Research, Society and Development, v. 11, n. 8, e31611830422, 2022
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i8.30422

Forensic palynology: computer vision and geotechnologies to support criminalistics expertise

Palinologia forense: visão computacional e geotecnologias para o apoio da perícia criminal

Palinología forense: visión computacional y geotecnologías para apoyar la experiencia criminal

Received: 05/16/2022 | Reviewed: 06/06/2022 | Accept: 06/10/2022 | Published: 06/20/2022

Ariadne Barbosa Gonçalves
ORCID: https://orcid.org/0000-0003-2496-5723
Universidade Federal de Mato Grosso do Sul, Brazil
E-mail: ariadne.gon@gmail.com
Pedro Lucas França Albuquerque
ORCID: https://orcid.org/0000-0002-8340-4187
Universidade Católica Dom Bosco, Brazil
E-mail: pedrolucas.025@gmail.com
Rodolfo de França Alves
ORCID: https://orcid.org/0000-0003-4081-5725
Instituto Federal de Sergipe, Brazil
E-mail: rodolfoalvesbot@gmail.com
Gilberto Astolfi
ORCID: https://orcid.org/0000-0003-2565-1822
Instituto Federal de Mato Grosso do Sul, Brazil
E-mail: gilbertoastolfi@gmail.com
Felipe Silveira Brito Borges
ORCID: https://orcid.org/0000-0002-8502-3505
University of Debrecen, Hungary
E-mail: eng.fe.silveira@gmail.com
Milena dos Santos Carmona
ORCID: https://orcid.org/0000-0003-3991-438X
Universidade Católica Dom Bosco, Brazil
E-mail: milenascarmona@gmail.com
Marney Pascoli Cereda
ORCID: https://orcid.org/0000-0002-8181-7754
Agro Laboratórios de Pesquisa, Processos e Produtos, Brazil
E-mail: mpcereda@gmail.com
Sérgio Augusto de Miranda Chaves
ORCID: https://orcid.org/0000-0002-5767-0713
Fundação Oswaldo Cruz, Brazil
E-mail: smiranda@ensp.fiocruz.br
Alessandro dos Santos Ferreira
ORCID: https://orcid.org/0000-0002-9588-5889
Universidade Federal de Mato Grosso do Sul, Brazil
E-mail: asf2005kn@hotmail.com
Raquel de Faria Godoi
ORCID: https://orcid.org/0000-0003-1869-1850
Universidade Federal de Mato Grosso do Sul, Brazil
E-mail: rachgodoi@gmail.com
Geazy Vilharva Menezes
ORCID: https://orcid.org/0000-0002-7057-5091
Universidade Federal de Mato Grosso do Sul, Brazil
E-mail: geazyme01@gmail.com
Wedney Rodolpho de Oliveira
ORCID: https://orcid.org/0000-0003-3636-6090
Universidade Católica Dom Bosco, Brazil
E-mail: wedney7@hotmail.com
Antonio Conceição Paranhos Filho
ORCID: https://orcid.org/0000-0002-9838-5337
Universidade Federal de Mato Grosso do Sul, Brazil
E-mail: antonio.paranhos@ufms.br
Arnildo Pott
ORCID: https://orcid.org/0000-0002-8599-9276
Universidade Federal de Mato Grosso do Sul, Brazil
E-mail: arnildo.pott@gmail.com

Karl Jan Reinhard
ORCID: https://orcid.org/0000-0003-1653-9329
University of Nebraska – Lincoln, United States
E-mail: kreinhard1@mac.com
Francisco de Assis Ribeiro dos Santos
ORCID: https://orcid.org/0000-0002-9246-3146
Universidade Estadual de Feira de Santana, Brazil
E-mail: f.a.r.Santos@gmail.com
Hongbo Su
ORCID: https://orcid.org/0000-0002-2147-3646
Florida Atlantic University, United States
E-mail: suh@fau.edu
Hemerson Pistori
ORCID: https://orcid.org/0000-0001-8181-760X
Universidade Católica Dom Bosco, Brazil
E-mail: pistori@ucdb.br

Abstract
Pollen grains can provide valuable information to forensic palynology, such as the time of death or the possible origin of a corpse. Forensic Palynology is a vital tool to be used in a criminal investigation because the different environment has distinct pollen signatures. Brazil has a rich and diversified flora that is suitable for the application of forensic palynology. The purpose of this research is to introduce palynology automation as a tool to improve the investigative method in forensic palynology and apply it to forensic palynology automation. The studied city has different vegetation types, in which we performed assessments to identify its correspondent pollen types. PALINOVIC algorithm was developed using computer vision and geotechnology techniques. Our results show that it is possible to correlate pollen grains found in forensic samples by automatic pollen identification and with a mapping of the likely vegetation. Our results show that it is possible to relate the presence of pollen grains found in forensic samples through the automatic identification of images together with a database of georeferenced plant species. It was possible to analyze the pollen grains collected in eight bodies, where the algorithm presented a performance of 90.51% in the pollen grain classification tests. Furthermore, pollen grains could be correlated with the type of vegetation where the body was found. Thus, the technique developed can be applied in other urban centers from a previous georeferencing of plants, as well as a pollen database.
Keywords: Machine learning; Geoprocessing; Homicide.

Resumo
Os grãos de pólen podem fornecer informações valiosas para a palinologia forense, como a hora da morte ou a possível origem de um cadáver. A Palinologia Forense é uma ferramenta vital a ser utilizada em uma investigação criminal, pois os ambientes possuem digital polínica distintas. A flora rica e diversificada do Brasil, é adequada para a aplicação dessa técnica. O objetivo desta pesquisa é apresentar a automação da palinologia como ferramenta para melhorar o método investigativo em palinologia forense. A cidade foi selecionada por apresentar diversidade de tipos de vegetação no ambiente urbano, que foi amostrada para identificar os tipos polínicos que ocorrem. O algoritmo PALINOVIC foi desenvolvido com técnicas de visão computacional e geotecnologias. Nossos resultados mostram que é possível relacionar a presença de grãos de pólen encontrados em amostras forenses por meio da identificação automática das imagens em conjunto com um banco de dados de espécies vegetais georrefenciadas. Foi possível analisar os grãos de pólen coletados em oito corpos, onde o algoritmo apresentou desempenho de 90.51% nos testes de classificação de grãos de pólen. Além do mais, os grãos de pólen puderam ser correlacionados com o tipo de vegetação onde o corpo foi encontrado. Assim, a técnica desenvolvida pode ser aplicada em outros centros urbanos a partir de um georreferênciamento prévio de plantas, bem como um banco de dados polínicos.
Palavras-chave: Aprendizado de máquina; Geoprocessamento; Homicídio.

Resumen
Los granos de polen pueden aportar información valiosa para la palinología forense, como mejorar la cínica hora de la muerte o señalar el circuito del vivo y su cadáver. En este aspecto, la Palinología Forense es una herramienta vital para ser utilizada en una investigación criminal, ya que los entornos tienen diferentes digitales de polen. La rica y diversa flora de Brasil es apta para la aplicación de esta técnica. El objetivo de esta investigación es mostrar cómo la automatización de la palinología como herramienta para mejorar el método investigativo en palinología forense. La ciudad fue seleccionada por presentar diversidad de tipos de vegetación en el medio urbano, la cual fue muestreada para identificar los tipos de polen que se presentan. El algoritmo PALINOVIC se desarrolló utilizando técnicas de visión artificial y geotecnologías. Nuestros resultados muestran que es posible relacionar la presencia de granos de polen encontrados en muestras forenses a través de la identificación automática de imágenes junto con una base de datos de especies vegetales georreferenciadas. Se logró establecer de manera rápida y confiable los granos de polen colectados en ocho cuerpos, donde el algoritmo presentó un desempeño de 90.51% en las pruebas de clasificación de granos de polen. Además, los granos de polen podrían correlacionarse con el tipo de vegetación donde se encontró el

cuerpo. Así, la técnica desarrollada puede ser aplicada en otros núcleos urbanos a partir de una georreferenciación previa de plantas, así como de una base de datos de polen.
Palabras clave: Aprendizaje automático; Geoprocesamiento; Asesinato.

1. Introduction

The importance of using pollen traces in criminal investigations is related to the fact that pollen grains are not a visually perceivable trait. In contrast to fingerprints, which can be wiped off or avoided using gloves, pollen grains are microscopic and difficult to eliminate from the crime scene. Pollen grains can also be trapped on clothes, shoes, tools, and other materials of criminal importance.

Even a mud sample from a shoe may contain information about where a person had been. Pollen morphology has structures that facilitate their adhesion to clothing (Zavada et al., 2007), objects (Boi, 2015; Kumari, 2017) and the human body (Wiltshire, 2006), such that almost all samples found at a crime scene may contain relevant pollen information to help to solve criminal cases. Various researches have shown the use of palynology to verify documentary fraud (Morgan, 2013) and the origin of illicit drugs (Milne et al., 2004), among other forensic applications.

Pollen is inherent to flowers and gymnosperm strobiles. Pollination can occur by animals, i.e., zoophilia, or wind, anemophilia, with higher numbers of pollen grains (Bryant and Holloway, 1983). Pollen grains are microscopic reproductive structures with an average size between 25-60 μm (Mildenhall, 2006), which can easily adhere to objects and surfaces. They are excellent botanical and geographical markers.

Botanical surveys enable the assessment of the type of vegetation likely to occur in each locality of forensic interest. This data can be used to establish the origin of the pollen grains found in a specific location. The plant coordinate is taken using GNSS receivers (Global Navigation Satellite System), aerial images, drones or small size airplanes. Besides, surveys can be done by satellite images with an excellent spectral resolution for the individualization of vegetation.

Forensic palynology has been used to solve crimes in the United States, New Zealand, and the United Kingdom (Bryant, 2014; Mildenhall, 1990; Wiltshire, 2015). In this context, the use of technologies such as machine learning to develop algorithms for pollen grains and vegetation recognition are essential tools in the development of computer programs for forensic palynology automation.

We review the literature that discuss pollen grain identification based on computer vision techniques. These techniques can be divided into shallow and deep learning methods. Concerning shallow learning methods, da Silva et al. (2014) used KNN (K-Nearest Neighbors), C4.5, and Support Vector Machines (SVM) to classify seven pollen types collected in the Brazilian Savannah. The proposed approach consists of three phases: segmentation, texture feature extraction, and classification. Both segmentation and feature extraction are used techniques based on the watershed. In classification, the best result was achieved by the SVM classifier, 79% of F-Score rate.

Travieso et al. (2011) achieved an accuracy of 93.8% using an approach based on the Hidden Markov Model (HMM) and SVM classifiers for classifying 47 pollen types of 22 different species from Costa Rica, Central America. García et al. (2012) also used an HMM classifier to classify 17 types of pollen grains from 11 different species from Costa Rica. The approach, employed binarization and contour identification to obtain features from the pollen grains, reaching a 98% accuracy.

Using the same set of images as García et al. (2012), Ticay-Rivas et al. (2011) achieved an accuracy of 96.49% by adopting a multilayer neural network as a classifier trained with features of the pollen grains obtained by mean of geometrical measures, Fourier descriptors of morphological details, and color information.

Some studies adopted state-of-the-art deep learning classification methods for classifying pollen grains. Sevillano and Aznarte (2018) reported an experiment with three deep learning models, whose best approach achieved 97% of correct classification on a dataset with 23 pollen types. The method proposed by Daood et al. (2016) based on deep learning achieved 94% accuracy using a dataset with 30 pollen types collected in the United States. Combining Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN), Daood et al. (2018) reported an accuracy of 100% on a dataset with ten pollen types.

Sevillano et al. (2020) achieved an accuracy higher than 97% on a dataset with 46 pollen types using Alexnet architecture (Holt and Bebbington, 2014). Astolfi et al. (2020) conducted an extensive experimental evaluation using 73 pollen types collected in the Brazilian Savannah to compare eight CNN architectures. The experiments showed that DenseNet-201 (Huang et al., 2017) and ResNet-50 (He et al., 2016) are superior performance, achieving precisions of 95.7% and 94.0%, respectively.

Regardless of the learning technique used, shallow or deep learning, we can note that the use of computer vision in palynology presents encouraging results. This endorses forensic palynology automation to improve and facilitate the investigative method since palynology can be employed as a powerful tool in solving criminal investigations (Alotaibi et al., 2020). In this context, technological applications such as machine learning to develop algorithms for pollen grains and vegetation recognition are essential tools to be applied to forensic palynology automation. Our study aims to develop an algorithm to automate pollen grain identification to improve forensic palynology studies.

2. Methodology

This study was authorized by the Committee of Ethics and Research of the Universidade Católica Dom Bosco (UCDB) for a partnership between UCDB and the Secretaria de Segurança Pública of Mato Grosso do Sul, together with the Geoprocessing Laboratory (LABGIS) of the Universidade Federal de Mato Grosso do Sul.

2.1 Study scope

We present the Palinovic software, not yet widely used. This software could be used by the enforcement agencies, which utilizes scientific methods, as forensic palynology to aid in solving crimes. Palinovic was developed to identify pollen grains and the vegetation of an urban center derived from satellite images geoprocessing.

Palinovic can also be useful in other areas of palynology such as public health application to allergenic pollen. It is necessary to have a survey of the pollen grains and plants present in the urban environment of study to apply the software. Thus, flower sampling is essential for an image dataset with information of pollen grain features.

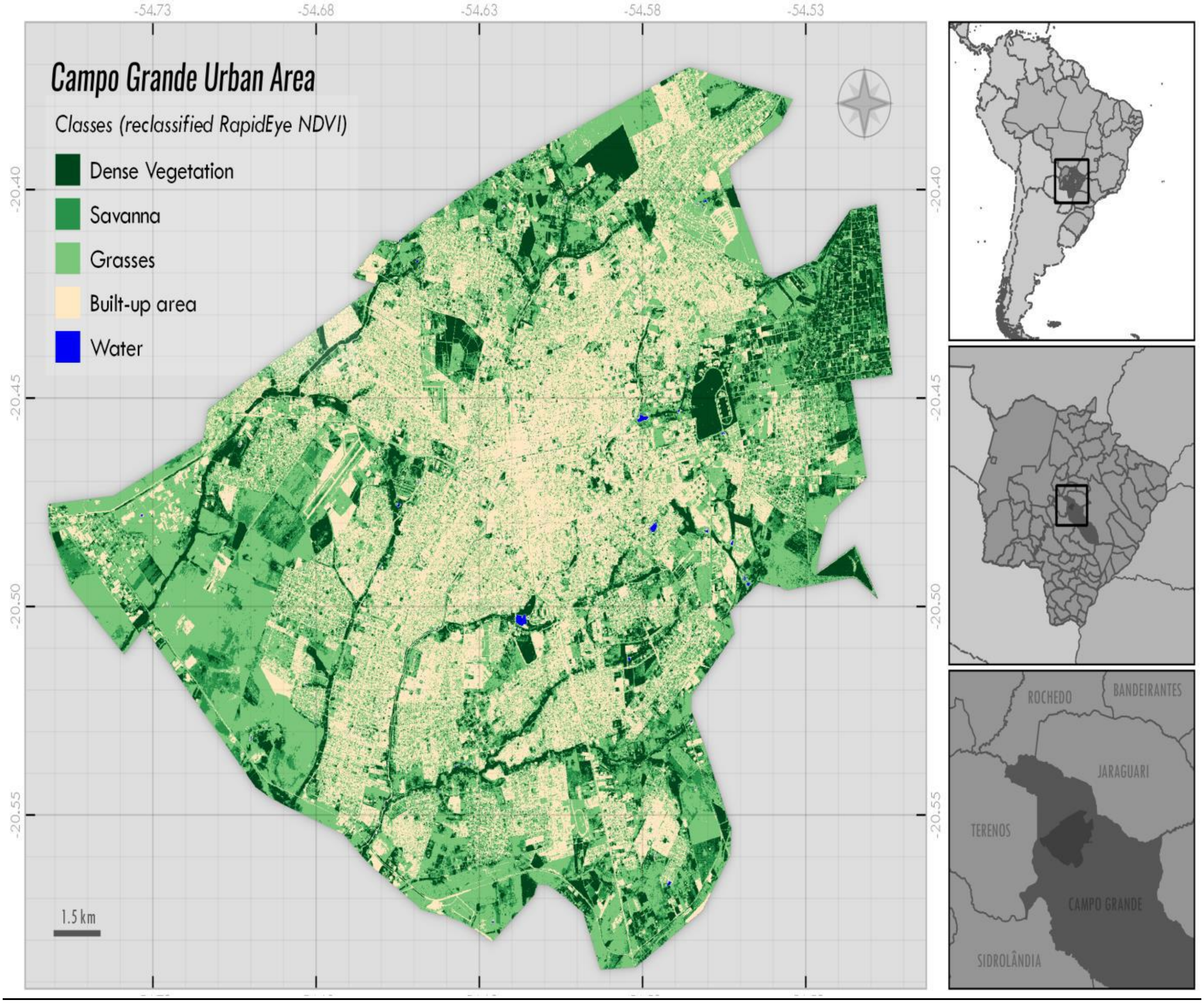
Given data on pollen samples, the system returns pollen samples are entered into the system, and it returns as information about the identification of the pollen grains found in the case and correlates them with the surrounding vegetation. To use the information of geographic coordinates and pollen of plants in a locality, the automation of both processes is performed using machine learning techniques and computer vision.

After the development of PALINOVIC, it is necessary to carry out experiments to verify its efficiency. In this context, we present the PALINOVIC operating modules and their application in eight concrete examples described in the following sections.

2.2 Description of Urban Vegetation Survey

We applied a part of the research in the urban area of Campo Grande, with nearly 900,000 inhabitants, capital of Mato Grosso do Sul State, Brazil. We analyzed RapidEye satellite images from the urban area (MMA, 2013) (Figure 1). We classified the terrain of the city in five classes: 1) water, 2) anthropized area, 3) dense vegetation, 4) herbaceous vegetation, and 5) open vegetation (Gonçalves et al., 2018).

Figure 1. Classification map containing the five classes found for the Campo Grande urban perimeter generated with RapidEye images and NDVI (Gonçalves et al. 2018). The green tones are dense, herbaceous or dense vegetation, the beige tone is anthropized areas and the blue color is the area with water resources.



Source: Authors.

2.3 Pollen grain images

We captured the pollen grain images from flowers collected in the urban region of Campo Grande, to prepare palynological slides. We captured the images under a light CARL ZEISS Axio Scope A1 microscope. Each of the 73 plants was photographed 35 times, resulting in 2555 pollen images (Astolfi, 2020).

The pollen types were organized in 35 families: Arecaceae, Anacardiaceae, Asteraceae, Bignoniaceae, Bombacaceae, Boraginaceae, Burseraceae, Cannabaceae, Caryocaceae, Combretaceae, Commelinaceae, Dilleniaceae, Euphorbiaceae, Fabaceae, Lamiaceae, Liliaceae, Magnoliaceae, Malvaceae, Melastomataceae, Myrtaceae, Ochnaceae, Oleaceae, Passifloraceae, Piperacea, Plantaginaceae, Poaceae, Rubiaceae, Sapindaceae, Solanaceae, Tiliaceae, Urticaceae, Vochysiaceae e Vitaceae.

We previously classified the species used to compose the pollen grain image according to the vegetation type occurrence dense, herbaceous or open. We established the dataset with the occurrence percentage of each family in the pollen, based on each species location and habit. The Brazilian savanna has diverse vegetation, characterized by tortuous trees, e.g., the Caryocaceae pollen is related to Cerradão woodland (Pott and Pott, 1994), therefore classified as dense vegetation. The classification of dense, herbaceous and open vegetation used in this work follows Gonçalves et al. (2018).

2.4 Machine learning techniques

The Palinovic modules consist of pollen grain identification and vegetation estimation of Campo Grande. For the first module, pollen identification, we applied the technique of deep convolutional neural networks on pollen images. Neural networks are mathematical models inspired by the reproduction of brain activity to solve a problem that, in this case, is linked to the recognition of pollen grains. Training, testing and validation were carried out with the pollen grain image bank containing 35 botanical families.

The architectures used for deep convolutional neural networks were ResNet-50 and Xception with 50 epochs and transfer of learning with weights from ImageNet, where 25% of the last layers of this network were retrained. We divided the dataset into 60% for the training set, 20% for the validation set, and 20% for the test. The network hyperparameters were input size of 256x256x3, the batch size of 16, stochastic descending gradient optimizer with a learning rate of 0.0001 and with a momentum of 0.9.

In the module of the vegetation estimation, the georeferenced image of Campo Grande was used with the five classes. The total amount of pixels within a radius estimates each vegetation type, whose center is a geographical coordinate. In this way, the proportion of each class, anthropized area, water, dense, herbaceous, or open vegetation within the radius is determined.

To generate a decision-making model with the data obtained from the Campo Grande vegetation mapping and the pollen grains found in each case, a distribution probability model was applied using the Kullback-Leibler divergence algorithm (Shalizi, 2006). The algorithm considers the size of the pixels in m2 that surround a known geographic coordinate and associates the vegetation to its surroundings with the pollen grains found in a sample.

The probability KL-divergence was performed to identify the relation between the distance from the vegetation and the probability of finding pollen grains from that location. Therefore, the map used to measure this distance is considered by the size of the pixels. The classification map encompassing the five classes of the Campo Grande urban perimeter contains a total of 14,306,457 pixels (100%), each corresponding to 1m2, 12,025 (0.1%) of these refer to water, 5,019,031 (35.1%) are anthropized areas, 5,926,273 (41.4%) are herbaceous, 1,583,166 (11.07%) are open and 1,765,962 (12.34%) are dense vegetation.

This probabilistic analysis verifies distances between vegetation to surroundings coordinates, with the proportion of the pollen in the cases. This analysis requires an assumption that any distribution of pollen grains is independent of any other factors beyond the distance to the position of the case. The KL-divergence (Shalizi, 2006) test correlates linearly to distances of each type of vegetation and the probability of finding a pollen grain of the vegetation in the case is demonstrated by Equation 1. The hypothesis analyzed is that the distance of the vegetation from the case is inversely proportional to the type of pollen grains' vegetation found in the sample, which is represented by:

$$D_{kl} = (P|Q) = -\sum_i P\,(i)\,log\,\frac{Q(i)}{P(i)}$$

Equation 1

Where, “P” with respect to “Q” is equals the sum of the indices i of the probability of the distribution “P” of index i, times the logarithm of the distribution “Q” in i. Where i is equivalent to each dense, herbaceous and open vegetation. “P” is the actual classification of the vegetation, ie, vegetation around the crime scene and “Q” is the presumed vegetation, that is, the vegetation of pollen grains origin found in the body (Equation 1).

Observe the following algorithm, indicating how this model is calculated, where "P" is the distribution of pollen, "D" is the vector of distances closest to each vegetation, "d" receives an element of vector "D". "S" is probability distributions of the models and "s" receives an element of the vector 'S'.

Algorithm (Maximum distance d_{max})

```
For each pollen in P
    Maximum distance d_max
    For each distance in D
        d = d_max – d
For each value in S
        s = s/max(S)
Returns S
```

2.5 Experiment: validation in forensic cases

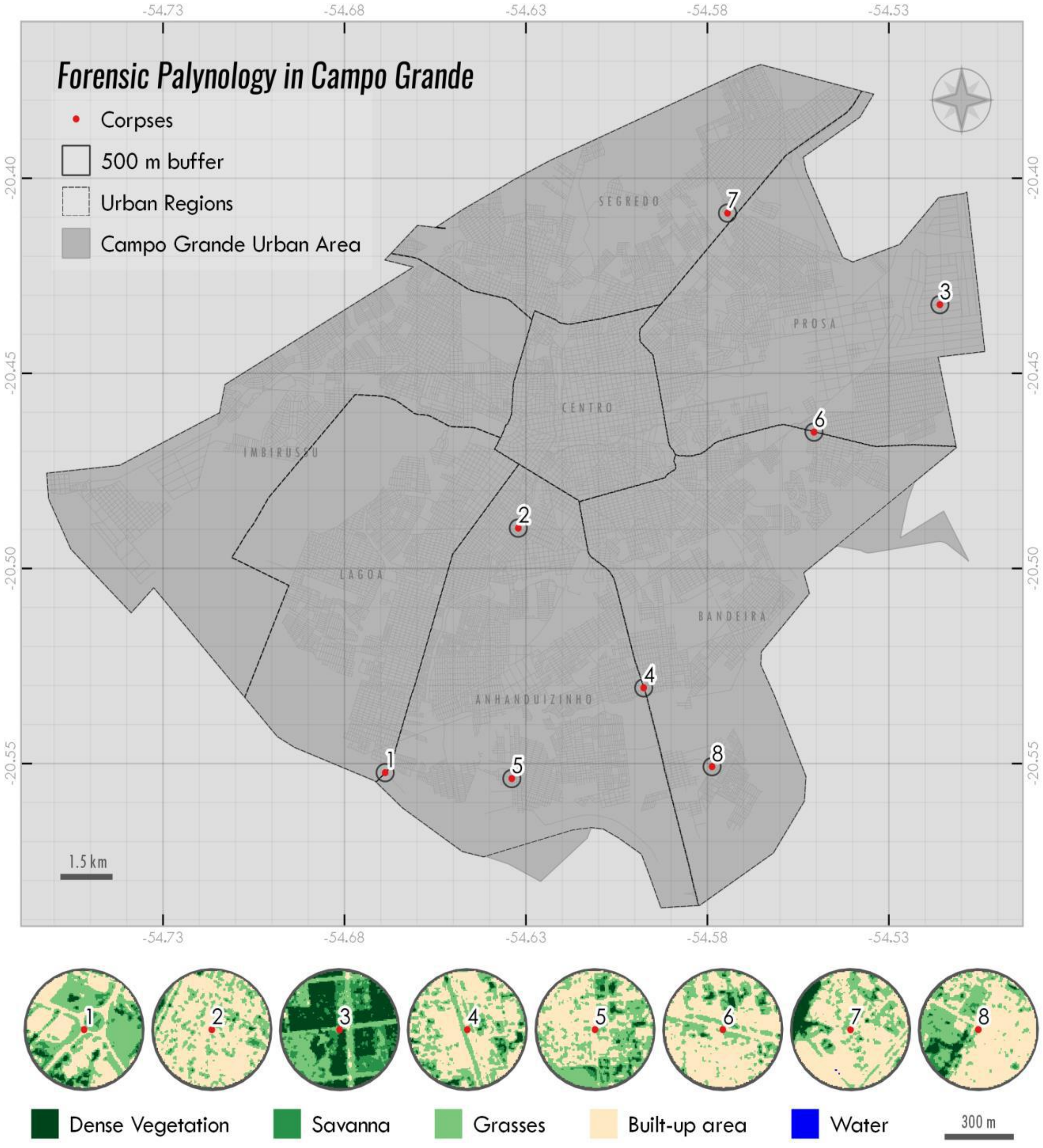
We took pollen grain samples in eight cases, utilizing the swab method, using cotton with a few drops of glycerin to collect the pollen grains from eight corpses at the morgue (Instituto de Medicina e Odontologia Legal) in Campo Grande. The sampled corpses had several causes of violent death: one traffic accident, two suicides and five homicides. The pollen grains were photographed and identified up to the family level, classified according to each vegetation and then we used deep convolutional neural networks for their automatic identification.

To estimate the probability of each pollen found in the cases occurring in the vegetation class found in the map of Campo Grande (Gonçalves et al., 2018), we evaluated the number of pixels per vegetation class relating the occurrence of each pollen family per vegetation class. It is possible to establish the probability of occurrence for each set of pollen grains using the distance of the cases from the vegetation class and the type of origin of vegetation class in each case.

We used a classified NDVI image from the urban area of Campo Grande (Gonçalves et al., 2018) as the entrance to the urban study center for the vegetation class tracing. By doing this, we could find coordinates of corpses, thus, able to establish the radius to surroundings of the case, which is defined by the user. This process is necessary to delimit the proportions of soil cover classes: water, paved area, and Open, Dense and Herbaceous vegetation to the site surroundings.

We used the 50 m radius to determine the type of dominant vegetation class of each case. The radius of 500 m (Figure 2) was used to determine the soil cover class patterns found in the site surroundings to explain possible influences of the environment on the pollen load. Also, for each case, we calculated the proportion of each vegetation class of the place and the correlation between vegetation and pollen grains found.

Figure 2. Map showing the area of 500 m radium used around the location of the bodies, in Campo Grande-MS, to estimate the proportion of each vegetation feature around the corpses. Shapefiles provided by the Ministry of Environment (MMA, 2013).



Source: Authors.

3. Results

We analyzed eight homicide cases in Campo Grande. The city has different types of native microhabitats that allow the application of forensic palynology. However, unlike the spatial distribution reported by Zavada et al. (2007), Campo Grande has ecosystems with high richness and diversity of arboreal to herbaceous plants, in many distinct vegetation classes, including cerrado *lato sensu*, gallery forest, paths, riparian forests, pasture, exotic ornamental plants, and sharing mosaics from rural to urban areas.

In case 1, the death occurred in a confrontation with the police. The body was without cadaveric rigidity, with post-death interval estimated under 2 hours. The place was characterized as Herbaceous vegetation in the expert report. We found pollen grains of Cannabaceae (33%), Fabaceae (33%) and Poaceae (33%). Within a radius of 50 m, 70% of the area had a predominance of Herbaceous vegetation. In a radius of 500 m, the predominance was 46% of Herbaceous vegetation.

In case 2, the reported death was presumably the execution of a passer-by. The death time was estimated less than 2 hours, as the body was without cadaveric rigidity. The place is of Herbaceous vegetation. Within a radius of 50 m, the only vegetation was Herbaceous 9% and the rest is a paved area. However, for a 500 m radius surroundings, the proportion of Herbaceous vegetation has an increased to 26% and classes Open 2% and Dense 1%. The pollen proportions we found: Arecaceae (20%), Asteraceae (20%), Liliaceae (20%) and Poaceae (40%). The pollen analysis indicates that it is improbable that the body comes from an area with Herbaceous vegetation, due to the predominance of pollen grains from Open and Dense vegetation in the sample since we detected pollen grains from families that occur in all vegetation classes.

Case 3 is of a missing person whose body was in advanced putrefaction, between 12 and 15 days. Presumably, it remained in the same place, as informed by the expert report. The soil cover data showed the highest proportion of Herbaceous vegetation (49%) in a 50 m radius, the proportions of Dense and Open vegetation were 29% and 23%, respectively, and no paved area. In 500 m the predominance is of Dense vegetation (41%), showing significantly most Open and Herbaceous vegetation within both radio. We found pollen grains from Open, Herbaceous and Dense vegetation, of Poaceae (75%) and Urticaceae (25%). The probable influence of the surroundings interfering with the sampled pollen grains is 26% of Open vegetation and 37% of Herbaceous vegetation (Table 1). It was possible to confirm the validity of this data by knowing that the location has dense vegetation.

Case 4 is a homicide, and the body was found in an advanced state of decomposition, in a place with grasses, as presented in the expert report. It occurred in an area where 49% was Herbaceous vegetation within a radius of 50 m. In a radius of 500 m, the proportion of Herbaceous vegetation was 34%, while other classes had 6%. We found pollen of Arecaceae (13%), Myrtaceae (13%), Malpighiaceae (61%) and Poaceae (13%), which may occur in Open, Dense and Herbaceous types.

Case 5 is a firearm homicide, the estimated death was less than 2 hours, and the place was characterized by Open vegetation in the expert report. The site had only Herbaceous vegetation within a radius of 50 m. Within 500 m the predominant vegetation class was the Herbaceous 42% and 7% Open. The highest proportion of pollen was 87% Malpighiaceae and 13% Poaceae. Both pollen types occur on public roads or on vacant land, but there is a chance that 100% of Malpighiaceae pollen belongs to Dense vegetation.

Case 6 is a traffic death that occurred approximately 1 hour before, in a location of Herbaceous vegetation, as stated in the expert report. The vegetation analysis by satellite image showed 38% of Herbaceous vegetation, the only class within 50 m, and the rest was pavement. At 500 m, the predominance was 27% Herbaceous vegetation and 2% Dense. Being a traffic crime, we did not find pollen grains matching the body location. We detected pollen grains of Malpighiaceae (66%) and Malvaceae (34%) on the body, consistent with the vegetation class of Open areas, not Herbaceous features where the death occurred.

Case 7 is suicide; the body was in decomposition conditions, with a post-mortem interval estimated between 2 to 6 days, on a paved site, accord the expert report. The Herbaceous vegetation class covered 27% within a 50 m radius, the same as within the 500 m radius; however, Open (3%) and Dense (7%) vegetation features appeared. We found pollen grains from Anacardiaceae (30%), Caryocaraceae (8%), Fabaceae (38%), Rubiaceae (8%), Sapindaceae (8%) and Urticaceae (8%). Of all probabilities presented in, the largest was this case with an 81% probability of pollen grains coming from Dense vegetation, corroborated by such a nearby area. To this location, yet within 500 m there were signs of water.

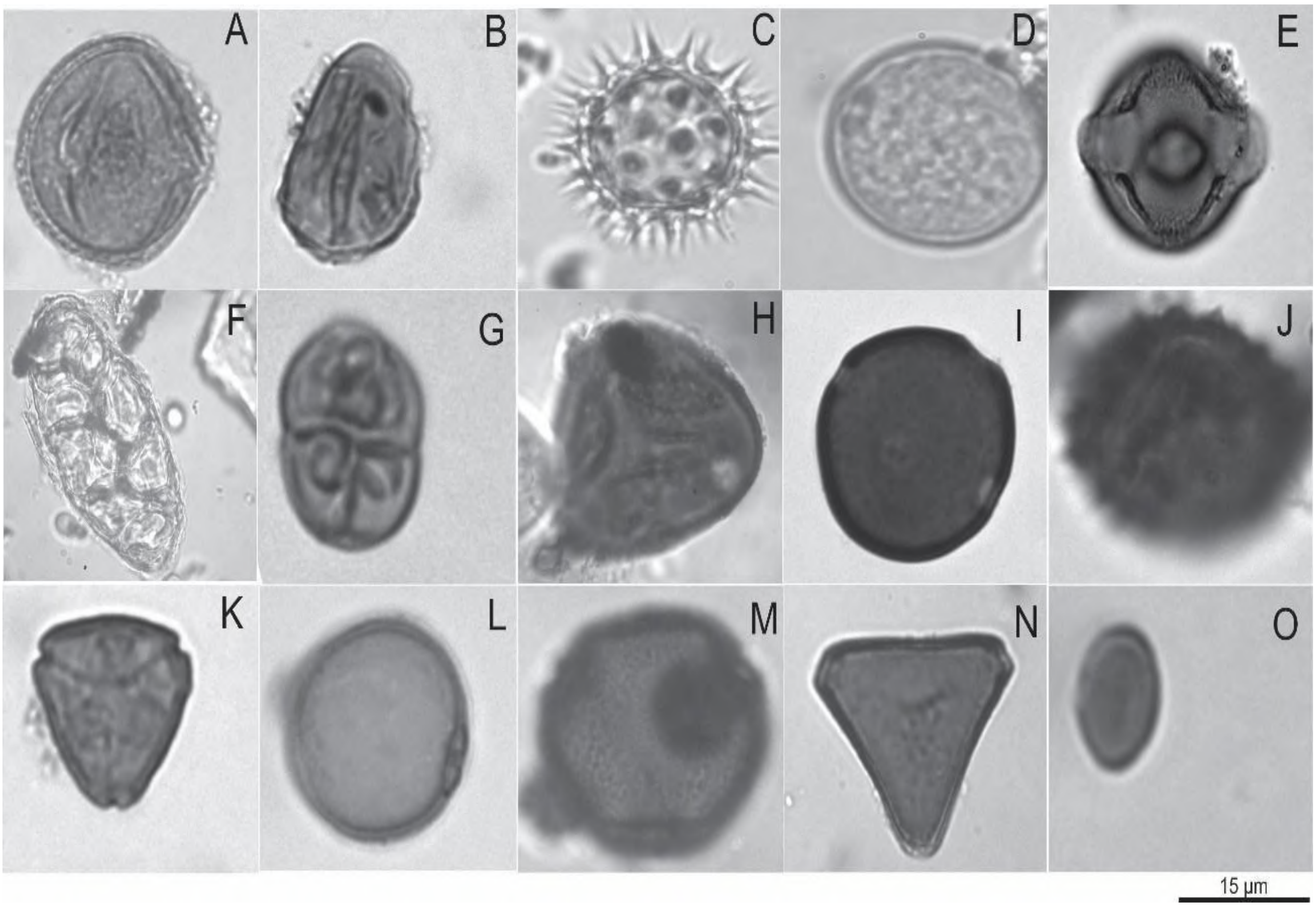
Case 8, as described in the expert report, is also suicide, the body had cadaveric stiffness, estimated death between 4

Research, Society and Development, v. 11, n. 8, e31611830422, 2022
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i8.30422

to 8 hours, on a paved area. The only vegetation class at 50 m was 44% Herbaceous, at 500 m occurred 30% Herbaceous, 4% Dense and 6% Open. We found pollen of Anacardiaceae (3%), Arecaceae (1%), Fabaceae (2%), Malvaceae (2%), Myrtaceae (88%) and Poaceae (4%).

In eight cases of violent deaths, we collected pollen grains from the clothes, shoes and skin of the dead body. Altogether, 134 pollen grains were recovered from the bodies, distributed into 14 pollen types and classified in families: Arecaceae (3), Asteraceae (1), Anacardiaceae (5), Cannabaceae (1), Caryocaraceae (1), Fabaceae (4), Liliaceae (1), Malpighiaceae (14), Malvaceae (1), Myrtaceae (74), Poaceae (14), Rubiaceae (1), Sapindaceae (1) and Urticaceae (2) (Figure 3). Only 10 remained unidentified.

Figure 3. Pollen grains found on corpses and their respective botanical families. A) Anacardiaceae; B) Arecaceae; C) Asteraceae; D) Cannabaceae; E) Caryocaraceae; F-G) Fabaceae; H) Liliaceae; I) Malpighiaceae; J) Malvaceae; K) Myrtaceae; L) Poaceae; M) Rubiaceae; N) Sapindaceae; O) Urticaceae.



Source: Authors.

3.1 Machine learning test

In the experiment with deep convolutional neural networks with the learning bank of 35 families of pollen grains and validation with the pollen dataset found in the cases, which has 14 botanical families, the results with the two ResNet50 and Xception architectures are presented in Table 1. The ResNet50 architecture had the best result for training, validation and testing.

Table 1: Accuracy performance using ResNet50 and Xception architectures in pollen grain identification.

Architecture	Training	Validation	Test
ResNet50	96.55	91.18	90.51
Xception	91.11	55.47	54.38

Source: Authors.

With identification data of pollen grains and information of their likely locations, the software calculates the distance of coordinates of the case to vegetation class and the probability of pollen grains observed in a case linked to the case. Probability is assessed by the distribution of each vegetation class to be interconnected with the pollen grains observed in the cases. The proportion of each type of pollen found in the cases expresses the surround vegetation influence (Table 2).

Table 2: Pollen grains proportion found in each criminal case. *Families with most representativity in each case.

Type of death	Pollen Grains	Pollen ratio
Case 1 (homicide)	Fabaceae	33%
	Cannabaceae	33%
	Poaceae	33%
Case 2 (homicide)	Arecaceae	20%
	Asteraceae	20%
	Liliaceae	20%
	Poaceae	*40%
Case 3 (homicide)	Poaceae	*75%
	Urticaceae	25%
Case 4 (homicide)	Arecaceae	14%
	Malpighiaceae	*72%
	Poaceae	14%
Case 5 (homicide)	Malpighiaceae	*87%
	Poaceae	13%
Case 6 (Traffic accident)	Malpighiaceae	*66%
	Malvaceae	34%
Case 7 (suicide)	Anacardiaceae	10%
	Caryocaraceae	10%
	Fabaceae	50%
	Rubiaceae	10%
	Sapindaceae	10%
	Urticaceae	10%
Case 8 (suicide)	Anacardiaceae	5%
	Arecaceae	1%
	Fabaceae	2%
	Malvaceae	2%
	Myrtaceae	*84%
	Poaceae	6%

Source: Authors.

Consequently, the pollen grains of families most alike to the cases are supposed to be observed to a greater extent if the corpse remained in place for a long time. Hence, from the family occurrence data per vegetation class together with the proportion of pollen grains per families found in the cases, it is possible to estimate the predominant pollen vegetation class in the body.

We determined the total of pixels found for the distribution of each vegetation class for the area of both rays of 50 m, which is close to the corpse, and 500 m, its surroundings. Therefore, the expected probability for each family was calculated from the classification of the three vegetation classes, Open, Dense and Herbaceous, within each area, equaling 100% of the distribution per family from the pollen identification observed in each case.

Considering the proportions of vegetation classes found per family of the image bank, the pollen grains observed in the cases were linked to the vegetation type. It was plausible to compare the type of vegetation class of origin of each case with the pollen types found in the cases. Table 2 exposes the proportion of families found per case from the pollen identification.

4. Discussion

The efficiency of deep neural networks technique concerning traditional methods such as wavelet transform (Silva et al., 2014) and visual word histogram (Rodrigues et al., 2015) is also confirmed by Daood et al. (2016). Thus, for studies aiming to automate the identification of pollen grains, it is suggested the application of deep convolutional neural networks.

We found this efficient for the recognition of pollen types, both in species and families, presented in this research relating pollen grains detected in cases of violent death with the three types of vegetation class for the urban region. It is the beginning of the use of environmental information with application in public safety. Zavada et al. (2007) were able to determine the distribution of pollen grains in urban, suburban, and rural areas of Rhode Island and the United States. They found exotic pollen grains in urban and native pollen in rural areas.

Ochando et al. (2018) applied a forensic pollen study in Cartagena where they sampled at four surfaces cloth, fabrics, footwear and soil. The authors collected 17 samples from three sites, they could recover more than 14 pollen types. The application of more refined geoprocessing and remote sensing techniques can distinguish more vegetation classes in urban regions. With only three vegetation classes in our work, it was already possible to filter the areas with different plant formations and five classes of soil cover.

Among the occurring families, only Poaceae is wind pollinated and was present in six of the eight analyzed cases. In Campo Grande, wasteland weeds are common, such as *Megathyrsus maximus* (Jacq.) B.K. Simon, Paspalum notatum Flüggé and Urochloa spp. The wind quickly disperses their pollen, thereby Poaceae are not reliable location indicators in forensics.
Plant pollen grains that are carried by insects, entomophilous, are transported or tend to remain around the parent plant. Bryant (2014) reports that such plants produce little pollen, approximately 1,000 grains per anther, unlike wind-pollinated, anemophilous plants, which produce 10,000 to 70,000 pollen per anther. In this context, the urban vegetation of Campo Grande is composed mostly of zoophilous plants, which can be critical geographic markers of pollen sources in the city.

In cases 1 and 2, considering the type of deaths and the bodies collected in a few hours, it was possible to analyze that the detected pollen grains did not match the local type of vegetation class. The vegetation class of the pollen group found on the bodies was different from the sites. In both cases, it is evident in the expert reports and combined with the pollen analyses that the bodies came from elsewhere.

Case 3 had a heavy presence of grass around the corpse, so perhaps it received a great deal of pollen load from the class of herbaceous vegetation represented by Poaceae. The other pollen grains found probably belong to another place where the victim was before the crime or the body surroundings, because as mentioned in the expert report, “it had dense vegetation around”. Thus, it did not correspond to the vegetation class of the site.

In case 4, we estimated that there is a chance that the body came from a region with open or dense tree vegetation, as the pollen grains found did not match the vegetation of the corpse last location, even though it laid the most time there. From the expert report, it is possible to link the pollen grains with Herbaceous and Open vegetation because the surrounding of the body site was regenerating vegetation.

In case 5, although the body was found in herbaceous vegetation, the pollen set indicated the probable origin from

Open or Dense vegetation. The expert report categorically stated probable crime dynamics that the body “was taken to the place”, thus corroborated by our pollen analysis.

The distance from the site of case 6 to the vegetation class closest to its surroundings confirmed the probable origin of the body from Open vegetation. Once again, the pollen data are in agreement with the expert report since the victim was displaced, as the pollen grains did not have the botanical fingerprint of the corpse discovery site.

Case 7 is 200 m near Dense vegetation which may explain the diversity of such pollen in the sample, especially Caryocaraceae from Cerradão fragments. The set of pollen types found occurs predominantly in native vegetation. Thus, the body site may be influenced by the proximity to the Dense area, or the person may have been there before committing suicide. The set of pollen types found does not match the location since Open and Dense vegetation is predominant among the pollen grains found; however, the Herbaceous vegetation is predominant around the site. The pollen grain distribution, in this case, indicates the highest probability of Open and Dense vegetation origin.

Not always the distance from the nearest vegetation class to the sample, in the case of Herbaceous vegetation, is more likely to indicate pollen origin. What also contributes to this analysis was the short time interval the body laid there, a few hours. Besides, the presence of water nearby is essential information to be used when exists suspicion if a body or suspect has passed through aquatic environments since they have distinct vegetation. So, it is advisable to investigate the closest water source.

Besides refining the floristic information of a city, knowing the spatial location of water resources can optimize the search work. We can analyze traces of aquatic vegetation adhered to forensic samples. For example, a plant found in the Mércia Nakashima case was decisive in the final sentence: “Mizael's shoe, which was regularly seized, had fragments of a freshwater underwater plant compatible with characteristics of the Nazaré Paulista Dam/SP” (Tjsp, 2014). Thus, the location of a water body combined with the surrounding vegetation class may point to likely places for a police search.

Given the floristic characteristics of each location, the criminal is not well aware of how efficient traces pollen grains are, and it is yet unlikely to prevent human body contact with them. Therefore, they are material evidence that the perpetrator does not give due importance to eliminate from the crime scene or himself or even try to eliminate since some pollen types resist washing or bathing, unlike removable fingerprint or fragments of firearm projectiles.

The pioneering application of forensic palynology to Campo Grande's expertise is the first step toward future applications of this science in many unsolved homicides. The local rate of homicides is high, according to the Department of Justice and Public Security, in 2016 occurred 561 such crimes. Forensic palynology is a tool that can assist in solving crimes by providing proof to find the perpetrator. Many cases lack evidence if the crime occurred in a particular location or the body was moved.

Palynology is a useful tool for linking or excluding a person or object from a particular location, and it should agree with the expert report, used not to defend his opinion but an analysis based on experience of facts interpretation (Hand, 1901). Thus, our presented results described ways of initiating forensic palynology applications by the Brazilian scientific police in urban regions. Forensic palynology may have even more application in comparing objects or suspects to link or exclude them from crime scenes.

Deep convolutional neural networks were also used by Daood et al. (2016) to identify 30 pollen species with a database containing 1000 images. Like those authors, in this paper, we also used ImageNet as a transfer of learning. Daood et al. (2016) obtained 92.4% in the pollen grain classification with the metric precision, while in this research, we obtained 96.55% accuracy with the pollen bank dataset used for ResNet50 architecture.

5. Conclusion

Palinovic and empirical knowledge were proven valid for automating forensic palynology-related information and can serve as a tool for expert use. The pollen information for each case together with the likelihood that they belong to the crime scene is information that may lead to criminal investigation in conjunction with the expert experience and other evidence and information relevant to the case resolution.

The Palinovic modules presented are an essential advance in Brazil in automating the identification and analysis of pollen grains, including relating a city's tree cover to palynology. Increasingly developed technologies and methods such as deep convolutional neural networks used for pollen grain recognition and mapping of urban centers by aerial or satellite imagery have been applied in the development of programs to assist in real cases such as the forensic palynology presented in this work.

Crime scenes contains pollen grains, tiny and easy to adhere. Thus, by analyzing pollen grains collected from clothes, it is possible to apply forensic palynology. Nevertheless, to link the pollen grains found on a corpse with their origin is essential a vegetation map. An image bank of pollen grains occurring in an urban area and periodically feeding it is necessary to keep Palinovic updated and to perform well in pollen recognition. Thorough identification of the pollen is essential for studies such as forensic palynology. For by knowing the geographical spatialization of plants and their pollen grains, it is possible to predict the pollen grains that people or objects passing by must contain. Furthermore, other essential tools for forensic palynology are improvements in pollen identification.

Our results show that it is possible to relate the presence of pollen grains found in forensic samples through the automatic identification of images together with a database of georeferenced plant species. It was possible to analyze the pollen grains collected in eight bodies, where the algorithm presented a performance of 90.51% in the pollen grain classification tests. Furthermore, pollen grains could be correlated with the type of vegetation where the body was found. The technique developed can be applied in other urban centers through a previous georeferencing of the plants to produce a pollen database.

Therefore, with the automation of vegetation location information, society will have tools to improve security, with the possibility of using computer programs to provide evidence to guide investigations. Besides, economically, the public security agents who use this can more quickly finish their expert examinations, saving public resources and time.

Our research is pioneering Brazil's criminal cases related to death using Forensic Palynology. We present the Palinovic to facilitate access to palynology by the scientific police. Moreover, the validation of the program in case studies with forensic palynology makes Mato Grosso do Sul state a pioneer in this area, representing a reference for future studies in other regions.

Acknowledgments

This work received financial support from the Dom Bosco Catholic University (UCDB), the Federal University of Mato Grosso do Sul (UFMS), the University of Nebraska - Lincoln (UNL) and the Foundation for the Support of the Teaching, Science and Technology Development of the State of Mato Grosso do Sul (FUNDECT) with Case nº 59/300.188/2016 and nº 71/700.082/2020. Scholarship was awarded by the Coordination for the Improvement of Higher Education Personnel (CAPES), of the National Council for Scientific and Technological Development (CNPq), ACPF scholarship (Process 304122/2015-7). Data access was granted by the Public Security Secretary of Mato Grosso do Sul. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

Alotaibi, S. S., Sayed, S. M., Alosaimi, M., Alharthi, R., Banjar, A., Abdulqader, N., & Alhamed, R. (2020). Pollen molecular biology: Applications in the forensic palynology and future prospects: A review. *Saudi Journal of Biological Sciences*, 27(5), 1185–1190. https://doi.org/10.1016/j.sjbs.2020.02.019

Astolfi, G., Gonçalves, A. B., Menezes, G. V., Borges, F. S. B., Astolfi, A. C. M. N., Matsubara, E. T., Alvarez, M., & Pistori, H. (2020). POLLEN73S: An image dataset for pollen grains classification. *Ecological Informatics*, 60(101165), 101165. https://doi.org/10.1016/j.ecoinf.2020.101165

Boi, M. (2015). Pollen attachment in common materials. *Aerobiologia*, 31(2), 261–270. https://doi.org/10.1007/s10453-014-9362-2

Bryant, V. M. (2014). *Pollen and Spore Evidence in Forensics*. In Wiley Encyclopedia of Forensic Science (pp. 1–16). John Wiley & Sons, Ltd.

Bryant, V. M., & Holloway, R. G. (1983). The role of palynology in archaeology. *Advances in Archaeological Method and Theory*, 6, 191–224. http://www.jstor.org/stable/20210068

Daood, A. I., Ribeiro, E., & Bush, M. (2018). Sequential recognition of pollen grain Z-stacks by combining CNN and RNN. *Proceedings of the International Florida Artificial Intelligence Research Society Conference*, FLAIRS, Melbourne, Florida, USA.

Daood, A., Ribeiro, E., & Bush, M. (2016). *Pollen grain recognition using deep learning. In Advances in Visual Computing* (pp. 321–330). Springer International Publishing.

García, N. M., Chaves, V. A. E., Briceño, J. C., & Travieso, C. M. (2012). Pollen grains contour analysis on verification approach. In Lecture Notes in *Computer Science* (pp. 521–532). Springer Berlin Heidelberg.

Gonçalves, A. B., Godoi, R. F., Paranhos, A. C., FILHO, Folhes, M. T., & Pistori, H. (2018). Urban phytophysiognomy characterization using NDVI from satellites images and free software. *Anuario Instituto de Geociencias*, 41(3), 24–36. https://doi.org/10.11137/2018_3_24_36

Hand, L. (1901). Historical and practical considerations regarding expert testimony. *Harvard Law Review*, 15(1), 40. https://doi.org/10.2307/1322532

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 770–778.

Holt, K. A., & Bebbington, M. S. (2014). Separating morphologically similar pollen types using basic shape features from digital images: A preliminary study(1.). *Applications in Plant Sciences*, 2(8), 1400032. https://doi.org/10.3732/apps.1400032

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2261–2269.

Kumari, M., Sankhla, M. S., Nandan, M., Sharma, K., & Kumar, R. (2017). Role of forensic palynology in crime investigation. *Journal of Social Relevance & Concern*, 5(3), 1-13.

Mildenhall, D. C. (1990). Forensic palynology in New Zealand. *Review of Palaeobotany and Palynology*, 64(4), 227–234.

Mildenhall, D. C., Wiltshire, P. E. J., & Bryant, V. M. (2017). Forensic palynology: Why do it and how it works. *Forensic Science International*, 163(3), 163–172.

Milne, L., Bryant, V. M., Jr, & Mildenhall, D. C. (2005). Forensic palynology. *In Forensic Botany:Principles and Applications to Criminal Casework* (pp. 217–252). CRC Press.

Ministério do Meio Ambiente. (2013). *Rapideye Satellite Constelation*. Santiago & Cintra Consultoria, São Paulo.

Morgan, R. M., Davies, G., Balestri, F., & Bull, P. A. (2013). The recovery of pollen evidence from documents and its forensic implications. *Journal of the Forensic Science Society*, 53(4), 375–384. https://doi.org/10.1016/j.scijus.2013.03.004

Ochando, J., Munuera, M., Carrión, J. S., Fernández, S., Amorós, G., & Recalde, J. (2018). Forensic palynology revisited: Case studies from semi-arid Spain. *Review of Palaeobotany and Palynology*, 259, 29–38. https://doi.org/10.1016/j.revpalbo.2018.09.015.

Pott, A. & Pott, V.J. (1994). *Plantas do Pantanal.* Embrapa.

Rodrigues, C.N.M., Gonçalves, A.B., Silva, G.G., & Pistori, H. (2015). Evaluation of Machine Learning and Bag of Visual Words Techniques for Pollen Grains Classification. *IEEE Latin America Transactions*, 13, 1-8.

Sevillano, V., & Aznarte, J. L. (2018). Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks. *PloS One*, 13(9), e0201807. https://doi.org/10.1371/journal.pone.0201807

Sevillano, V., Holt, K., & Aznarte, J. L. (2020). Precise automatic classification of 46 different pollen types with convolutional neural networks. *PloS One*, 15(6), e0229751. https://doi.org/10.1371/journal.pone.0229751

Soares Da Silva, D., Nara Balta Quinta, L., Gonçalves, A. B., Pistori, H., & Borth, M. R. (n.d.). Application of wavelet transform in the classification of pollen grains. *African Journal of Agricultural Research*. https://doi.org/10.5897/AJAR2013.7495

Shalizi, C. (2006). Shannon Entropy and Kullback-Leibler Divergence. *In: Shalizi C, Advanced Probability II*, 189-196.

Ticay-Rivas, J. R., del Pozo-Baños, M., Travieso, C. M., Arroyo-Hernández, J., Pérez, S. T., Alonso, J. B., & Mora-Mora, F. (2011). Pollen classification based on geometrical, descriptors and colour features using decorrelation stretching method. *In IFIP Advances in Information and Communication Technology* (pp. 342–349). Springer Berlin Heidelberg.

Travieso, C. M., Briceno, J. C., Ticay-Rivas, J. R., & Alonso, J. B. (2011). Pollen classification based on contour features. *IEEE International Conference on Intelligent Engineering Systems*, 17–21.

Tribunal de Justiça de São Paulo. (2014). *Vara do Tribunal do Júri da Comarca de Guarulhos/SP*, Processo nº 572/10.

Wiltshire, P. E. J. (2006). Hair as a source of forensic evidence in murder investigations. *Forensic Science International*, 163(3), 241–248. https://doi.org/10.1016/j.forsciint.2006.06.070

Wiltshire, P. E. J., Hawksworth, D. L., & Edwards, K. J. (2015). A rapid and efficient method for evaluation of suspect testimony: Palynological scanning. *Journal of Forensic Sciences*, 60(6), 1441–1450. https://doi.org/10.1111/1556-4029.12835

Zavada, M. S., McGraw, S. M., & Miller, M. A. (2007). The role of clothing fabrics as passive pollen collectors in the north-eastern United States. *Grana*, 46(4), 285–291. https://doi.org/10.1080/00173130701780104