Article

Alnus Airborne Pollen Trends during the Last 26 Years for Improving Machine Learning-Based Forecasting Methods

María Novo-Lourés [1,2], María Fernández-González [3,*], Reyes Pavón [1,2], Kenia C. Sánchez Espinosa [3], Rosalía Laza [1,2], Guillermo Guada [3], José R. Méndez [1,2], Florentino Fdez-Riverola [1,2] and Francisco Javier Rodríguez-Rajo [3]

[1] CINBIO, Department of Computer Science, ESEI—Escuela Superior de Ingeniería Informática, University of Vigo, 32004 Ourense, Spain; manovo@uvigo.es (M.N.-L.); pavon@uvigo.es (R.P.); rlaza@uvigo.es (R.L.); moncho.mendez@uvigo.es (J.R.M.); riverola@uvigo.es (F.F.-R.)

[2] SING Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, 36213 Vigo, Spain

[3] Department of Plant Biology and Soil Sciences, Faculty of Sciences, University of Vigo, 32004 Ourense, Spain; ksanchez8909@gmail.com (K.C.S.E.); guillermo.guada@uvigo.gal (G.G.); javirajo@uvigo.es (F.J.R.-R.)

* Correspondence: mfgonzalez@uvigo.es

Abstract: Black alder (*Alnus glutinosa* (L.) Gaertn.) is a species of tree widespread along Europe and belongs to mixed hardwood forests. In urban environments, the tree is usually located along watercourses, as is the case in the city of Ourense. This taxon belongs to the betulaceae family, so it has a high allergenic potential in sensitive people. Due to the high allergenic capacity of this pollen type and the increase in global temperature produced by climate change, which induces a greater allergenicity, the present study proposes the implementation of a Machine Learning (ML) model capable of accurately predicting high-risk periods for allergies among sensitive people. The study was carried out in the city of Ourense for 28 years and pollen data were collected by means of the Hirst trap model Lanzoni VPPS-2000. During the same period, meteorological data were obtained from the meteorological station of METEOGALICIA in Ourense. We observed that *Alnus* airborne pollen was present in the study area during winter months, mainly in January and February. We found statistically significant trends for the end of the main pollen season with a lag trend of 0.68 days per year, and an increase in the annual pollen integral of 112 pollen grains per year and approximately 12 pollen grains/m^3 per year during the pollen peak. A Spearman correlation test was carried out in order to select the variables for the ML model. The best ML model was Random Forest, which was able to detect those days with medium and high labels.

Keywords: *Alnus*; pollen; machine learning; random forest; trends

1. Introduction

Alnus glutinosa (L.) Gaertn., commonly known as black alder, is naturally widespread across Europe, from mid-Scandinavia to the Mediterranean countries, including northern Morocco and Algeria [1,2]. It is usually part of mixed hardwood forests and represents less than 1% of the forest cover in most countries [3]. Despite this low percentage of representation in forests, black alder has a good potential for timber production [1]. On the other hand, black alder is important in open landscapes; for example, in the city of Ourense, *Alnus* is located along the riverbank, as in other European cities where alders are equally represented in woodland and in linear features along watercourses [4,5]. As the Miño river flows through the city, this species approaches the sensitive population that lives in the city, causing many people to present allergic symptoms to this pollen type.

Nowadays, allergies have become an important public health problem in the urban environment of industrialised countries. This sickness is considered a global pandemic with a great impact on the worldwide economy [6,7]. In general, a higher prevalence of cases

of sensitization related to pollen was detected in urban environments compared to rural areas [8,9]. In addition to individual genetic predispositions, a recent study has identified environmental factors, such as climate change [10], that contribute to allergic reactions. An increase in temperature will modify the flowering process, inducing changes in the beginning of flowering, as earlier blooms will modify the duration of the pollen season [11]. The expected amount of pollen production will depend on the plant species, the time of the year and the region in question, since climate change produces different effects depending on the bioclimatological region [11,12]. Associated with climate change, pollen allergies are expected to change their pattern and/or intensity due to early flowering, changes in pollen production and the invasion of new allergenic plants [13]. Several authors around the world have pointed out that in the long term, there may be large variations in the local pollen season, as well as in the onset of the season, seasonal pollen integral (SPIn) and season duration, which fluctuate between years depending on local climatic variations and climate change that affects the growth of species [14–25]. In addition, meteorological variables can produce a direct effect on the airways and induce asthma, and an indirect effect through airborne allergens and pollutant levels. For this reason, the climatic change and greenhouse gases influence the frequency and severity of respiratory allergies [26]. Several researchers found that the sensitivity of a population depends on the local flora (and pollen), concluding that if a plant species is widespread in a geographical area, the population can tolerate a higher concentration of this pollen before the symptoms appear [27,28]. Different authors pointed out that the presence of a certain plant species affects the tolerance of the population and the appearance of symptoms, such as *Betula* [29,30], *Phleum pratense* [30] or *Dactylis glomerata* [31].

In order to inform people who manifest allergic symptoms of possible risks of an allergic response, different groups of scientists have proposed threshold values of pollen concentration that causes symptomatology. The European Academy of Allergy and Clinical Immunology (EAACI) determined the high pollen days for five different pollen types [32] based on a review of threshold values for pollen-induced symptoms [33]. The first alder pollen concentration threshold study was carried out in Sweden in 1970; Swedish doctors established four pollen concentration risk thresholds (low: 0–10 pollen grains/m^3; middle: 10–30 grains/m^3; high: 30–100 grains/m^3 and very high: >100 grains/m^3) [34]. A later study carried out in Denmark identified four risk thresholds (low risk: 0–30 grains/m^3; middle: 30–50 grains/m^3; high: 50–150 grains/m^3 and very high: >150 grains/m^3) [35]. Recent studies have indicated that these threshold values may be highly dependent on regional conditions, such as vegetation and climate [33].

Due to the growing global problem of increasing pollen allergenicity, wide varieties of studies have addressed the prediction of allergic risk periods caused by high pollen concentrations [36–42]. In such situations, it is important to find a set of variables that enable the early prediction of pollen counts through the application of intelligent models [43]. A survey written by Suanno et al. [43] highlights the utility of aeropalynological data (pollen counts) to accurately predict pollen trends. However, pollen counts taken on previous days are mainly used to run kriging methods, which allow for obtaining approximate pollen counts in areas where there are no pollen collectors by using pollen counts from other nearby locations. Furthermore, pollen counts on previous days incorporate information on the timing of the pollen season and summarize information provided by other variables that have affected the plants during previous days. Therefore, we believe (hypothesize) that this type of information can complement other variables to accurately forecast pollen release.

In the present study, an analysis of the *Alnus* pollen trends in recent years was carried out, as well as the assessment of the influence exerted by meteorological variables on pollen concentrations in the atmosphere. Finally, we also implemented a Machine Learning (ML) model able to accurately predict the periods of allergy risk in a sensitive population by successfully combining meteorological and aeropalynological data.

2. Materials and Methods

2.1. Characterization and Location of Study Area

The study was carried out in the city of Ourense (42°20′ N; 7°52′ W) located in the northwest of Spain (Figure 1). The city is located at 139 m above sea level and its climate has a marked Mediterranean tendency, with warm temperatures, low humidity, an annual average temperature around 14.9 °C and a total annual rainfall of 811 mm [44].
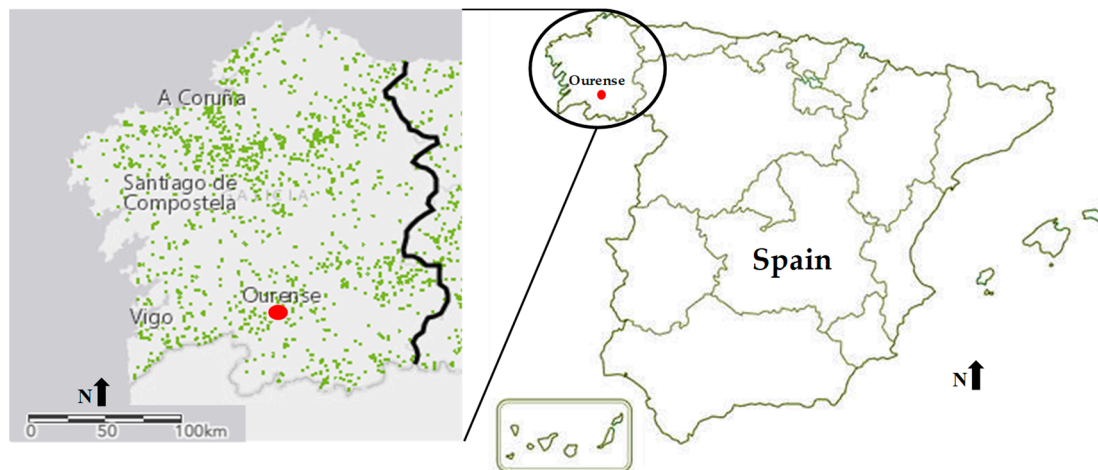

Figure 1. Sampling location area; green zone indicates the *Alnus glutinosa* tree distribution in Spain. To locate *Alnus* trees, we used the Spanish Forest Species Inventory (http://especiesforestales.com/fichas.html, accessed on 23 May 2023).

2.2. Airborne Pollen

The monitoring of airborne pollen was carried out with a Lanzoni VPPS-2000 volumetric trap [45] from 1993 to 2021. Pollen data were counted following the standardised protocol of the Spanish Aerobiological Network [46]. The results obtained were expressed as the seasonal pollen integral (SPIn) (pollen grains), and daily mean values were expressed as pollen/m^3 of air [47].

2.3. Main Pollen Season and ML Pollen Period

In order to calculate the main pollen season (MPS), an AeRobiology (2.0.1) software package was applied [48]. The MPS was calculated using a percentage method based on eliminating a certain percentage at the beginning and end of the pollen season. The present study applied the pollen season based on 95% of the total annual pollen [49]. Therefore, the MPS start date was considered as the day in which 2.5% of the total pollen was registered, and the end date of the pollen season was marked as the day in which 97.5% of the total pollen was registered.

In order to build and adjust a successful ML model, we avoided discarding information about some days with a low pollen emission (MPS disregarded 5% of the data) and tried to preserve some data on days when no pollen grains were widespread. This idea led us to create the new concept of the "ML pollen period". Moreover, the MPS was different every year and must be calculated. However, the ML pollen period was defined using fixed dates.

The ML pollen period was additionally defined to include a pollen concentration greater than zero—useful for building the different ML models studied in this work. The ML pollen period covers the same set of months in which any pollen grain was observed in any previous year. Therefore, the ML pollen period includes the main pollen season, but does not discard any days with observations greater than 0 grain/m^3. In addition, this decision does not imply the use of an excess of zero-value observations (which would be the case if we took all days of the year), which could lead to an increase in the effects produced by unbalanced data [50].

Throughout the 26-year period under study (1993 to 2018), there were very few days with no data due to failure of the collectors or power outages. Missing data were calculated using the AeRobiology software package with the "interpollen" function using the "linear" method, and the interpolation was performed by drawing a straight line between the ends of the gap [48].

2.4. Meteorological Data

The meteorological variables, such as maximum, minimum and average temperatures (°C) and rainfall (mm), as well as data from 1 to 7 days before, were monitored by the meteorological station of the Galician Institute for Meteorology and Oceanography ME-TEOGALICIA (https://www.meteogalicia.gal, accessed on 16 January 2023), located 300 m from the pollen trap.

2.5. Statistical Analysis

2.5.1. Correlation Analysis

A linear regression analysis was conducted in order to estimate the increasing or decreasing trends for airborne pollen, MPS characteristics (start, end and length dates of the MPS, pollen integral, pollen peak concentration and pollen peak day) and meteorological parameters. Moreover, a Spearman correlation analysis was performed between the airborne pollen and different variables (meteorological and pollen counts of previous days) by using IBM SPSS Statistics 24.0 software.

2.5.2. Machine Learning Models

We took advantage of different ML models to predict whether the *Alnus* pollen concentration in the air would be high (>50 grains/m^3), medium (30–50 grains/m^3) or low (0–30 grains/m^3) following the EAACI criterion. To this end, we selected different ML algorithms to compare their achieved performance including Random Forests (RF), Support Vector Machines (SVM), Gaussian Naïve Bayes (GNB) and Multi-Layer Perceptron (MLP).

RF is an ensemble approach that is mainly used for classification (label assignment) purposes. It is based on the combination of multiple weak classifiers (usually denoted as base classifiers/estimators) built on different columns. This ML scheme has been widely used to tackle different classification and regression problems including pollen forecasting [36,37]. For our experimentation, we selected DecisionTree classifiers [51] as base estimators.

SVM comprises a family of geometric-inspired algorithms commonly used to accurately solve classification and regression problems. When the objective is to assign labels (the current problem), SVMs are able to find parameters of a kernel to transform each input observation into a point in an *n*-dimensional space. Then, they compute a hyperplane for the space that is able to keep observations connected with a label (positive) separated from the remaining ones (negative). SVM optimises both the transformation and the hyperplane generated to maximize the distance between positive and negative instances with regard to one label. This kind of model has been successfully applied to solve a wide variety of problems including pollen forecasting [39]. In our experiments, we used a radial basis function (RBF) kernel.

GNB is a probabilistic-based model, which forms part of the Naïve Bayes family. GNB takes advantage of Gaussian distributions for estimating the probability of the occurrence of each label when the input variable takes a specific value. This approach is appropriate for the resolution of different kinds of problems represented by float variables, also being applied in the context of pollen estimation [40].

MLP represents a widely used neural-based model able to perform label assignment and time series prediction [52] and is particularly applied for pollen forecasting [41,42]. Generally speaking, an MLP classifier consists of a neural network that has an input layer containing as many neurons as input variables, and an output layer with as many neurons as output values. Following this scheme, the output layer in our experiments contains

3 neurons that correspond to 'Low', 'Medium' and 'High' labels. In addition, an MLP classifier also comprises one or more hidden layers (2 in our configuration) made of a certain number of neurons (100 per layer in our case), with a configurable activation function [53] (ReLU in our configuration). However, as the labels defined for the current approach are exclusive, the output layer was configured with a Softmax activation function that causes only the neuron with the highest output value to be activated. The parameters (weights) of the MLP model are typically optimized by using a software algorithm such as Adam [54] (https://ieeexplore.ieee.org/document/8624183/metrics#metrics, accessed on 11 April 2023).

The dataset was built from *Alnus* pollen concentration data (in grains/m^3) collected from 1 December 1993 to 30 April 2021 in the city of Ourense. Also, for this period, the Ourense meteorological station of METEOGALICIA supplied the data of average temperature (°C), maximum temperature (°C), minimum temperature (°C) and rainfall (mm) registered. Moreover, for the training, testing and validation of the different ML models, the dataset was split: (i) data within the period 1993–2013 were used for the construction of the model; (ii) for testing purposes, we used the information available about the period 2014–2018; and finally, (iii) we preserved the data between 2019 and 2021 for validation purposes.

In the framework of the proposed research, pollen observations taken outside the ML pollen periods were discarded. Moreover, all variable values were scaled between 0 and 1, considering their respective maximum and minimum values. To globally assess the performance of the different classifiers, we selected Cohen's kappa coefficient [55]. Kappa allows for calculating the probability that the rankings made by two evaluators will overlap. We used the proposed model and the manually labelling method (data included in the original dataset) as evaluators to assess the performance of the former. Kappa was selected as the evaluation metric because it takes into account imbalance in class distribution.

3. Results

Alnus airborne pollen was present in the study area during the winter months, mainly in January and February (Figure 2).
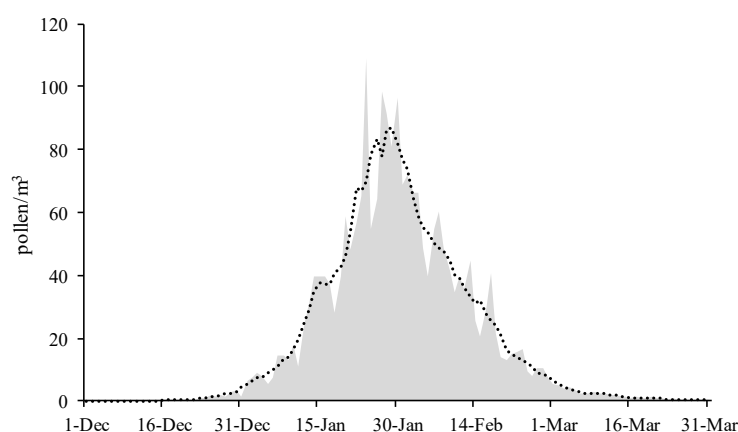

Figure 2. Average daily values of *Alnus* pollen concentration in the atmosphere of the city of Ourense from 1993 to 2018. The black dotted line represents the 5-day running mean.

Over the 26 years under study, the MPS average start date took place on January 8th and the end date was on 1 March (Table 1). We obtained a major standard deviation from the end date (11.5%) compared to the start date (6.4%). The earliest start date took place on December 29 in the year 1994, while the most delayed end date was observed on 6 April 2015. The *Alnus* main pollen season covered a long period with an average duration of 53 days during the entire period under analysis, ranging from 35 days in 2008 to 82 days in 2015 (Table 1).

Table 1. Characteristics of the *Alnus* MPS over the 26 years under study: start date, end date, length of the MPS, annual pollen integral (pollen grains), daily pollen peak (pollen/m^3) and date of pollen peak of each year. Average value (mean), maximum value (max.), minimum value (min.), standard deviation (SD) and relative standard deviation (RSD, %) in all study years.

Year	Start MPS	End MPS	Length MPS	Annual Pollen	Pollen Peak	Pollen Peak Date
1993	8-Jan	8-Mar	60	1414	75	29-Jan
1994	29-Dec	19-Feb	53	1098	163	15-Jan
1995	10-Jan	21-Feb	43	1147	171	25-Jan
1996	3-Jan	5-Mar	62	1387	91	14-Feb
1997	5-Jan	12-Feb	39	3113	534	27-Jan
1998	7-Jan	23-Feb	48	809	72	21-Jan
1999	12-Jan	26-Feb	46	1369	161	1-Feb
2000	11-Jan	15-Mar	64	666	113	30-Jan
2001	1-Jan	19-Feb	50	630	94	8-Jan
2002	15-Jan	20-Feb	37	1246	163	28-Jan
2003	31-Dec	15-Feb	47	2132	400	28-Jan
2004	1-Jan	24-Feb	55	1114	79	18-Jan
2005	16-Jan	27-Feb	43	1937	261	24-Jan
2006	20-Jan	1-Mar	41	2754	353	3-Feb
2007	9-Jan	2-Mar	53	1695	159	8-Feb
2008	19-Jan	22-Feb	35	4319	499	30-Jan
2009	5-Jan	13-Mar	68	1294	208	30-Jan
2010	18-Jan	5-Mar	47	1724	156	24-Jan
2011	9-Jan	23-Feb	46	3516	345	20-Jan
2012	7-Jan	11-Mar	64	4351	305	20-Jan
2013	2-Jan	4-Mar	62	2253	192	16-Jan
2014	10-Jan	22-Mar	72	1216	96	27-Jan
2015	15-Jan	6-Apr	82	1137	147	17-Feb
2016	31-Dec	8-Mar	68	5317	617	24-Jan
2017	10-Jan	2-Mar	52	2692	408	30-Jan
2018	17-Jan	26-Feb	41	6441	867	24-Jan
Mean	8-Jan	1-Mar	53	2184	259	26-Jan
Max.	20-Jan	6-Apr	82	6441	867	24-Jan
Min.	29-Dec	12-Feb	35	630	72	21-Jan
SD	6.45	11.49	11.96	1503.84	197.19	8.73
RSD (%)	0.01	0.03	22.57	68.87	76.19	0.02

The annual pollen integral evidenced higher fluctuations, ranging between 630 pollen grains in 2001 and 6441 pollen grains in 2018, with an average value of 2184 pollen grains during the analysed period. These variations were also observed in the pollen peak concentrations. In general, the peak was registered during the second fortnight of January or first fortnight of February, with an average value of 259 pollen/m^3 registered on 26 January, a maximum pollen peak of 867 pollen/m^3 on 24 January in 2018 and a minimum pollen peak of 72 pollen/m^3 on 21 January in 1998 (Table 1).

The percentage of the total annual pollen registered in the whole period suffered severe fluctuations between different years. In detail, we observed a significant and positive trend that led to a 19% increase in the percentage of pollen grains by year (Figure 3).

When we analysed the trends of the characteristics of the airborne *Alnus* MPS for all the years, we observed that the variables showing a significant trend were the MPS end, annual pollen and pollen peak. The MPS end exhibited a significant and positive delay trend of approximately 0.68 days per year, an increase in the annual pollen integral of 112 pollen grains per year and an increasing trend in the pollen peak of approximately 12.4 pollen grains/m^3 per year (Figure 4).
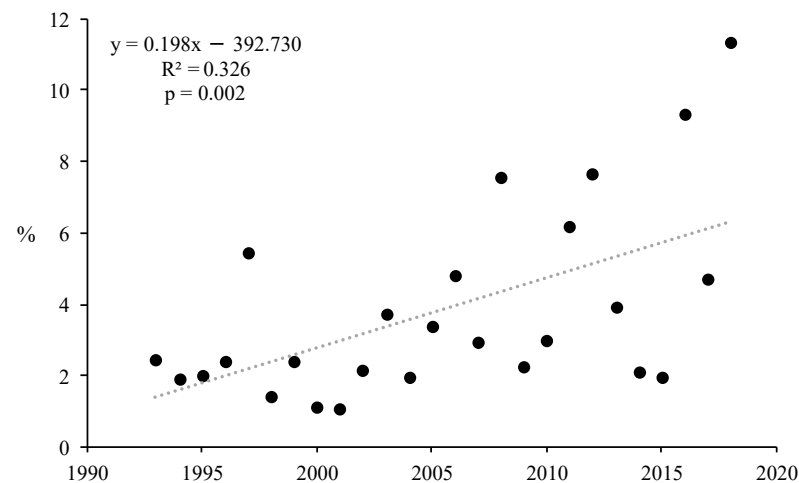
Figure 3. Trend of the airborne *Alnus* pollen concentration percentage regarding the total pollen in the period under study (1993–2018).
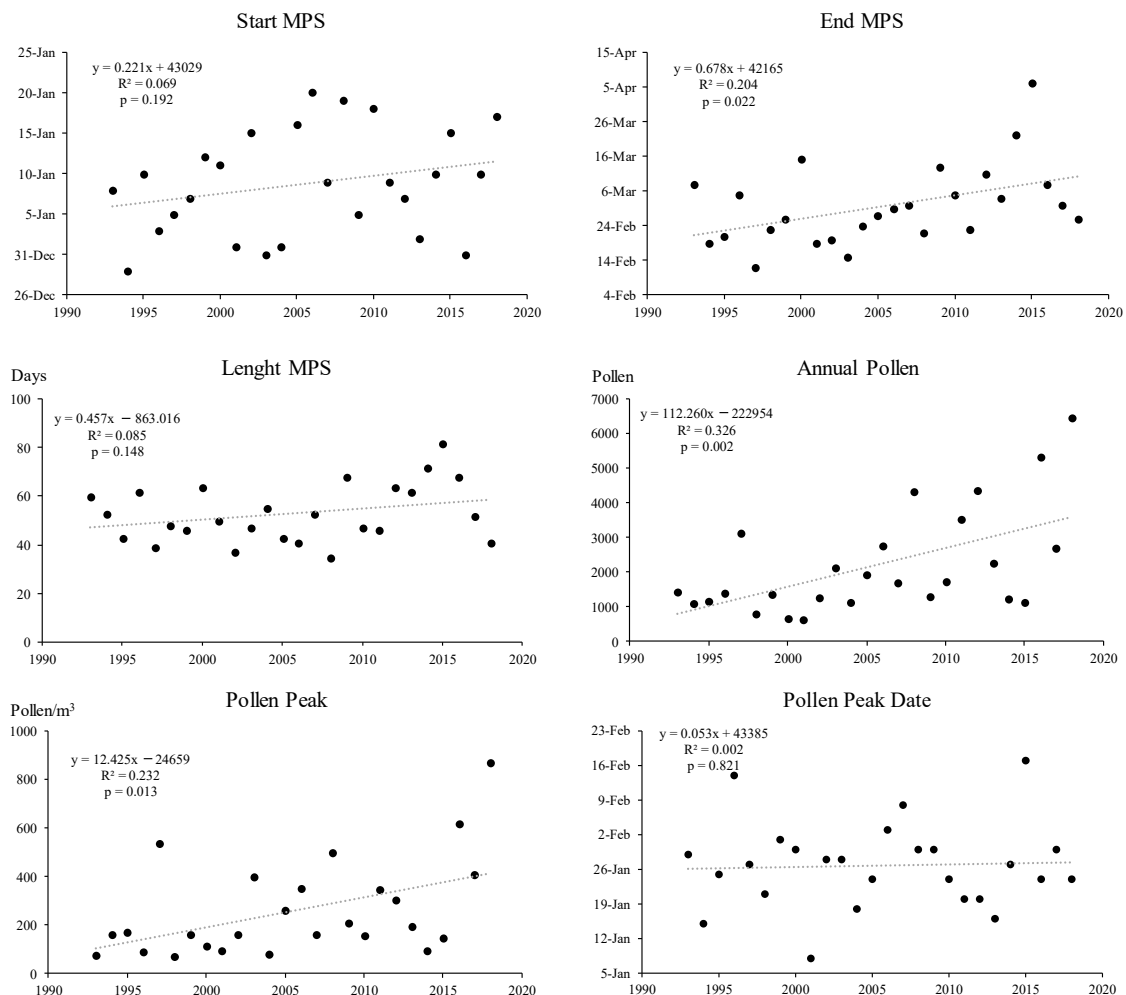

Figure 4. Trends of the airborne *Alnus* MPS during the period under study (1993–2018): start and end dates, length, annual pollen integral, pollen peak and pollen peak date.

In order to apply the different ML models as pollen predictions, we used the information included in Table 1 to compute the ML pollen period for the *Alnus* species. By applying

the ML pollen period definition, we found that ML models should only consider the *Alnus* pollen concentration series between 1 December and 30 April.

To decide which variables to use for the construction of the different ML models, we performed a Spearman's correlation analysis between all the available input variables and the pollen concentration in a specific day, i.e., $cAp(d)$. We preserved the data between 2019 and 2021 for evaluation purposes. Therefore, the correlation was computed using data between 1993 and 2018. In the analysis, we included the following combination of variables: (i) the concentration of Alnus pollen in the previous 7 days, i.e., $cAp(d-1)$, $cAp(d-2)\ldots$, $cAp(d-n)$, n = 7, where cAp(d) indicates the concentration of Alnus pollen in a given day, d; (ii) the rainfall on the previous days, $rain(d-1)$, $rain(d-2)\ldots$, $rain(d-n)$, n = 7, where the rain(d) function is used for indicating the precipitation amount for a specific day, d; (iii) the maximum temperature reached in the previous days, i.e., $maxt(d-1)$, $maxt(d-2),\ldots$, $maxt(d-n)$, n = 7, where maxt(d) indicates the maximum temperature achieved for a certain day, d; (iv) minimum temperature recorded on previous days, i.e., $mint(d-1)$, $mint(d-2),\ldots$, $mint(d-n)$, n = 7, where mint(d) stands for the minimum temperature achieved in a specific day, d; and (v) the average temperature on previous days, $avgt(d-1)$, $avgt(d-2),\ldots$, $avgt(d-n)$, n = 7, where avgt(d) represents the average temperature of a specific day, d. Table 2 shows the correlation found between all the variables.

Table 2. Spearman's correlation test coefficients (R) in the 1993–2018 study period, applied to daily pollen concentration variables. cAp(d) indicates the concentration of *Alnus* pollen in a given day, d. maxt(d), mint(d) and avgt(d) indicate the maximum, minimum and average temperature achieved for a certain day, respectively, d; the rain(d) function is used for indicating the precipitation amount for a specific day, d. The expression (d-i) is used to indicate the value of the variable in the previous i days. Statistical significance was considered at 95% (* $p \leq 0.05$) and 99% (** $p \leq 0.01$) confidence level.

Alnus	R	p	*Alnus*	R	p	*Alnus*	R	p
cAp(d)	1.000		maxt(d − 6)	−0.276 **	0.000	avgt(d − 4)	−0.275 **	0.000
cAp(d − 1)	0.803 **	0.000	maxt(d − 7)	−0.289 **	0.000	avgt(d − 5)	−0.287 **	0.000
cAp(d − 2)	0.770 **	0.000	mint(d)	−0.214 **	0.000	avgt(d − 6)	−0.300 **	0.000
cAp(d − 3)	0.756 **	0.000	mint(d − 1)	−0.193 **	0.000	avgt(d − 7)	−0.315 **	0.000
cAp(d − 4)	0.740 **	0.000	mint(d − 2)	−0.189 **	0.000	rain(d)	−0.065 **	0.000
cAp(d − 5)	0.727 **	0.000	mint(d − 3)	−0.186 **	0.000	rain(d − 1)	−0.059 **	0.000
cAp(d − 6)	0.714 **	0.000	mint(d − 4)	−0.189 **	0.000	rain(d − 2)	−0.012	0.475
cAp(d − 7)	0.690 **	0.000	mint(d − 5)	−0.184 **	0.000	rain(d − 3)	0.012	0.460
maxt(d)	−0.146 **	0.000	mint(d − 6)	−0.188 **	0.000	rain(d − 4)	0.032	0.051
maxt(d − 1)	−0.168 **	0.000	mint(d − 7)	−0.202 **	0.000	rain(d − 5)	0.030	0.062
maxt(d − 2)	−0.200 **	0.000	avgt(d)	−0.238 **	0.000	rain(d − 6)	0.031	0.060
maxt(d − 3)	−0.221 **	0.000	avgt(d − 1)	−0.239 **	0.000	rain(d − 7)	0.039 *	0.017
maxt(d − 4)	−0.237 **	0.000	avgt(d − 2)	−0.253 **	0.000			
maxt(d − 5)	−0.257 **	0.000	avgt(d − 3)	−0.265 **	0.000			

Table 2 shows how *Alnus* concentrations pertaining to the last days were the most correlated variables, although rainfall and temperature were also related. On this basis, we optimised the parameters used for the construction of the proposed ML models by using only the values belonging to the ML pollen periods included in the period 1993–2013 for training purposes, leaving the period 2014–2018 for testing.

We ran the models incorporating *n* previous *Alnus* pollen concentration days as independent variables (that is, cAp(d − i) with i = [1...n] and n = [1...7]), achieving the best results for n = 4. Using this configuration as a baseline, we applied different schemes

based on the idea that there is a (biological) connection between meteorological variables (e.g., precipitation and/or temperature) and pollen concentration. For this purpose, we introduced rainfall in three different ways: (i) precipitation modelled with independent variables with the rainfall of the previous n days (i.e., $\text{rain}(d-i)$, with $i = [1\ldots n]$ and $n = [1\ldots 7]$), (ii) cumulative precipitation during the previous n days (i.e., $\sum_{i=1}^{n} \text{rain}(d-i)$) and (iii) the precipitation estimated for the target day, using $\text{rain}(d)$ to model this value.

We also incorporated average temperature, $(\text{avg}(d))$, in different forms: (i) $\text{avgt}(d-i)$ with $i = [1\ldots n]$ and $n = [1\ldots 7]$, (ii) combined average temperature, $\text{avg}_{i=1}^{n}(\text{avgt}(d-i))$, where $n = [1\ldots 7]$ and $\text{avg}_{j=1}^{m}(x_j)$ indicates the average of x_j and (iii) the average temperature on the target day, $(\text{avgt}(d))$. Maximum and minimum temperatures were also tested with similar configurations. Taking into consideration the results achieved using the previous 4 days of pollen (i.e., $\text{cAp}(d-1)$, $\text{cAp}(d-2)$, $\text{cAp}(d-3)$, $\text{cAp}(d-4)$), only the inclusion of the variable $\text{rain}(d)$ made it possible to reach better results. This fact seems to support the idea that pollen recordings from past days (used for model construction) may indirectly contain information about temperatures and precipitation from previous days.

Starting from the best configuration achieved in the testing phase, we carried out a validation step with those values of ML pollen periods included in the years 2019, 2020 and 2021 (previously unseen by the models). The experiments were carried out year by year, retraining the models to incorporate the new available data. Table 3 shows the kappa values together with the median/difference of the analysed model.

Table 3. Kappa values obtained by the models analysed during the validation period (2019–2021).

Classifier Year	RF	SVM	GNB	MLP
2019	0.680	0.668	0.592	0.749
2020	0.570	0.586	0.551	0.569
2021	0.571	0.426	0.312	0.479
Median/Difference	0.607/±0.073	0.56/±0.134	0.485/±0.173	0.599/±0.150

As indicated in Table 3, the models showed an adequate performance. However, although the MLP model achieved an excellent result for the year 2019, it may probably be affected by overfitting, since the kappa index worsened in the upcoming years. On the other hand, although the RF model attained a slightly lower evaluation for the first year, it was able to maintain its accuracy in subsequent years. This observation suggests that this model is best suited to address changes in pollen distribution trends. Then, using the RF model, we calculated the underline confusion matrix [56] (Table 4). The confusion matrix allows for a numerical comparison between the predictions (classes) made with the model and the real situation. In particular, it allows us to numerically assess the probability of specific errors (e.g., the probability of classifying a day labelled with high pollen liberation as low). In each cell included in our confusion matrix, we include the percentage of successful predictions as compared to the real classification of data.

As seen in Table 4, despite the very imbalanced data (more than 90% of the data are labelled as low), the Random Forest classifier is able to easily detect those days with medium and high labels. Similarly, it is also possible to observe how, on many occasions (especially in the years 2020 and 2021), an error occurs in classifying a medium label as a high one, which is not a problem from a medical point of view. From another perspective, the number of days with a high incidence of pollen being misclassified is quite low (which only happens in the year 2019).

Table 4. Confusion Matrix for Random Forest.

Year			Predicted Labels		
			low	medium	high
2019	Real labels	low	95.35	0.0	4.65
		medium	33.33	0.0	66.67
		high	6.25	0.0	93.75
			low	medium	high
2020	Real labels	low	90.22	0.0	9.78
		medium	0.0	0.0	100.0
		high	0.0	0.0	100.0
			low	medium	high
2021	Real labels	low	96.55	0.0	3.45
		medium	0.0	33.33	66.67
		high	0.0	0.0	100.0

4. Discussion

During the last decades, several studies have noted that different atmospheric biological particles, including alder pollen grains, cause human health problems such as allergies and infections [32,57–60]. In the present study, we found that, on average, the *Alnus* MPS starts on 8 January and ends on 1 March with an average length of 53 days. These results were similar to those pointed out by several authors from areas in northern Spain [61]. Studies carried out in Central Spain indicated that the onset of *Alnus* MPS was delayed to mid-January and the end to the final days of March [62]. Research conducted in Central Europe noted the MPS was advanced around the Christmas season, e.g., Switzerland or Austria [63,64], whereas southern European, as in Italy or Turkey, areas showed a later MPS start in the month of February [61,65,66]. In addition, we found an average of pollen grains during the study period (1993–2018). The value is similar to that noted in [61] for the same study area, and higher than the data pointed out for southern Europe locations [66]. On the other hand, northern Europe locations such as Poland registered higher annual values of 7050 pollen grains [65]. The start, end and length features of the MPS and even the pollen concentration are highly influenced by meteorological factors; hence, in recent years, different authors have studied how climate change will influence plant behaviour and its impact on people's health [67,68]. In the present study, we calculate the trends for the main MPS characteristics, registering statistically significant linear trends ($p < 0.01$) for the end of MPS, the annual pollen integral (SPIn) and the pollen peak. The detected trends indicate that the pollen concentration will rise 112 pollen grains per year and the pollen peak will increase by 12.4 grains of pollen/m^3. These data are consistent with a study carried out in Bratislava, which detected increases in the pollen peak of 16.3 pollen grains/m^3 [69]. As a general form, these trends were also similar to the data pointed out for North America over the period 1990–2018, with significant trends in different pollen metrics, such as the seasonal features and pollen integral values [70].

Different research has shown that plant flowering, and therefore airborne pollen concentration, may be influenced by meteorological parameters [71–73]. In particular, a study conducted in Poland noted that the onset and the length of the pollen season depend on the weather variables prevailing before and during the release of alder pollen grains [65]. Moreover, several researchers have pointed out that the variations in alder airborne pollen concentrations happened due to temperature [71–73]. The Spearman correlation analysis carried out between the *Alnus* pollen and weather variables showed a negative and significative ($p < 0.01$) correlation between *Alnus* pollen and temperatures (maximum, minimum and average) for the same day and the 7 previous days, as well as the rain of the same day and 1 day before. With a lower significance level ($p < 0.05$), there is a positive correlation with the rain of the 7 previous days. These data are consistent with a study conducted in Malaga (southern Spain), in which the authors pointed out significant and negative correlations between average and minimum temperatures [74]. In

addition, other investigations conducted in Poland showed a negative correlation between alder pollen and maximum temperature and a positive correlation with the rainfall [65]. Different long-term studies on selected plant taxa pointed to an increase in the pollen concentrations as well as the length of the pollen season in recent years, often modulated by temperature [17,75,76]. Moreover, we also found a positive correlation with a high significance level ($p < 0.01$) between the *Alnus* pollen count of a day and each of the pollen counts made on the 7 previous days.

Once the correlation analysis was performed, we carried out an analysis using ML techniques in order to predict the *Alnus* pollen concentration. In recent years, ML models have been successfully implemented to predict air quality indices in Smart cities [77,78]. In the present work, four different classifiers were tested: Random Forests (RF), Support Vector Machines (SVM), Gaussian Naïve Bayes (GNB) and Multi-Layer Perceptron (MLP). The best classifier was RF, showing a kappa value of 0.680, 0.570 and 0.571 in 2019, 2020 and 2021, respectively. In addition, the most informative variables were the combination of pollen counts in the previous 4 days and the rainfall of the same day. This finding is consistent with several studies pointing out that RF is the model with a higher r^2 value for *Alnus* in nine Polish cities, with r^2 value oscillations between 0.220 and 0.480 [79]. This study also pointed out that the variables that most influenced the RF model were temperature and rainfall. A previous study conducted in Germany using autoregression and neural network approaches to predict *Betula* and Poaceae airborne pollen registered an r^2 value ranging from 0.13 to 0.62 for *Betula* and 0.03 to 0.55 for Poaceae [80]. These results are slightly lower than those obtained in the present work.

Finally, a confusion matrix was analysed to know the RF model sensitivity on each validation year under study (i.e., 2019, 2020 and 2021). From an overall perspective, it is possible to state that the model accurately predicts high or low values (e.g., in 2019, the sensitivity of the model for days with a high pollen concentration was 93.75%, and only 6.25% of those cases were misclassified as a low level). In contrast, the sensitivity of the model with days classified as a low concentration was 95.35% (with 4.65% of the instances being misclassified as a high level). In 2020 and 2021, the sensitivities achieved by the model for the label "High" were 100%, while in the case of the label "Low", the values were 90.22% and 96.55%, respectively (9.78% and 3.45% of the occurrences were misclassified as a high level). All cases from 2019 with medium labels were misclassified as high (66.67%) or low (33.33%). In 2020, 100% of cases having medium labels were misclassified as a high value. In 2021, the model attained a sensibility of 33.33% for medium-labelled instances while the remaining 66.66% were misclassified as high. As long as most cases with the medium label were classified as high, the impact caused by the errors is reduced. Similar results were pointed out in a study conducted in Poland for *Alnus* airborne pollen [37]. In this study, the authors only considered two labels (i.e., high and low) and two tests were applied. The percentage of sensitivity registered was lower than that obtained in the current study, with 70.0% in the case of days with a high pollen concentration in the first test, and a percentage of 59.0% in the second test.

Alnus pollen is small in size and this feature allows it to be transported over long distances by the wind [81]. Therefore, the transport of *Alnus* pollen from distant locations may contribute to its high concentrations, in some years, far from the area of emission [65]. Specific episodes of the long-range transport of pollen could hamper the ML model's accuracy for pollen concentration predictions. Events of pollen transport were previously detected with other taxa in our study area [82], which reinforces the robustness and accuracy of the proposed ML models.

5. Conclusions

This study introduces an analysis of pollen seasonal changes in *Alnus* over the last 26 years (1993–2018) and documents inner distribution changes: the start/end dates and the amount of pollen grains registered yearly. In addition, the work incorporates an analysis of different variables biologically connected with the liberation of pollen from *Alnus* species.

We found that, on average, the *Alnus* MPS starts on 8 January and ends on 1 March with an average length of 53 days. An annual average of 2184 pollen grains were detected during the study period. These parameters can be influenced by meteorological factors, so climate change has modified plant behaviour in recent years and its impact on people's health. Significant trends in MPS characteristics indicated that pollen concentration will increase by 112 pollen grains per year and the pollen peak will increase by 12 pollen/m^3. In the correlation analysis, the most correlated variables (i.e., pollen observations in the previous 7 days, min./max./avg. temperature and rainfall) were chosen to build and analyse the performance of different state-of-the-art ML models able to accurately estimate *Alnus* pollen liberation. As a result, it was determined that pollen counts taken in the previous 4 days together with the rainfall forecast of the current day are the best attributes to make the prediction. Finally, the Random Forests (RF) model proved to be a suitable model to carry out the prediction of *Alnus* pollen concentration in the atmosphere by using rainfall and previous pollen concentrations as features, being able to classify the amount of grains as low (0–30 grains/m^3), medium (30–50 grains/m^3) or high (>50 grains/m^3) following the EAACI criterion. On the basis of the above, the Random Forests (RF) model was built by using these attributes to achieve short-term (1–5 days) predictions of *Alnus* pollen concentration in the atmosphere, which makes it possible for people to make informed decisions prior to suffering allergic symptoms, such as the use of prophylactics to reduce the severity of reactions caused by pollen. The reason for predicting in a short period of time is because foreseen meteorological data are used. Due to the weather variability in the region of Galicia, the prediction of *Alnus* pollen concentration will be much more accurate in a short period of time. As long as our model requires the pollen amount of 4 previous days and the rainfall forecast for the target day, we plan to obtain the rainfall forecast from the Meteogalicia website and the pollen grain forecast using the AeRobiology software package [48].

References

1. Claessens, H.; Oosterbaan, A.; Savill, P.; Rondeux, J. A Review of the Characteristics of Black Alder (*Alnus glutinosa* (L.) Gaertn.) and Their Implications for Silvicultural Practices. *Forestry* **2010**, *83*, 163–175. [CrossRef]
2. Kajba, D.; Gračan, J. *Technical Guidelines for Genetic Conservation and Use Black Alder (Alnus glutinosa)*; IPGRI: Rome, Italy, 2003; p. 4.
3. Josef, T.; Gösta, E.; Jochen, K.; Sonja, C. *Noble Hardwoods Network. Report of the First Meeting*; European Forest Genetic Resources Programme (EUFORGEN): Rome, Italy, 1996; ISBN 92-9043-291-8.
4. Lecomte, H.; Florkin, P.; Morimont, J.-P. *La Forêt Wallonne: État de la Ressource à la fin du 20ème Siècle*; Ministère de la Région wallonne. Direction Générale des Ressources Naturelles et de L'environnement: Jambes, Belgium, 2002.
5. Debruxelles, N.; Chandelier, A.; Dufays, E.; Claessens, H.; Cavelier, M.; Rondeux, J. Le Dépérissement de l'aulne En Wallonie. *Silva Belg.* **2007**, *114*, 2–5.

6. Damialis, A.; Traidl-Hoffmann, C.; Treudler, R. Climate Change and Pollen Allergies. In *Biodiversity and Health in the Face of Climate Change*; Marselle, M.R., Stadler, J., Korn, H., Irvine, K.N., Bonn, A., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 47–66. ISBN 978-3-030-02318-8.

7. Traidl-Hoffmann, C. Allergie—Eine Umwelterkrankung! *Bundesgesundheitsblatt Gesundheitsforsch. Gesundheitsschutz* **2017**, *60*, 584–591. [CrossRef]

8. Nicolaou, N.; Siddique, N.; Custovic, A. Allergic Disease in Urban and Rural Populations: Increasing Prevalence with Increasing Urbanization. *Allergy* **2005**, *60*, 1357–1360. [CrossRef]

9. Bosch-Cano, F.; Bernard, N.; Sudre, B.; Gillet, F.; Thibaudon, M.; Richard, H.; Badot, P.-M.; Ruffaldi, P. Human Exposure to Allergenic Pollens: A Comparison between Urban and Rural Areas. *Environ. Res.* **2011**, *111*, 619–625. [CrossRef]

10. D'Amato, G.; Chong-Neto, H.J.; Monge Ortega, O.P.; Vitale, C.; Ansotegui, I.; Rosario, N.; Haahtela, T.; Galan, C.; Pawankar, R.; Murrieta-Aguttes, M.; et al. The Effects of Climate Change on Respiratory Allergy and Asthma Induced by Pollen and Mold Allergens. *Allergy* **2020**, *75*, 2219–2228. [CrossRef] [PubMed]

11. D'Amato, G.; Cecchi, L. Effects of Climate Change on Environmental Factors in Respiratory Allergic Diseases. *Clin. Exp. Allergy* **2008**, *38*, 1264–1274. [CrossRef]

12. Weber, R.W. Meteorologic Variables in Aerobiology. *Immunol. Allergy Clin. N. Am.* **2003**, *23*, 411–422. [CrossRef]

13. Frei, T.; Gassner, E. Climate Change and Its Impact on Birch Pollen Quantities and the Start of the Pollen Season an Example from Switzerland for the Period 1969–2006. *Int. J. Biometeorol.* **2008**, *52*, 667–674. [CrossRef]

14. Adams-Groom, B.; Selby, K.; Derrett, S.; Frisk, C.A.; Pashley, C.H.; Satchwell, J.; King, D.; McKenzie, G.; Neilson, R. Pollen Season Trends as Markers of Climate Change Impact: Betula, Quercus and Poaceae. *Sci. Total Environ.* **2022**, *831*, 154882. [CrossRef]

15. Ariano, R.; Canonica, G.W.; Passalacqua, G. Possible Role of Climate Changes in Variations in Pollen Seasons and Allergic Sensitizations during 27 Years. *Ann. Allergy Asthma Immunol.* **2010**, *104*, 215–222. [CrossRef] [PubMed]

16. van Vliet, A.J.H.; Overeem, A.; De Groot, R.S.; Jacobs, A.F.G.; Spieksma, F.T.M. The Influence of Temperature and Climate Change on the Timing of Pollen Release in the Netherlands. *Int. J. Climatol.* **2002**, *22*, 1757–1767. [CrossRef]

17. Ziska, L.H.; Makra, L.; Harry, S.K.; Bruffaerts, N.; Hendrickx, M.; Coates, F.; Saarto, A.; Thibaudon, M.; Oliver, G.; Damialis, A.; et al. Temperature-Related Changes in Airborne Allergenic Pollen Abundance and Seasonality across the Northern Hemisphere: A Retrospective Data Analysis. *Lancet Planet. Health* **2019**, *3*, e124–e131. [CrossRef]

18. Bruffaerts, N.; De Smedt, T.; Delcloo, A.; Simons, K.; Hoebeke, L.; Verstraeten, C.; Van Nieuwenhuyse, A.; Packeu, A.; Hendrickx, M. Comparative Long-Term Trend Analysis of Daily Weather Conditions with Daily Pollen Concentrations in Brussels, Belgium. *Int. J. Biometeorol.* **2018**, *62*, 483–491. [CrossRef]

19. de Weger, L.A.; Bruffaerts, N.; Koenders, M.M.J.F.; Verstraeten, W.W.; Delcloo, A.W.; Hentges, P.; Hentges, F. Long-Term Pollen Monitoring in the Benelux: Evaluation of Allergenic Pollen Levels and Temporal Variations of Pollen Seasons. *Front. Allergy* **2021**, *2*, 676176. [CrossRef] [PubMed]

20. Frenguelli, G. Interactions between Climatic Changes and Allergenic Plants. *Monaldi Arch. Chest Dis. = Arch. Monaldi Mal. Torace* **2002**, *57*, 141–143.

21. Gehrig, R.; Clot, B. 50 Years of Pollen Monitoring in Basel (Switzerland) Demonstrate the Influence of Climate Change on Airborne Pollen. *Front. Allergy* **2021**, *2*, 677159. [CrossRef]

22. Glick, S.; Gehrig, R.; Eeftens, M. Multi-Decade Changes in Pollen Season Onset, Duration, and Intensity: A Concern for Public Health? *Sci. Total Environ.* **2021**, *781*, 146382. [CrossRef]

23. Jochner-Oette, S.; Menzel, A.; Gehrig, R.; Clot, B. Decrease or Increase? Temporal Changes in Pollen Concentrations Assessed by Bayesian Statistics. *Aerobiologia* **2019**, *35*, 153–163. [CrossRef]

24. Lind, T.; Ekebom, A.; Alm Kübler, K.; Östensson, P.; Bellander, T.; Lõhmus, M. Pollen Season Trends (1973–2013) in Stockholm Area, Sweden. *PLoS ONE* **2016**, *11*, e0166887. [CrossRef]

25. Ruiz-Valenzuela, L.; Aguilera, F. Trends in Airborne Pollen and Pollen-Season-Related Features of Anemophilous Species in Jaen (South Spain): A 23-Year Perspective. *Atmos. Environ.* **2018**, *180*, 234–243. [CrossRef]

26. D'Amato, G.; Cecchi, L.; D'Amato, M.; Liccardi, G. Urban Air Pollution and Climate Change as Environmental Risk Factors of Respiratory Allergy: An Update. *J. Investig. Allergol. Clin. Immunol.* **2010**, *20*, 95–102; quiz following 102.

27. Buters, J.T.M.; Thibaudon, M.; Smith, M.; Kennedy, R.; Rantio-Lehtimäki, A.; Albertini, R.; Reese, G.; Weber, B.; Galan, C.; Brandao, R.; et al. Release of Bet v 1 from Birch Pollen from 5 European Countries. Results from the HIALINE Study. *Atmos. Environ.* **2012**, *55*, 496–505. [CrossRef]

28. Buters, J.; Alberternst, B.; Nawrath, S.; Wimmer, M.; Traidl-Hoffmann, C.; Starfinger, U.; Behrendt, H.; Schmidt-Weber, C.; Bergmann, K.-C. Ambrosia Artemisiifolia (Ragweed) in Germany—Current Presence, Allergological Relevance and Containment Procedures. *Allergo J. Int.* **2015**, *24*, 108–120. [CrossRef] [PubMed]

29. Bahbah, F.; Gentil, C.; Sousa-Pinto, B.; Canonica, G.W.; Devillier, P.; Pfaar, O.; Bousquet, J. Patient-reported Outcome Measures in Birch Pollen Allergic Patients Treated with Sublingual Immunotherapy Reflect Real Life. *Allergy* **2023**, *78*, 1113–1116. [CrossRef]

30. Berge, M.; Bertilsson, L.; Hultgren, O.; Hugosson, S.; Saber, A. Qualitative and Quantitative Comparison of Allergen Component-Specific to Birch and Grass Analyzed by ImmunoCAP Assay and Euroline Immunoblot Test. *Eur. Ann. Allergy Clin. Immunol.* **2023**, *55*, 68. [CrossRef]

31. Fernández-González, M.; Ribeiro, H.; Rodríguez-Rajo, F.J.; Cruz, A.; Abreu, I. Short-Term Exposure of Dactylis Glomerata Pollen to Atmospheric Gaseous Pollutants Is Related to an Increase in IgE Binding in Patients with Grass Pollen Allergies. *Plants* **2022**, *12*, 76. [CrossRef] [PubMed]

32. Pfaar, O.; Karatzas, K.; Bastl, K.; Berger, U.; Buters, J.; Darsow, U.; Demoly, P.; Durham, S.R.; Galán, C.; Gehrig, R.; et al. Pollen Season Is Reflected on Symptom Load for Grass and Birch Pollen-induced Allergic Rhinitis in Different Geographic Areas—An EAACI Task Force Report. *Allergy* **2020**, *75*, 1099–1106. [CrossRef] [PubMed]

33. de Weger, L.A.; Bergmann, K.C.; Rantio-Lehtimäki, A.; Dahl, Å.; Buters, J.; Déchamp, C.; Belmonte, J.; Thibaudon, M.; Cecchi, L.; Besancenot, J.-P.; et al. Impact of Pollen. In *Allergenic Pollen*; Springer: Dordrecht, The Netherlands, 2013; pp. 161–215.

34. Steckling-Muschack, N.; Mertes, H.; Mittermeier, I.; Schutzmeier, P.; Becker, J.; Bergmann, K.-C.; Böse-O'Reilly, S.; Buters, J.; Damialis, A.; Heinrich, J.; et al. A Systematic Review of Threshold Values of Pollen Concentrations for Symptoms of Allergy. *Aerobiologia* **2021**, *37*, 395–424. [CrossRef]

35. Kiotseridis, H.; Cilio, C.M.; Bjermer, L.; Tunsäter, A.; Jacobsson, H.; Dahl, Å. Grass Pollen Allergy in Children and Adolescents-symptoms, Health Related Quality of Life and the Value of Pollen Prognosis. *Clin. Transl. Allergy* **2013**, *3*, 19. [CrossRef]

36. Lo, F.; Bitz, C.M.; Hess, J.J. Development of a Random Forest Model for Forecasting Allergenic Pollen in North America. *Sci. Total Environ.* **2021**, *773*, 145590. [CrossRef] [PubMed]

37. Nowosad, J. Spatiotemporal Models for Predicting High Pollen Concentration Level of Corylus, Alnus, and Betula. *Int. J. Biometeorol.* **2016**, *60*, 843–855. [CrossRef] [PubMed]

38. Scheifinger, H.; Belmonte, J.; Buters, J.; Celenk, S.; Damialis, A.; Dechamp, C.; García-Mozo, H.; Gehrig, R.; Grewling, L.; Halley, J.M.; et al. Monitoring, Modelling and Forecasting of the Pollen Season. In *Allergenic Pollen*; Springer: Dordrecht, The Netherlands, 2013; pp. 71–126.

39. Zewdie, G.K.; Liu, X.; Wu, D.; Lary, D.J.; Levetin, E. Applying Machine Learning to Forecast Daily Ambrosia Pollen Using Environmental and NEXRAD Parameters. *Environ. Monit. Assess.* **2019**, *191*, 261. [CrossRef]

40. Seka, D.; Bonny, B.S.; Yoboué, A.N.; Sié, S.R.; Adopo-Gourène, B.A. Identification of Maize (*Zea mays* L.) Progeny Genotypes Based on Two Probabilistic Approaches: Logistic Regression and Naïve Bayes. *Artif. Intell. Agric.* **2019**, *1*, 9–13. [CrossRef]

41. Valencia, J.A.; Astray, G.; Fernández-González, M.; Aira, M.J.; Rodríguez-Rajo, F.J. Assessment of Neural Networks and Time Series Analysis to Forecast Airborne Parietaria Pollen Presence in the Atlantic Coastal Regions. *Int. J. Biometeorol.* **2019**, *63*, 735–745. [CrossRef]

42. Astray, G.; Fernández-González, M.; Rodríguez-Rajo, F.J.; López, D.; Mejuto, J.C. Airborne Castanea Pollen Forecasting Model for Ecological and Allergological Implementation. *Sci. Total Environ.* **2016**, *548–549*, 110–121. [CrossRef]

43. Suanno, C.; Aloisi, I.; Fernández-González, D.; Del Duca, S. Pollen Forecasting and Its Relevance in Pollen Allergen Avoidance. *Environ. Res.* **2021**, *200*, 111150. [CrossRef] [PubMed]

44. *Guía Resumida Del Clima En España (1981–2010)*; Agencia Estatal de Meterorología: Madrid, Spain, 2012; Available online: http://www.aemet.es/ (accessed on 1 June 2023).

45. Hirst, J.M. An automatic volumetric spore trap. *Ann. Appl. Biol.* **1952**, *39*, 257–265. [CrossRef]

46. Galán, C.; Cariñanos, P.; Alcázar, P.; Domínguez, E. *Spanish Aerobiology Network: Management and Quality Manual*; Servicio de publicaciones de la Universidad de Córdoba: Córdoba, Spain, 2007.

47. Galán, C.; Ariatti, A.; Bonini, M.; Clot, B.; Crouzy, B.; Dahl, A.; Fernandez-González, D.; Frenguelli, G.; Gehrig, R.; Isard, S.; et al. Recommended Terminology for Aerobiological Studies. *Aerobiologia* **2017**, *33*, 293–295. [CrossRef]

48. Rojo, J.; Picornell, A.; Oteros, J. AeRobiology: The Computational Tool for Biological Data in the Air. *Methods Ecol. Evol.* **2019**, *10*, 1371–1376. [CrossRef]

49. Andersen, T.B. A Model to Predict the Beginning of the Pollen Season. *Grana* **1991**, *30*, 269–275. [CrossRef]

50. Haibo, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [CrossRef]

51. Breiman, L. *Classification and Regression Trees*; Routledge: London, UK, 2017.

52. Ramchoun, H.; Amine, M.; Idrissi, J.; Ghanou, Y.; Ettaouil, M. Multilayer Perceptron: Architecture Optimization and Training. *Int. J. Interact. Multimed. Artif. Intell.* **2016**, *4*, 26. [CrossRef]

53. Szandała, T. Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks. In *Bio-inspired Neurocomputing*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 203–224.

54. Zhang, Z. Improved Adam Optimizer for Deep Neural Networks. In Proceedings of the 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 4–6 June 2018; IEEE: New York, NY, USA, 2018; pp. 1–2.

55. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]

56. Heydarian, M.; Doyle, T.E.; Samavi, R. MLCM: Multi-Label Confusion Matrix. *IEEE Access* **2022**, *10*, 19083–19095. [CrossRef]

57. Perez-Badia, R.; Rapp, A.; Morales, C.; Sardinero, S.; Galan, C.; Garcia-Mozo, H. Pollen Spectrum and Risk of Pollen Allergy in Central Spain. *Ann. Agric. Environ. Med.* **2010**, *17*, 139–151. [PubMed]

58. Oteros, J.; Galán, C.; Alcázar, P.; Domínguez-Vilches, E. Quality Control in Bio-Monitoring Networks, Spanish Aerobiology Network. *Sci. Total Environ.* **2013**, *443*, 559–565. [CrossRef]

59. De Linares, C.; Alcázar, P.; Valle, A.M.; Díaz de la Guardia, C.; Galán, C. Parietaria Major Allergens vs Pollen in the Air We Breathe. *Environ. Res.* **2019**, *176*, 108514. [CrossRef]

60. Lara, B.; Rojo, J.; Costa, A.R.; Burgos-Montero, A.M.; Antunes, C.M.; Pérez-Badia, R. Atmospheric Pollen Allergen Load and Environmental Patterns in Central and Southwestern Iberian Peninsula. *Sci. Total Environ.* **2023**, *858*, 159630. [CrossRef] [PubMed]

61. Jato, M.V.; Rodríguez-Rajo, F.J.; Aira, M.J.; Tedeschini, E.; Frenguelli, G. Differences in Atmospheric Trees Pollen Seasons in Winter, Spring and Summer in Two European Geographic Areas, Spain and Italy. *Aerobiologia* **2013**, *29*, 263–278. [CrossRef]

62. Rojo, J.; Fernández-González, F.; Lara, B.; Bouso, V.; Crespo, G.; Hernández-Palacios, G.; Rodríguez-Rojo, M.P.; Rodríguez-Torres, A.; Smith, M.; Pérez-Badia, R. The Effects of Climate Change on the Flowering Phenology of Alder Trees in Southwestern Europe. *Mediterr. Bot.* **2021**, *42*, e67360. [CrossRef]

63. Gassner, M.; Gehrig, R.; Schmid-Grendelmeier, P. Hay Fever as a Christmas Gift. *N. Engl. J. Med.* **2013**, *368*, 393–394. [CrossRef] [PubMed]

64. Bastl, K.; Kmenta, M.; Berger, U. Unusually Early Flowering of Alder in Vienna: First Report of Alnus × Spaethii in Austria, Combined LM and SEM Study of Alder Species and Impact on Pollen Allergy Sufferers. *Aerobiologia* **2015**, *31*, 515–524. [CrossRef]

65. Ślusarczyk, J.; Kopacz-Bednarska, A.; Posłowska, J. Influence of Meteorological Factors on the Dynamics of Hazel, Alder, Birch and Poplar Pollen in the 2021 Season in Kielce, Poland. *Ann. Agric. Environ. Med.* **2022**, *29*, 502–512. [CrossRef] [PubMed]

66. Uğuz, U.; Güvensen, A.; Tort, N.Ş.; Dereboylu, A.E.; Baran, P. Volumetric Analysis of Airborne Pollen Grains in the City of Uşak, Turkey. *Turk. J. Bot.* **2018**, *42*, 57–72. [CrossRef]

67. Kim, K.R.; Oh, J.-W.; Woo, S.-Y.; Seo, Y.A.; Choi, Y.-J.; Kim, H.S.; Lee, W.Y.; Kim, B.-J. Does the Increase in Ambient CO_2 Concentration Elevate Allergy Risks Posed by Oak Pollen? *Int. J. Biometeorol.* **2018**, *62*, 1587–1594. [CrossRef]

68. Ziska, L.H.; Caulfield, F.A. Rising CO_2 and Pollen Production of Common Ragweed (*Ambrosia artemisiifolia* L.), a Known Allergy-Inducing Species: Implications for Public Health. *Funct. Plant Biol.* **2000**, *27*, 893. [CrossRef]

69. Ščevková, J.; Dušička, J.; Hrabovský, M.; Vašková, Z. Trends in Pollen Season Characteristics of Alnus, Poaceae and Artemisia Allergenic Taxa in Bratislava, Central Europe. *Aerobiologia* **2021**, *37*, 707–717. [CrossRef]

70. Anderegg, W.R.L.; Abatzoglou, J.T.; Anderegg, L.D.L.; Bielory, L.; Kinney, P.L.; Ziska, L. Anthropogenic Climate Change Is Worsening North American Pollen Seasons. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2013284118. [CrossRef]

71. Frenguelli, G.; Spieksma, F.T.M.; Bricchi, E.; Romano, B.; Mincigrucci, G.; Nikkels, A.H.; Dankaart, W.; Ferranti, F. The Influence of Air Temperature on the Starting Dates of the Pollen Season of Alnus and Populus. *Grana* **1991**, *30*, 196–200. [CrossRef]

72. Myszkowska, D.; Jenner, B.; Puc, M.; Stach, A.; Nowak, M.; Malkiewicz, M.; Chłopek, K.; Uruska, A.; Rapiejko, P.; Majkowska-Wojciechowska, B.; et al. Spatial Variations in the Dynamics of the Alnus and Corylus Pollen Seasons in Poland. *Aerobiologia* **2010**, *26*, 209–221. [CrossRef]

73. Puc, M.; Kasprzyk, I. The Patterns of Corylus and Alnus Pollen Seasons and Pollination Periods in Two Polish Cities Located in Different Climatic Regions. *Aerobiologia* **2013**, *29*, 495–511. [CrossRef]

74. Picornell, A.; Recio, M.; Ruiz-Mata, R.; García-Sánchez, J.; Cabezudo, B.; Trigo, M.d.M. Medium- and Long-Range Transport Events of Alnus Pollen in Western Mediterranean. *Int. J. Biometeorol.* **2020**, *64*, 1637–1647. [CrossRef] [PubMed]

75. Ziello, C.; Sparks, T.H.; Estrella, N.; Belmonte, J.; Bergmann, K.C.; Bucher, E.; Brighetti, M.A.; Damialis, A.; Detandt, M.; Galán, C.; et al. Changes to Airborne Pollen Counts across Europe. *PLoS ONE* **2012**, *7*, e34076. [CrossRef] [PubMed]

76. Zhang, Y.; Bielory, L.; Mi, Z.; Cai, T.; Robock, A.; Georgopoulos, P. Allergenic Pollen Season Variations in the Past Two Decades under Changing Climate in the United States. *Glob. Chang. Biol.* **2015**, *21*, 1581–1589. [CrossRef] [PubMed]

77. Mishra, A.; Jalaluddin, Z.M.; Mahamuni, C.V. Air Quality Analysis and Smog Detection in Smart Cities for Safer Transport Using Machine Learning (ML) Regression Models. In Proceedings of the 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, 23 April 2022; IEEE: New York, NY, USA, 2022; pp. 200–206.

78. Mohammed, W.; Adamescu, A.; Neil, L.; Shantz, N.; Townend, T.; Lysy, M.; Al-Abadleh, H.A. Application of Machine Learning and Statistical Modeling to Identify Sources of Air Pollutant Levels in Kitchener, Ontario, Canada. *Environ. Sci. Atmos.* **2022**, *2*, 1389–1399. [CrossRef]

79. Nowosad, J.; Stach, A.; Kasprzyk, I.; Chłopek, K.; Dąbrowska-Zapart, K.; Grewling, Ł.; Latałowa, M.; Pędziszewska, A.; Majkowska-Wojciechowska, B.; Myszkowska, D.; et al. Statistical Techniques for Modeling of Corylus, Alnus, and Betula Pollen Concentration in the Air. *Aerobiologia* **2018**, *34*, 301–313. [CrossRef]

80. Muzalyova, A.; Brunner, J.O.; Traidl-Hoffmann, C.; Damialis, A. Forecasting Betula and Poaceae Airborne Pollen Concentrations on a 3-Hourly Resolution in Augsburg, Germany: Toward Automatically Generated, Real-Time Predictions. *Aerobiologia* **2021**, *37*, 425–446. [CrossRef]

81. Stępalska, D.; Myszkowska, D.; Piotrowicz, K.; Kasprzyk, I. The Phenological Phases of Flowering and Pollen Seasons of Spring Flowering Tree Taxa against a Background of Meteorological Conditions in Kraków, Poland. *Acta Agrobot.* **2016**, *69*. [CrossRef]

82. Garrido, A.; Fernández-González, M.; Álvarez-López, S.; González-Fernández, E.; Rodríguez-Rajo, F.J. First Phenological and Aerobiological Assessment of Olive Orchards at the Northern Limit of the Mediterranean Bioclimatic Area. *Aerobiologia* **2020**, *36*, 641–656. [CrossRef]