Associate, IERSD/NOA
Georgios Kyros, Research Associate, IERSD/NOA
Konstantinos Lagouvardos, Research Director, IERSD/NOA

ATHENS

JULY 2023

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Αλγόριθμοι Μηχανικής Μάθησης για την Ανίχνευση Χιονόπτωσης σε Δορυφορικές Εκτιμήσεις Υετού

Ιωάννης Θ. Δραβίλας

Επιβλέποντες: **Μανόλης Κουμπαράκης,** Καθηγητής
Σταύρος Ντάφης, Επιστημονικός Συνεργάτης, ΙΕΠΒΑ/ΕΑΑ
Γεώργιος Κύρος, Επιστημονικός Συνεργάτης, ΙΕΠΒΑ/ΕΑΑ
Κωνσταντίνος Λαγουβάρδος, Διευθυντής Ερευνών, ΙΕΠΒΑ/ΕΑΑ

ΑΘΗΝΑ

ΙΟΥΛΙΟΣ 2023

BSc THESIS

Machine Learning Snowfall Retrieval Algorithms for Satellite Precipitation Estimates

Ioannis Th. Dravilas

S.N.: 1115201900053

SUPERVISORS: **Manolis Koubarakis,** Professor
Stavros Dafis, Research Associate, IERSD/NOA
Georgios Kyros, Research Associate, IERSD/NOA
Konstantinos Lagouvardos, Research Director, IERSD/NOA

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Αλγόριθμοι Μηχανικής Μάθησης για την Ανίχνευση Χιονόπτωσης σε Δορυφορικές Εκτιμήσεις Υετού

Ιωάννης Θ. Δραβίλας

Α.Μ.: 1115201900053

ΕΠΙΒΛΕΠΟΝΤΕΣ: **Μανόλης Κουμπαράκης,** Καθηγητής
Σταύρος Ντάφης, Επιστημονικός Συνεργάτης, ΙΕΠΒΑ/ΕΑΑ
Γεώργιος Κύρος, Επιστημονικός Συνεργάτης, ΙΕΠΒΑ/ΕΑΑ
Κωνσταντίνος Λαγουβάρδος, Διευθυντής Ερευνών, ΙΕΠΒΑ/ΕΑΑ

ABSTRACT

Remote sensing of snowfall has been proved to be a significant challenge since the start of the satellite era. Several techniques have been applied to satellite data, in order to estimate the fraction of frozen precipitation that reaches the surface. This thesis aims at investigating the efficacy of different Machine Learning (ML), and especially Deep Learning (DL) algorithms, in estimating the precipitation phase of NASA's Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM-IMERG). To achieve that, a training phase with hourly high-resolution numerical model outputs and in-situ observational data is chosen for the period of late-2020 and 2021. Results show that ML and DL models can estimate precipitation phase with relatively high accuracy, when compared to traditional methods, based on several case studies. The findings suggest that ML models offer a promising approach for advancing the nowcasting of snowfall and building a long-term archive dataset of IMERG-based snowfall, utilizing conventional near real-time data.

SUBJECT AREA: Machine Learning

KEYWORDS: Machine Learning, Deep Learning, Snowfall, Satellite Precipitation, Precipitation Phase

ΠΕΡΙΛΗΨΗ

Η αναγνώριση της χιονόπτωσης με τηλεπισκόπηση έχει αποδειχθεί μια δύσκολη πρόκληση ήδη από τα πρώτα στάδια χρήσης δορυφόρων στην ανθρώπινη ιστορία. Στο παρελθόν έχουν εφαρμοστεί ποικίλες τεχνικές σε δορυφορικά δεδομένα, με σκοπό την εκτίμηση του ποσοστού των στέρεων κατακρημνισμάτων που φτάνουν στην επιφάνεια του εδάφους. Η παρούσα εργασία στοχεύει στη διερεύνηση της αποτελεσματικότητας ποικίλων αλγόριθμων Μηχανικής Μάθησης, καθώς και Νευρωνικών Δικτύων, για την εκτίμηση της φάσης του υετού στα δεδομένα του αλγόριθμου Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM-IMERG) της NASA. Προς αυτήν την κατεύθυνση, επιλέχθηκε μια φάση εκπαίδευσης κατά την οποία χρησιμοποιούνται τα αποτελέσματα ενός αριθμητικού μοντέλου πρόγνωσης καιρού σε ωριαίο χρονικό βήμα, μαζί με επιτόπιες παρατηρήσεις εδάφους από μετεωρολογικούς σταθμούς για τις τελευταίες εβδομάδες του 2020 και το 2021. Τα αποτελέσματα δείχνουν πως τα μοντέλα Μηχανικής Μάθησης, και ειδικότερα τα Νευρωνικά Δίκτυα, μπορούν να εκτιμήσουν τη φάση του υετού με σχετικά υψηλή ακρίβεια, σε σύγκριση με αντίστοιχες παραδοσιακές μεθόδους, σε πολλές από τις μελετηθείσες περιπτώσεις. Τα ευρήματα υποδεικνύουν πως η Μηχανική Μάθηση αποτελεί μια υποσχόμενη μέθοδο για τη βελτίωση της εκτίμησης της χιονόπτωσης από δορυφορικά δεδομένα, αλλά και για τη δημιουργία ενός μακροχρόνιου αρχείου χιονοπτώσεων, που βασίζεται στον αλγόριθμο IMERG, αξιοποιώντας συμβατικά δεδομένα σχεδόν πραγματικού χρόνου.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Μηχανική Μάθηση

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Μηχανική Μάθηση, Βαθιά Μηχανική Μάθηση, Χιονόπτωση, Δορυφορική Εκτίμηση Υετού, Φάση Υετού

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Professor Manolis Koubarakis, for his invaluable guidance and support. His expertise and encouragement helped me complete this research and write this thesis.

I would also like to thank Stavros Dafis and Georgios Kyros, researchers at the Institute for Environmental Research and Sustainable Development (IERSD) of the National Observatory of Athens (NOA), as well as Konstantinos Lagouvardos, Research Director at IERSD/NOA, for providing helpful feedback and suggestions, along with the IERSD meteorological datasets and computational resources.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Precipitation

Precipitation is defined as ”all liquid or solid phase aqueous particles that originate in the atmosphere and fall to the Earth’s surface” [1]. It plays a key role in the worldwide water cycle and is of significant importance for meteorology, climatology and hydrology. Accurate measurement of precipitation and its distribution at both regional and global scales has always been a challenge for the scientific community. However, continuous advancements in satellite technology have siginficantly improved precipitation-retrieval methods over the last half century [2].

1.2 Precipitation Phase

Deriving the phase of precipitation and distinguishing between its liquid and frozen state, is of major importance for human activities, hydrological processes and climate change studies [3, 4]. For example, in a drainage basin (”an area of land where water from rain or snow melt drains downhill into a body of water such as a river, lake, wetland or ocean” [5]) the response time is different for rainfall and snowfall, while for remotely sensed observations a misclassification of precipitation phase can result in significant errors in the estimated precipitation rate [4, 6].

Towards this direction, a plethora of techniques is being used today to detect snowfall. Some of the most successful methods include using in-situ observations, remote sensing through dual-wavelength radars, or using numerical weather models [7, 8]. However, none of those approaches has been proved fully reliable, while for the most accurate ones, such as the measurements from in-situ instruments, the available data are generally sparse or even absent, for example in mountainous or sparsely populated areas [4, 9]. The use of satellite data to obtain precipitation estimates has been one of the most used methods for measuring precipitation so far, giving both satisfactory and continuous results with almost no missing values or temporal and spatial gaps [10]. Nonetheless, the estimation of precipitation phase based solely on satellites is still of mediocre performance for various reasons [11].

1.3 Machine Learning

Rebala et al. (2019) [12] described Machine Learning (ML) as a computer science field focusing on enabling computers to learn and improve their performance without requiring explicit programming. Traditional programming methods involve creating a detailed design and implementing it as a program, but this can be challenging for problems like detecting handwritten characters, due to the difficulty in designing rules for such variations.

ML can be applied to various types of problems such as: classification, where data are categorized into different classes, like will it rain or snow tomorrow; clustering, where data points are grouped into clusters based on shared properties; and regression, which uses historical data to forecast a continuous range of values, like temperature.

ML algorithms can be classified into various categories, based on the learning techniques

they use [13, 14]:

- Supervised ML, that includes training on labeled target data.
- Unsupervised ML, that does not involve labeled target data.
- Semi-supervised ML, which uses both labeled data and unlabeled data for training.
- Reinforcement ML, where models learn by interacting with an environment.

Supervised Machine Learning is a subset of Artificial Intelligence that uses algorithms which can learn a model or a set of rules from an existing labeled dataset, with the goal of correctly predicting new and unlabeled data. A common aspect of all supervised ML methods is the utilization of training data in order to optimize a set of weights that facilitate predictions. The determination of these weights involves either minimizing the error (i.e. loss) of the ML prediction (for regression problems) or maximizing the probability of a class label (for classification problems). Even though ML algorithms are more accurate than human-created rules for some types of problems, some ML algorithms, particularly those based on neural networks, lack interpretability, which is crucial in certain domains [12].

The use of Machine Learning in Meteorology has been constantly increasing during the last few years (Chase et al. 2022) [15], and is predicted to increase even more, as the volume of meteorological data that can be used to train ML models grows.

1.4 Machine Learning Models Used in Meteorology

According to Chase et al., some of the most common ML algorithms used in meteorology are Linear Regression, Logistic Regression, Naive Bayes, Decision Trees, Random Forest and Support Vector Machines.

1.4.1 Linear Regression

Linear Regression is a simple and computationally efficient method, which works by fitting the weight terms (w_i) in the following equation:

$$y = \sum_{i=0}^{N} w_i x_i, \tag{1.1}$$

where N is the number of features.

The loss function commonly used for Linear Regression is the Residual Summed Squared error (RSS), given by the following formula:

$$RSS = \sum_{j=1}^{N} (y_j - \hat{y_j})^2, \tag{1.2}$$

where N is the number of features, y_j is the actual value and $\hat{y_j}$ is the predicted value.

1.4.2 Logistic Regression

Logistic Regression is a classification method that applies the logistic function (also known as the sigmoid function) to the output of equation 1.1, in order to transform it to probability.

$$S(\hat{y}) = \frac{1}{1 + e^{(-\hat{y})}} \tag{1.3}$$

The loss function used for Logistic Regression is the Cross Entropy Loss, which for a binary classification problem can be written as follows:

$$CE = \sum_{i=0}^{N} -y_i \cdot log(S(\hat{y})) + (1 - y_i) \cdot log(1 - S(\hat{y})) \tag{1.4}$$

1.4.3 Naive Bayes

One more probabilistic classification method is Naive Bayes, which uses Bayes' theorem to calculate the conditional probability of the input vector x belonging to each class.

$$P(y|x) = \frac{P(y)P(x|y)}{P(x)} \tag{1.5}$$

The term ”naive” in the Naive Bayes algorithm refers to its assumption that input features x are independent of one another. That means, the algorithm treats each feature as if it has no relationship with the others. This simplification allows the algorithm to model $P(x|y)$ by assuming a distribution, such as the normal distribution.

Figure 1.1. by Chase et al. shows a visualization of the probability distribution across a range of temperatures, given the precipitation phase class ”snow” or ”rain” for a weather station. The data are presented in a histogram, showing the relative frequency of observations in temperature bins. Smooth curves are overlaid on the histogram, representing normal distribution fits to the data. Red colour is used to denote instances with rain, while blue colour indicates instances with snow.

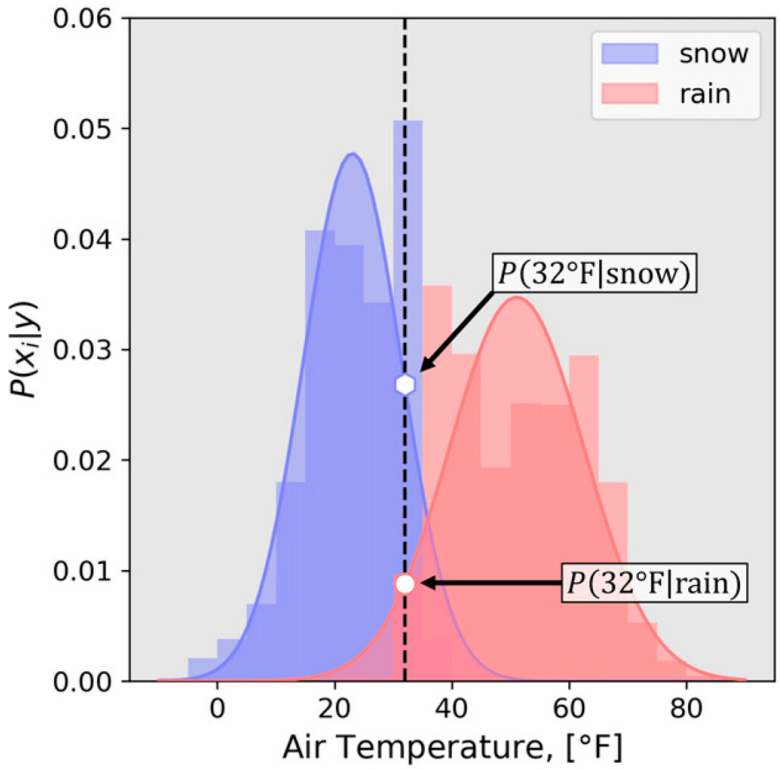

Figure 1.1: Probability distribution of snow and rain across temperature vales. Reprinted from "A Machine Learning Tutorial for Operational Meteorology. Part I: Traditional Machine Learning" by R. J. Chase, Weather and Forecasting, vol. 37, no. 8, pp. 1509–1529, Aug. 2022

1.4.4 Decision Trees

Decision trees are a widely used method for decision-making, working like flow charts, which are especially valuable for classification tasks. Those trees split the data in sub-groups, in a way that reduces Gini impurity or entropy after each split.

Gini impurity and entropy are both metrics that gauge the coherence of data labels within each group. The objective is to select branches that lead to leaves with minimal Gini impurity or entropy, resulting in distinct and, at the same time, consistent subgroups.

$$Gini = 1 - \sum_{i=0}^{k} {p_i}^2 \tag{1.6}$$

$$Entropy = -\sum_{i=0}^{k} p_i log_2(p_i), \tag{1.7}$$

where for both metrics p_i denotes the probability of class i in the training dataset.

1.4.5 Random Forest

Random Forest is a collection of Decision Trees, with each tree trained on a subset of the input data or the input features. For classification tasks, the final result is usually produced by counting the predicted classes of the tress and selecting the class with the highest number of appearances.

1.4.6 Gradient Boosted Trees

Gradient Boosted Trees are produced using an initial Decision Tree, called a weak learner, which is iteratively trained in order to reduce the error remaining from the previous trees.

1.4.6.1 XGBoost

Chen et al. in 2016 [16] introcuded XGBoost, a tree boosting system improving the traditional tree boosting technique mentioned above.

XGBoost introduces a regularization function that penalizes the complexity of the model and helps avoid overfitting. The regularized objective that is minimized is the following:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \tag{1.8}$$

where

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{1.9}$$

and

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), f_k \in F, \tag{1.10}$$

because K additive functions are used to produce the output of the tree ensemble model.

Here, l represents a convex loss function, which quantifies the dissimilarity between the predicted value $\hat{y}_i$ and the actual target value y_i. The second term Ω serves as a penalty for the model's complexity, ensuring that the final weights are smooth enough to prevent overfitting. Essentially, this regularized objective will favour models that utilize simple and predictive functions.

XGBoost also uses an exact greedy algorithm for split finding, which explores all possible split points for each feature in order to determine the best one. However, for large datasets, where a greedy search becomes very slow, the splits are approximated.

2. BACKGROUND AND RELATED WORK

The problem of deriving precipitation phase has been extensively studied before, both with conventional and Machine Learning methods.

2.1 Conventional Methods

Matsuo et al. in 1981 [17] showed that liquid water content and fall velocity of snowflakes, and therefore the depth of the layer below freezing level where melting does not occur, were dependent on surface air temperature, relative humidity, and snowflake mass.

In 2014, Lee et al. [18] presented the enhanced Matsuo scheme, an algorithm devised by meteorologists to forecast winter precipitation types in South Korea. The scheme utilizes air temperature, relative humidity, and the thickness of the 1000-850 hPa layer as input parameters.

Sims and Liu in 2015 [4] studied the effect of multiple geophysical parameters on precipitation phase and developed a parameterization scheme that utilizes 2-m temperature, relative humidity, low-level vertical lapse rate, surface skin temperature and surface type to calculate the conditional probability of solid precipitation to occur. Surface pressure is also used in order to calculate wet-bulb temperature (Tw).

2.2 Machine Learning Methods

In 2018, Behrangi et al. [9] reported that near-surface air temperature is usually used to derive precipitation phase. To distuingish between liquid and solid phase, the most used methods are a static threshold, a linear transition from all-rain to all-snow temperatures and a sigmoidal relationship produced by a logistic regression model, that can map values to probability of snow or rain to occur. It was found that relative humidity, wind speed and air pressure can also affect the melting of snowflakes and thus more variables should be used together with air temperature to determine precipitation phase. Even though among all single predictors Tw yields the highest score, the authors concluded that the use of logistic regression to combine the previously mentioned variables produces even better results.

Tang et al. in 2018 [19] combined passive microwave, infrared and environmental data using Deep Neural Networks to produce an improved snow estimation compared to GPM's passive microwave algorithm.

In 2020, Moon et al. [20] utilized correlation-based feature selection along with multinomial logistic regression to determine precipitation phase. Model data from the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Regional Data Assimilation and Prediction System (RDAPS) were used.

Shin et al. in 2022 [21] developed an ML-based winter precipitation type classification algorithm, testing three popular supervised Machine Learning Algorithms, Decision Trees, Random Forest and Support Vector Machines. The Random Forest algorithm exhibited the most outstanding performance, having the highest Probability Of Detection and the lowest False Alarm Ratio scores across all precipitation types. The results of the study

indicate that Tw and 850-1000 hPa layer thickness are the two most important variables for determining precipitation phase.

A paper presented at the AGU Fall Meeting 2022 by Bédard-Therrien et al. [22] introduced an ML approach for partitioning precipitation phase, using air temperature, relative humidity and disdrometer data, along with a Random Forest Regression Model. The precipitation phase used as target for the model, was measured with the WS100 radar disdrometer, which identifies solid, liquid and mixed precipitation phases.

Sanò et al. in 2022 [23] developed a new ML-based algorithm for snowfall retrieval, named SLALOM-CT (Snow retrievaL ALgorithm fOr gpM–Cross Track), using the Advanced Technology Microwave Sounder (ATMS) and Cloud Profiling Radar (CPR) observations. The Machine Learning algorithms that were compared, are Random Forest, Gradient Boosting and Neural Networks. The most accurate model for snowfall detection was a Convolutional Neural Network (CNN).

3. MATERIALS AND METHODS

In this study, Machine Learning and especially Deep Learning (DL) algorithms are used along with numerical weather data and in-situ observational data to classify the phase of precipitation acquired by the Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM-IMERG) operated by the National Aeronautics and Space Administration (NASA) [24].

3.1 Description of the Acquired Data

3.1.1 In-situ Observations From Ground Stations

During the past 15 years, the Institute for Environmental Research and Sustainable Development of the National Observatory of Athens (NOA/METEO) has established and is currently managing a dense network of automated weather stations throughout Greece (NOAAN) [25]. The in-situ variables measured by NOAAN stations that are used in this study include the air temperature, relative humidity and atmospheric pressure. Data from a diverse range of 480 stations, spanning various geographical areas and altitudes, are incorporated in the analysis.

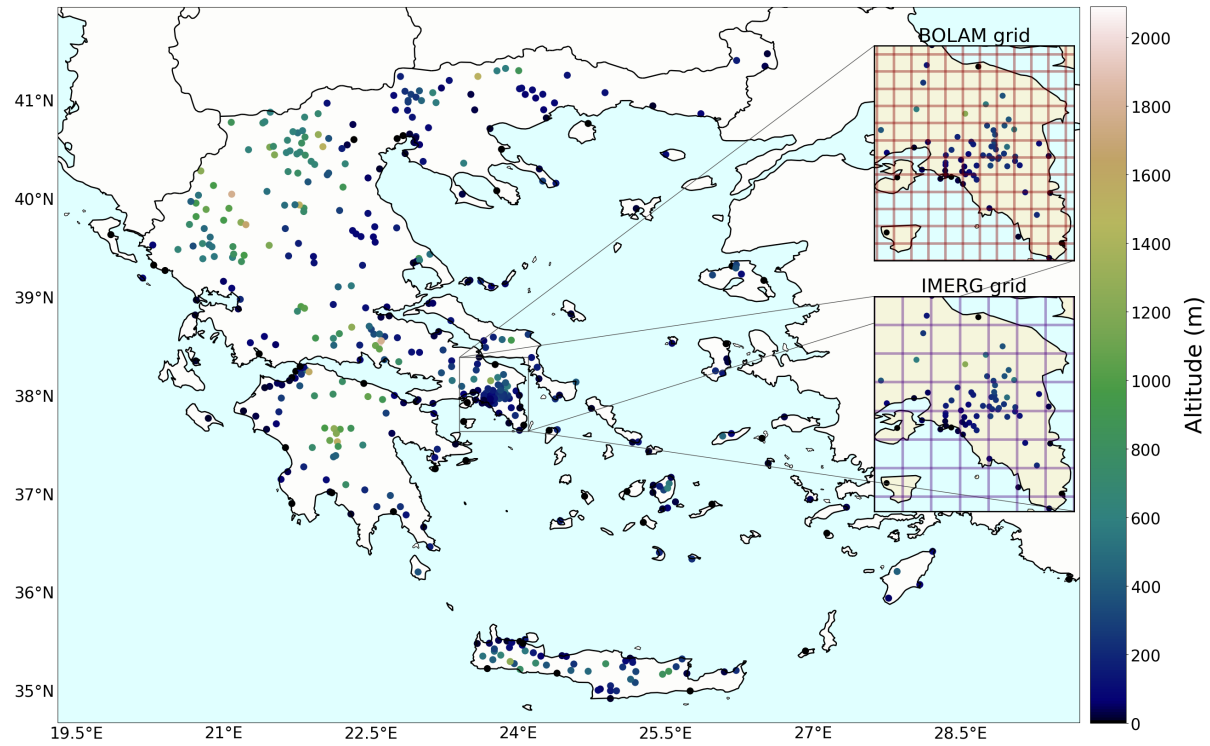

Figure 3.1: The NOAAN weather stations used, coloured by altitude, and the corresponding grids of BOLAM and IMERG over the Attica region.

3.1.2 Numerical Weather Model Data

The National Observatory of Athens also runs the hydrostatic meteorological Bologna Limited-Area Model (BOLAM) in operational mode. The current set-up of BOLAM at NOA/METEO involves a grid covering Europe, the Mediterranean basin and North Africa. The grid consists of 770 x 702 points with a 0.06° horizontal grid interval (≈6 km) and 40 vertical levels [26].

Numerous studies have consistently demonstrated that precipitation phase is influenced by a combination of meteorological variables, with air temperature being a key determinant. Another crucial variable that plays a vital role in precipitation phase determination is relative humidity. Relative humidity quantifies the moisture content present in the air and can aid in discerning the particle size distribution of precipitation. [21].

Geopotential Height is obtained by normalizing Geopotential, which represents the incremental work performed by shifting a unit mass from a surface to another, by the acceleration of gravity ($= 9.80665 m/s^2$) [27]. It helps identify the altitude corresponding to each pressure level, thereby revealing the specific height at which variables like temperature and relative humidity of an isobaric surface are observed.

Precipitation phase is also influenced by solid precipitation through the process of sublimation, particularly when the melting layer is in a sub-saturated state. This phenomenon can lead to an extended rain-snow transition zone, affecting the depth and distribution of different precipitation types [28]. Additionally, cooling occurs when raindrops evaporate, contributing to temperature changes in the atmospheric column [29]. In the context of convective precipitation, downdrafts play a significant role in temperature discontinuity. As colder air descends towards the ground, it can cause abrupt changes in temperature profiles, influencing precipitation type on the ground [30].

Considering the above-mentioned factors that can influence precipitation phase, the following variables of BOLAM are used in the current study:

- Temperature at 1000 hPa, 950 hPa, 900 hPa, 850 hPa, 700 hPa, 500 hPa
- Geopotential Height at 850 hPa, 500 hPa
- Wind Direction at 1000 hPa, 900 hPa, 700 hPa, 500 hPa
- Specific Humidity at 1000 hPa, 950 hPa, 900 hPa, 850 hPa, 700 hPa, 500 hPa
- Relative Humidity at 1000 hPa, 950 hPa, 900 hPa, 850 hPa, 700 hPa, 500 hPa
- Precipitation
- Convective Precipitation
- Temperature 2 m
- Relative Humidity 2 m
- Atmospheric Pressure
- Model Terrain Height

3.1.3 Satellite Precipitation Estimates

The Integrated Multi-satellitE Retrievals for GPM (IMERG) algorithm is designed to intercalibrate, combine and interpolate microwave precipitation measurements, along with microwave calibrated infrared (IR) satellite measurements, precipitation gauge analyses and possibly other precipitation estimators, at fine time and space scales worldwide. The system runs multiple times for each observation time, providing a first estimation and successively generating more accurate estimates as more data become available. IMERG covers a global grid with a spatial resolution of 0.1° and its early products are available with a delay of approximately 4 hours, in 30-minute time intervals [24].

Precipitation phase in IMERG is currently computed diagnostically, based on the Liu scheme [4]. The Liu scheme used by NASA calculates the Probability of Liquid Precipitation Phase (PLPP) based solely on data from a numerical weather model or model analysis, relying on surface wet-bulb temperature values. For the purpose of this study, data from IMERG Early Run are used. The precipitation estimates used are uncalibrated and IMERG PLPP is based on wet-bulb temperature data from the Japan Meteorological Agency (JMA) Numerical Weather Prediction (NWP) forecast [31].

3.2 Creation of a Custom Dataset

Snowfall observations in Greece are sparse in time and space, and usually only available in airports [32]. As a consequence, the availability of such data is insufficient to adequately train an ML Model and an alternative approach needs to be explored. Sims and Liu (2015) showed that Tw is a better indicator than ambient air temperature for separating solid and liquid precipitation [4]. Since air temperature, relative humidity and atmospheric pressure data are available from the dense NOAAN weather station network, a dataset classifying whether favorable conditions for snowfall were present for each station observation can be created, based on Tw.

Using NOAAN observations for air temperature, relative humidity and atmospheric pressure, BOLAM's nowcast (namely the first 12 hours after model initialization time excluding a spin-up period of 12 hours) and the 1.1°C Tw threshold chosen by NASA for IMERG V06 over land as the value corresponding to a Probability of Liquid Precipitation Phase equal to 50%, a new dataset is created. For each in-situ observation, the new dataset contains information about whether conditions were favorable for snowfall according to the surface Tw<1.1°C threshold, the corresponding numerical weather model data for the nearest grid point, as well as the station metadata such as latitude, longitude and altitude. The dataset comprises data collected from 480 locations across Greece, covering the time period of late-2020 and 2021. The temporal resolution of the dataset is set at 30 minutes, providing detailed and frequent measurements that match the temporal resolution of IMERG V06.

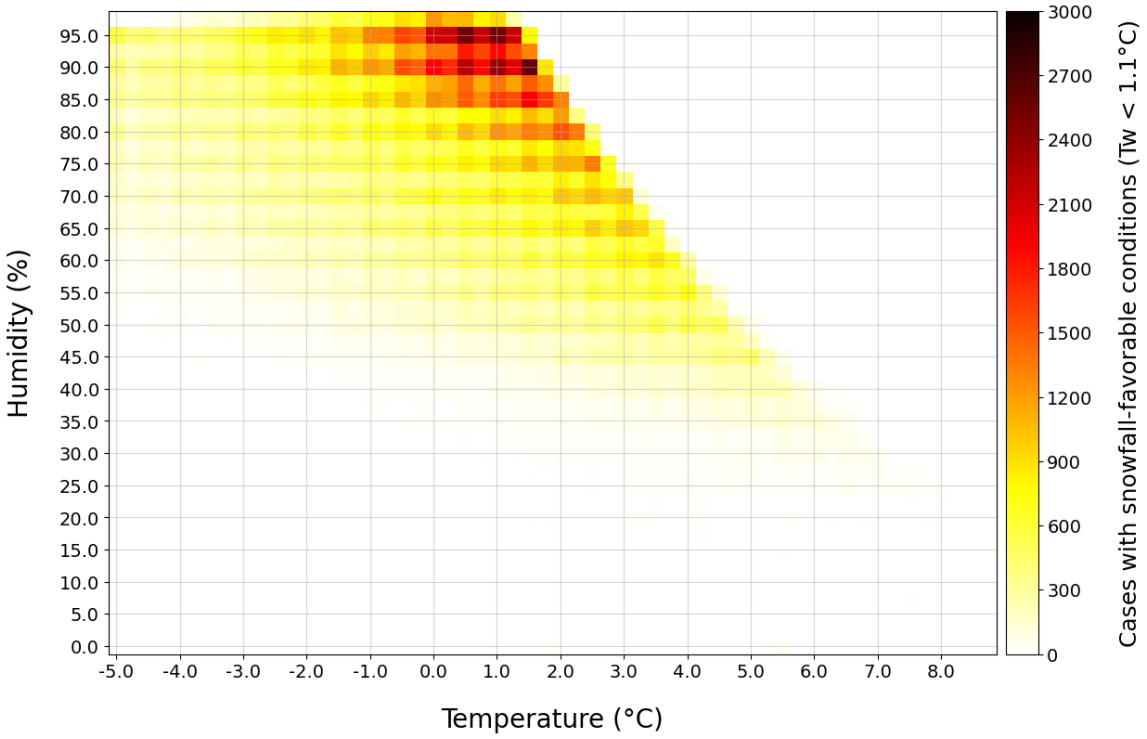

Figure 3.2: A heatmap of surface air temperature and relative humidity records by NOAAN stations when conditions were favorable for snowfall (Tw<1.1°C).

A separate testing dataset is also created in the same way, using 3-month-long data from early-2022 for 11 locations across Greece, with a temporal resolution of 1 hour.

In our datasets, Tw is calculated using the Newton's Iteration Method as employed by Sims and Liu (2015). Notably, this very method is also currently used by NASA for determining PLPP in IMERG V06 data.

In Machine Learning, the models undergo a three-step process: training, validation, and testing. During training, the model learns from a designated training dataset. The validation dataset is then used to fine-tune the model, and finally, the model is tested on a testing dataset that contains data it has not encountered before. However, constructing these datasets can be challenging in the case of weather data, as spatial and temporal overlaps may occur, compromising the integrity of the testing dataset.

For the purpose of this study, spatial independence is achieved by building training, validation and testing datasets that do not contain data from identical station locations.

Moreover, temporal independence is achieved by constructing training and testing datasets using different time periods.

3.3 Machine Learning Models Used

Our goal is to train ML models that, given only the numerical weather model's data and metadata for a specific location, are able to solve the classification task of predicting whether conditions favorable for snow are present near the surface or not. This layer

can then be used as a mask on IMERG precipitation products to create an IMERG-Snow dataset.

The three types of ML models used to solve this classification problem are a Random Forest Classifier, a Gradient Boosting Classifier and a Feedforward Neural Network. The latter is a subset of Deep Learning, where multiple layers are used to extract information from data.

3.3.1 Random Forest

Random Forest was first introduced in 2001 and is consisted of many basic classifiers in the form of Decision Trees [33]. Each Decision Tree makes a prediction, which is completely independent from the decisions of the other Decision Trees, and, in classification tasks, the final result is produced by a voting procedure, resembling an ensemble technique.

3.3.2 Gradient Boosting

Gradient Boosting is a popular ML technique used, among others, in classification tasks. It works by creating multiple weak models, which often are Decision Trees, and combining them to form a better-performing model. This is usually done by building an initial weak model, then a second model aiming to more accurately predict the cases where the first one performs poorly, etc. Each new model created, targets minimizing the error of the loss function; thus, the gradient of the loss function is calculated in every step of the algorithm [34].

3.3.3 Feedforward Neural Network

Feedforward Neural Networks are the simplest type of artificial neural networks, where information moves only in a forward direction, from the input nodes, to the hidden nodes and to the output nodes [35]. Here, a multi-layer neural network, also called a Multilayer Perceptron is used. A Multilayer Perceptron consists of multiple layers of computational units, containing neurons that are connected to the neurons of the next layer. These models are trained using back-propagation, a technique utilized to adjust the weight values of each connection, in a way that minimizes the error between predictions and actual values [36].

3.4 Training And Testing Process

3.4.1 Data Splitting and Cross-Validation

Data are divided into training and testing datasets based solely on station locations, using an 80:20 ratio.

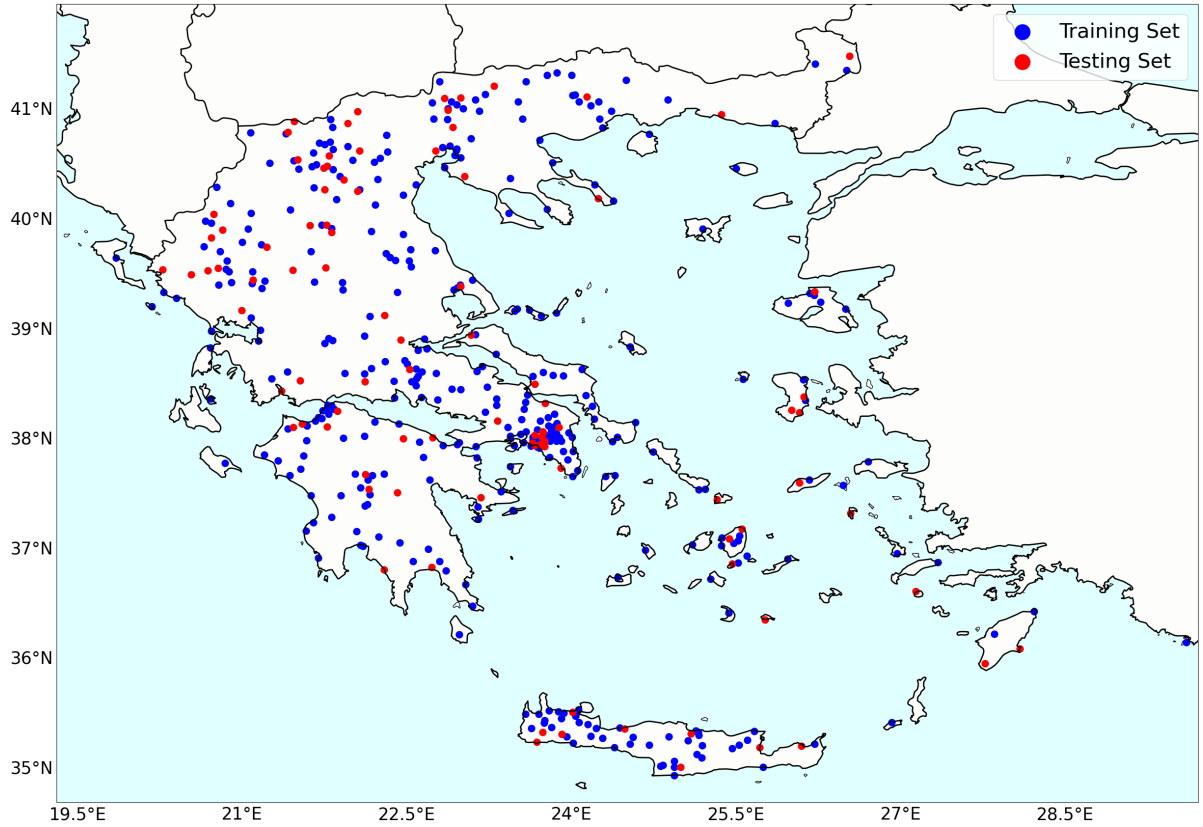

Figure 3.3: The NOAAN weather stations divided into the training and the testing dataset.

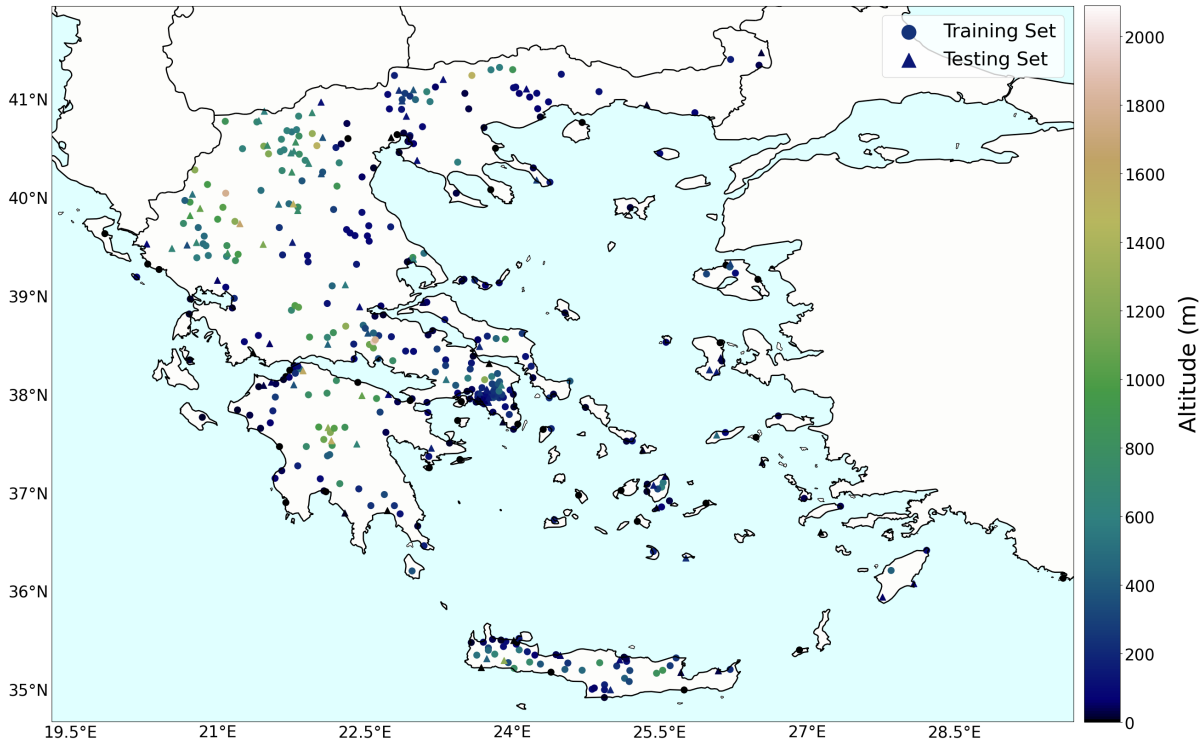

Figure 3.4: The NOAAN weather stations divided into the training and the testing dataset, coloured by altitude.

For each of the three models described above, the best architecture is determined through a 5-fold cross-validation process for hyperparameter tuning, ensuring that data for each station location are exclusively present in only one of the training and validation datasets during each iteration. The best hyperparameters are determined after averaging the F1-score (a measure of the harmonic mean of Precision and Recall) derived by training each architecture on each of the five splits of the training dataset, and validating it on the remaining 20% of the training dataset.

Finally, the trained models are tested on the corresponding testing dataset, comprising 20% of the initial dataset.

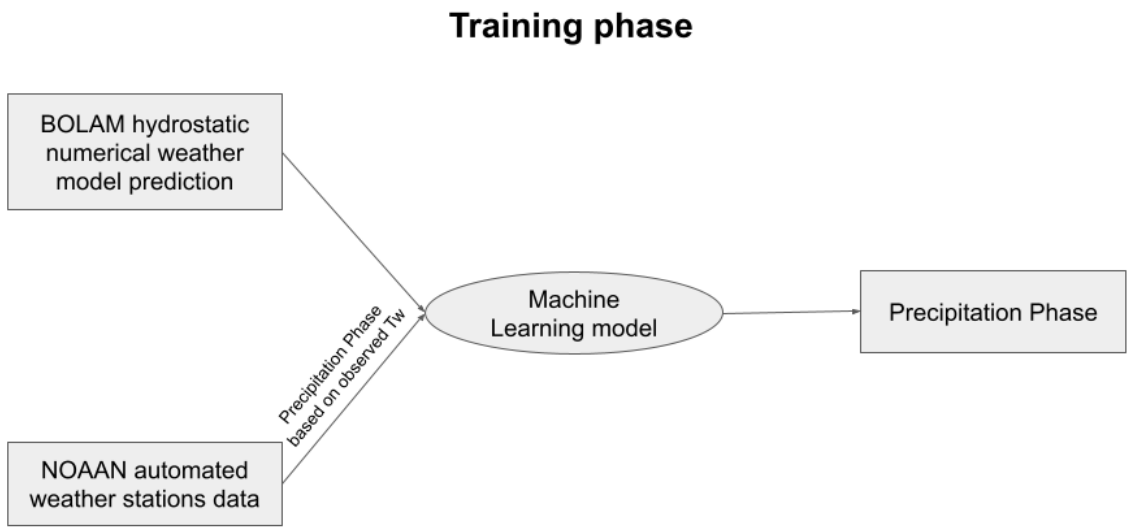

Figure 3.5: Schematic representation of the training phase.

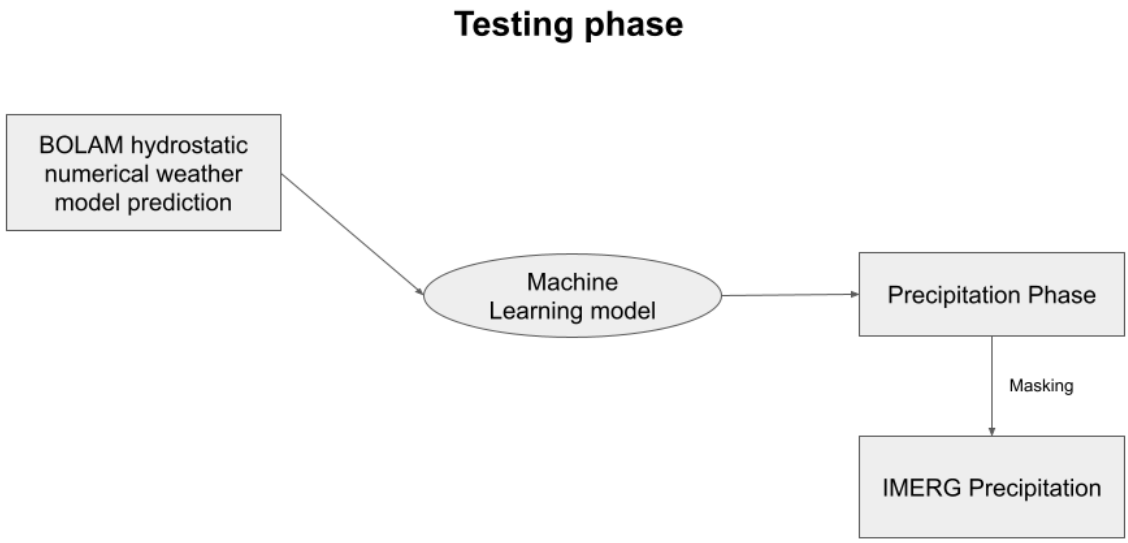

Figure 3.6: Schematic representation of the testing phase.

3.4.2 Hyperparameter Optimization

3.4.2.1 Random Forest

For the Random Forest model, the RandomForestClassifier from scikit-learn library is used. The best-performing model is selected among the following grid of parameters, using GridSearchCV, which performs an exhaustive search over the specified parameter values:

- Maximum Tree Depth: 5, 7 or 10
- Number of Estimators (number of different Decision Trees in the ensemble method): 100, 200 or 400
- Class Weight (the weights associated with each class): None (same weight for each class) or ”balanced” (automatic weights adjustment, based on the class frequencies of the input dataset)

3.4.2.2 Gradient Boosting

For the Gradient Boosting model, the XGBClassifier from the XGBoost library is used. The best-performing model is selected among the following grid of parameters, using GridSearchCV, which performs an exhaustive search over the specified parameter values:

- Learning Rate (the pace at which the algorithm updates its parameters to reach a local minimum): 0.1 or 0.2
- Maximum Tree Depth: 5, 7 or 10

- Number of Estimators (number of additive trees): 100, 200 or 400
- Scale Pos Weight (the weights associated with each class): 1 (same weight for each class) or 22 (the number of non-snow labels in our dataset is approximately 22 times larger than the number of snow lables)
- Objective: ”binary:hinge” (hinge loss for binary classification) or ”reg:squarederror” (regression with squared loss)

3.4.2.3 Feedforward Neural Network

For the Feedforward Neural Network, the PyTorch Machine Learning framework is utilized. The architecture of the model contains linear layers with 512 neurons, as well as non-linear activation functions.

To introduce non-linearity into the model, a non-linear activation function is applied to the output of each linear layer, with the exception of the last one, while a dropout layer is incorporated after each of the previously-mentioned non-linear activation functions.

The final output of the model is transformed into probabilities using a Sigmoid activation function applied to the output of the last linear layer. This step ensures that the model outputs values within the range $[0, 1]$, representing the probabilities of the input belonging to each class of ”rain” or ”snow”.

The Adam optimizer is selected as the minimizer of the loss function during the training process.

The best-performing model is selected among the following grid of parameters, using Optuna, a hyperparameter optimization framework, which decreases execution time and enhances performance, compared to traditional methods such as GridSearch:

- Number of Linear Layers: 10 or 15
- Non-Linear Activation Function: ReLU or Tanh
- Dropout Rate (ignoration of randomly selected neurons to avoid overfitting): 0, 0.25 or 0.5
- Batch Size (the number of samples processed before a single update of the model): 128 or 256
- Number of Epochs (the number of complete passes through the training data): 10 or 20
- Criterion (the loss function): MSELoss or BCELoss
- Learning Rate of the Optimizer (the pace at which the optimizer updates the parameters to reach a local minimum): 1e-5, 2e-5, or 3e-5

For the training of the Feedforward Neural Network, due the compute-intensive nature of the operations, CUDA programming model is utilized in order to parallelize the training across GPUs and train the models faster.

3.5 Evaluation and Metrics

The metrics used to evaluate the results of the models are Precision, Recall (also called Probability of Detection - POD), F1-score, Critical Success Index (CSI), False Alarm Ratio (FAR) and Heidke Skill Score (HSS).

3.5.1 Precision

Precision is a statistical metric that measures the proportion of the True Positive (TP) results among all the Positive results predicted by the model, showing how many of the Positive results the model predicted are actually correct and not False Positive (FP). A value of 0 indicates complete disagreement between forecast and observations, while a value of 1 indicates a perfect forecast.

$$Precision = \frac{TP}{TP + FP} \tag{3.1}$$

3.5.2 Recall

Recall is a statistical metric that measures the proportion of the TP results among all the actual Positive cases of a dataset (TP and False Negative - FN). It represents the ability of the model to identify Positive cases and is also called Probability of Detection (POD). As in Precision, a value of 0 indicates complete disagreement between forecast and observations, while a value of 1 indicates a perfect forecast.

$$Recall = \frac{TP}{TP + FN} \tag{3.2}$$

3.5.3 F1-score

F1-score is calculated as the harmonic mean of precision and recall, skillfully merging these two competing metrics. As in Precision and Recall, a value of 0 indicates complete disagreement between forecast and observations, while a value of 1 indicates a perfect forecast.

$$F1 = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{3.3}$$

3.5.4 Critical Success Index

Critical Success Index (CSI), also called Threat Score (TS), is a verification measure of categorical forecast performance, which does not take into account the number of forecasts belonging in the negative class, that verify. Again, a value of 0 indicates complete disagreement between forecast and observations, while a value of 1 indicates a perfect forecast.

$$CSI = \frac{TP}{TP + FP + FN} \tag{3.4}$$

3.5.5 False Alarm Ratio

False Alarm Ratio (FAR) measures forecast performance by dividing the number of false alarms by the total number of event forecasts. It indicates the percentage of negative class events which were falsely forecast as positive events. A value of 1 indicates complete disagreement between forecast and observations, while a value of 0 indicates a perfect forecast.

$$FAR = \frac{FP}{TP + FP} \tag{3.5}$$

3.5.6 Heidke Skill Score

Heidke Skill Score is a skill-corrected verification metric which also takes into account the number of correct random forecasts. It is derived by dividing, the total number of correct forecasts minus the number of correct random forecasts, by the total number of forecasts minus the number of correct random forecasts. It measures the improvement of the forecast over a chance forecast. Here, negative values indicate a forecast worse than random chance, a value of 0 means no skill level, while a value of 1 is a perfect forecast [37].

$$HSS = 2 \cdot \frac{TP \cdot TN - FP \cdot FN}{(TP + FN) \cdot (FN + TN) + (TP + FP) \cdot (FP + TN)} \tag{3.6}$$

4. RESULTS

For each of the Random Forest, Gradient Boosting and Feedforward Neural Network models, the best architecture is selected after hyperparamter tuning, and is then evaluated on the testing dataset.

4.1 Best Hyperparameters

The best hyperparameters found for each model are the following:

4.1.1 Random Forest

- Maximum Tree Depth: 10
- Number of Estimators (number of different Decision Trees in the ensemble method): 100
- Class Weight (the weights associated with each class): None (same weight for each class)

4.1.2 Gradient Boosting

- Learning Rate (the pace at which the algorithm updates its parameters to reach a local minimum): 0.1
- Maximum Tree Depth: 10
- Number of Estimators (number of additive trees): 400
- Scale Pos Weight (the weights associated with each class): 1 (same weight for each class)
- Objective: ”reg:squarederror” (regression with squared loss)

4.1.3 Feedforward Neural Network

- Number of Linear Layers: 15
- Non-Linear Activation Function: ReLU
- Dropout Rate (ignoration of randomly selected neurons to avoid overfitting): 0.25
- Batch Size (the number of samples processed before a single update of the model): 128
- Number of Epochs (the number of complete passes through the training data): 20
- Criterion (the loss function): BCELoss
- Learning Rate of the Optimizer (the pace at which the optimizer updates the parameters to reach a local minimum): 1e-5

4.2 Feature Importance

Feature importance serves as a scoring mechanism, providing insights into the usefulness of each variable during the model's construction.

For the Gradient Boosting model, we gain an understanding of feature importance through two different calculations. The first approach, known as "weight," calculates the number of occurrences of each feature within the tree. The second method, "gain," evaluates the average gain achieved across all splits that involve the utilization of that specific feature [38].

The visual representations in the following figures illustrate the most critical features of the model using the two calculation methods. Notably, both Temperature at 2 m and altitude emerge as highly influential variables, making appearances in both types of importance metrics. Additionally, Temperature and Specific Humidity at the 1000 hPa isobaric surface stand among the top three features, completing the groups of the three most important contributors.

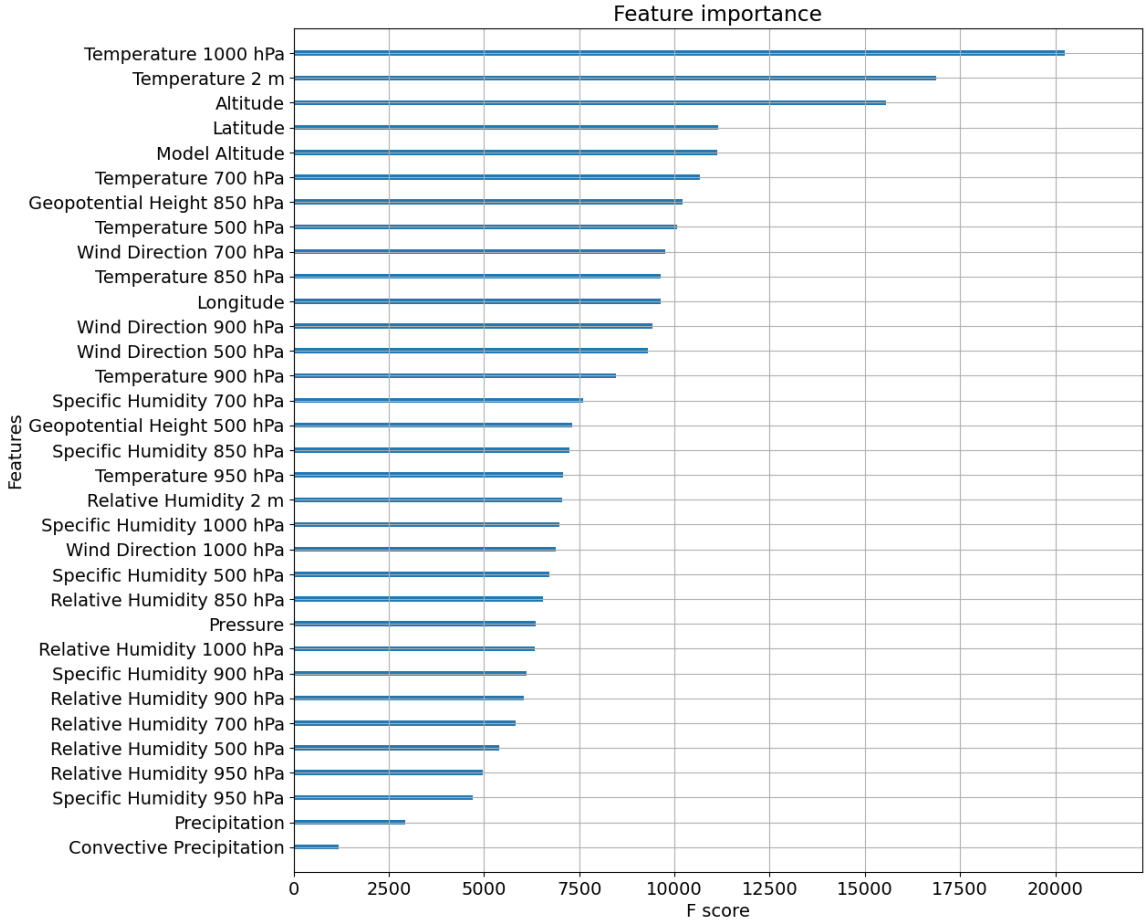

Figure 4.1: The "weight" importance of each feature in constructing the Gradient Boosting model.

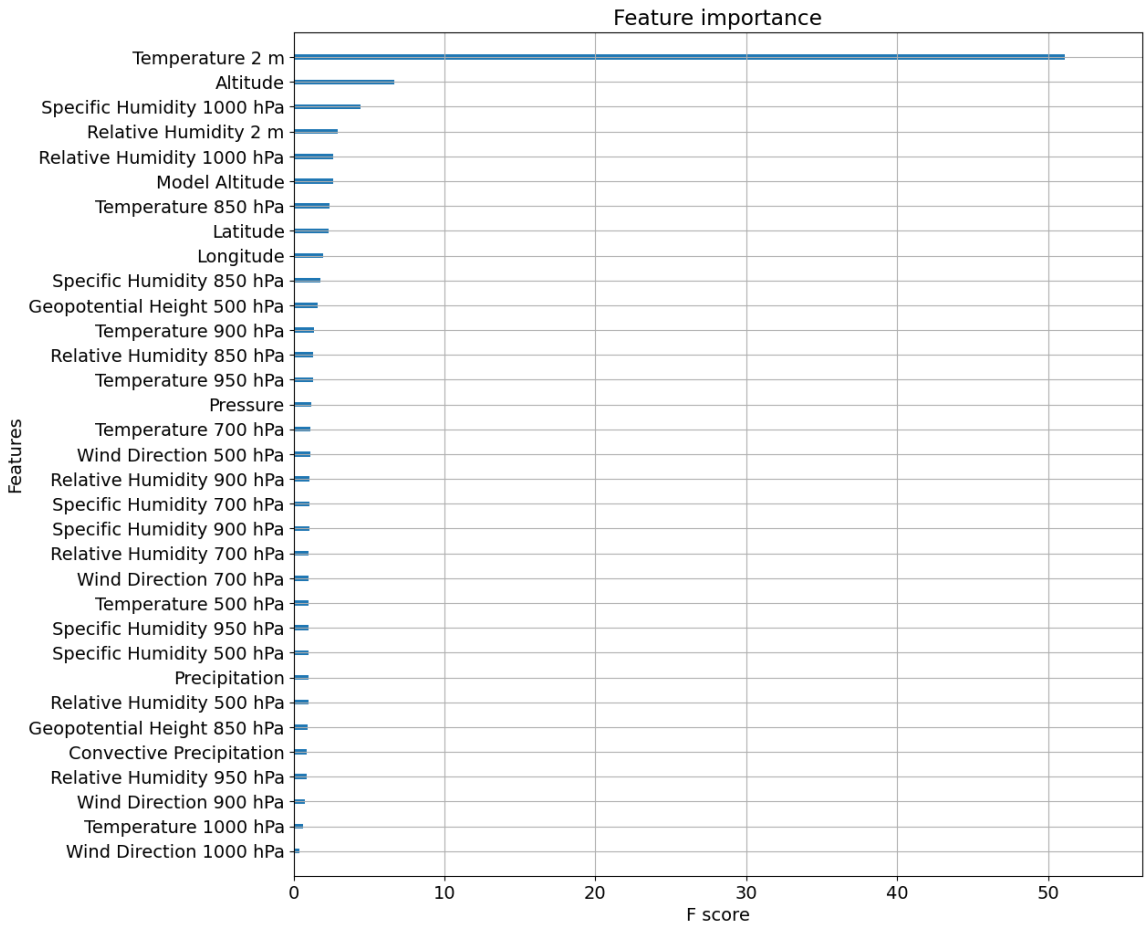

Figure 4.2: The ”gain” importance of each feature in constructing the Gradient Boosting model.

4.3 Evaluation on the First Testing Dataset

The testing dataset used here is the one containing data for late-2020 and 2021 from 96 station locations in a 30-minute time-step.

4.3.1 Machine Learning Models

The six score metrics were calculated on the testing dataset for all the Random Forest, Gradient Boosting and Feedforward Neural Network models, with regard to their ability of predicting cases with conditions favorable for snowfall (Tw<1.1°C). Results are shown in Table 4.1.

Table 4.1: Scores of the 3 ML models for predicting cases with conditions favorable for snowfall, evaluated on the first testing dataset.

Model	Precision	Recall (POD)	F1-score	CSI	FAR	HSS
Random Forest	0.87	0.72	0.79	0.65	0.13	0.78
Gradient Boosting	0.87	0.81	0.84	0.72	0.13	0.83
Feedforward Neural Network	0.85	0.80	0.82	0.70	0.15	0.81

4.3.2 Conventional Methods

A comparison with traditional precipitation phase derivation techniques is also made on the same testing dataset.

The first method uses the Probability of Liquid Precipitation Phase field currently available in IMERG V06. This field is produced using a simple look-up table, which calculates PLPP as a function of wet-bulb temperature. To achieve this for the Early Run dataset which is used in our study, the surface temperature, humidity and pressure forecasts are taken from the JMA NWP model [31]. According to NASA, the 1.1°C threshold for Tw is used to distinguish between less and more than 50% PLPP over land.

The second method uses the raw nowcast data acquired from BOLAM for near-surface air temperature, relative humidity and atmospheric pressure, in order to calculate Tw and classify each case using the 1.1°C threshold. It is important to note here, that no corrections are made to account for the difference between the model grid point altitude and the actual station altitude.

Results for the two conventional methods are shown in Table 4.2.

Table 4.2: Scores of the 2 conventional methods for predicting cases with conditions favorable for snowfall, evaluated on the first testing dataset.

Model	Precision	Recall (POD)	F1-score	CSI	FAR	HSS
IMERG V06 PLPP	0.61	0.61	0.61	0.44	0.39	0.59
BOLAM nowcast	0.66	0.81	0.73	0.57	0.34	0.71

4.4 Application and Evaluation on 2022 Data

The data employed in both the training and the testing phases for the aforementioned methods span from late-2020 to 2021.

In this section, a new dataset comprising data from the initial three months of 2022 is introduced. The purpose is to assess the predictive capability of the Gradient Boosting model, which exhibited superior performance compared to all other methods, in predicting the precipitation phase for data outside the training dataset's time span. This is accomplished using two methods.

Firstly, the model is applied to a grid encompassing data for various cold waves that impacted Greece in 2022. The output is then used as a mask on IMERG V06 Early Run precipitation estimates and the results are plotted, allowing us to visualize and comprehend

the model's effectiveness in determining precipitation phase during real-world conditions.

Secondly, the model's predictions for the initial three months of 2022 are compared with the NOAAN weather stations measurements for 11 carefully selected locations.

These steps aim to evaluate and validate the model's performance under real-world conditions, both during specific weather events and in comparison to ground truth measurements from established weather stations.

4.4.1 Examples of Application on IMERG Precipitation Estimates

During the initial three months of 2022, Greece experienced several cold waves of varying intensity. To assess the effectiveness of the Gradient Boosting model in identifying the precipitation phase during these events, BOLAM nowcast data are used as input. The precipitation phase predicted by the model is then utilized to mask the corresponding IMERG V06 Early Run uncalibrated precipitation data for the same time period.

4.4.1.1 10th of January 2022

On January 10th of 2022, a cut-off low in the upper/mid troposphere moved from Italy towards the Ionian Sea in Greece. This weather system was accompanied by a mild cold air advection from the Balkans towards Northern Greece.

As a result of these weather conditions, snowfall was expected in the mountains of Mainland Greece and in some lower altitude areas of Western Macedonia.

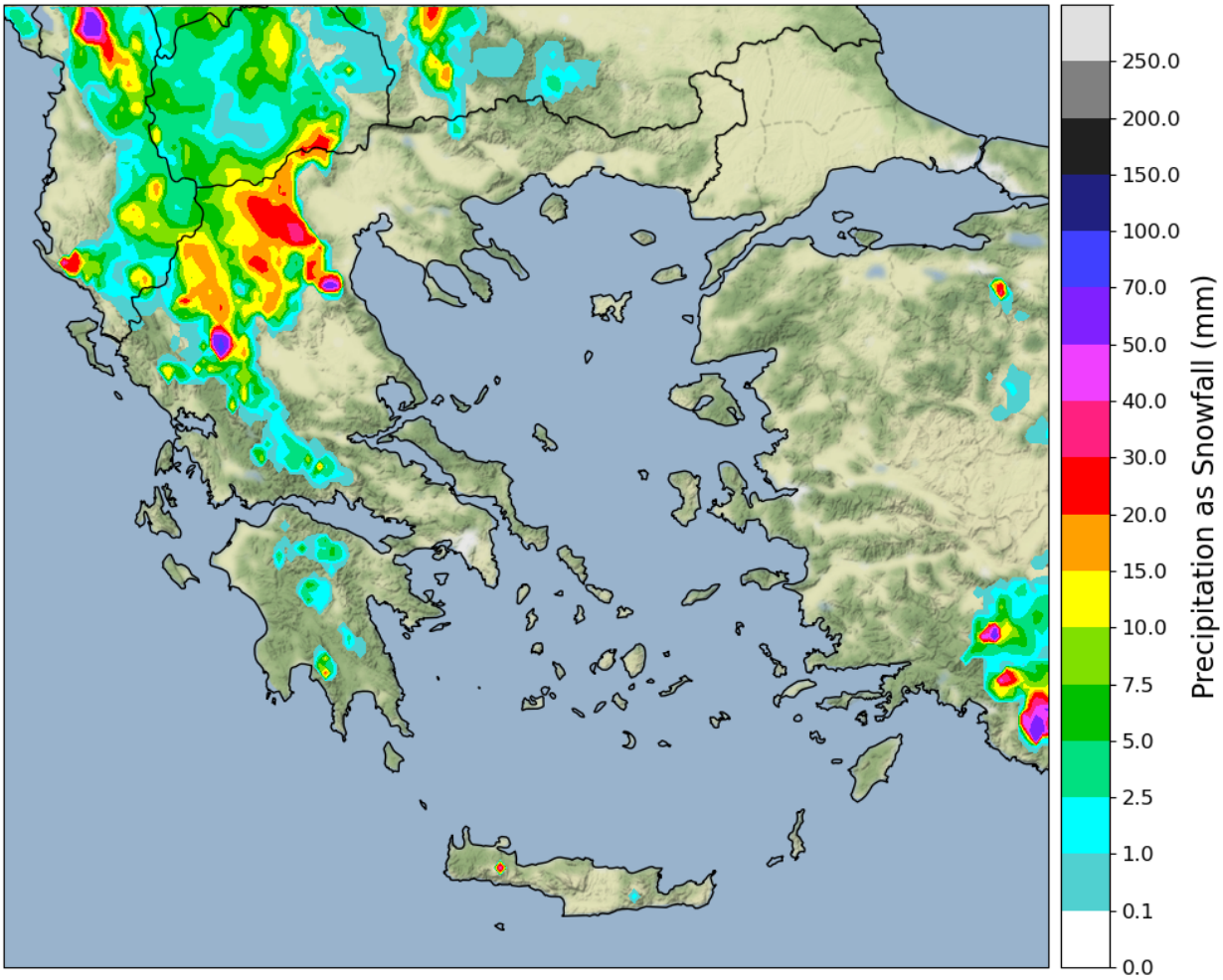

Figure 4.3: The 24-hour accumulated precipitation from IMERG that fell as snowfall during the 10th of January 2022 in Greece, as indicated by the Gradient Boosting model.

4.4.1.2 24th of January 2022

On January 24th of 2022, an upper level closed low over the Eastern Mediterranean resulted in a significant cold air advection, leading to heavy snowfall in the eastern regions of Greece, including Attica and the capital city, Athens, as well as the Aegean Islands. The snowfall was so intense that it caused disruptions and paralysis in the affected areas of Athens, and a two-day public holiday was declared by the Greek government after the snowstorm [39].

Moreover, the adverse weather conditions also impacted Turkey, with heavy snowfall affecting Istanbul. As a consequence, Istanbul Airport had to suspend its operations temporarily.

The storm was named ”Elpis” by the Hellenic National Meteorological Service.

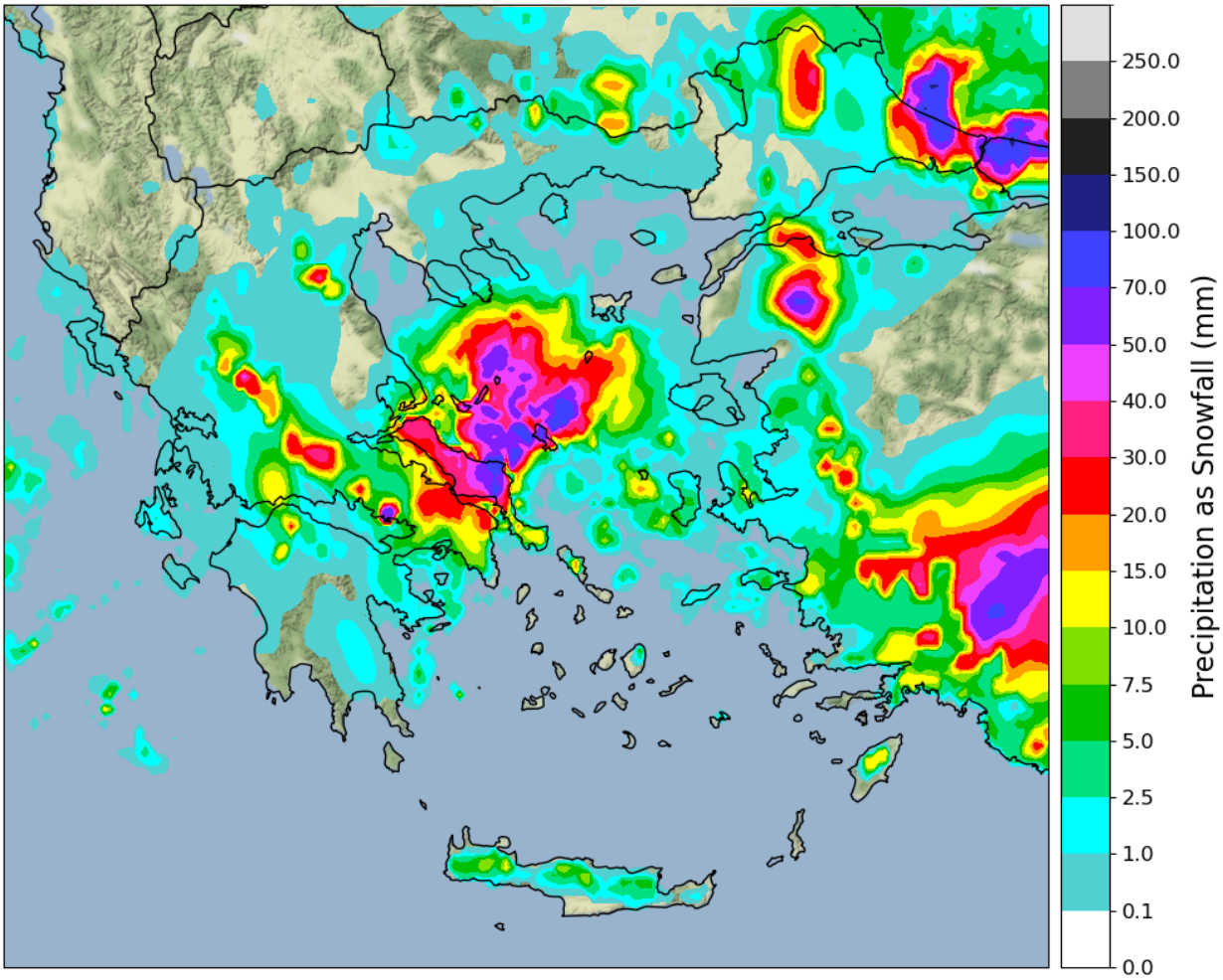

Figure 4.4: The 24-hour accumulated precipitation from IMERG that fell as snowfall during the 24th of January 2022 in Greece, as indicated by the Gradient Boosting model.

4.4.1.3 29th of January 2022

On January 29th of 2022, a barometric low-pressure system developed in the southern parts of the Ionian Sea and moved towards Crete, triggering precipitation in the regions of Sterea Ellada and Peloponnese in Greece. Due to the presence of cold air masses over Mainland Greece, the precipitation fell as snowfall in the aforementioned areas.

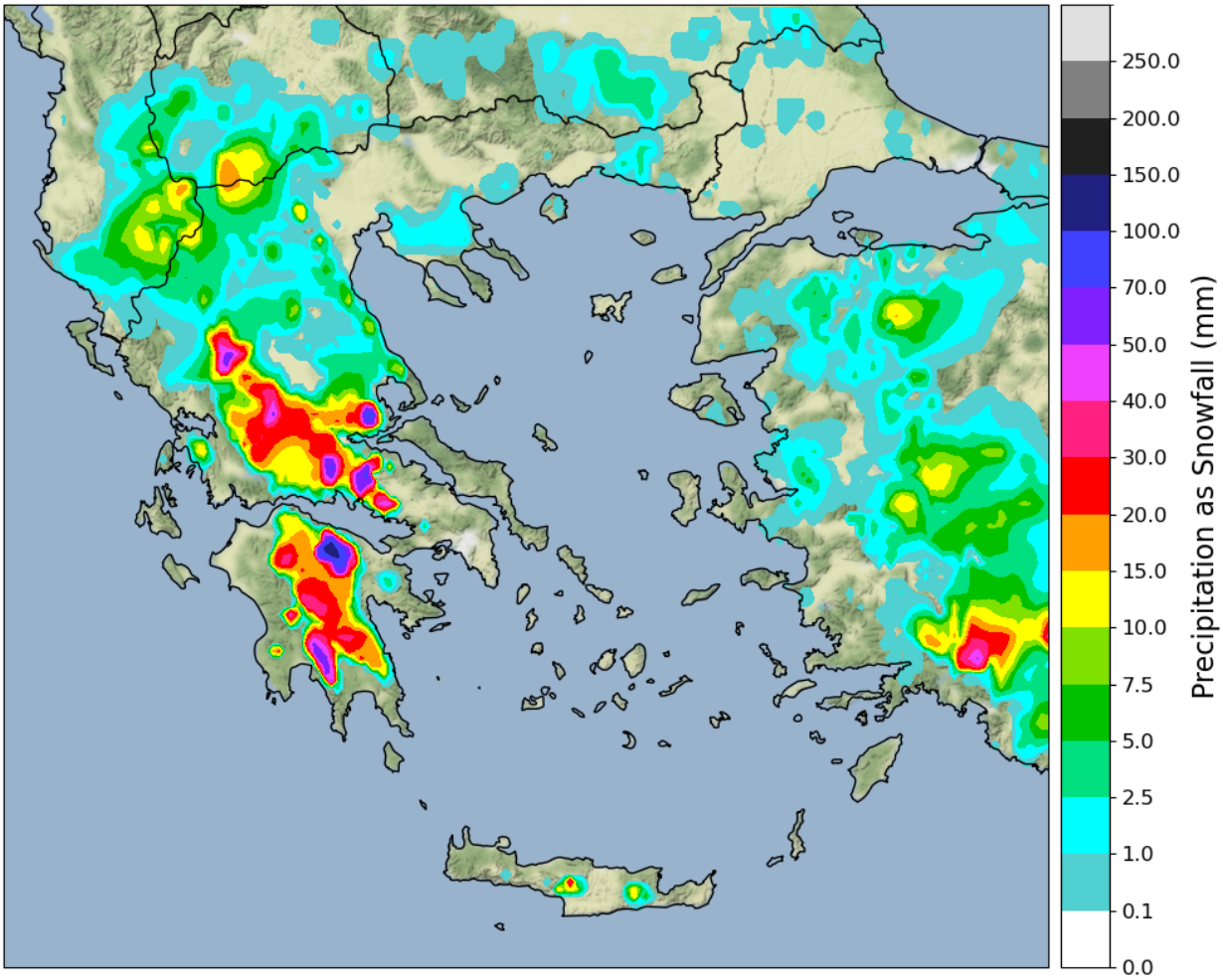

Figure 4.5: The 24-hour accumulated precipitation from IMERG that fell as snowfall during the 29th of January 2022 in Greece, as indicated by the Gradient Boosting model.

4.4.1.4 8th of February 2022

On February 8th of 2022, a high-pressure system over Central and Western Europe led to the formation of a trough, influencing the weather in Greece. The trough was accompanied by cold air masses, causing snowfall in many areas of Mainland Greece.

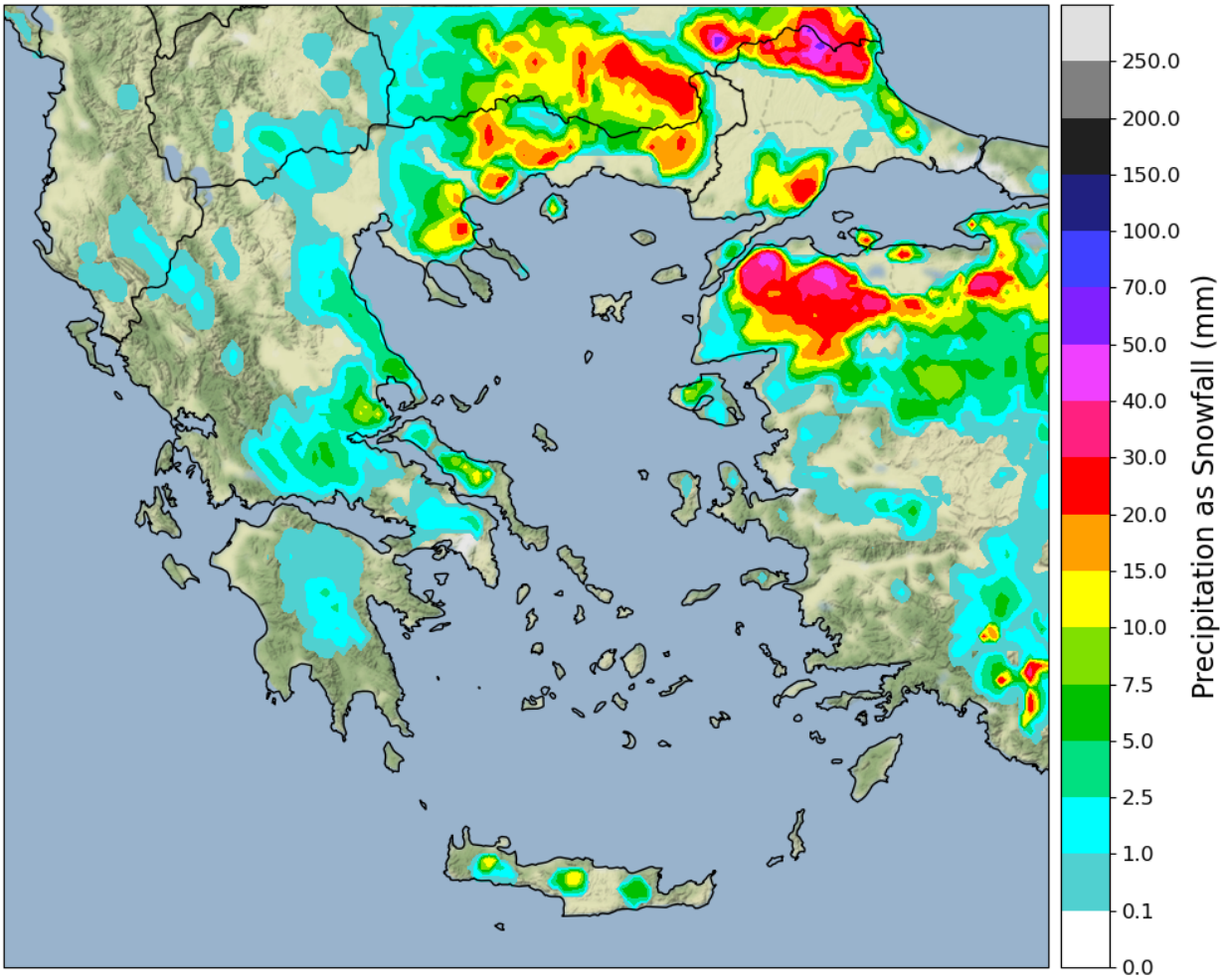

Figure 4.6: The 24-hour accumulated precipitation from IMERG that fell as snowfall during the 8th of February 2022 in Greece, as indicated by the Gradient Boosting model.

4.4.1.5 27th of February 2022

On February 27th of 2022, a storm named ”Bianca” by the Hellenic National Meteorological Service, affected Greece, causing heavy snowfall in Northern Greece. Suburbs of Thessaloniki, the second largest city of Greece, were also impacted by snowfall.

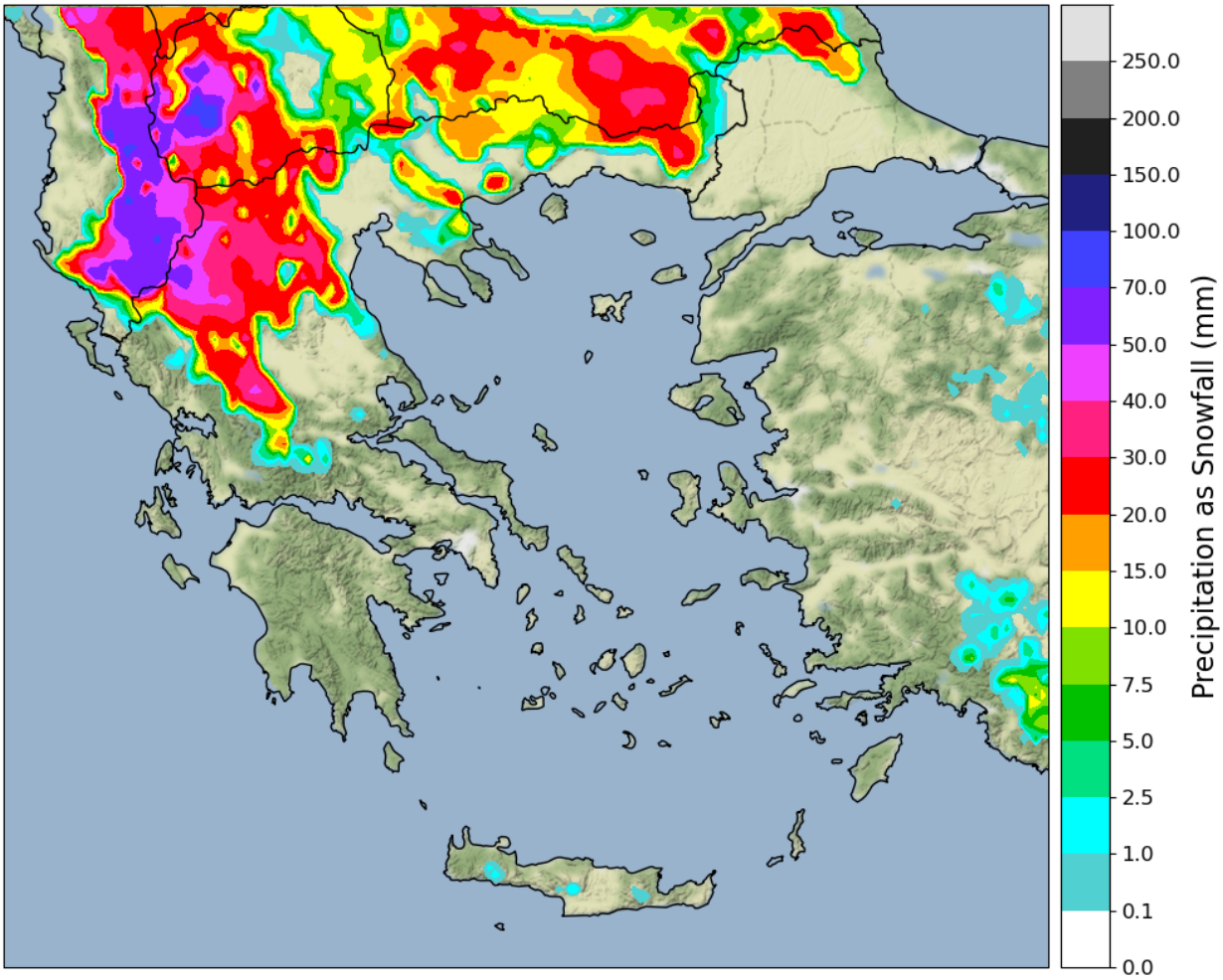

Figure 4.7: The 24-hour accumulated precipitation from IMERG that fell as snowfall during the 27th of February 2022 in Greece, as indicated by the Gradient Boosting model.

4.4.1.6 10th of March 2022

On March 10th of 2022, a long-wave trough affected Greece from the northeast, causing snowfall in the eastern regions of the country, including some of the Aegean Islands, Euboea and Athens.

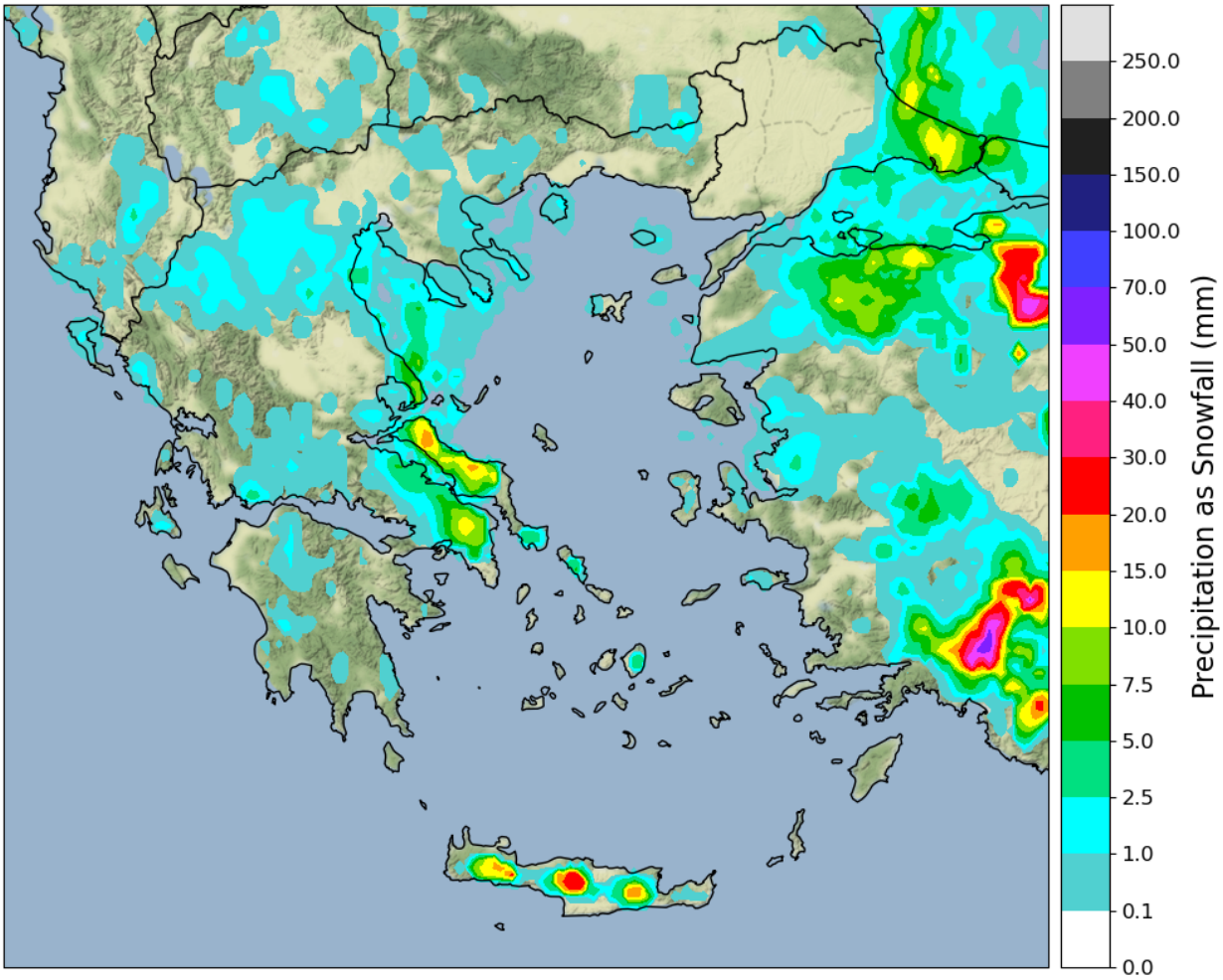

Figure 4.8: The 24-hour accumulated precipitation from IMERG that fell as snowfall during the 10th of March 2022 in Greece, as indicated by the Gradient Boosting model.

4.4.2 Evaluation on 2022 Data From 11 Ground Stations

Air temperature, relative humidity and atmospheric pressure data were also acquired from 11 NOAAN weather stations for the period between the 1st of January 2022 and the 31st of March 2022 in an hourly time-step. These data were utilized to calculate Tw and subsequently compare the predictions of each model with the observed values.

The locations of the 11 ground stations where the evaluation was performed, are displayed in the accompanying figure. It is noteworthy that most of these stations are situated at relatively high altitudes.

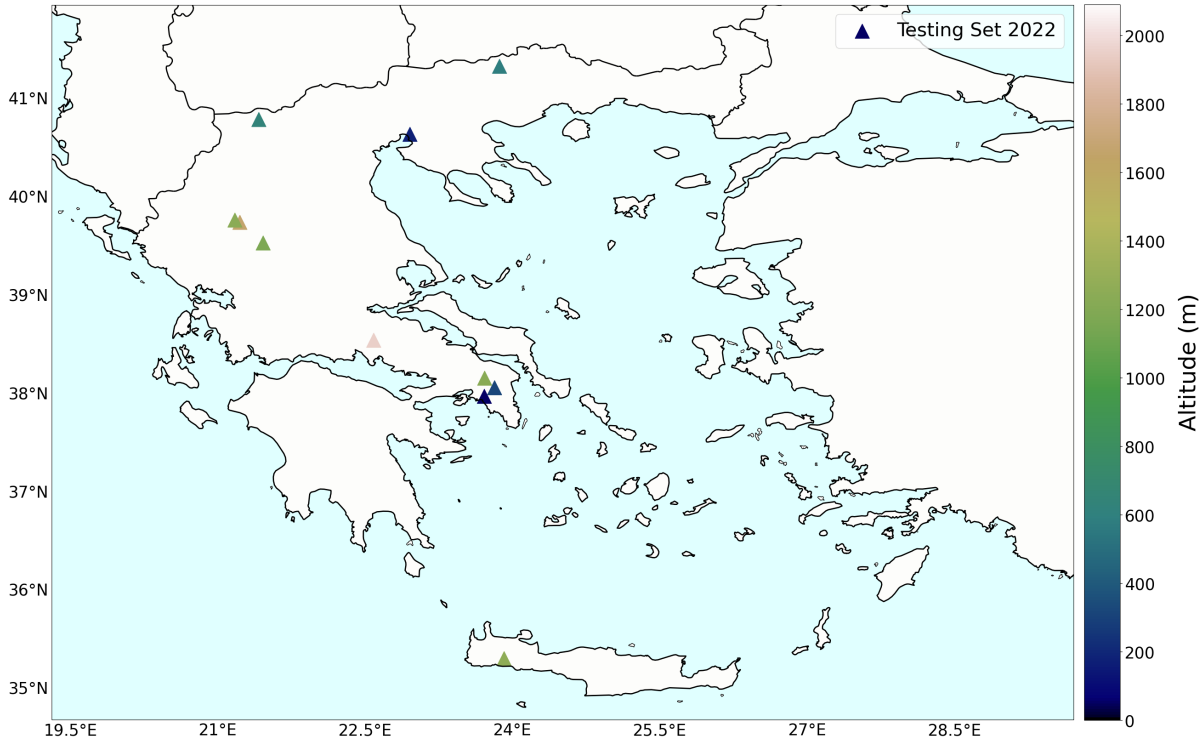

Figure 4.9: The NOAAN weather stations used for the 2022 period, coloured by altitude.

During this period, a total of 12300 cases with conditions favorable for snowfall (Tw<1.1°C) were identified in the collected data. The table presented below, illustrates the capability of various traditional techniques and ML models used in this study, in predicting these cases.

Table 4.3: Scores of the the traditional techniques and the 3 ML models for predicting cases with conditions favorable for snowfall, evaluated on the 2022 testing dataset.

Model	Precision	Recall (POD)	F1-score	CSI	FAR	HSS
IMERG V06 PLPP	0.87	0.58	0.70	0.53	0.13	0.48
BOLAM nowcast	0.89	0.84	0.87	0.76	0.11	0.73
Random Forest	0.96	0.85	0.90	0.83	0.04	0.81
Gradient Boosting	0.95	0.92	0.93	0.87	0.05	0.86
Feedforward Neural Network	0.95	0.92	0.93	0.88	0.05	0.86

5. DISCUSSION

In this study, several Machine Learning techniques, including Deep Learning, are leveraged to identify conditions favorable for snowfall from numerical weather data. A training process is performed using wet-bulb temperature data from in-situ observations of ground weather stations as target data. Our primary goal is to create a reliable precipitation phase mask, that can be employed in conjuction with satellite precipitation estimates from IMERG to create an IMERG-Snow dataset.

5.1 Comparison With Previous Work

Moon et al. in 2020 [20] achieved an HSS of 73% in determining precipitation type for snow cases, using ML models trained with short-range forecasts from numerical models of the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Regional Data Assimilation and Prediction System (RDAPS). The individual HSS values for ECMWF and RDAPS alone were comparatively lower at 52% and 55% respectively, while the improved Matsuo scheme [18] used operationally at the time by the Korea Meteorological Administration (KMA) exhibited an HSS of 71%.

In 2022, Shin et al. [21] trained and then fine-tuned ML models using dual-polarization radar measurements and thermodynamic fields derived from the very-short-range forecast system of KMA, which is updated by multiquadric interpolation of observations. A CSI of 96% was achieved, also outperforming the improved Matsuo scheme which scored 77%.

Sanò et al. in 2022 [23] developed a snowfall retrieval algorithm for the Advanced Technology Microwave Sounder (ATMS), using Cloud Profiling Radar (CPR) on board CloudSat snowfall products as a reference. Regarding Snowfall Detection, a POD of 83% was achieved, with a FAR of 18%. HSS and CSI scores reached 68% and 70% respectively.

In 2022, Xiong et al. [40] evaluated IMERG and ERA5 precipitation phase partitiniong on a global scale, using target data from ground observations. POD of snowfall over land was 87% for IMERG and 91% for ERA5. CSI was 67% for IMERG and 81% for ERA5, while FAR was 16% for IMERG and 7% for ERA5.

In our study, IMERG exhibited a POD of 61%, CSI of 44% and FAR of 39%, results far worse than the ones in the study of Xiong et al. It should be noted that the target variable in our study is based on ground observations of Tw, and not on ground observations of weather type. The predictor used by IMERG for rain-snow classification, however, is Tw.

The Gradient Boosting model used in this study achieved a POD of 81%, CSI of 72%, HSS of 85% and FAR of 13%. This is a significant improvement compared to most of the previously-described methods.

More specifically, compared to the PLPP product of IMERG as evaluated in our study, the Gradient Boosting model achieves an improvement of HSS by 24 percentage points. POD is improved by 20 percentage points and FAR by 26 percentage points.

Compared to the PLPP product of IMERG as evaluated by Xiong et al., POD of the Gradient Boosting model is worse by 6 percentage points, CSI is improved by 5 percentage points and FAR is improved by 3 percentage points. It should be noted, however, that the two studies use disparate target datasets.

In comparison with the results by Moon et al. and Sanò et al., HSS is improved by 10 and

15 percentage points respectively, however the target datasets between the three studies are also different. Moon et al. use ground observations of snowfall, Sanò et al. use Cloud Profiling Radar products, while this study uses ground observations of Tw.

The CSI score of the Gradient Boosting model is worse by 24 percentage points compared with the results by Shin et al. However, the latter study uses radar measurements and products with blended observations as input data for the ML model. Such an accurate illustration of the atmosphere is not available in our study, as the input data only consist of a short-term forecast from a numerical weather model.

The results of our study show that ML-based models, including DL ones, trained on numerical weather model data, offer a very promising approach for rain-snow partitioning. Using the Gradient Boosting model, 81% of the total favorable for snowfall conditions are identified, while 87% of such predicted cases are proved correct. The results present a significant improvement when compared to solely relying on the near-surface forecast of the numerical weather model (without altitude correction). In such instances, only 66% of the predicted cases are proven correct.

5.2 Insights From the Application on IMERG Precipitation Estimates

The map examples from real-world cases in 2022, generated by overlaying the ML model's output onto IMERG precipitation estimates, vividly demonstrate the effectiveness of the Gradient Boosting model in accurately depicting areas impacted by snowfall during cold waves.

For instance, in the case of January 24th, the model precisely identifies the extensive snowfall that occurred in Northeast Attica and Istanbul, creating a faithful illustration of the cold wave's impact on these regions.

Similarly, on January 29th, the model successfully retrieves the significant amounts of snowfall that affected the mountains of Peloponnese, providing an accurate representation of the weather event.

These compelling examples highlight the model's capability to precisely capture and visualize the spatial distribution of snowfall during cold waves, when combined with satellite precipitation estimates. The accuracy and reliability showcased in these real-world cases reinforce the model's practical utility for snowfall identification and its potential to enhance our understanding of weather patterns and their implications in specific regions.

5.3 Evaluation on 2022 Data From 11 Ground Stations

Utilizing a dataset spanning the first 3 months of 2022, we determine the precipitation phase for data beyond the training dataset's time span. This section predominantly focuses on station locations situated at relatively high altitudes, which experienced frequent snow-favorable conditions during this period. Among the models evaluated, the Gradient Boosting model and the Feedforward Neural Network exhibit the most promising performance, yielding nearly equal scores. These models demonstrate an impressive ability to detect 92% of the actual snow-favorable cases, with a striking 95% of all the predicted snow-favorable cases proved correct. This significant improvement stands in stark contrast to the performance of IMERG V06 PLPP, which detected only 58% of the actual

snow-favorable cases and 87% of all the predicted snow-favorable cases were proved correct.

Although BOLAM standalone outperforms IMERG V06 PLPP, it still lags behind the performance of the ML models. Notably, IMERG V06 PLPP seems to exhibit very poor performance when estimating precipitation phase in the areas of Samaria and Parnitha, potentially due to the complex topography in these regions, leading to grid points being located hundreds of meters lower than the actual station altitude. Remarkably, the ML models have learned to address this consistent underestimation, resulting in impressive performance for these specific locations too.

5.4 Limitations and Applications

Despite the promising results, we encountered certain limitations and challenges. One of the key challenges is the absence of a robust dataset of snowfall observations for Greece. To address this issue, near-surface Tw was used as the predictor determining precipitation phase, in the same way that it is currently utilized in IMERG V06. However, a different approach than the one currently in use by IMERG V06 was taken to derive whether the near-surface Tw exceeded a certain threshold or not.

In terms of applications, the developed snowfall mask holds promise for improving satellite precipitation estimates. By accurately identifying snowfall events, precipitation rate can be better estimated, while comprehensive snowfall datasets can be developed for various regions and time spans, helping provide response to snowfall events in near real-time, as well as better understand hydrological processes.

6. CONCLUSIONS

During this study an algorithm that is able to identify the precipitation phase of IMERG precipitation data was developed, leveraging Machine Learning models based on Random Forest and Gradient Boosting, and a Deep Learning model employing a Feedforward Neural Network. The 1.1°C wet-bulb temperature was used as an upper threshold for solid precipitation to occur over land. The results of our analysis indicate that the use of Machine Learning models is a very promising approach for estimating precipitation phase. Specifically, it was found that 81% of the actual snow-favorable conditions can be identified, while 87% of all the predicted snow-favorable conditions are proved correct. Application of the best-performing model's output on IMERG precipitation estimates from real-world cases, also shows that rain-snow partitioning on IMERG data yields comprehensive and reliable results.

The developed model's capability to accurately determine precipitation phase on satellite data, holds tremendous potential for near-real-time snowfall monitoring, providing valuable insights for emergency responses and aid distribution in areas affected by severe weather. Furthermore, this model opens up new possibilities for creating a thorough and enduring snowfall dataset, significantly enhancing our understanding of hydrological processes, supporting various water resource management initiatives and contributing to a deeper understanding of climate change impacts.

While our study has several strengths, it is not without limitations. For example, the wet-bulb temperature threshold applied to distinguish between solid and liquid precipitation on in-situ observational data, is not the optimal indicator for the actual precipitation phase. It is planned to make use of additional snowfall in-situ data from NOAAN in order to further evaluate the developed models.

ABBREVIATIONS - ACRONYMS

ML	Machine Learning
DL	Deep Learning
IMERG	Integrated Multi-satellitE Retrievals for GPM
PLPP	Probability of Liquid Precipitation Phase
NASA	National Aeronautics and Space Administration
Tw	Wet-Bulb Temperature
JMA	Japan Meteorological Agency
NWP	Numerical Weather Prediction
NOA	National Observatory of Athens
NOAAN	National Observatory of Athens Automatic Network
BOLAM	Bologna Limited-Area Model
POD	Probability Of Detection
CSI	Critical Success Index
FAR	False Alarm Ratio
HSS	Heidke Skill Score
ECMWF	European Centre for Medium-Range Weather Forecasts
RDAPS	Regional Data Assimilation and Prediction System
KMA	Korea Meteorological Administration
ERA5	ECMWF Reanalysis v5

REFERENCES

[1] S. Michaelides, V. Levizzani, E. Anagnostou, P. Bauer, T. Kasparis, and J. Lane, "Precipitation: Measurement, remote sensing, climatology and modeling," *Atmospheric Research*, vol. 94, no. 4, pp. 512–533, Dec. 2009.

[2] "Precipitation," in *Advanced Remote Sensing*. Elsevier, 2020, pp. 621–647.

[3] G. Skofronick-Jackson, M. Kulie, L. Milani, S. J. Munchak, N. B. Wood, and V. Levizzani, "Satellite Estimation of Falling Snow: A Global Precipitation Measurement (GPM) Core Observatory Perspective," *Journal of Applied Meteorology and Climatology*, vol. 58, no. 7, pp. 1429–1448, Jul. 2019.

[4] E. M. Sims and G. Liu, "A Parameterization of the Probability of Snow–Rain Transition," *Journal of Hydrometeorology*, vol. 16, no. 4, pp. 1466–1477, Jul. 2015.

[5] "Drainage Basins — cteco.uconn.edu," https://cteco.uconn.edu/guides/Basin.htm, [Accessed 01-08-2023].

[6] S. Froidurot, I. Zin, B. Hingray, and A. Gautheron, "Sensitivity of Precipitation Phase over the Swiss Alps to Different Meteorological Variables," *Journal of Hydrometeorology*, vol. 15, no. 2, pp. 685–696, Apr. 2014.

[7] L. Liao, R. Meneghini, T. Iguchi, and A. Detwiler, "Use of Dual-Wavelength Radar for Snow Parameter Estimates," *Journal of Atmospheric and Oceanic Technology*, vol. 22, no. 10, pp. 1494–1506, Oct. 2005.

[8] "How do the various forms of precipitation map into the IMERG "probabilityLiquidPrecipitation" data field? — gpm.nasa.gov," https://gpm.nasa.gov/resources/faq/how-do-various-forms-precipitation-map-imerg-probabilityliquidprecipitation-data, [Accessed 01-08-2023].

[9] A. Behrangi, X. Yin, S. Rajagopal, D. Stampoulis, and H. Ye, "On distinguishing snowfall from rainfall using near-surface atmospheric information: Comparative analysis, uncertainties and hydrologic importance," *Quarterly Journal of the Royal Meteorological Society*, vol. 144, no. S1, pp. 89–102, Aug. 2018.

[10] R. K. Pradhan, Y. Markonis, M. R. V. Godoy, A. Villalba-Pradas, K. M. Andreadis, E. I. Nikolopoulos, S. M. Papalexiou, A. Rahim, F. J. Tapiador, and M. Hanel, "Review of GPM IMERG performance: A global perspective," *Remote Sensing of Environment*, vol. 268, p. 112754, Jan. 2022.

[11] Y. You, C. Peters-Lidard, S. Ringerud, and J. M. Haynes, "Evaluation of Rainfall-Snowfall Separation Performance in Remote Sensing Datasets," *Geophysical Research Letters*, vol. 48, no. 21, Oct. 2021.

[12] G. Rebala, A. Ravi, and S. Churiwala, "Machine Learning Definition and Basics," in *An Introduction to Machine Learning*. Springer International Publishing, 2019, pp. 1–17.

[13] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, "A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science," in *Unsupervised and Semi-Supervised Learning*. Springer International Publishing, Sep. 2019, pp. 3–21.

[14] T. Qin, "Machine Learning Basics," in *Dual Learning*. Springer Singapore, 2020, pp. 11–23.

[15] R. J. Chase, D. R. Harrison, A. Burke, G. M. Lackmann, and A. McGovern, "A Machine Learning Tutorial for Operational Meteorology. Part I: Traditional Machine Learning," *Weather and Forecasting*, vol. 37, no. 8, pp. 1509–1529, Aug. 2022.

[16] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2016.

[17] T. Matsuo and Y. Sasyo, "Melting of Snowflakes below Freezing Level in the Atmosphere," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 59, no. 1, pp. 10–25, 1981.

[18] S. M. Lee, S. U. Han, H. Y. Won, J. C. Ha, Y. H. Lee, J. H. Lee, and J. C. Park, "A Method for the Discrimination of Precipitation Type Using Thickness and Improved Matsuo's Scheme over South Korea," *Atmosphere*, vol. 24, no. 2, pp. 151–158, Jun. 2014.

[19] G. Tang, D. Long, A. Behrangi, C. Wang, and Y. Hong, "Exploring Deep Neural Networks to Retrieve Rain and Snow in High Latitudes Using Multisensor and Reanalysis Data," *Water Resources Research*, vol. 54, no. 10, pp. 8253–8278, Oct. 2018.

[20] S. H. Moon and Y. H. Kim, "An improved forecast of precipitation type using correlation-based feature selection and multinomial logistic regression," *Atmospheric Research*, vol. 240, p. 104928, Aug. 2020.

[21] K. Shin, K. Kim, J. J. Song, and G. Lee, "Classification of Precipitation Types Based on Machine Learning Using Dual-Polarization Radar Measurements and Thermodynamic Fields," *Remote Sensing*, vol. 14, no. 15, p. 3820, Aug. 2022.

[22] D. Nadeau, "Operational Partitioning of Precipitation Phase Using Machine Learning — ui.adsabs.harvard.edu," https://ui.adsabs.harvard.edu/abs/2022AGUFM.H22J..06B/abstract, [Accessed 01-08-2023].

[23] P. Sanò, D. Casella, A. Camplani, L. P. D'Adderio, and G. Panegrossi, "A Machine Learning Snowfall Retrieval Algorithm for ATMS," *Remote Sensing*, vol. 14, no. 6, p. 1467, Mar. 2022.

[24] G. Huffman, E. Stocker, D. Bolvin, E. Nelkin, and J. Tan, "GPM IMERG Early Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06," 2019, [Accessed 01-08-2023].

[25] K. Lagouvardos, V. Kotroni, A. Bezes, I. Koletsis, T. Kopania, S. Lykoudis, N. Mazarakis, K. Papagiannaki, and S. Vougioukas, "The automatic weather stations NOANN network of the National Observatory of Athens: operation and database," *Geoscience Data Journal*, vol. 4, no. 1, pp. 4–16, Apr. 2017.

[26] K. Lagouvardos, V. Kotroni, A. Koussis, H. Feidas, A. Buzzi, and P. Malguzzi, "The Meteorological Model BOLAM at the National Observatory of Athens: Assessment of Two-Year Operational Use," *Journal of Applied Meteorology and Climatology*, vol. 42, no. 11, pp. 1667–1678, Nov. 2003.

[27] R. Minzner, C. Reber, L. Jacchia, F. Huang, A. Cole, A. Kantor, T. Keneshea, S. Zimmerman, and J. Forbes, "Defining constants, equations, and abbreviated tables of the 1975 US Standard Atmosphere — ntrs.nasa.gov," https://ntrs.nasa.gov/citations/19760017709, 1976, [Accessed 01-08-2023].

[28] É. Poirier, J. M. Thériault, and M. Leriche, "Role of sublimation and riming in the precipitation distribution in the Kananaskis Valley, Alberta, Canada," *Hydrology and Earth System Sciences*, vol. 23, no. 10, pp. 4097–4111, Oct. 2019.

[29] P. E. Taylor and H. Jonsson, "Thunderstorm asthma," *Current Allergy and Asthma Reports*, vol. 4, no. 5, pp. 409–413, Sep. 2004.

[30] H. R. Byers, "Structure and Dynamics of the Thunderstorm," *Science*, vol. 110, no. 2856, pp. 291–294, Sep. 1949.

[31] G. Huffman, "Integrated Multi-satellitE Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD) — gpm.nasa.gov," https://gpm.nasa.gov/resources/documents/algorithm-information/IMERG-V06-ATBD, 2020, [Accessed 01-08-2023].

[32] S. Dafis, C. J. Lolis, E. E. Houssos, and A. Bartzokas, "The atmospheric circulation characteristics favouring snowfall in an area with complex relief in Northwestern Greece," *International Journal of Climatology*, vol. 36, no. 10, pp. 3561–3577, Dec. 2015.

[33] A. Parmar, R. Katariya, and V. Patel, "A Review on Random Forest: An Ensemble Classifier," in *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*. Springer International Publishing, Dec. 2018, pp. 758–763.

[34] J. H. Friedman, "Greedy function approximation: A gradient boosting machine." *The Annals of Statistics*, vol. 29, no. 5, Oct. 2001.

[35] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43–62, Nov. 1997.

[36] E. B. Baum, "On the capabilities of multilayer perceptrons," *Journal of Complexity*, vol. 4, no. 3, pp. 193–215, Sep. 1988.

[37] "NOAA Forecast Verification Glossary — https://www.swpc.noaa.gov/," https://www.swpc.noaa.gov/sites/default/files/images/u30/Forecast%20Verification%20Glossary.pdf, [Accessed 01-08-2023].

[38] X. Shi, Y. D. Wong, M. Z. F. Li, C. Palanisamy, and C. Chai, "A feature learning approach based on XGBoost for driving assessment and risk prediction," *Accident Analysis & Prevention*, vol. 129, pp. 170–179, Aug. 2019.

[39] P. Patlakas, I. Chaniotis, M. Hatzaki, J. Kouroutzoglou, and H. A. Flocas, “The eastern Mediterranean extreme snowfall of January 2022: synoptic analysis and impact of sea-surface temperature,” *Weather*, Apr. 2023.

[40] W. Xiong, G. Tang, T. Wang, Z. Ma, and W. Wan, “Evaluation of IMERG and ERA5 Precipitation-Phase Partitioning on the Global Scale,” *Water*, vol. 14, no. 7, p. 1122, Mar. 2022.