

## Article

# Application of Photo Texture Analysis and Weather Data in Assessment of Air Quality in Terms of Airborne PM<sub>10</sub> and PM<sub>2.5</sub> Particulate Matter

Monika Chuchro <sup>1,\*</sup> , Wojciech Sarlej <sup>1</sup> , Marta Grzegorzczuk <sup>1</sup> and Karolina Nurzyńska <sup>2</sup> 

<sup>1</sup> Department of Geoinformatics and Applied Computer Science, Faculty of Geology, Geophysics and Environment Protection, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Krakow, Poland; wsarlej@agh.edu.pl (W.S.); marta.manka.95@gmail.com (M.G.)

<sup>2</sup> Institute of Informatics, Faculty of Automatic Control, Electronics, and Computer Science, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland; Karolina.Nurzynska@polsl.pl

\* Correspondence: chuchro@agh.edu.pl

**Abstract:** The study was undertaken in Krakow, which is situated in Lesser Poland Voivodeship, where bad PM<sub>10</sub> air-quality indicators occurred on more than 100 days in the years 2010–2019. Krakow has continuous air quality measurement in seven locations that are run by the Province Environmental Protection Inspectorate. The research aimed to create regression and classification models for PM<sub>10</sub> and PM<sub>2.5</sub> estimation based on sky photos and basic weather data. For this research, one short video with a resolution of 1920 × 1080 px was captured each day. From each film, only five frames were used, the information from which was averaged. Then, texture analysis was performed on each averaged photo frame. The results of the texture analysis were used in the regression and classification models. The regression models' quality for the test datasets equals 0.85 and 0.73 for PM<sub>10</sub> and 0.63 for PM<sub>2.5</sub>. The quality of each classification model differs (0.86 and 0.73 for PM<sub>10</sub>, and 0.80 for PM<sub>2.5</sub>). The obtained results show that the created classification models could be used in PM<sub>10</sub> and PM<sub>2.5</sub> air quality assessment. Moreover, the character of the obtained regression models indicates that their quality could be enhanced; thus, improved results could be obtained.

**Keywords:** classification; particulate matter; regression; texture analysis



**Citation:** Chuchro, M.; Sarlej, W.; Grzegorzczuk, M.; Nurzyńska, K. Application of Photo Texture Analysis and Weather Data in Assessment of Air Quality in Terms of Airborne PM<sub>10</sub> and PM<sub>2.5</sub> Particulate Matter. *Sensors* **2021**, *21*, 5483. <https://doi.org/10.3390/s21165483>

Academic Editors: Changho Lee and Changhan Yoon

Received: 5 July 2021

Accepted: 11 August 2021

Published: 14 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Air quality in cities and suburban areas is a crucial and emerging problem for governments. Among various air pollutants, airborne particulate matter (PM) with diameters less than 10 micrometers (PM<sub>10</sub>) and less than 2.5 micrometers (PM<sub>2.5</sub>) are the most common pollutants in Polish cities. PM is a complex mixture of extremely small particles and liquid droplets made up of acids, organic chemicals, metal, soil, and dust particles. Sources of PM are both natural and anthropogenic. Man-made sources of PM include combustion in mechanical and industrial processes, vehicle emissions, and even tobacco smoke. Natural sources include volcanoes, fires, dust storms, and aerosolized sea salt [1].

PM has a considerable negative effect on human health, including increased rates of cardiovascular, cerebrovascular, and respiratory diseases [2,3]. Puentes et al. in [4] showed that every 50 µg/m<sup>3</sup> increase in PM<sub>10</sub> caused 4–12% growth of hospital visits for children with respiratory syndromes. Many techniques are available to measure the mass concentration of PM in air. The most popular methods include filter-based gravimetric methods [5], tapered element oscillating microbalance [6], beta attenuation monitoring [7], optical analysis [8–10], and black smoke measurement [11]. All these methods require sophisticated equipment, space to install, and staff to maintain the equipment and evaluate the data. A simple, fast, and cheap method of monitoring PM in the air has the potential to increase public awareness, alert those with respiratory diseases to take proper prevention measures, and provide local air quality data that are not otherwise available. Easy access

to cheap cameras makes the described method easily accessible. The current methods, although undoubtedly more precise, are burdened with a very high cost of purchasing the equipment required for PM analysis. As indicated by the authors of the evaluation, the cost of purchasing a professional measuring device is so high that it becomes inaccessible for ordinary people [12].

The literature mentions the use of artificial neural networks, SVM, spatial interpolation models, and statistical models for air quality modeling. In neural network-based methods, Vahdatpour et al. proposed a method to estimate pollution forecast with Convolutional Neural Network based on sky images and Gabor transform [13,14]. Yang et al. presented shallow ResNet with layer enhancement for PM<sub>2.5</sub> index prediction [15] based on image data from Beijing, Shanghai in China. The usage of MLP was presented by the authors in [16,17]. Their results demonstrated that the MLP approach obtains good PM levels prediction quality.

Liu et al. in [18] proposed PM<sub>2.5</sub> level prediction with the SVR method based on six features (including textures) and weather data, time, and geographical location. The research was based on outdoor photographs from Beijing, Shanghai in China and Phoenix in US. In [19], the authors proposed a system to estimate air quality based on publicly available sky photos from Flickr and public webcams and statistics computed from sky pixels color values. The research of Sajjadi et al. in [20] presented PM<sub>2.5</sub> and PM<sub>10</sub> assessment in Sabzevar in Iran based on spatial interpolation models. Models of Radial Basis Functions, Inverse Distance Weighting, Ordinary Kriging, and Universal Kriging were used on 48 PM station data. The use of neural networks and public cameras or even smartphones with an integrated full-HD camera can provide an effective tool for assessing air quality [21,22]. Several forecasting models have been developed to assess levels of PM in atmospheric air without using photographs. In paper [23], the authors presentend generalized models with gamma distribution to predict daily average PM<sub>10</sub> in Brno, Czech Republic. Models were created for daily averaged data from two stations using weather data and additional seasonal informations. A similar approach was presented by the authors of [4]. The authors used a bivariate predictive model based on GBS distribution to predict next-day maximum PM<sub>10</sub> and PM<sub>2.5</sub> levels. In paper [24], the authors analyzed the impact of weather on PM level using generalized additive models. Finally, in [25], the authors presented a system to estimate PM<sub>2.5</sub>/haze level based on a single photograph. Experiments were performed both on synthetic and real datasets with depth trans and jcsb2014 methods.

The aim of the project was assessment of PM<sub>10</sub> and PM<sub>2.5</sub> pollution using image texture analysis and commonly available weather data (air temperature, precipitation, average wind speed). The authors aimed to verify whether it is possible to correctly forecast the pollution having information obtained from a simple camera, an image sensor that everyone has at home, and weather forecast one can obtain online. The task used the signal processing approach to derive discriminative features from images that are not visible with the naked eye. These data were combined with information collected by sensors built into commercial devices (Vaisala WXT520, Plantower PMS5003) that were measuring parameters such as wind, rain, pressure, temperature, and relative humidity. The assessment of air pollution with PM<sub>10</sub> and PM<sub>2.5</sub> was carried out quantitatively and qualitatively with the use of multilayer perceptron (MLP) artificial neural networks. Despite the fact that air quality in Krakow has improved in recent years, a tool that could quickly assess air quality anywhere and at any time of day would be useful. Such a tool could be based on regression or classification models that assess air quality based on texture and weather information. For example, the entire model could be integrated into a smartphone or tablet application.

## 2. Air Quality Assessment for the City of Krakow

Krakow agglomeration is located in the Lesser Poland Voivodeship; it is inhabited by around 770,000 people and covers an area of 327 km<sup>2</sup>. Air quality assessment for the city is based on the Regulation of the Minister of the Environment from 24 August 2012 on the

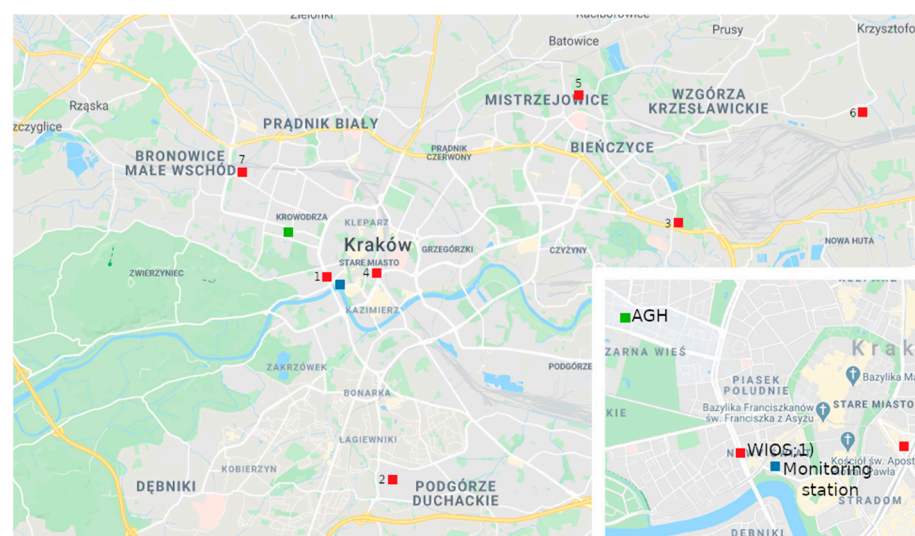
topic of levels of certain substances in the air and EU directives for the protection of human health and plant protection [26–29] in which acceptable levels of atmospheric pollution are specified. The maximum yearly average concentration of PM<sub>10</sub> equals 40 µg/m<sup>3</sup>, but the aforementioned regulation [28] also allows the daily average concentration to be exceeded (Table 1). The norm for PM<sub>2.5</sub> concentration equals 25 µg/m<sup>3</sup> without any exceptions.

**Table 1.** Acceptable daily and annual maximum levels of PM<sub>10</sub> concentration in the air.

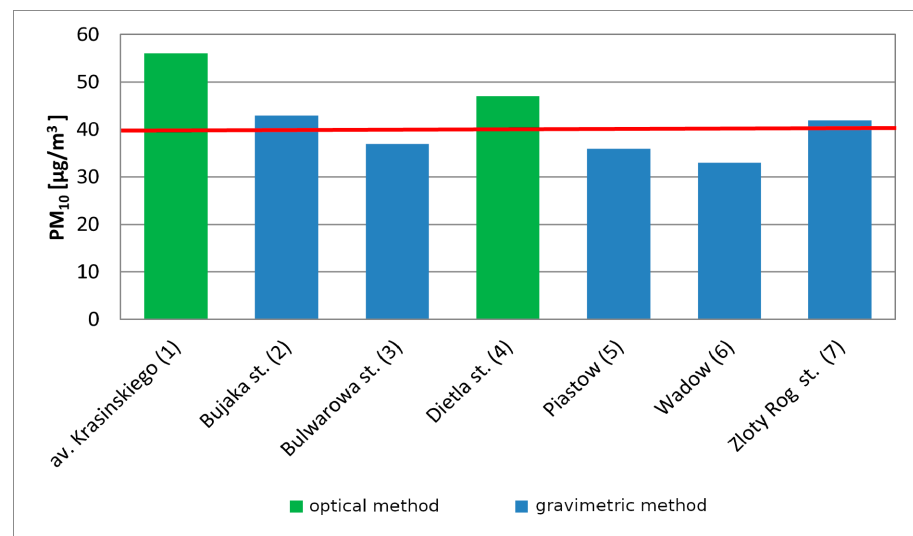
Average Period of Concentration	Permissible Level of PM <sub>10</sub> in the Air [µg/m <sup>3</sup> ]	Allowed Frequency of Exceeding the Permissible Level of PM <sub>10</sub>
24 h	50	35 times
Calendar year	40	N/A

Measurements of PM<sub>10</sub> and PM<sub>2.5</sub> content were carried out every day using continuous, high-quality, automatic or manual methods. The main factor affecting air quality in Lesser Poland Voivodeship is emissions from the municipal and household sectors. In the structure of pollutants, these emissions account for approximately 77% PM<sub>10</sub>, 88% PM<sub>2.5</sub>, 97% BaP, 14% NO<sub>x</sub>, and 65% SO<sub>x</sub>. An increased pollution level is especially visible in winter, when fuel consumption for heating increases due to low temperatures. In summer, emissions from the municipal and household sectors decrease; therefore, emissions only come from households that use solid-fuel furnaces to heat water. As in other highly polluted cities, such as Santiago, Chile, it should be noted that topographic (poor ventilation), meteorological (low temperature at winter), and socio-economic (coal-fired home heating) conditions negatively affect air quality in the city, especially during the winter period [4]. Another source of emissions that is visible especially in large cities and agglomerations is transport, which accounts for around 5% of PM<sub>10</sub> emissions, 4% of PM<sub>2.5</sub> emissions, and 44% of NO<sub>x</sub> emissions for the whole voivodeship [30].

To assess air quality in Krakow for PM<sub>10</sub>, measurement series from seven stations (two automatic optical and five manual gravimetric) were collected and analyzed by the Province Environmental Protection Inspectorate. The locations of all stations are presented in Figure 1. The most frequently exceeded norm was observed at Krasinski Avenue [30], and the highest yearly PM<sub>10</sub> number is also observed there (Figure 2).



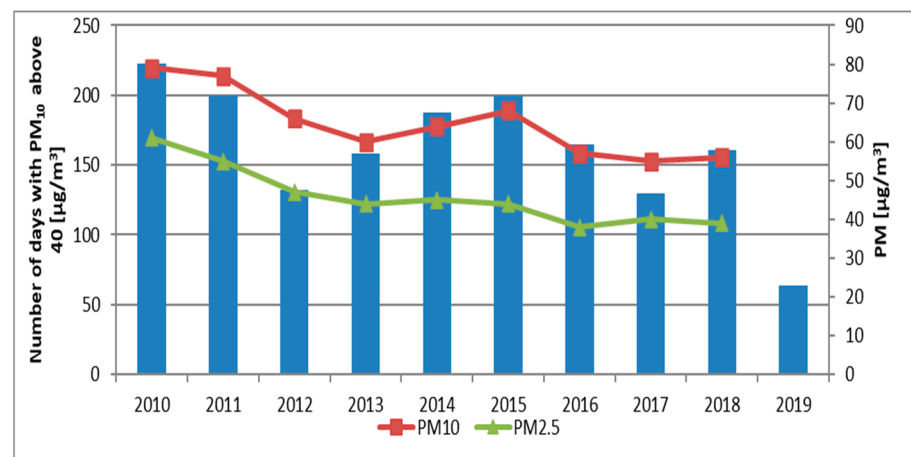
**Figure 1.** WIOS air quality station locations in Krakow.



**Figure 2.** Concentration of PM<sub>10</sub> at WIOS stations in Krakow in 2017.

To assess air quality in terms of PM<sub>2.5</sub> particulate matter, measurement series from three stations (two automatic and one manual) were analyzed. The average annual concentrations ranged from 39 µg/m<sup>3</sup> at Krasynski Avenue up to 27 µg/m<sup>3</sup> at Bulwarowa Street in 2017 [30].

The analysis of the number of exceedances of PM<sub>10</sub> levels showed that the worst year in terms of air quality was 2010, which had 223 exceedances. The data show a decreasing trend, with a period of increased values in 2014–2015 (Figure 3) and a lower exceedance value in 2012. The concentrations of PM<sub>10</sub> and PM<sub>2.5</sub> show a clear decreasing trend (Figure 3), but a slight increase in concentrations was visible in the years 2014–2015.



**Figure 3.** Days per year with exceeded norms of PM<sub>10</sub> for Krakow agglomeration and PM<sub>10</sub>, PM<sub>2.5</sub> concentrations at Krasynskiego.

### 3. Theory

In this paper, texture analysis using First-order Features, Gray-level Co-occurrence Matrix, and Grey Tone Difference Matrix methods was used to extract information from images. On the basis of the obtained results and weather data, regression and classification neural networks were created using multilayer perceptron. The task of the neural networks was to estimate PM<sub>10</sub> and PM<sub>2.5</sub> in atmospheric air and exceedances of air quality. In this section, the theory of used methods is presented.

### 3.1. Texture Analysis

The image processing literature illustrates multiple methods of describing the characteristics of a texture. The simplest method is image histogram analysis (e.g., First-order Features). More sophisticated approaches additionally analyze local changes in pixel intensities (e.g., Gray-level Co-occurrence Matrix); others try to mimic the way the human visual system works (e.g., Gray Tone Difference Matrix) [31,32]. Since the texture operator has already proved its sensitiveness for tiny changes resulting from noise introduction [33], it is believed the pollution recorded on images may also be visible. Moreover, in the presented research, a texture operator, which returns a short vector of values that could describe the image quality, was needed (which in assumption should vary depending on air condition). As a consequence, the authors have concentrated on early approaches of texture description and neglected the more up-to-date methods, e.g., Local Binary Patterns or Histogram of Oriented Gradients, that describe the image with extremely long feature vectors. Details of the methods chosen for this research are described in the following paragraphs.

The First-order Features method (FOF) is based on statistical information derived from a normalized image histogram. It denotes a gray-scale image whose resolution is  $H \times W$ , and there are  $G$  pixel intensities. The normalized histogram is formulated as:

$$p(i) = 1/WH \sum_{x=1}^W \sum_{y=1}^H I(x, y) == i. \quad (1)$$

From such a histogram, the following features can be derived: mean, variance, kurtosis, skewness, energy, and entropy; their formulas are given in Equation (2) [34].

$$\text{FOF1 : mean : } \mu = \sum_{i=0}^{G-1} ip(i)$$

$$\text{FOF2 : variance : } \sigma^2 = \sum_{i=0}^{G-1} (i - \mu^2)p(i)$$

$$\text{FOF3 : kurtosis : } \mu_3 = \sigma^{-3} \sum_{i=0}^{G-1} (i - \mu^3)p(i)$$

$$\text{FOF4 : skewness : } \mu_4 = \sigma^{-4} \sum_{i=0}^{G-1} (i - \mu^4)p(i) - 3$$

$$\text{FOF5 : energy : } E = \sum_{i=0}^{G-1} [p(i)]^2$$

$$\text{FOF6 : entropy : } H = - \sum_{i=0}^{G-1} p(i) \log_2[p(i)] \quad (2)$$

In the Neighborhood Gray-tone Difference Matrix method (GTDM) [35], higher-order parameters are based on the histogram of differences between the intensity of the central pixel and its eight-sided neighborhood average. Such a definition mimics the way humans perceive brightness. From this data structure, five features are derived, which describe the general image quality: thickness, contrast, business value, complexity, and endurance. Equation (3) defines these features.

$$\text{GTDM1 : coarseness : } \left[ \varepsilon + \sum_{i=0}^{G_h} p_i s(i) \right]^{-1}$$

$$\text{GTDM2 : contrast : } \left[ \frac{1}{N_g(N_g) - 1} \sum_{i=0}^{G_h} \sum_{j=0}^{G_h} p_i p_j (i - j)^2 \right] \left[ \frac{1}{n^2} \sum_{i=0}^{G_h} s(i) \right]$$

$$\begin{aligned}
 \text{GTDM3 : business value : } & \frac{\left[ \sum_{i=0}^{G_h} p_i s(i) \right]}{\sum_{i=0}^{G_h} \sum_{j=0}^{G_h} i p_i - j p_j} \\
 \text{GTDM4 : complexity : } & \left[ \frac{\sum_{i=0}^{G_h} \sum_{j=0}^{G_h} (\text{mod}(i, j))}{n^2 (p_i + p_j)} (p_i s(i) + p_j s(j)) \right] \\
 \text{GTDM5 : strength : } & \sum_{i=0}^{G_h} \sum_{j=0}^{G_h} (p_i + p_j) \frac{(i - j)^2}{\varepsilon - \sum_{i=0}^{G_h} s(i)} \quad (3)
 \end{aligned}$$

where

$p_i$ —the probability of the dependence of a pixel on intensity  $i$ ;

$s(i)$ —NGTDM intensity value calculated as  $\sum |i - A_i|$ ;

$A_i$ —the average intensity of the surrounding pixels without considering the center voxel (calculated from the intensity  $i$ ).

The Gray-Level Co-occurrence Matrix (COM) is the third texture method used in this paper. This method was introduced by Haralick [36] and is based on a symmetrical co-occurrence matrix. Each matrix entry stores information about the number of occurrences of two-pixel intensities that index this element in an image. This definition makes it possible to memorize not only an image's intensity distribution but also its spatial relation to the image. The matrix dimensions are related to the number of gray-level pixel values,  $G$ , and may be reduced to a smaller number to optimize the calculation time. In the presented research,  $G$  equals 64. To assure the rotational invariance of the method, Haralick et al. suggested incorporating into the co-occurrence matrix the information that is obtained from analyzing the pixel adjacency in four directions [36]. Moreover, it is possible to define the distance between pixels, which are assumed to be adjacent. In the presented experiments, this distance equals one. These authors also suggested deriving 14 features that describe various qualities of a texture. The detailed formulation is presented in Equation (4).

$$\begin{aligned}
 \text{COM1 : Angular Second Moment : } & \sum_i \sum_j p(i, j)^2 \\
 \text{COM2 : Contrast : } & \sum_{n=0}^{N_p-1} n^2 \left\{ \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} p(i, j) \right\}, \text{ mod}(i - j) = n \\
 \text{COM3 : Correlation : } & \frac{\sum_i \sum_j (ij) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \\
 \text{COM4 : Sum of Squares : Variance : } & \sum_i \sum_j (i - \mu)^2 p(i, j) \\
 \text{COM5 : Inverse Difference Moment : } & \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j) \\
 \text{COM6 : Total Average : } & \sum_{i=2}^{2N_p} i p_{x+y}(i) \\
 \text{COM7 : Sum of Variance : } & \sum_{i=2}^{2N_p} (i - f_s)^2 p_{x+y}(i) \\
 \text{COM8 : Sum of Entropy : } & - \sum_{i=2}^{2N_p} p_{x+y}(i) \log \{ p_{x+y}(i) \} \\
 \text{COM9 : Entropy : } & - \sum_i \sum_j p(i, j) \log(p(i, j))
 \end{aligned}$$

$$\text{COM10 : Difference Variance : } \sum_{i=0}^{N_p-1} i^2 p_{x+y}(i)$$

$$\text{COM11 : Difference Entropy : } - \sum_{i=0}^{N_p-1} i^2 p_{x-y}(i) \log\{p_{x-y}(i)\}$$

$$\text{COM12 : Information Measures of Correlation 1 : } \frac{HXY - HXY1}{\max\{HX, HY\}}$$

$$\text{COM13 : Information Measures of Correlation 2 : } (1 - \exp[-2(HXY2 - HXY)])^{1/2}$$

$$\text{COM14 : Max. Correlation Coefficient : } (\text{sec ond biggest value of } Q)^{1/2}$$

$$HXY = - \sum_i \sum_j p(i, j) \log(p(i, j))$$

$HX, HY$ —entropies  $p_x, p_y$

$$HXY1 = - \sum_i \sum_j p(i, j) \log(p_x(i) p_y(j))$$

$$HXY2 = - \sum_i \sum_j p_x(i) p_y(j) \log(p_x(i) p_y(j))$$

$$Q(i, j) = \sum_k \frac{P(i, k)P(j, k)}{P_x(i)P_y(k)}. \quad (4)$$

where

$\mu$ —average value,  $\sigma$ —standard deviation;

$p_x$ —partial probability density function;

$x, y$ —input coordinates in the co-occurrence matrix;

$p_{x+y}(i)$ —probability of the sum of the  $x$  and  $y$  coordinates from the matrix of co-occurrences;

### 3.2. Correlation

The correlation coefficient describes the statistical relationship between features. Pearson's correlation is used to measure the statistical relationship or association between two continuous variables. It provides information about the magnitude of the association and the direction of the relationship [37]. Calculation of Pearson's correlation coefficient requires the assumption that two samples are normally distributed. If normality is violated, Pearson's correlation coefficient gives unreliable results. Hence, an alternative to Pearson's correlation coefficient may be Spearman's rank correlation. The dependence of ordinal variables is designated as rank correlation, and their intensity is represented by correlation coefficients [37]. Correlations with  $p \leq 0.05$  are considered statistically significant, which corresponds to correlation values greater than 0.195.

### 3.3. Artificial Neural Networks

Artificial neural networks are a powerful data modeling tool with a high proven efficiency in dealing with nonlinearity in a dataset as well as complex problems in the classification, regression, and clustering fields [17,37,38]. An extensive description of neural networks has been provided by the authors in [15,17,37–39]. For the regression problem, a neural network with a multilayer perceptron (MLP) was selected [16,17,40]. Network training was performed with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, which is an iterative method for solving unconstrained nonlinear optimization problems. The BFGS is a memory-efficient training algorithm usually used for nonlinear least squares and may require in a smaller number of iterations to train a neural network [38]. The sum of squares (SOS) was selected as the error function. Initially, the number of neurons in the input layer depended on statistically significant correlations: one neuron for each quantitative variable. The optimal number of variables used in the model was based on sensitivity analysis, so the number of neurons in the input layer was changeable. Variables

that did not improve the quality of the obtained networks were removed from the neural network models. The constructed neural networks included one layer of hidden neurons with five to 25 neurons and one neuron in the output layer [17,38].

Originally, two types of neural networks were considered for the classification problem: a multilayer perceptron and a network with a radial basis function (RBF) [39]. After preliminary analyses, it turned out that the quality of the RBF network was lower than that of MLP; therefore, the multilayer perceptron was chosen. Moreover, for neural networks classification, sensitivity analysis was used to select the optimal number of variables, and the number of neurons in the hidden layer was from five to 25 neurons. The output layer contained as many neurons as the air quality classes (for  $PM_{10}$ —3, for  $PM_{2.5}$ —2).

The quality of the regression neural networks was expressed as the value of Pearson's correlation coefficient between the data and the model, and the Coefficient of Determination ( $R^2$ ). Mean Error (ME), and Mean Absolute Percentage Error (MAPE) were used additionally for test datasets. The quality of the classification networks was expressed as the percentage of correct classifications, which corresponds with accuracy \* 100%. Additionally, Area Under Curve (AUC) was used to assess the ability of a classifier to distinguish between classes.

#### 4. Proposed Method

For this project dataset of images, numerical datasets of weather data and air pollution with  $PM_{10}$  and  $PM_{2.5}$  data have been used. Proposed method for estimating  $PM_{10}$  and  $PM_{2.5}$  in atmospheric air and exceedances of air quality was based on the Cross-Industry Standard Process for Data Mining (CRISP-dm) methodology. Figure 4 illustrates the block diagram of the proposed method. The authors described the phases presented in Figure 4 in the following parts of this paper. The proposed method is a transitive method between the use of neural networks to predict air quality based on photos presented in publications [13–15] and the approach based on modeling numerical data presented by the authors in [16,23,24]. In the first step, texture analysis for each of the average photo frames has been performed, using three complementary methods. Thanks to the selected methods, a single value for each texture feature was obtained. The next step was to combine the results from the texture analysis, weather, and air pollution data. Then, regression and classification MLP neural networks were performed for numerical data. The selected method allowed for the creation and preliminary evaluation of a single neural network in less than 1 s. In the last step, the quality of selected neural network models was assessed.

##### 4.1. Data Collection

Data analysis was performed on good-quality image data that were acquired daily. Image data acquisition was performed periodically for 100 days (from 21 November 2018 to 28 February 2019) at sunrise (between 7 and 7.30 AM, UTC+1) from the “Wawel castle and the Vistula bend” city monitoring station, whose coordinates are 50°03' N 19°55' [41]. The automated process was downloading videos from the camera according to the aforementioned schedule. Short \*.mp4 files (movies with resolution 1920 × 1080 px) were saved in the disk space, from which five random frames were extracted and averaged. Each frame contains the same view presented in Figure 5. Images carry unstructured sets of pixels; however, each image has a texture that has been analyzed in further steps. On the other hand, weather parameters collected by sensors have been provided as structured \*.csv files. The combination of both sources done by the automated process gave the results presented below. To ensure the best possible convergence of parameters, weather and air quality data were collected from two weather stations located in the neighborhood of the image acquisition location, which is presented in Figure 1.



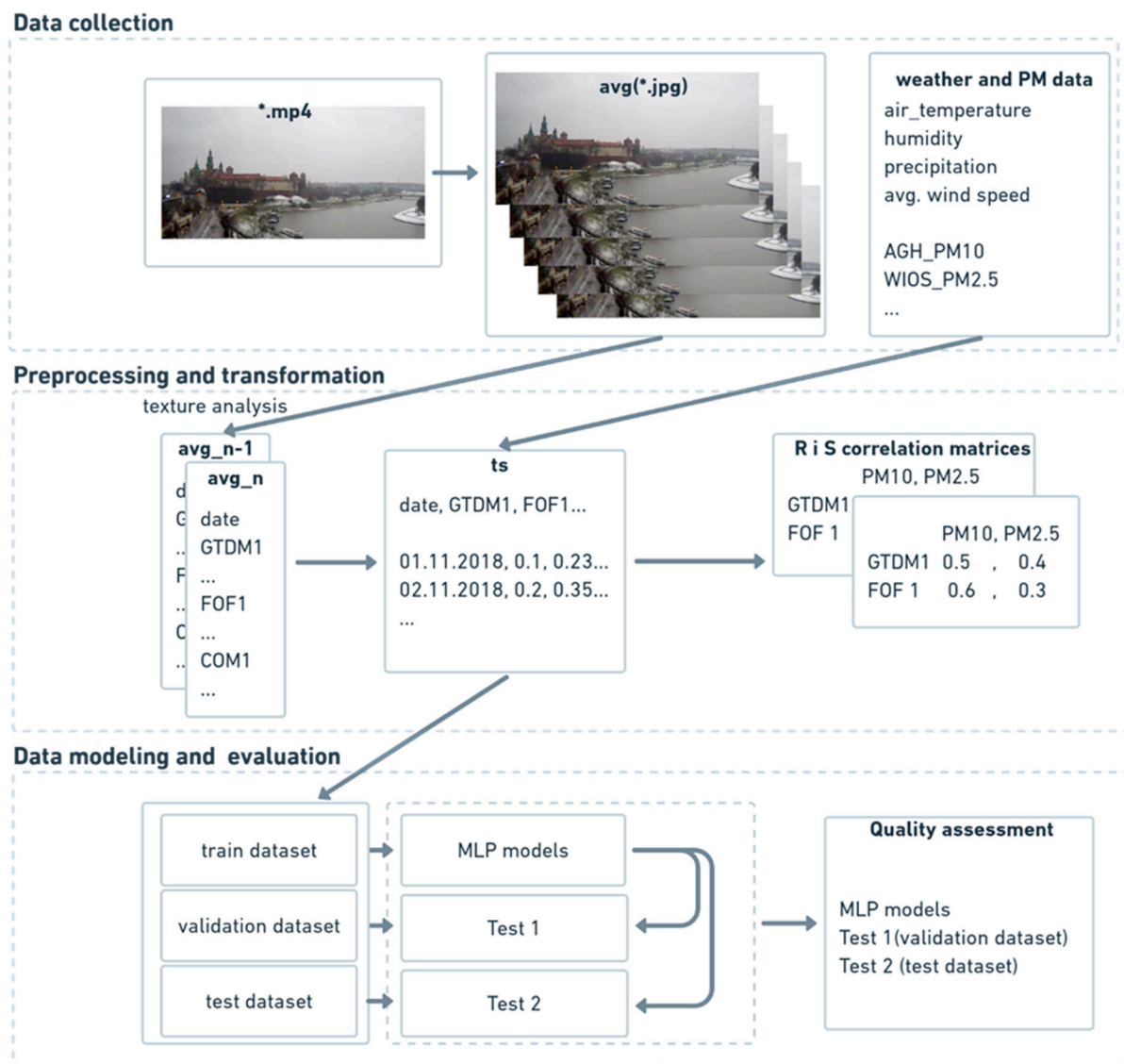


Figure 4. Block diagram of the proposed method.



Figure 5. Single frame used in visual inspection of air quality.

Weather data were obtained from the weather station run by the Environmental Physics Group of AGH University of Science and Technology in Krakow (AGH), which is shown on the map (Figure 1) as a green square [42]. The averaged results of parameter

measurements from a full hour were registered. For this period (21 November 2018 to 28 February 2019), 7:00 a.m. was established as a measuring point. The location of the station is 50°04' N 19°55' E, and its foundation height is 220 m a.s.l. The weather station provided meteorological information, including average air temperature, average relative humidity, average atmospheric pressure, average wind speed, maximum speed of wind gusts, precipitation and air quality status assessed using the PM<sub>10</sub> indicator. Weather and airborne particulate matter measurements were provided by a Vaisala WXT520 automatic weather station and a Plantower PMS5003 [42].

Additional air quality status information was obtained from the station of the Voivodeship Inspectorate of Environmental Protection in Krakow (WIOS), located at Krasinski Avenue (Figure 1) [43]. The measuring point coordinates are 50°03' N 19°55' E; its foundation height is 207 m a.s.l. The station is located close to the “Wawel Castle and Vistula bend” monitoring point, which provided images from webcams. The measuring station at Krasinski Avenue recorded the following parameters: nitrogen oxide content, carbon monoxide content, nitrogen dioxide content, benzene content, PM<sub>10</sub>, and PM<sub>2.5</sub> indicators.

#### 4.2. Preprocessing

Every object has a texture that can be used to characterize it. Even if one considers an object to have no texture, image processing methods consider it to have a plain texture. An example of a plain texture is a photograph of a clear sky on a sunny day. When clouds are visible, each of them may have a different texture. Changes in air pollution are also noticeable in photographs and can be reflected as changes in texture. The project used three complementary texture methods presented in Section 3 on average images from the camera. The textures features were saved into structured \*.csv file. In the next step, files with weather data, texture features, and PM data were combined into one file with a time index (see Figure 4). Such a prepared file was used in correlation analysis and modeling.

#### 4.3. Modeling and Evaluation

Regression and classification of the neural network models were used to assess the possibility of predicting the content of PM<sub>10</sub> and PM<sub>2.5</sub> in atmospheric air. Moreover, based on image texture parameters and basic meteorological data (air temperature, average wind speed, humidity, precipitation), they were also used to predict the exceedances of air quality standards.

The analyzed particulate matter data from each of the measuring stations were divided into 3 subsets. Since the analyzed data are a time series, the data were not randomly divided into subsets. The oldest 70% of the observations were assigned to the training set, another 15% was assigned to the validation set, and the newest observations were assigned to the test set. One thousand different regression neural networks with different architectures were made for each parameter. The neural network was created using a training dataset with the automatic Data Mining toolbox in Statistica 13.1. The created neural networks had a three-layer structure with one hidden layer. In the hidden layer, we tested from 5 to 25 neurons. The maximum number of epochs was 200, and the stop criterion was 0.00001. The initial weights were random using normally distributed values within a range whose mean is zero and standard deviation is equal to one. These networks differed also in the activation functions of neurons in the hidden layer and the output layer. Each neural network was created based on a training dataset. Successively, the fitted model is used to predict the responses for the observations in a second dataset called the validation dataset. The validation dataset provides an unbiased evaluation of a model fit on the training dataset. In the last step, an unbiased evaluation of a final model fit on the training dataset is performed with the test dataset.

Of all the created networks, the one with the highest quality values for all three datasets and the lowest number of neurons in the hidden layer was selected for each predicted PM dataset, to present in this paper.

Classification methods are sensitive to the unevenness of the occurrence of classes in subsets. Therefore, for the classification methods, only a random selection of observations for individual sets was used to better balance the observations belonging to individual classes. For the classification models, the quantitative data were changed to 3 PM<sub>10</sub> classes. The good air quality class was defined as PM<sub>10</sub> lower than 50 µg/m<sup>3</sup>; the poor air quality class was PM<sub>10</sub> between 50 and 100 µg/m<sup>3</sup>; and the very poor-quality class was PM<sub>10</sub> higher than 100 µg/m<sup>3</sup>. This division was made based on the ordinance of the Polish Minister of the Environment [28,29]. For suspended particle matter with a diameter smaller than 2.5 micrometers, two classes were distinguished: air quality is considered good if dust content is lower than 25 µg/m<sup>3</sup>, and it is considered poor for higher values [28]. The same as for the regression models, one thousand models were created automatically for each PM dataset, with the same conditions as presented in previous paragraph; from these, only the ones with the highest quality were used in the tests.

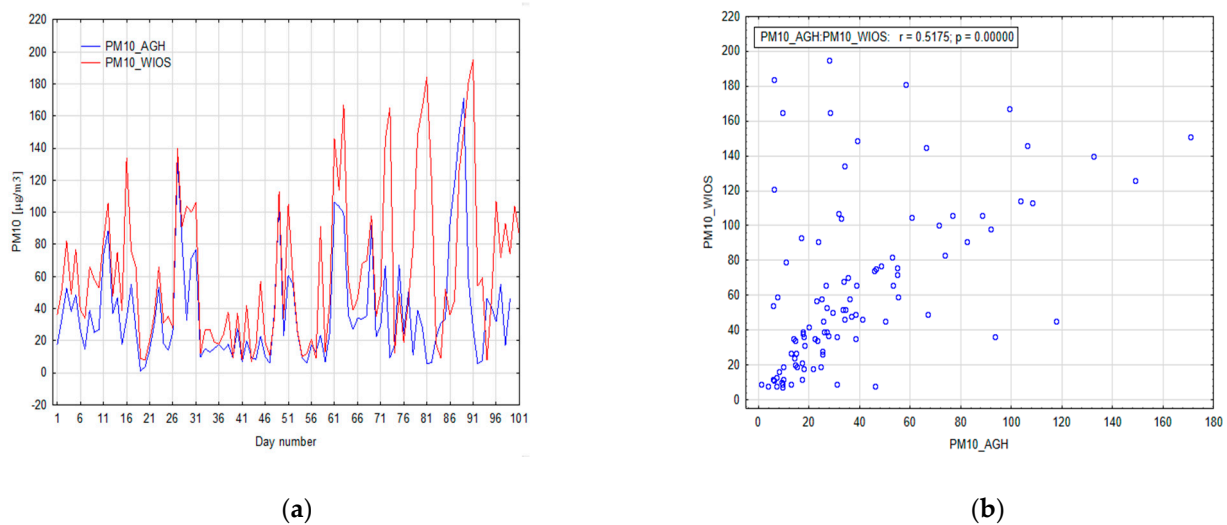
#### 4.4. Software

The project was implemented on computers with a Windows 10 × 64-bit system with an Intel Core i7-3630QM CPU 2.4 GHz processor, 16 GB RAM, and a Windows 10 × 64-bit system with an Intel Core i7-10710U CPU 1.10 GHz processor, 16 GB RAM. In this project, we used specialized software, computing platforms, and programming languages for statistical computing and graphics. For texture analysis, we used MATLAB 2016a, which is a programming and numeric computing platform. Correlation matrices and quality assessment were prepared in the R × 64 4.0.2 language with function `cor` from `stats` package version 3.6.2, and `caret` package version 6.0–88. Neural networks were created in the Statistica 13.1 64-bit program with SANN toolbox for automatic neural network building.

## 5. Results

Data analysis was performed to assess the relationship between dust suspended in atmospheric air and the texture parameters calculated from the averaged film frames recorded at sunrise and the basic weather data: air temperature, humidity, precipitation, average wind speed. The following analyses were carried out in sequence: Pearson's linear correlation, Spearman's rank nonlinear correlation, MLP regression neural network model, and MLP neural network classification models.

In the first step, the relationship between the measurement stations for the recorded suspended PM<sub>10</sub> dust measurements was determined. The linear correlation coefficient between PM<sub>10</sub> at AGH station and PM<sub>10</sub> at WIOS is 0.52. The value of the Spearman's rank correlation is slightly higher, which indicates the presence of nonlinear relationships between these stations (0.6). This is a relatively low similarity when it is considered that the stations are only about one kilometer apart. The analysis of the line graph of both PM<sub>10</sub> datasets shows that the data are positively correlated with each other (Figure 6a), but there are also significant differences in mean values and variation between those two stations (Figure 6a). Differences between measurements are smaller for low PM<sub>10</sub> values (Figure 6b). As the measured values increase, the difference between the PM<sub>10</sub> datasets also increases. In Figure 6a, two periods are visible with very high differences between the PM<sub>10</sub> WIOS station and PM<sub>10</sub> AGH station values. The first period is 70–73 observations with a maximum of 73 observations. The difference between measurements is 155.6 µg/m<sup>3</sup>. The second period is 76–81 observations with a maximum of 81 observations. The difference between measurements is 178 µg/m<sup>3</sup>. Both days with very high differences in measurements are days with negative air temperature lower than −5 °C, with an average wind speed lower than 1.5 m/s and without precipitation. Both periods are in a validation dataset. This may be related to the fact that the WIOS station is located in a green area in the middle of a busy road, while the AGH station is located near a less busy road on the AGH campus.



**Figure 6.** Correlation between AGH and WIOS stations for PM10: (a) time series comparison, (b) correlation plot.

Based on the correlation analysis, we can conclude that there is a significant variation in PM content even over short distances; therefore, the models will be local in nature and will only be accurate in the immediate vicinity. In the next step, the relationship between the measurements for the WIOS station was computed. The Pearson's correlation coefficient between  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  is 0.89. The value of the Spearman's rank correlation is slightly lower, which indicates the presence of strong linear relationships between these stations (0.87).

The correlations between  $\text{PM}_{10}$ ,  $\text{PM}_{2.5}$ , and texture parameters as well as air temperature, precipitation, and air humidity were calculated. Pearson's linear correlation coefficient was used to assess linear relationships; Spearman's rank coefficient was used to assess nonlinear relationships. In the Pearson's correlation coefficients, the results with the highest correlation coefficient values for texture parameters (Table 2) were obtained using the GTDM method. All texture parameters are statistically significant with the  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  datasets. The highest correlation coefficient values were observed for  $\text{PM}_{10}$  from the WIOS station. In the FOF method, only two parameters correlated at a significant level for  $\text{PM}_{10}$  from the WIOS station (FOF1, FOF3).  $\text{PM}_{10}$  from the AGH station and  $\text{PM}_{2.5}$  were not significantly correlated with First-order Feature texture data. In Haralick's method, almost all texture parameters correlated with  $\text{PM}_{10}$  WIOS; only correlations with COM1, COM8, and COM14 were not statistically significant ( $p > 0.05$ ). Pearson's correlation coefficients obtained for  $\text{PM}_{2.5}$  at WIOS and  $\text{PM}_{10}$  at AGH were weaker: 4% and 10%, respectively. Furthermore, the number of statistically significant correlations decreased: 20 for  $\text{PM}_{10}$  WIOS, 15 for  $\text{PM}_{2.5}$  WIOS, and 12 for  $\text{PM}_{10}$  AGH (Table 2). The highest Pearson's correlation coefficient value was observed for the correlation between average wind speed and  $\text{PM}_{10}$  at WIOS station ( $-0.626$ ). Moreover, Pearson's correlation coefficient between  $\text{PM}_{2.5}$  from WIOS station and the average wind speed equals  $-0.58$ . It is visible that a medium-strength negative relationship exists between PM data and average wind speed. The highest correlation between PM datasets and texture parameters was observed between  $\text{PM}_{2.5}$  from WIOS station and GTDM2 ( $-0.485$ ).

When comparing the results of the two correlation methods, the Spearman correlation coefficient was found to have higher values. The number of statistically significant correlations increases for  $\text{PM}_{10}$  from the AGH station. For all stations, the FOF4 and COM1 Spearman rank correlation coefficients are higher than Pearson's correlation coefficients and are statistically significant.

**Table 2.** Pearson’s (R) and Spearman (S) correlation coefficients between PM<sub>10</sub> and PM<sub>2.5</sub> texture parameters and weather data. Statistically significant correlations with  $p \leq 0.005$  are marked in bold.

Parameter	Cor.	PM <sub>10</sub> _AGH	PM <sub>10</sub> _WIOS	PM <sub>2.5</sub> _WIOS
Average hourly air temperature	R	<b>−0.224</b>	<b>−0.300</b>	<b>−0.284</b>
	S	<i>−0.270</i>	<i>−0.294</i>	<b>−0.238</b>
Average hourly wind speed	R	<b>−0.487</b>	<b>−0.626</b>	<b>−0.580</b>
	S	<i>−0.526</i>	<i>−0.628</i>	<i>−0.550</i>
Average hourly relative humidity	R	−0.016	−0.112	−0.047
	S	0.004	−0.169	−0.068
GTDM1	R	<b>0.297</b>	<b>0.451</b>	<b>0.441</b>
	S	<i>0.338</i>	<i>0.479</i>	<i>0.454</i>
GTDM2	R	<b>−0.401</b>	<b>−0.483</b>	<b>−0.485</b>
	S	<i>−0.457</i>	<i>−0.516</i>	<i>−0.498</i>
GTDM3	R	<b>−0.286</b>	<b>−0.379</b>	<b>−0.405</b>
	S	<i>−0.339</i>	<i>−0.394</i>	<i>−0.424</i>
GTDM4	R	<b>−0.301</b>	<b>−0.448</b>	<b>−0.413</b>
	S	<i>−0.372</i>	<i>−0.497</i>	<i>−0.458</i>
GTDM5	R	<b>0.250</b>	<b>0.479</b>	<b>0.423</b>
	S	<i>0.314</i>	<i>0.445</i>	<i>0.392</i>
FOF1	R	−0.064	<b>−0.220</b>	−0.139
	S	−0.008	−0.162	−0.083
FOF2	R	−0.020	0.156	0.076
	S	−0.051	0.159	0.079
FOF3	R	0.062	<b>0.249</b>	0.153
	S	0.011	<i>0.238</i>	0.142
FOF4	R	0.175	0.126	0.162
	S	<i>0.225</i>	<i>0.259</i>	<i>0.256</i>
FOF5	R	0.025	0.130	0.106
	S	0.119	0.124	0.124
FOF6	R	−0.053	−0.049	−0.06
	S	−0.122	−0.045	−0.084
COM1	R	0.059	0.153	0.129
	S	<i>0.210</i>	<i>0.211</i>	<b>0.221</b>
COM2	R	<b>−0.292</b>	<b>−0.416</b>	<b>−0.383</b>
	S	<i>−0.392</i>	<i>−0.518</i>	<i>−0.471</i>
COM3	R	0.174	<b>0.309</b>	<b>0.262</b>
	S	<i>0.303</i>	<i>0.447</i>	<i>0.384</i>
COM4	R	−0.175	<b>−0.249</b>	−0.192
	S	−0.156	<b>−0.229</b>	−0.152
COM5	R	<b>0.296</b>	<b>0.464</b>	<b>0.429</b>
	S	<i>0.353</i>	<i>0.465</i>	<i>0.429</i>
COM6	R	−0.064	<b>−0.219</b>	−0.139
	S	−0.009	−0.162	−0.083
COM7	R	−0.172	<b>−0.244</b>	−0.187
	S	−0.155	<i>−0.227</i>	−0.149
COM8	R	−0.105	−0.129	−0.14
	S	−0.181	−0.148	−0.168
COM9	R	<b>−0.24</b>	<b>−0.366</b>	<b>−0.344</b>
	S	<i>−0.324</i>	<i>−0.399</i>	<i>−0.371</i>

Table 2. Cont.

Parameter	Cor.	PM <sub>10</sub> _AGH	PM <sub>10</sub> _WIOS	PM <sub>2.5</sub> _WIOS
COM10	R	<b>−0.214</b>	<b>−0.279</b>	<b>−0.256</b>
	S	<i>−0.396</i>	<i>−0.52</i>	<i>−0.474</i>
COM11	R	<b>−0.298</b>	<b>−0.472</b>	<b>−0.433</b>
	S	<i>−0.357</i>	<i>−0.476</i>	<i>−0.433</i>
COM12	R	<b>−0.278</b>	<b>−0.467</b>	<b>−0.427</b>
	S	<i>−0.332</i>	<i>−0.472</i>	<i>−0.432</i>
COM13	R	0.155	<b>0.293</b>	<b>0.247</b>
	S	<i>0.286</i>	<i>0.461</i>	<i>0.433</i>
COM14	R	0.109	0.102	0.109
	S	0.110	0.087	0.109

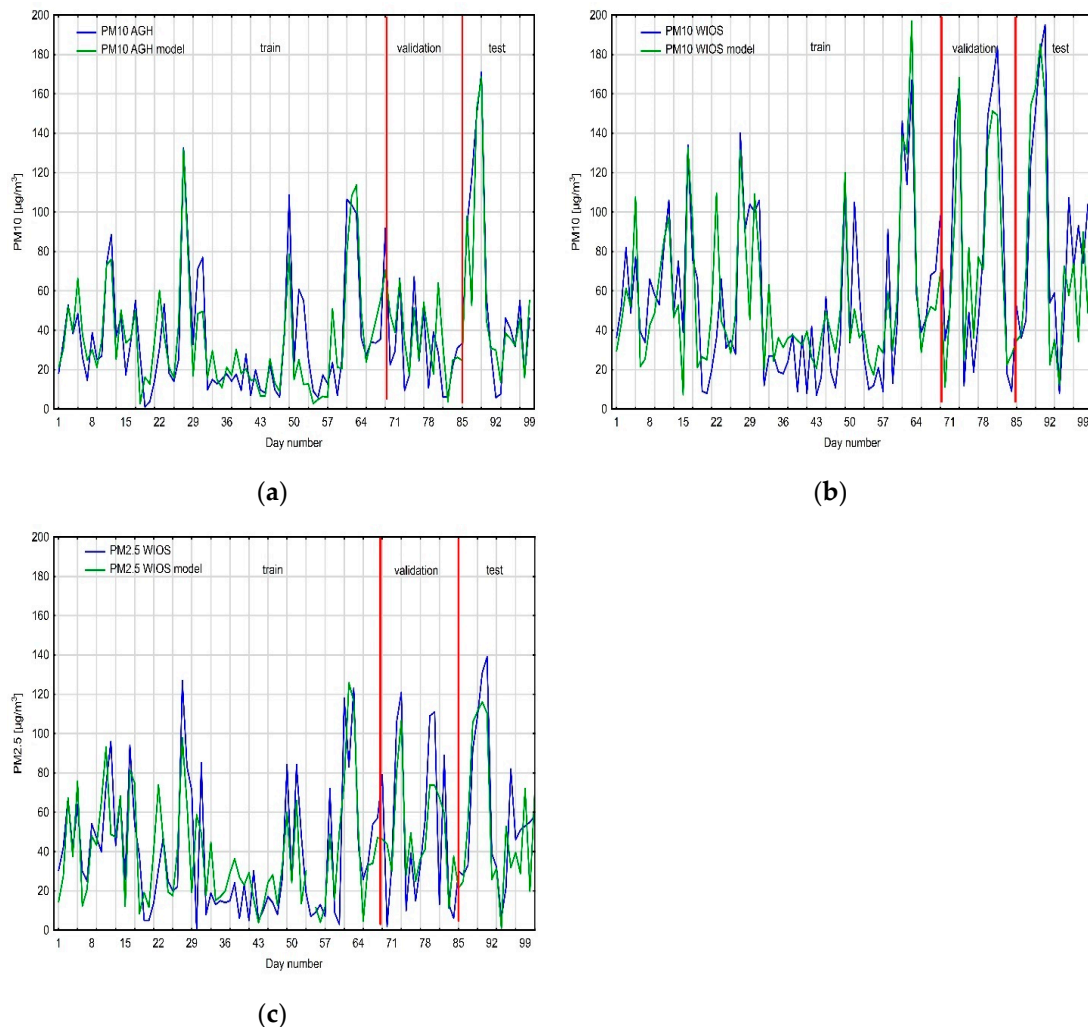
For PM<sub>10</sub> AGH, four more statistically significant correlations were measured with Spearman rank correlation than with the Pearson correlation coefficient: FOF4, COM1, COM3, and COM13. The highest changes in correlation values between the Pearson and Spearman correlation coefficients are visible for Haralick's texture parameters: 0.241 (correlation was measured between COM10 and PM<sub>10</sub> WIOS). When comparing the results of the two correlation methods, higher values for Spearman's correlation coefficient are visible. The number of statistically significant correlations increases especially for PM<sub>10</sub> from AGH station. For all stations, the FOF4 and COM1 correlation values increased and became statistically significant compared to the Pearson correlation coefficient. For PM<sub>10</sub> from AGH station, four more statistically significant correlations were measured with Spearman rank coefficient (FOF4, COM1, COM3, COM13).

Due to the nonlinear relationships between PM data and texture parameters, a multi-layer perceptron was used as a regression model. Texture parameters and weather data were chosen as independent variables, and they showed a statistically significant correlation with PM data. The number of neurons in the input layer equals the number of variables in the model. The variables used in the models are marked in dark blue italics in Table 2. From the 1000 neural networks that were created for each PM dataset, only the one with highest quality was chosen for tests.

The regression neural network prediction for PM<sub>10</sub> values from the AGH measuring station consisted of 17 neurons in the input layer. The best network presented in the results had one hidden layer consisting of seven neurons, and an output layer with one neuron. An exponential function was chosen as the hidden layer activation function, and the neuron activation function in the output layer was linear. Learning, validation and testing qualities are high: 0.9, 0.8, and 0.85, respectively. The neural network algorithm required 57 epochs. The neural network that predicted PM<sub>10</sub> values from WIOS station had two more neurons in the input layer and two fewer neurons in the hidden layer. The activation functions are the same as for the first neural network. Learning quality is similar to the first neural network (0.89); validation quality is higher (0.9); however, testing quality is lower (0.73). This neural network required 34 epochs. The last neural network, which predicted PM<sub>2.5</sub> values, had 14 neurons in the input and hidden layers, and one neuron in the output layer. Its activation functions differ from the previous networks: in the hidden layer, we used the logistic function; in the output layer, we used the hyperbolic tangent; the learning process takes 26 epochs. The learning, validation, and testing qualities are the lowest from all the three neural networks: 0.8, 0.84, and 0.63, respectively. The coefficient of determination ( $R^2$ ) was computed for learning, validation, and test datasets. For PM<sub>10</sub>, AGH equals 0.81, 0.64, and 0.72, respectively. The  $R^2$  for PM<sub>10</sub> WIOS are 0.79, 0.81, and 0.53. Lower  $R^2$  values were computed for PM<sub>2.5</sub> WIOS: 0.64, 0.70, and 0.40.

Moreover, the models' qualities were checked with mean error (ME) in the test dataset and mean absolute percentage error (MAPE). The MAPE values for the test dataset equal 27% (PM<sub>10</sub> AGH), 13% (PM<sub>10</sub> WIOS), and 19% (PM<sub>2.5</sub> WIOS). The MEs for the test datasets

equal 3.7, 5.5, and 3.9, respectively. The high ex-post measure values for the created models relate to differences in the dataset. The test datasets had higher variability than the training datasets for all PM data (Figure 7).



**Figure 7.** Comparison of the obtained models against real data: (a) PM<sub>10</sub> AGH, (b) PM<sub>10</sub> WIOS, and (c) PM<sub>2.5</sub> WIOS.

Air quality assessment of PM<sub>10</sub> and PM<sub>2.5</sub> exceedances was performed based on the MLP network. According to the ordinance of the Minister of the Environment, the PM data were divided into three classes (PM<sub>10</sub>) and two classes (PM<sub>2.5</sub>). The number of classes determined the number of neurons in the created networks' output layers.

The best three neural networks were tested. The architecture differences between these neural networks are significant. The PM<sub>10</sub> AGH neural network has 17 neurons in the input layer and nine neurons in the hidden layer. All statistically significantly correlated variables were used in the model (Table 2). The activating functions of the neurons were hyperbolic tangent and softmax. A more extensive network was obtained for PM<sub>10</sub> WIOS, in which the input layer had one more neuron and the same number of neurons in the hidden layer as PM<sub>10</sub> AGH. This network is additionally distinguished by the functions of the activating neurons in the hidden and output layers: logistics and hyperbolic tangent, respectively. The last PM<sub>2.5</sub> WIOS neural network had 17 neurons in the input layer and 12 neurons in the hidden layer, and the activation functions were hyperbolic tangent and softmax, respectively.

The obtained percentage of correct classifications (accuracy \* 100%) in the learning datasets varies from 87.3% for PM<sub>10</sub> AGH to 94.4% for PM<sub>2.5</sub> WIOS. For PM<sub>10</sub> WIOS, the

percentage of correct classification is 92%. The percentage of misclassifications is evenly distributed over the classes. The percentage of correct classifications in the validation datasets are 92% for PM<sub>10</sub> AGH, 71% for PM<sub>10</sub> WIOS, and 85% for PM<sub>2.5</sub> WIOS. The quality of the neural networks in the test dataset is lower than in the learning dataset, but it is still high, and in the worst case (PM<sub>2.5</sub> WIOS), it achieves 80% of correct classifications. The best results were obtained by the network that predicts PM<sub>10</sub> AGH class. The test set achieved 92.9% correct classifications. PM<sub>10</sub> WIOS obtained 85.7% correct classifications in the test dataset. The Area under Curve (AUC) values computed for whole datasets are 0.792 for PM<sub>10</sub> AGH, 0.802 for PM<sub>10</sub> WIOS, and 0.928 for PM<sub>2.5</sub> WIOS.

The created confusion matrix for the test dataset shows that no error repeats constantly (Table 3). All the neural networks have one wrong classification, where poor air quality was classified as good. Two observations classified as “good” were classified as poor in PM<sub>2.5</sub> WIOS, thus giving a total of only 80% of correct classifications. In addition, for PM<sub>10</sub> WIOS, one additional wrong classification was observed: poor air quality was classified as very poor (Table 3).

**Table 3.** Neural networks’ confusion matrices for the testing datasets.

Class	PM <sub>10</sub> AGH (MLP 17–9–3)			PM <sub>10</sub> WIOS (MLP 18–9–3)			PM <sub>2.5</sub> WIOS (MLP 17–12–2)	
	Good	Poor	Very Poor	Good	Poor	Very Poor	Good	Poor
good	11	0	0	10	0	0	2	2
poor	1	2	0	1	1	1	1	10
very poor	0	0	0	0	0	1	-	-
Quality	92.9%			85.7%			80%	

## 6. Conclusions, Limitations, and Future Research

The issue of air pollution, including the particulate matter, has been extensively presented in the literature [1–3,24]. This is a key issue for most urbanized areas, especially for big agglomerations [4,14,16,23,24,44]. Air pollution in Poland is a big problem, especially in urbanized areas [22]. In the analyzed period in the city of Krakow, exceedances of the permissible standards presented in Table 1 were observed over 100 times a year (Figure 3). Due to frequency with which air quality limits were exceeded, cheap and effective tools are needed to quantify and qualify air quality.

The assessment of PM air pollution is commonly carried out using various types of methods discussed in the literature [5–9]. Carretero-Peña et al. (2019) and Sarimveis et al. (2006) proposed assessing air quality using image analysis. The assessment of air pollution with particulate matter is often based on images and images analysis. Papers presented models based on images analysis with Gabor filter [13], conversion to gray scale, and the Otsu method [18]. Liu et al. also proposed using six image features: transmission, whole image and local image contrast, entropy, sky smoothness, and color. The authors in [25] used Dark Channel Prior for transmission matrix estimation from multiple scene images. In paper [45], the authors used the Gray-level Co-occurrence Matrix method (COM) for texture feature extraction in PM assessment. In this paper, we used COM and two additional complementary textures methods: FOF and GTDM. Additionally, in previous research [20], RBF neural networks were used for the quantitative estimation of air quality, which was assessed with quality measures of ex-post forecast assessment. This paper also proposes the use of artificial neural networks; however, in the work of Ordieres et al., the MLP multilayer perceptron proved to be a better solution [17]. The use of MLP neural networks to predict air pollution on the basis of numerical data was presented by the authors in [16,17]. In paper [16], the authors use numerical variables and a K-mean algorithm for PM<sub>10</sub> prediction with MLP and MLR. The R was between 0.67 and 0.77, depending on measurement localization for the regression model. Better regression results were obtained with CNN models with the PMIE method: R<sup>2</sup> was between 0.68 (multiple-scene images) and 0.91 (single-scene images) for PM<sub>2.5</sub> prediction [15]. The main



limitation in using CNN is a small dataset of images. In paper [13], the authors presented shallow (Random Forest) and deep classifiers (CNN) for five-class air quality assessment for an Air Quality Index (AQI). The authors used a method for multiple-scene images and obtained an AUC of 0.6 and an accuracy of 0.53. A very similar approach to modeling was used in this article by adding texture features as additional variables. Additionally, a similar application of texture analysis in  $PM_{2.5}$  prediction was presented in [18] for images from China and United States. The main difference between [18] and this paper is in the used methods: six texture features versus three texture methods. Quality SVR models for Beijing were between 0.68 and 0.7  $R^2$ , for Shanghai, between 0.72 and 0.76  $R^2$ . The air quality analysis was based on analysis of textures on averaged images with full HD resolution:  $1920 \times 1080$  px. The texture features were measured using three methods: The First-order Features, the Neighborhoods Gray-tone Difference Matrix method, and the Gray-Level Co-occurrence Matrix. For each of the three methods, statistically significant nonlinear correlations were demonstrated between the texture parameters and  $PM_{10}$  AGH,  $PM_{10}$  WIOS, and  $PM_{2.5}$  WIOS data. The correlations are weak to moderate. The highest correlation value is  $-0.518$  between COM2 and  $PM_{10}$  WIOS. Based on the results obtained from the correlation analysis, two types of neural networks were created: regression and classification. Both types of networks had a satisfactory quality. The regression models obtained quality (R) in the 0.63–0.85 range for the test sets. However, the  $R^2$  vary from only 0.4 for  $PM_{10}$  AGH in the test dataset to 0.81 for  $PM_{10}$  AGH in the learning dataset, and  $PM_{10}$  WIOS in the validation dataset. Relatively low  $R^2$  values for  $PM_{10}$  AGH show model weakness, especially in validation and test datasets. What is more, the calculated MAPE value exceeds 10% for all the regression networks, which indicates that they will probably not be good enough to use, despite the relatively low average errors of these networks:  $PM_{10}$  AGH—3.7;  $PM_{10}$  WIOS—5.5;  $PM_{2.5}$  WIOS—3.9. The datasets used in the tests have high variation and higher mean values than in the learning datasets, so the neural network algorithms did not correctly predict data variation. This could be changed by using longer time series in each dataset. Better results were obtained with the classification models than with the regression models. The worst quality was obtained for the  $PM_{2.5}$  WIOS classification in the test dataset: 80% (Table 3). The qualities of other neural networks are equal and higher than 85%. The higher the AUC, the better the performance of the model at distinguishing between classes. The lowest AUC was calculated for  $PM_{10}$  AGH (0.792). The highest AUC equals 0.928 and was calculated for  $PM_{2.5}$  WIOS. High AUC values indicate a good fit of the classification models to the data.

The results of this study indicate that photo texture analysis could be useful in air quality assessment. All chosen texture methods were useful in the performed analysis. Additionally, it is possible to predict air quality exceedances by analyzing textures in HD photos with basic weather data as additional information. It is also possible to predict the values of  $PM_{10}$  and  $PM_{2.5}$  in atmospheric air, but these results have a greater error. This error could probably be minimized by using a longer period of data in the learning dataset that contains all seasons, or higher resolution photos, e.g., 4 K.

The obtained models could be used as part of an application for air quality control using smartphone camera photographs, especially for checking air quality because of the high quality and sensitivity of the  $PM_{10}$  model (higher than 90%). Currently, image sensors are widely accessible, for example, in online cameras and smartphones [21]. Existing infrastructure could be used to provide more photos to create better air quality models. Classification models could be even implemented in mobile phone applications, which will make analysis more accessible for end users. Future work should focus on building a process that will continuously examine images captured at different locations. Such a constant stream of input data combined with weather data will give a much better model that will assess air quality over a larger area. The limitation of this approach is not only the quality of the camera sensor but also the registered image itself. Night-time photos are not providing enough information for the model; the same applies to photos taken in bad weather conditions. Both of these aspects can be eliminated with the use of appropriate

lighting and background, but at the same time, it would make this method much more difficult and more expensive to implement.

**Author Contributions:** Conceptualization, M.C.; Data curation, M.G.; Formal analysis, M.C., W.S. and K.N.; Methodology, M.C.; Writing—original draft, M.C., W.S. and K.N.; Writing—review and editing, M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by AGH University of Science and Technology as part of a scientific project of the Faculty of Geology, Geophysics and Environmental Protection 16.16.140.315/06.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets from The Voivodeship Inspectorate of Environmental Protection in Krakow were analyzed in this study. This data can be found here: ([https://powietrze.gios.gov.pl/pjp/current/station\\_details/chart/400](https://powietrze.gios.gov.pl/pjp/current/station_details/chart/400), accessed on 1 June 2021).

**Acknowledgments:** The authors would like to thank WebCamera.pl for providing camera images, and Environmental Physics Group, AGH University of Science and Technology for providing PM<sub>10</sub> and weather data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Anderson, J.O.; Thundiyil, J.G.; Stolbach, A. Clearing the air: A review of the effects of particulate matter air pollution on human health. *J. Med. Toxicol.* **2012**, *8*, 166–175. [[CrossRef](#)]
- Samet, J.M.; Dominici, F.; Currier, I.; Coursac, I.; Zeger, S.L. Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. *N. Engl. J. Med.* **2000**, *343*, 1742–1749. [[CrossRef](#)]
- Pope, A.C.I.; Dockery, D.W. Health effects of fine particulate air pollution: Lines that connect. *J. Air Waste Manag. Assoc.* **2006**, *56*, 709–742. [[CrossRef](#)]
- Puentes, R.; Marchant, C.; Leiva, V.; Figueroa-Zúñiga, J.I.; Ruggeri, F. Predicting PM<sub>2.5</sub> and PM<sub>10</sub> levels during critical episodes management in Santiago, Chile, with a bivariate birnbaum-saunders log-linear model. *Mathematics* **2021**, *9*, 645. [[CrossRef](#)]
- Hauck, H.; Berner, A.; Gomiscek, B.; Stopper, S.; Puxbaum, H.; Kundi, M.; Preining, O. On the equivalence of gravimetric PM data with TEOM and beta-attenuation measurements. *J. Aerosol Sci.* **2004**, *35*, 1135–1149. [[CrossRef](#)]
- Ruppecht, E.; Meyer, M.; Patashnick, H. The tapered element oscillating microbalance as a tool for measuring ambient particulate concentrations in real time. *J. Aerosol Sci.* **1992**, *23*, 635–638. [[CrossRef](#)]
- Macias, E.S.; Husar, R.B. Atmospheric particulate mass measurement with beta attenuation mass monitor. *Environ. Sci. Technol.* **1976**, *10*, 904–907. [[CrossRef](#)]
- Anderson, T.L.; Ogren, J.A. Determining aerosol radiative properties using the TSI 3563 integrating nephelometer. *Aerosol Sci. Technol.* **1998**, *29*, 57–69. [[CrossRef](#)]
- Smith, J.D.; Atkinson, D.B. A portable pulsed cavity ring-down transmissometer for measurement of the optical extinction of the atmospheric aerosol. *Analyst* **2001**, *126*, 1216–1220. [[CrossRef](#)]
- Shang, P.; Shen, F. Atmospheric correction of satellite GF-1/WFV imagery and quantitative estimation of suspended particulate matter in the Yangtze estuary. *Sensors* **2016**, *16*, 1997. [[CrossRef](#)]
- Muir, D.; Laxen, D.P.H. Black smoke as a surrogate for PM<sub>10</sub> in health studies? *Atmos. Environ.* **1995**, *29*, 959–962. [[CrossRef](#)]
- Air Quality Sensor Performance Evaluation Center (AQ-SPEC). Field Evaluation RTI microPEM PM<sub>2.5</sub> Sensor. Available online: <http://www.aqmd.gov/docs/default-source/aq-spec/field-evaluations/rti-micropem---field-evaluation.pdf?sfvrsn=2> (accessed on 1 June 2021).
- Vahdatpour, M.S.; Sajedi, H.; Ramezani, F. Air pollution forecasting from sky images with shallow and deep classifiers. *Earth Sci. Informatics* **2018**, *11*, 413–422. [[CrossRef](#)]
- Zhang, C.; Yan, J.; Li, C.; Rui, X.; Liu, L.; Bie, R. On estimating air pollution from photos using convolutional neural network. In Proceedings of the 24th ACM International Conference on Multimedia, New York, NY, USA, 15–19 October 2016; pp. 297–301.
- Yang, W.; Feng, J.; Bo, Q.; Yang, Y.; Jiang, B. A shallow resnet with layer enhancement for image-based particle pollution estimation. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision, Guangzhou, China, 23–26 November 2018; pp. 381–391.
- Cortina-Januchs, M.G.; Quintanilla-Dominguez, J.; Vega-Corona, A.; Andina, D. Development of a model for forecasting of PM<sub>10</sub> concentrations in Salamanca, Mexico. *Atmos. Pollut. Res.* **2015**, *6*, 626–634. [[CrossRef](#)]
- Ordieres, J.B.; Vergara, E.P.; Capuz, R.S.; Salazar, R.E. Neural network prediction model for fine particulate matter (PM<sub>2.5</sub>) on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environ. Model. Softw.* **2005**, *20*, 547–559. [[CrossRef](#)]

18. Liu, C.; Tsow, F.; Zou, Y.; Tao, N. Particle pollution estimation based on image analysis. *PLoS ONE* **2016**, *11*, e0145955. [CrossRef] [PubMed]
19. Spyromitros-Xioufis, E.; Moumtzidou, A.; Papadopoulos, S.; Vrochidis, S.; Kompatsiaris, Y.; Georgoulas, A.K.; Alexandri, G.; Kourtidis, K. Towards improved air quality monitoring using publicly available sky images. In *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*; Joly, A., Vrochidis, S., Karatzas, K., Karpinen, A., Bonnet, P., Eds.; Springer: Cham, Switzerland, 2018; pp. 67–92, ISBN 978-3-319-76445-0.
20. Sajjadi, S.A.; Zolfaghari, G.; Adab, H.; Allahabadi, A.; Delsouz, M. Measurement and modeling of particulate matter concentrations: Applying spatial analysis and regression techniques to assess air quality. *MethodsX* **2017**, *4*, 372–390. [CrossRef] [PubMed]
21. Budde, M.; Leiner, S.; Köpke, M.; Riesterer, J.; Riedel, T.; Beigl, M. FeinPhone: Low-cost smartphone camera-based 2D particulate matter sensor. *Sensors* **2019**, *19*, 749. [CrossRef] [PubMed]
22. Carretero-Peña, S.; Calvo Blázquez, L.; Pinilla-Gil, E. Estimation of PM10 levels and sources in air quality networks by digital analysis of smartphone camera images taken from samples deposited on filters. *Sensors* **2019**, *19*, 4791. [CrossRef]
23. Huebnerova, Z.; Michalek, J. Analysis of daily average PM10 predictions by generalized linear models in Brno, Czech republic. *Atmos. Pollut. Res.* **2014**, *5*, 471–476. [CrossRef]
24. Yáñez, M.A.; Baettig, R.; Cornejo, J.; Zamudio, F.; Guajardo, J.; Fica, R. Urban airborne matter in central and southern Chile: Effects of meteorological conditions on fine and coarse particulate matter. *Atmos. Environ.* **2017**, *161*, 221–234. [CrossRef]
25. Li, Y.; Huang, J.; Luo, J. Using user generated online photos to estimate and monitor air pollution in major cities. In Proceedings of the 7th International Conference on Internet Multimedia Computing and Service, Zhangjiajie, China, 19–21 August 2015.
26. The European Parliament and the Council of the European Union. Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 Relating to Arsenic, Cadmium, Mercury, Nickel and Polycyclic Aromatic Hydrocarbons in Ambient Air. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32004L0107&from=PL> (accessed on 1 June 2021).
27. The European Parliament and the Council of the European Union. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on Ambient Air Quality and Cleaner air for Europe. Available online: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32008L0050> (accessed on 1 June 2021).
28. Minister of the Environment. Rozporządzenie Ministra Środowiska z dnia 24 Sierpnia 2012 r. w Sprawie Poziomów Niektórych Substancji w Powietrzu. Available online: <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20120001031> (accessed on 1 June 2021).
29. Minister of the Environment. Rozporządzenie Ministra Środowiska z dnia 8 Października 2019 r. Zmieniające Rozporządzenie w Sprawie Poziomów Niektórych Substancji w Powietrzu. Available online: <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190001931> (accessed on 1 June 2021).
30. The Voivodeship Inspectorate of Environmental Protection in Krakow. Air Quality Monitoring. Available online: <http://krakow.pios.gov.pl/stan-srodowiska/monitoring-powietrza/wyniki-pomiarow/> (accessed on 1 June 2021).
31. Strzelecki, M.; Materka, A. *Tekstura Obrazów Biomedycznych. Metody Analizy Komputerowej*; Wydawnictwo Naukowe PWN: Warsaw, Poland, 2017; ISBN 9788301189778.
32. Bėbas, E.; Borowska, M.; Derlatka, M.; Oczeretko, E.; Hładuński, M.; Szumowski, P.; Mojsak, M. Machine-learning-based classification of the histological subtype of non-small-cell lung cancer using MRI texture analysis. *Biomed. Signal Process. Control* **2021**, *66*, 102446. [CrossRef]
33. Obuchowicz, R.; Piórkowski, A.; Urbanik, A.; Strzelecki, M. Influence of acquisition time on MR image quality estimated with nonparametric measures based on texture features. *Biomed Res. Int.* **2019**, *2019*, 3706581. [CrossRef]
34. Materka, A.; Strzelecki, M. *Texture Analysis Methods—A Review*; Technical University of Lodz: Lodz, Poland, 1998.
35. Amadasun, M.; King, R. Textural features corresponding to textural properties. *IEEE Trans. Syst. Man. Cybern.* **1989**, *19*, 1264–1274. [CrossRef]
36. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man. Cybern.* **1973**, SMC-3, 610–621. [CrossRef]
37. Hand, D.J.; Smyth, P.; Mannila, H. *Principles of Data Mining*; MIT Press: Cambridge, MA, USA, 2001; ISBN 026208290X.
38. Bishop, C.M. *Neural Network for Pattern Recognition*; Oxford University Press: New York, NY, USA, 1995.
39. Sarimveis, H.; Doganis, P.; Alexandridis, A. A classification technique based on radial basis function neural networks. *Adv. Eng. Softw.* **2006**, *37*, 218–221. [CrossRef]
40. Murtagh, F. Multilayer perceptrons for classification and regression. *Neurocomputing* **1991**, *2*, 183–197. [CrossRef]
41. WebCamera.pl. Krakow City Monitoring. Available online: [https://www.krakow.pl/odwiedz\\_krakow/1849,artykul,wawel\\_i\\_zakole\\_wisly.html](https://www.krakow.pl/odwiedz_krakow/1849,artykul,wawel_i_zakole_wisly.html) (accessed on 1 June 2021).
42. Environmental Physics Group. Meteo and Air Pollution Station AGH. Available online: <http://meteo.ftj.agh.edu.pl> (accessed on 1 June 2021).
43. The Inspectorate of Environmental Protection. Air Pollution Monitoring al. Krasinskiego. Available online: [https://powietrze.gios.gov.pl/pjp/current/station\\_details/chart/400](https://powietrze.gios.gov.pl/pjp/current/station_details/chart/400) (accessed on 1 June 2021).

- 
44. Sotomayor-Olmedo, A.; Aceves-Fernández, M.A.; Gorrostieta-Hurtado, E.; Pedraza-Ortega, C.; Ramos-Arreguín, J.M.; Vargas-Soto, J.E. Forecast urban air pollution in Mexico city by using support vector machines: A kernel performance approach. *Int. J. Intell. Sci.* **2013**, *3*, 126–135. [[CrossRef](#)]
  45. Manivannan, K.; Aggarwal, P.; Devabhaktuni, V.; Kumar, A.; Nims, D.; Bhattacharya, P. Particulate matter characterization by gray level co-occurrence matrix based support vector machines. *J. Hazard. Mater.* **2012**, *223*, 94–103. [[CrossRef](#)] [[PubMed](#)]