Memorias de SICEL 2023
UNIVERSIDAD NACIONAL DE COLOMBIA

Predicción de días de tormenta en el territorio chileno usando inteligencia artificial

Prediction of thunderstorm days in Chilean territory using machine learning techniques

Johny Montaña[1], Sergio Rosales-Baros[2], Carlos Valle[3]
Sergio Zumaran-Rivera [4]

[1]Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, Chile. Orcid: 0000-0002-9999-2366. correo electrónico: johny.montana@usm.cl
[2]Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, Chile. Orcid: 0009-0006-7821-6102. correo electrónico: sergio.rosales.12@sansano.usm.cl
[3]Departamento de Ciencia de Datos e Informática, Universidad de Playa Ancha, Chile. Orcid: 0000-0001-7158-2069. correo electrónico: carlos.valle@upla.cl
[4]Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, Chile. Orcid: 0000-0002-8742-8688 correo electrónico: sergio.zumaran.12@sansano.usm.cl

Recibido: 08/06/2023. Aceptado: 24/08/2023. Versión final: 15/09/2023

Resumen

Este estudio presenta el diseño de un modelo que permite la predicción de días de tormenta eléctrica. El estudio se desarrolló en una localidad dentro Chile, con la restricción de que sea una ubicación que haya registrado una alta cantidad de descargas atmosféricas y cuente con los suficientes registros meteorológicos entre los años 2012 y 2021 para poder entrenar una máquina de aprendizaje automático. La localidad escogida fue Visviri, región de Arica y Parinacota. Se empleó una metodología que consistió en crear un conjunto de datos, la realización de un Análisis Exploratorio de Datos, la imputación de datos no disponibles, la selección de atributos, la reducción de la dimensional, búsqueda de hiper parámetros y un análisis de sensibilidad para el modelo que presente el mejor rendimiento basado en la puntuación F1. El modelo realizado consistió en una red neuronal multicapa con función de activación ReLU y una tasa de dropout del 20% que alcanzó un desempeño del 74.68% en la puntuación F1 para el año 2021.

Palabras clave: días de tormenta, inteligencia artificial, redes neuronales, análisis inteligente de datos.

Abstract

This study presents the design of a model that allows for the prediction of thunderstorm days. The study was conducted in a location within Chile, with the requirement that it be an area that has recorded a high number of lightning and has sufficient meteorological records between the years 2012 and 2021 to train a machine learning model. The chosen location was Visviri, in the Arica and Parinacota region. A methodology was employed, which included creating a dataset, conducting Exploratory Data Analysis, missing data imputation, feature engineering, feature selection, hyperparameter tuning, and a sensitivity analysis to find the best-performing model based on the F1 score. The model

How to cite: J. Montaña, S. Rosales, C. Valle, S. Zumarán, "Prediction of thunderstorm days in Chilean territory using machine learning techniques," in *XI Simposio Internacional de Calidad de la Energía Eléctrica*, Valledupar: Universidad Nacional de Colombia, Nov. 2023. doi: https://doi.org/10.15446/sicel.v11.109897

developed was a multilayer neural network with a ReLU activation function and a 20% of dropout rate, achieving a performance of 74.68% in F1 score for the year 2021.

Keywords: Thunderstorm days, machine learning, neural networks, intelligent data analysis.

1. Introduction

The study of lightning produced during a storm is a topic that has captured the interest of several researchers worldwide. In this context, predicting this natural phenomenon is a multidisciplinary problem that, although traditionally approached through the analysis of the physical model, an alternative that has gained recent relevance in recent years due to computational advancements is the use of machine learning techniques using artificial intelligence. These machines can address a problem without having to model the physical phenomenon.

During a review of the current state of research, it was found that studies have already been conducted in other countries. This, to be able to develop a model for Chilean territory. One such study was conducted in Sri Lanka at the University of Moratuwa, where a low-cost alert system was developed [1]. The system used a perceptron with two inputs: inter-cloud radiofrequency signals emitted by lightning and static electric field measurements obtained using a homemade electric field mill. The output of the perceptron determined the level of threat (high, medium, or none), and the fully trained model achieved satisfactory results. However, the study did not provide further information about the activation function of the perceptron.

In Malaysia, several models have been developed, including a neural network designed for the city of Subang Jaya [2]. This network took as input 24 variables, such as wind, dew point, humidity, pressure, and others, along with one indicator for each month and season of the year. It had two hidden layers (8 neurons in the first layer and 5 in the second) using a logarithmic sigmoid activation function, and a linear output indicating the occurrence or non-occurrence of lightning. The Levenberg-Marquardt algorithm was used for training, with a learning rate of 0.4819 and a momentum constant of 0.0577. The training dataset consisted of 378 data points, while 197 data points were used for testing. The training error was calculated using the root mean square (RMS) error, which yielded an error rate of 0.41%. The fully trained model achieved a performance of 99.997% in terms of correlation between the expected value and the output of the network. However, the time window for data measurements or the imbalance percentage of the target variable was not provided.

Another Malaysian study focused on a Multilayer Perceptron (MLP) developed for the Kuala Lumpur International Airport [3]. This MLP took as input 5 variables: dry air temperature (Ts), relative humidity (HR), sea-level air pressure (QFF), wind speed (ff), and precipitation (RRR). It had one hidden layer using 35 neurons and a hyperbolic tangent sigmoid activation function, while the output represented the number of lightning occurrences. The Levenberg-Marquardt algorithm was also used for training, with a learning rate of 0.08 and a momentum constant of 0.95. The training dataset consisted of 288 data points (from January 2010 to December 2013), and 72 data points were used for testing (from January 2015 to December 2015). All the data were obtained from the Malaysian Meteorological Department. The training error, calculated using the RMS error, was 0.0786%. The correlation between the expected value and the network's output was 99.999%, and the fully trained model achieved a performance of 94.64% in terms of correlation. However, the imbalance percentage of the target variable was not provided.

Another Malaysian study involved another MLP developed for the Sultan Abdul Aziz Shah Airport located in Subang [4]. This MLP took as input 5 variables (Ts, HR, QFF, ff, and RRR), had 1 hidden layer, and had an output indicating the number of lightning occurrences. The training dataset consisted of 360 data points (from January 2010 to December 2014), and 72 data points were used for testing (from January 2015 to December 2015). The training error, calculated using the RMS error, was 11.05%. The correlation between the expected value and the network's output was 99.990%, and the fully trained model achieved a performance of 98.718% in terms of correlation. However, the confusion matrix results was not presented.

In a different study, a machine learning model was developed for 12 locations in Switzerland, including Säntis and Monte San Salvatore mountains [5]. This model took as input 4 meteorological variables (station-level air pressure (QFE), air temperature at 2 meters above ground level, HR, and ff) and had an output indicating the occurrence of lightning. Data recorded every 10 minutes from 2006 to 2017 were used for training. This machine learning model was compared to a persistence forecasting method, a model based on the electrostatic field method, and a scheme based on the threshold of available potential energy for convection. The machine learning model yielded the best results among the compared methods. Unfortunately, no additional information was provided regarding the network architecture, the source of meteorological data, the training dataset, compilation, optimizers, or the loss function.

Another study involved an ensemble of neural networks and decision trees developed for the cities of Mashhad, Neyshabur, and Quchan in the Razavi Khorasan province, Iran [6]. The dataset used in this study was characterized by class imbalance, which was addressed through under sampling techniques. The network took as input temporal data (year, month, day, and hour) and meteorological data (visibility, cloudiness, wind direction, wind speed, Ts, dew point temperature (Td), QFE, QFF, RRR, Ts, HR, low cloudiness, type of low clouds, among others). It consisted of one hidden layer with a sigmoid activation function for the neural network and a hard-limit activation function for the output, which indicated the presence of lightning. The data used for training and testing ranged from 1992 to 2018. The dataset was randomly split, with 85% for training and validation (70% training + 15% validation), and the remaining 15% for testing. All the data was obtained from the Iranian Meteorological Organization, with a time window of 3 hours. According to the results obtained, the decision tree outperformed the neural networks in all the datasets. The best fully trained model achieved an F1 score performance of 86.8% for the Mashhad dataset, 85.6% for Neyshabur, and 85.6% for Quchan. Unfortunately, no further information was provided about the network compilation, optimizers, or loss function.

2. Data and Methodology

2.1. Meteorological data

The Meteorological Directorate of Chile has 1423 automatic weather stations throughout Chile. These stations capture different meteorological variables, such as QFF, HR, Ts, radiation, among others. Considering that there is higher electrical activity in the northern part of the Chilean territory, the stations in that region with the highest amount of meteorological information were considered. The variables measured for a station are listed in Table 1.

2.2. Exploratory Data Analysis

It is common practice to perform a descriptive analysis of the data to obtain an overview of the distribution of features, check for outliers, missing values, or other anomalies, for discover patterns or relationships among variables in the dataset that can aid in predicting stormy or non-stormy days for the selected location within the Chilean territory.

The target values are called "Strokes". This feature should only record 2 possible values, indicating the occurrence or non-occurrence of a stroke. For binary classification problems, it is common to use 1 and -1 as values, respectively. Subsequently, a descriptive analysis of the numerical characteristics of two variables is performed separately for cases of stormy weather and clear skies, aiming to describe the numerical data for these scenarios.

An explanatory analysis of the numerical features is conducted to identify cause-effect relationships among the observed characteristics. Multivariate analysis and correlation between variables are fundamental statical tools for developing and understanding of the numerical features.

Table 1. Meteorological variables available for a station.

Meteorological variables	
Date and time, Datetime	(dd/mm/yyyy h)
Accumulated precipitation in 1 hour	RRR6 (mm)
Relative humidity	HR (%)
Atmospheric pressure at station level	QFE (hPa)
Sea-level atmospheric pressure	QFF (hPa)
Sea-level atmospheric pressure using International Civil Aviation Organization (ICAO) Standard Atmosphere	QNH (hPa)
Instantaneous Global Solar Radiation	RadGInst (W/m2)
Dew point temperature	Td (°C)
Dry air temperature	Ts (°C)
Wind direction at 10m height	dd10m (°)
Wind speed at 10m height	ff10m (kt)
Average wind direction every 2 minutes	dd2min (°)
Average wind speeds every 2 minutes	ff2min (kt)
Wind direction at 2m height	dd2m (°)
Wind speed at 2m height	ff2m (kt)

Source: Own elaboration.

2.3. Preprocessing the dataset

Preprocessing the dataset transforms the data into a format that machine learning algorithms can understand. Raw data is not directly comparable, except within the same feature. It may contain errors, inconsistencies, or missing values.

2.3.1. Outliers

A time series plot of the records was created, and any values outside the normal range were identify.

2.3.2. Feature engineering

This section uses existing features to create new features that provide new information and help with the imputation process, taking into account the context of the dataset (domain knowledge). For example, since Chile is subject to constant westward winds, it is important to generate a new feature for this. When the remaining data can fill in the gaps, the consensus value is usually the mean of these recorded values. Additionally, it was observed that the data was collected throughout the year, so it would be interesting to capture each season by approximating the dates of solstices and equinoxes to the nearest hour. Considering the week of the year and the time of day as features may also improve the model's performance. The last 2 new features have both high cardinality and periodicity between their extremes, so they were decomposed into a Cartesian plane. Finally, a feature cleaning process was performed. Some features can be contained within others, so it would generate double correlation during the algorithm training. These redundant features were removed, leaving only the higher-quality ones. In general, Td is composed of Ts and HR. When Td is equal to Ts, it means that HR is 100%. To avoid duplicity of information, Td was not used whenever HR could be retained.

2.3.3. Unavailable data

A hierarchical order was established for filling in the missing values. First, for values that were unavailable for more than one consecutive week in the middle of a datasets, the average value between the data from the previous and following year was used. Second, for almost all features, the value from the previous day at the same time was used, except for those where persistence can compromise the quality of the dataset. Finally, the k-nearest neighbors imputation method was used to complete the filling of missing values. This method is useful when there is no guarantee of correlation between data [7].

2.3.4. Data standardization

The objective of this section is to establish a criterion for normalizing the data to avoid overfitting. A heuristic approach was used to evaluate the characteristics of the created dataset. It was decided to apply a standard normalization to features whose extreme values do not exceed 3 times their standard deviation. The remaining features were normalized using the min-max method. It was done to ensure that all features are on a similar scale, as many machine learning algorithms are sensitive to the scale of the input features.

2.4. Dataset

Due to the amount of data, the last year (2021) was used to test the model, and the remaining years were used to train the model. The strategy was to first test 2 libraries: Scikit-learn and TensorFlow. The best model was then chosen using stratified k-fold cross-validation for different lag quantities. These models were called base models, and their hyperparameters were the adjusted. The loss function chosen was mean squared error, which measures how closely the model approximates the data.

The performance of each model was evaluated based on its errors, F1 scores, and total training time. The best model was selected based on the average F1 scores from stratified k-fold cross-validation due to is a useful technique for evaluating models when the data is imbalanced. [8]. For TensorFlow models, a maximum of 100 training epochs was defined and an EarlyStopping callback was used to prevent overfitting [9]. After developing the model, a search for hyperparameters that improve the algorithm's success was performed using the F1 metric through Greed search. Greed search is a technique that searches for hyperparameters that improve the model's performance one at a time. The search was limited to a small set of hyperparameters due to hardware limitations and the time constraints.

After that, an ablation or sensitivity analysis was performed to fine-tune and create a better model. The year 2021 was used to test the model. Additionally, since the goal is to identify thunderstorm days, the information was post-processed. This involved resampling the test set every 24 hours, taking the maximum value, and evaluating the model using the confusion matrix and F1 score.

Table 2. Descriptive statistics for Visviri

Statistics									
	RRR6	HR	QFE	QFF	QNH	RadGInst	Td	Ts	Cos dd10m
Quantity	74352	71925	71926	57790	58861	73417	58864	71924	58864
Mean	0.03	38.18	627.21	1028.21	1036.46	286.66	-9.8	6.18	-0.15
Deviation	0.17	25.49	1.53	13.26	2.41	391.04	9.18	7.12	0.65
Minimum	0.0	1.0	617.3	830.8	845.2	0.0	-38.7	-19.1	-1.0
25%	0.0	17.0	626.2	1017.8	1035.0	0.0	-16.6	1.5	-0.75
50%	0.0	31.0	627.3	1028.8	1036.5	5.5	-10.0	6.0	-0.28
75%	0.0	56.6	628.3	1037.2	1038.0	574.0	-1.8	11.9	0.42
Maximum	3.7	99.0	632.7	1060.0	1044.7	1634.0	8.6	23.2	1.0
	sin dd10m	ff10m	cos dd2min	sin dd2min	ff2min	cos dd2m	sin dd2m	ff2m	Strokes
Quantity	58864	58864	60295	60295	60295	24889	24889	24889	75844
Mean	-0.33	6.21	-0.11	-0.32	6.15	0.22	-0.32	4.34	-0.93
Deviation	0.66	5.41	0.68	0.65	5.53	0.62	0.68	3.8	0.37
Minimum	-1.0	0	-1.0	-1.0	0	-1.0	-1.0	0	-1
25%	-0.88	2	-0.75	-0.87	2	-0.28	-0.95	2	-1
50%	-0.59	4	-0.22	-0.56	4	0.21	-0.56	3	-1
75%	0.03	9	0.52	0.0	9	0.91	0.0	7	-1
Maximum	1.0	29	1.0	1.0	33	1.0	1.0	21	-1

Source: Own elaboration.

3. Results

3.1. Selected location

The best location for this study it was Visviri Station, which is located at an altitude of 4084 meters above mean sea level. This station has meteorological information available since 2013.

3.2. Descriptive analysis of temporal features

In Table 2, it can observe the descriptive statistics that summarize the quantity, central tendency, and variability of the available data for the location of Visviri. Notice that low values of HR are reflected in negative values of Td. In Table 3, you can see a summary with the cardinality, skewness, and kurtosis values to describe the distribution shape of each recorded feature. Figure 1 shows a heatmap of the correlations between the recorded features for the location of Visviri.

The results of a bivariate analysis show that, storms are more likely to occur in the last quarter of the day. Clear skies can occur at any time of day and in any week of the year. Summer is the season with the highest occurrence of storms, with 10.01% of days having storms. The remaining percentages are 1.98% for autumn, 0.47% for winter, and 2.56% for spring.

Table 3. Cardinality, skewness, and kurtosis of the features in Visviri

Statistics			
	Cardinality	Skewness	Kurtosis
RRR6	171	9.26	111.58
HR	2292	0.66	-0.75
QFE	283	-0.17	-0.03
QFF	639	0.03	0.23
QNH	162	-8.59	669.58
RadGInst	14530	1.08	-0.23
Td	446	-0.18	-0.82
Ts	903	-0.24	-0.46
Cos dd10m	323	0.39	-1.24
sin dd10m	330	0.84	-0.75
ff10m	30	1.31	1.09
cos dd2min	323	0.34	-1.33
sin dd2min	330	0.79	-0.74
ff2min	34	1.33	1.29
cos dd2m	323	-0.22	-1.13
sin dd2m	330	0.61	-0.99
ff2m	22	0.90	0.13
Strokes	2	5.02	23.19

Source: Own elaboration.

3.3. Results of preprocessed data

Many recorded values have zero magnitude, so the standard deviation had to be increased several times to isolate the outliers. The values that were kept were those within the 0.1% and 99.9% intervals. In this case, there was only 1 outlier value that was replaced with the mean value of the surrounding hours.

Additionally, there are 3 subscripts associated with wind features. Wind data obtained at higher altitudes is of higher quality, so a good strategy is to give more weight to this higher-quality data than to the other two.

Table 4 shows the percentage of missing data for each features in the Visviri location. And the final features amount to 19, resulting in a dataset split in 67,084 values for training and 8,760 for testing.

3.4. Cause-effect relationship

After performing this adjustment in the feature engineering, the new correlation map is shown in Figure 2. The strongest correlations are between sin_hour and Temperature, sin_hour and Radiation, and Temperature and Radiation. This is because the Temperature increases when sunlight appears during the day. The best correlations of storms are with summer, sin_dd, temperature, and pressure.

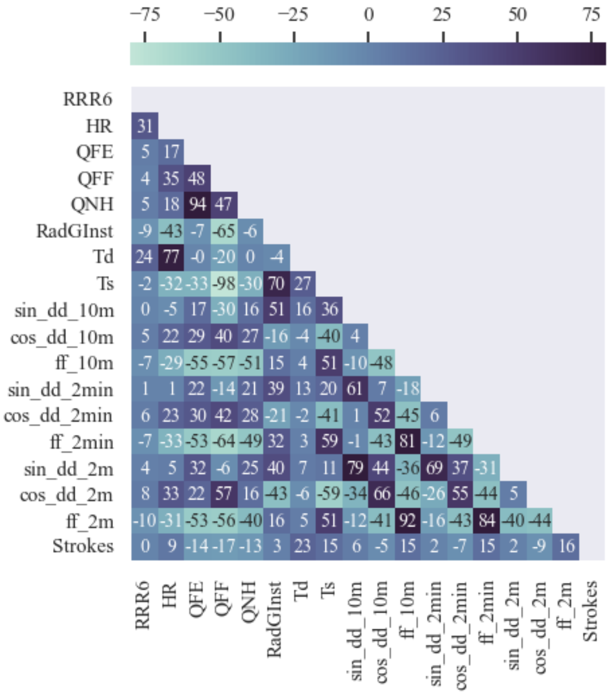


Figure 1. Heatmap for Visviri. Source: Own elaboration.

Table 4. Amount of missing data for Visviri

Unavailable data		
	Data	%
QFF	18054	23.8
QNH	16983	22.4
Td	16980	22.4
Ts	3920	5.2
HR	3919	5.2
QFE	3918	5.2
EO	3917	5.2
NS	3917	5.2
ff	3917	5.2
cos_dd	3917	5.2
sin_dd	3917	5.2
RadGInst	2427	3.2
RRR6	1492	2.0

Source: Own elaboration.

3.5. Temporal series analysis

The Dickey-Fuller test showed that all the features in the dataset are at least 99% stationary, which makes them suitable for models that learn from the past. Autocorrelation analysis revealed that the best correlations occur every 24 hours and then approximately every 8760 hours (1 year).

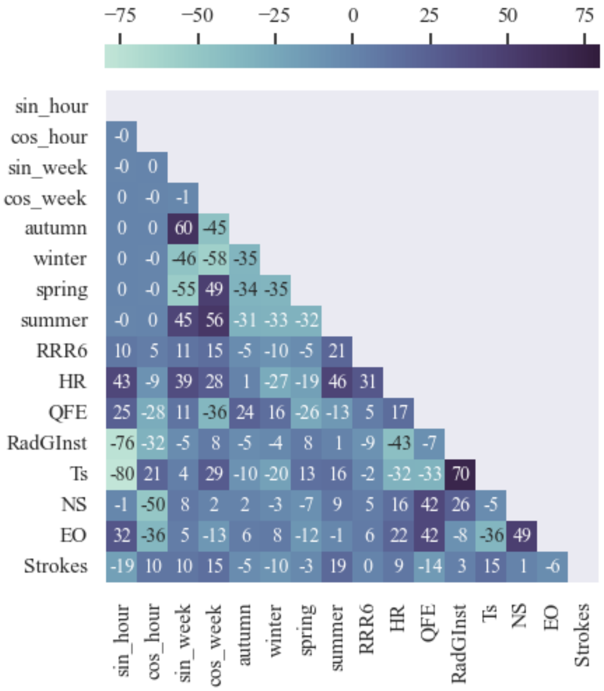


Fig. 2. Final correlations for the dataset. Source: Own elaboration.

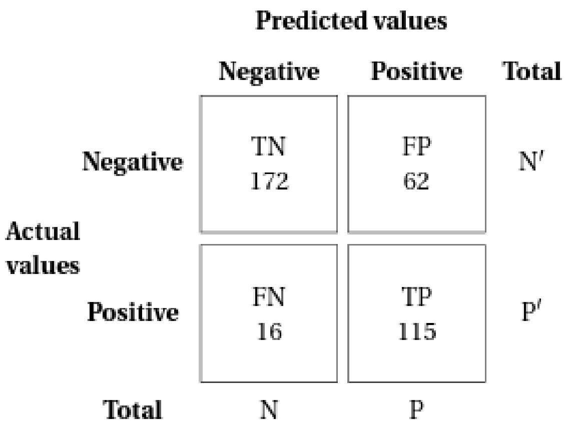
3.6. Model selection and hyperparameter tuning

Machine learning models from Scikit-learn and TensorFlow were trained using stratified k-fold cross-validation for different lag quantities (0h, 24h, 48h, 72h, 96h). The best-performing model, on average, was a multilayer classifier without lag in its climatic features. Hyperparameter tuning was performed on this model, and the results showed that the model with the lowest training error had 2 hidden layers, a hyperbolic tangent activation function, and 500 neurons.

To improve the performance of this model, a dropout layer was added, and the tuning process was repeated. Dropout layers randomly remove neurons from layers to avoid overfitting [10]. This meant moving away from Scikit-learn and using TensorFlow's dense layers. The sensitivity analysis showed that the best model had 5 dense layers with a 10% dropout rate, 500 neurons per layer, and a ReLU activation function. This model had an F1 score of 0.6218. Finally, a model composed of 30 combinations of dense layers with a 20% dropout rate, 100 neurons per layer, and a ReLU activation function was obtained. This model had an F1 score of 0.8046.

3.7. Final results

The model requires post-processing to identify thunderstorm days. A day is considered a thunderstorm day if at least one hour in that day is predicted to be a thunderstorm. To do this, the predicted values are resampled every 24 hours, and the maximum value is taken. The confusion matrix results for the test set shown in Figure 3. A confusion matrix is a table that summarizes the performance of a classification model [11].



		Predicted values		
		Negative	Positive	Total
Actual values	Negative	TN 172	FP 62	N′
	Positive	FN 16	TP 115	P′
	Total	N	P	

Fig. 3. Confusion matrix. TP: true positive, TN: true negative, FP: false positive, FN: false negative. Source: Own elaboration.

Finally, after an ablation analysis, this network was trained on temporal and meteorological input variables. The best model had 30 combinations of dense layers with 100 neurons and a ReLU activation function, a dropout layer with a 20% dropout rate, and an output layer with a sigmoid function. The best fully trained model achieved a performance of 41.99% in F1 score with raw data and 80.46% with post-processed data on the validation set. For the test set, the results were 33.47% in F1 score with raw data and 74.68% with post-processed data.

4. Discussion

4.1. Descriptive analysis of meteorological features

The section describes the meteorological features that contribute to the formation of storms. There is no clear relationship between precipitation and the conditions that favor storms or clear skies. However, relative air humidity is higher during storms (50.27% in average) than during clear skies (37.37% in average). This is because humidity is a relevant factor for cloud formation. Pressure is lower during storms (626.08 hPa in average) than during clear skies (627.27 hPa in average). This is because pressure decreases due to the presence of clouds, which is a factor for storm occurrence.

The average radiation during storms is 348.14 W/m2, while during clear skies is 281.64 W/m2. These may seem contradictory to cloud formation, but the nighttime hours, which have zero values, skew the results. When nighttime hours are excluded, the average radiation during storms is 388.20 W/m2, and during clear skies it is 522.49 W/m2.

A similar phenomenon occurs with temperature. The average temperature during storms is 11.78 °C, while during clear skies, it is 5.79°C. However, when nighttime hours are excluded, the average temperature during storms is 12.42°C, and during clear skies it is 9.61 °C.

The average wind speed during storms is 9.70 kt, with a southwest direction. During clear skies, the average wind speed is 5.69 kt, with a west direction. This is expected as wind is a contributing factor to storm formation. The direction of the wind may not change significantly, the intensity does.

4.2. Temporal series analysis of dataset

This means that each feature can be shifted by multiples of 24-hour before the next best correlation occurs around 8760 hours. Although using shifts that occur after a year allows training the annual patterns in the data, the size of the dataset become larger. The limit before the next best correlation occurs at approximately one year for each feature were as follows: RRR6 no more than 4 days, HR no more than 25 days, QFE no more than 60 days, RadGInst could reach up to 365 days, Ts no more than 66 days, NS no more than 60 days, and EO no more than 60 days.

4.3. Improving the best model

The key features that significantly influenced the model's performance metrics were relative humidity, station-level pressure, radiation, dry air temperature, easterly wind strength with 24- and 48-hour delays (lags), and Cartesian decomposition of day and time. Analyzing the confusion matrix in Figure 3 is evident that the number of true negatives (TN) is consistently higher than the number of true positives (TP), reflecting the class imbalance where non-thunderstorm days outnumber thunderstorm days. However, the model's false negatives (FN) rate is also high, which is a critical problem. Predicting no thunderstorm when one occurs can lead to negligent planning of outdoor activities and pose risks to public health and emphasize the importance of accurate predictions to avoid compromising public health and ensure proper planning and precautionary measures.

5. Conclusions

In this study, two libraries with several artificial intelligence models were compared. The hyperparameters were adjusted to optimize their performance. The dense neural network model outperformed all the other models. To further enhance the performance of the dense network model, an ablation technique was applied, with an improved performance and a reduction in critical failures.

Based on these findings, an algorithm was developed to predict thunderstorm days. The algorithm achieved an approximate 74.68% performance metric on the test dataset spanning the entire year 2021. The confusion matrix exhibited a low number of failures, particularly in the critical failure category (FN), which was relatively low compared to the overall success rate. The predictor algorithm shows promising results, as it provides an estimation of thunderstorm occurrences with a minimal percentage of failures. Additionally, the algorithm ensures physical coherence by establishing relationships between the predicted thunderstorm days.

Acknowledgments

The authors would like to thank to Universidad Técnica Federico Santa María for supporting this study through the research project PI-LIR- 2022-14.

References

[1] G. Jayendra, R. Lucas, S. Kumarawadu, L. Neelawala, C. Jeevantha, P. Dharmapriya, “Intelligent lightning warning system”. Third international conference on information and automation for sustainability, Melbourne, 2007, p.19-24. doi:10.1109/ICIAFS.2007.4544774

[2] D. Johari, T. Rahman, I. Musirin, “Artificial neural network based technique for lightning prediction”. 5th student conference on research and development, Malaysia, 2007, p.1-5. doi:10.1109/SCORED.2007.4451448

[3] M. Ramzi, R. Adnan, A. M. Samad, F. Ruslan, “Lightning prediction modelling using mlpnn structure. case study: Kuala lumpur international airport (klia)”. IEEE international conference on automatic control and intelligent systems (i2cacis), Malaysia ,2018, p. 63-66. doi:10.1109/I2CACIS.2018.8603704

[4] N. H. Abdullah, R. Adnan, A. M. Samad, F. Ahmat Ruslan, “Lightning forecasting modelling using artificial neural network (ann): Case study sultan Abdul Aziz Shah airport or skypark subang”, IEEE conference on systems, process and control (icspc), Malaysia, 2018, p.1-4. doi:10.1109/SPC.2018.8704147.

[5] A. Mostajabi, D. L.Finney, M. Rubinstein, F. Rachidi, “Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques”, NPJ Clim. Atmos. Sci, 2, 41, 2019. From: https://www.nature.com/articles/s41612-019-0098-0/ doi:https://doi.org/10.1038/s41612-019-0098-0.

[6] M. Pakdaman, S. Naghab, L. Khazanedari, S. Malbousi, Y. Falamarzi, “Lightning prediction using an ensemble learning approach for northeast of Iran”, Journal of Atmospheric and Solar-Terrestrial Physics, vol. 209, no. 105417, 2020. doi:https://doi.org/10.1016/j.jastp.2020.105417

[7] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, R. Altman, “Missing value estimation methods for DNA microarrays Bioinformatics”, vol. 17, no. 6, pp. 520-525, 2001. doi: 10.1093/bioinformatics/17.6.520

[8] S. Widodo, H. Brawijaya, S. Samudi, “Stratified k-fold cross validation optimization on machine learning for prediction”, Sinkron : Jurnal dan Penelitian Teknik Informatika, vol. 7, no. 4, pp. 2407-2414, 2022. doi: 10.33395/sinkron.v7i4.11792

[9] F. Girosi, M. Jones, T. Poggio, “Regularization theory and neural networks architectures”, Neural Computation, vol. 7, no. 2, pp. 219-269, 1995. doi: 10.1162/neco.1995.7.2.219

[10] X. Liang, L. Wu, J. Li, Y. Wang, Q. Meng, T. Qin, T. Liu, “R-drop: Regularized dropout for neural networks”, 35th Conference on Neural Information Processing Systems, Virtual, 2021.

[11] C. C. Aggarwal, Data classification: Algorithms and applications. CRC Press, 2015.