Air Microbial Pollution in Public Places Based on Convolutional Neural Network

Asian Ullah[*]

American University of Afghanistan, Afghanistan

[*]*corresponding author*

Abstract: Air microbial pollution in public places is particularly harmful to children who are in the growth and development period and the elderly with reduced physical function. Therefore, it is urgent to improve the ecological environment and give mankind a healthy earth. The purpose of this paper is to study air microbial pollution in public places based on convolutional neural networks. The relevant theoretical knowledge of microbial species identification based on CNN model is introduced in detail. The identification of bacterial and fungal concentrations of microorganisms in medical buildings in colleges and universities is analyzed. The species identification method based on hierarchical CNN is compared with the existing species identification tools RDPClassifier and 16SClassifier based on machine learning algorithms in terms of classification accuracy and identification speed. By comparison, the feasibility and efficiency of the hierarchical CNN-based species identification method are verified.

1. Introduction

With the continuous pursuit of people's quality of life, while having a comfortable indoor environment, it also provides a suitable living environment for microorganisms, so that infectious pathogens and some outdoor bacteria and fungi can multiply and spread indoors, while modern people generally live indoors. Spending most of the day, people are increasingly concerned about indoor air quality (IAQ) [1]. In addition to particulate aerosol pollution in indoor air pollution, biological aerosols also account for a certain proportion. Inhalation of bioaerosols can be detrimental to health and can even cause serious casualties when pathogenic microorganisms are involved [2].

Microbial contamination is a key factor in indoor air pollution [3]. Bacteria and fungi that grow indoors often contribute to indoor air pollution when there is enough moisture. Landry K G

assessment of microbial concentrations and main bacteria and fungi in the indoor environment of the Central Library of the University of Yaounde. 76 samples were drawn from indoor air, surfaces, and moldy books. Bioaerosol sampling and air concentrations were performed using passive air sampling techniques using petri dishes containing different media and exposures for 30, 60 and 90 min in the morning and afternoon. Sampling by wiping surfaces and moldy books with sterile cotton swabs. The identification of isolated microorganisms is based on macroscopic, microscopic and biochemical characteristics [4]. Odonkor ST assessment of microbial air quality and associated environmental health hazards at landfill sites in selected areas of the Greater Accra region of Ghana. A random sampling method was used to select sampling sites for dry and rainy seasons from the landfill site and its corresponding communities. The results showed that the total number of bacteria and fungi (CFU/m 3) in the air of the landfill was higher than that of the community. Statistically significant changes in bacterial and fungal concentrations between the two seasons p = 0.05 for landfills and communities [5]. In summary, it is a very meaningful work to propose an easy-to-operate and low-cost air quality monitoring method [6].

The sources of indoor air microorganisms are diverse and related to various environmental parameters. This paper intends to study the parameters of microbial concentration in the indoor environment of the school hospital and establish an indoor microbial identification model. Using the identification model studied in this paper, some parameters related to monitoring air are indirectly detected, and then the indoor microbial concentration is calculated in real time, so that the real-time control of indoor microbial pollution can be quickly realized at a low cost. Ventilation operating strategy or air purifier volume. This has theoretical significance and practical application value for improving indoor environment quality and protecting people's health.

2. Research on Air Microbial Pollution in Public Places Based on Convolutional Neural Networks

2.1. Air Microorganisms

Airborne microorganisms refer to living organisms such as bacteria, molds and actinomycetes in the air. The main sources are soil, water, animals, plants and humans. In addition, sewage treatment, animal feeding, fermentation processes and agricultural activities are also airborne microorganisms. important source. As one of the important members of the biosphere, microorganisms are indispensable in the cycle of natural organic matter mineralization, nitrogen, carbon, sulfur, phosphorus and other important elements. The spread and transfer of aerosols can not only affect the ecological balance of the biosphere, but also cause environmental pollution and outbreaks of infectious diseases [7-8]. High concentrations of airborne microbial particles can cause various animal and plant diseases, human acute and chronic diseases, such as respiratory tract infections, asthma, wheezing, skin allergies, chronic lung diseases and cardiovascular diseases. They multiply in the air and spread to the surrounding environment, causing disease transmission and allergic reactions, which can easily cause serious harm to the health of people with low immunity. In addition, some bacteria can also cause the deterioration of daily necessities and the decay of buildings. Therefore, widely distributed airborne microbial particles are closely related to human society, and not only have important ecosystem functions, but are also one of the important indicators of ambient air quality [9-10].

2.2. Bioaerosols

Biological aerosol is a kind of aerosol containing microorganisms or other macromolecular life forms. The common "smoke, fog, haze, dust" and so on belong to the category of aerosols. The particle size of atmospheric aerosols is between 0.01 μm and 100 μm. According to the aerodynamic diameter (da), it includes total suspended particulate matter (TSP), inhalable particulate matter (PM10) and fine particulate matter (PM2.5) [11-12]. Since the particles below 10 μm can enter the nasal cavity, and the particles below 2.5 μm can enter the human blood circulation from the lungs, so in aerosol monitoring, generally focus on particles with da<10 μm [13].

2.3. Deep Learning

Deep learning, also known as deep structural learning or hierarchical learning, is one of the machine learning methods [14]. The deep learning method has a multi-layer neural network, each layer of neurons will use the information vector output by the previous layer as input, coupled with the nonlinear processing unit for feature extraction and conversion, so that the deep learning method has a wider range of data learning expressions ability [15]. The feature extraction of traditional machine learning methods requires manual work, which not only requires a lot of professional domain knowledge, but also cannot adapt to a variety of tasks, making the model inefficient and not universal. Compared with deep learning methods, which do not require manual feature extraction, each layer of neural network transfers the learned rules to the next layer, and combines simple models to learn more complex features. The feature learning of deep learning focuses on finding the optimal feature subset and eliminating irrelevant or redundant features of the data set, thereby reducing the number of features, improving the accuracy of prediction results, and reducing computing time.

2.4. Convolutional Neural Networks

Convolutional Neural Network (CNN) is a deep feedforward neural network, suitable for processing image data, and has achieved great success in the field of computer vision. Convolutional neural networks usually consist of input and output layers and several hidden layers [16]. The input layer is usually a two-dimensional vector. Hidden layers can be divided into convolutional layers, pooling layers and fully connected layers [3<)1. The convolutional layer is the core of CNN. As a feature extractor, it performs local feature extraction on the input two-dimensional vector to maintain the spatial continuity of the image. The pooling layer is divided into a maximum pooling layer and an average pooling layer, which reduces the dimension of the intermediate hidden layer through downsampling, reduces the amount of data operations, and maintains the usefulness of image data. The fully connected layer is a conventional artificial neural network that connects the vector features output by multiple convolutional layers and pooling layers. The output layer calculates the predicted value of the result. If it is a classification problem, the predicted probability is obtained through the Softmax function [17-18]. Convolutional neural networks are characterized by local field-of-view perception, weight sharing in the same convolutional layer, and multiple convolutional layers extracting different features. The pooling operation reduces the model parameters, reduces the complexity of the network, and also provides translation, rotation, and scaling invariance of the image.

3. Investigation and Research on Air Microbial Pollution in Public Places Based on Convolutional Neural Networks

3.1. Testing Site

The hospital of the tested school is located on the first floor of the building, and its internal space is small. The room is equipped with independent air-conditioning equipment and is in a normally closed state. The ventilation mode is mainly natural ventilation, and it is regularly cleaned every day. The patients are mainly students and teachers. At 11 am From 00:00 to 2:00 pm, there are a large number of patients, and the staff stays for a short time. The microbial aerosol concentration level data in the school hospital are shown in Table 1.

Table 1. Data of microbial aerosol concentration levels in school hospitals

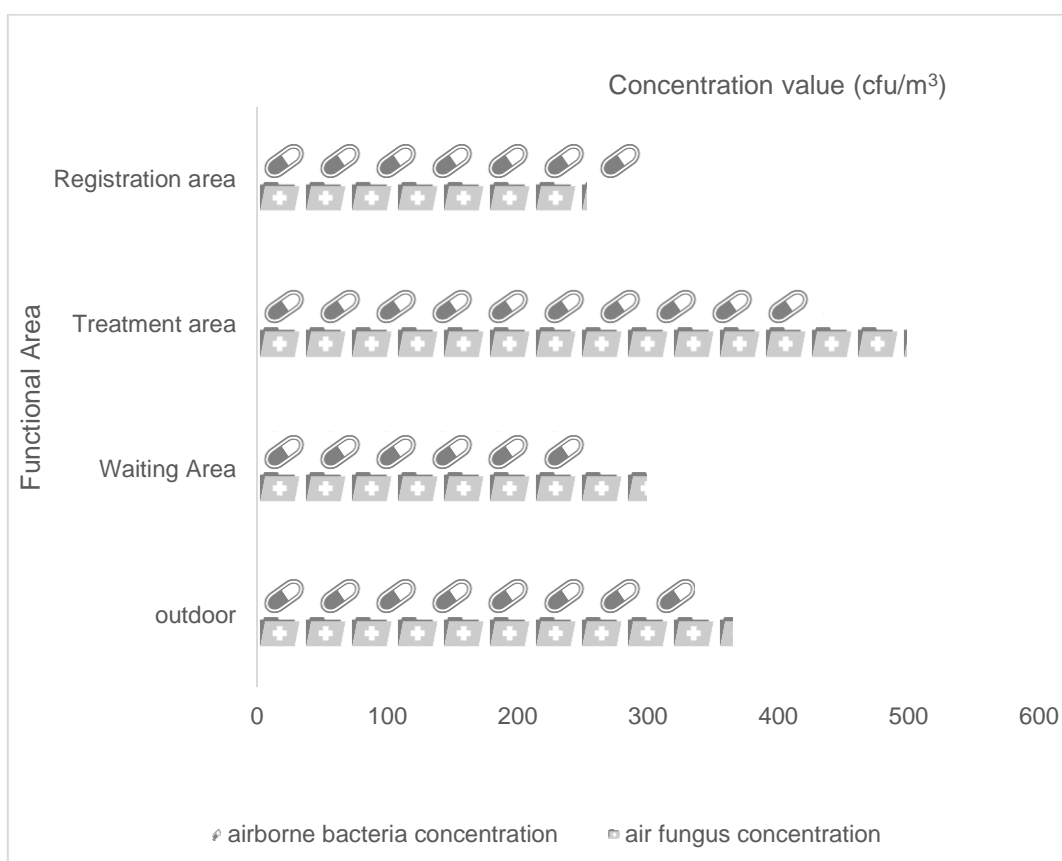| Functional area	Airborne bacteria concentration	Air fungus concentration
Registration area	302	253
Treatment area	435	498
Waiting area	257	299
Outdoor	336	365

Figure 1. Microbial aerosol concentration data at the testing site

It can be seen from Figure 1 that the bacterial aerosol concentration level in the school hospital is relatively low. The main reason is that the stagnation time of the personnel in the school hospital is short, and the indoor hygiene is kept clean and tidy, which effectively reduces the bacterial aerosol concentration.

3.2. The Training Process of CNN Network

Assuming that the lth layer is a convolutional layer, the calculation of this layer includes convolution calculation and activation function nonlinear transformation calculation. The calculation formula is shown in formula (1), where z' represents the output after convolution, and al-1 represents The output of the l-1th layer, that is, the input of the lth layer, wl represents the weight matrix of the convolution kernel of the lth layer, the symbol * represents the operation of convolution, σ represents the activation function, this paper uses ReLU activation function.

$$a^l = \sigma(z^l) = \sigma(a^{l-1} * W^l + b^l) \tag{1}$$

Assuming that the first layer is a pooling layer, its error calculation formula is shown in formula (2). Upsample means that the matrix is up-sampling so that the size of the matrix is the same as the size before pooling.

$$\delta^l = upsample(\delta^{l+1})\Theta\sigma'(z^l) \tag{2}$$

Unlike fully connected layers and convolutional layers, pooling layers do not require backpropagation to modify parameters. The convolutional layer in this experiment adopts maximum subsampling, and the convolution kernel of this subsampling is to take the maximum value of the corresponding position.

4. Analysis and Research of Air Microbial Pollution in Public Places Based on Convolutional Neural Network

4.1. Identification of Air Microorganisms Based on CNN

We draw the model of the convolutional neural network in text classification as shown in Figure 2. The network contains a single convolutional layer and a pooling layer. The input of this paper is not an English word, but a base sequence substring, which is obtained by means of a sliding window.
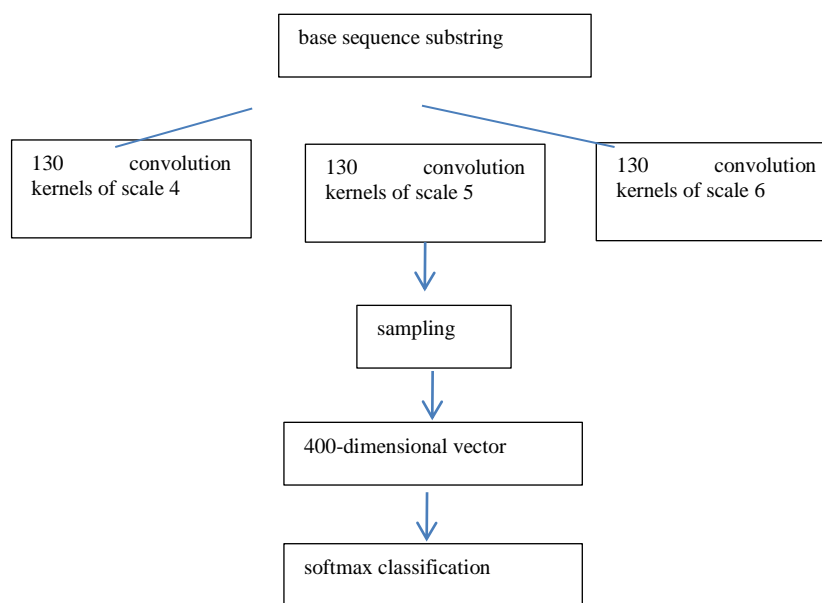
Figure 2. Model diagram of the CNN network

The first layer in the convolutional neural network is the input layer. It first uses the sliding window method to divide all 16SrRNA sequences into substrings of a specific length to construct the vocabulary. Due to the inconsistent length of the sentences, fill in the blanks. The character method makes all sequences have the same length, and then the embedding layer is used to map the index of the sequence substring in the thesaurus into a vector with a fixed size. The second layer is the convolution layer. The convolution layer uses a combination of convolution kernels, that is, the convolution layer contains three different scales of convolution kernels. The scales of the convolution kernels are 4, 5, and 6, respectively. There are 130 scale convolution kernels, which means that each scale convolution kernel will generate 130 feature maps after convolving the input data, because each convolution will generate tensors of different shapes, so we need to traverse each The results of each convolution kernel are then combined into a total feature vector. The third layer is the downsampling layer, that is, the pooling layer. The downsampling method used in this paper is the maximum subsampling method, which combines the tensors after the subsampling operation to obtain a vector with a length of 400 dimensions. The combined vector is input into the next layer, the next layer is the fully connected layer and the classification layer, the drop-out method can be used for the fully connected layer to prevent over-fitting, and the classifier used in the classification layer is the softmax classifier. The characteristic of the device is that the output is the probability that the sequence belongs to a certain class, and finally the class with the highest probability is used as the classification result of the unknown species sequence.

4.2. Performance Comparison with Other Microbial Identification Algorithms

Compared with the existing microbial identification tools RDPClassifier and 16SClassifier based on machine learning algorithms, the effectiveness of the CNN-based microbial identification method is verified. In order to further verify the efficiency of the model in this paper, 1000 sequences were randomly selected from the V2 data for testing. The time spent by each tool is shown in Table 2. RDP# in the table represents RDPClassifier, 16S# represents 16SClassifier, CNN

$ denotes a CNN-based microbial identification method. The CNN-based microbial identification method took about 80 seconds with 16SClassifier and about 360 seconds with RDPClassifier, as shown in Figure 3.

Table 2. Comparison of identification speed between CNN-based species identification and RDPClassifier and 16SClassifier

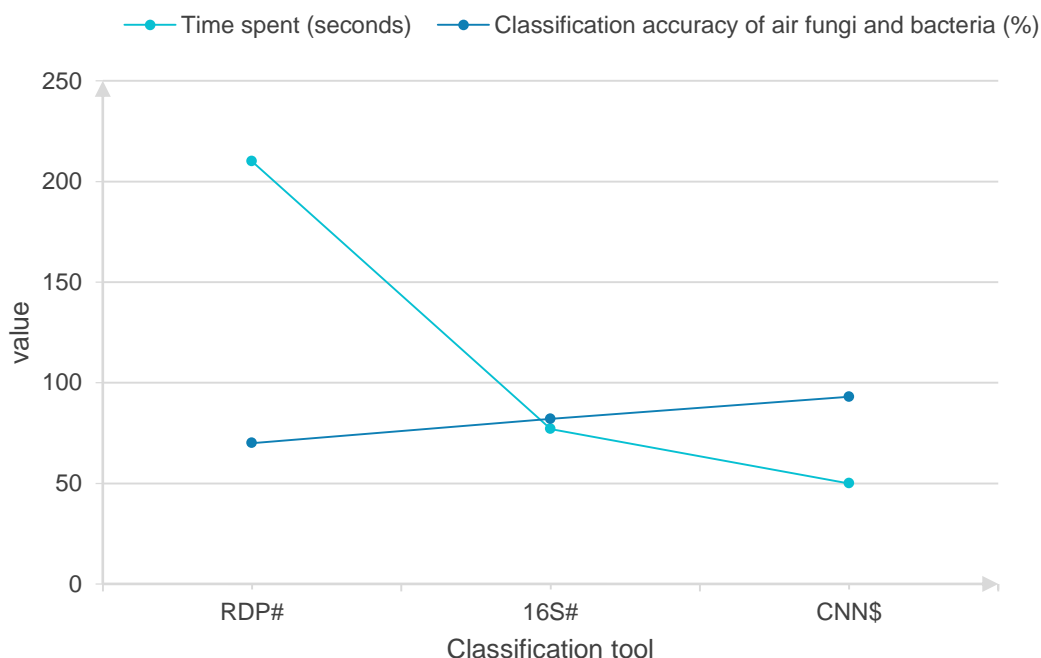| Classification tool	Time spent (seconds)	Classification accuracy of air fungi and bacteria (%)
RDP#	210	70
16S#	77	82
CNN$	50	93

Figure 3. Comparison results

Compared with RDPClassifier and 16SClassifier, the efficiency and feasibility of the CNN-based microbial identification method are verified from the perspectives of the classification accuracy of air fungi and bacteria and the speed of microbial identification.

5. Conclusion

In recent years, with the continuous progress of the city and the continuous growth of the population, the air pollution problem has caused a great negative impact on the health and daily life of the residents, and the air is closely related to the life of each of us. The air quality algorithm

based on convolutional network proposed in this paper effectively improves the air quality detection performance. However, there is still room for further expansion in future research work, and in-depth research can be carried out in the following directions: applying the convolutional network proposed in this paper to other spatiotemporal sequence prediction problems, using more real-world datasets for Research. The algorithm model has been continuously improved, which can be better applied to problems in different fields but similar in nature, making the model more generalizable and universal. In terms of model training and parameter optimization, consider trying more diverse and advanced optimization algorithms, which can be improved for different problems and have better results.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Burris H H, Just A, Riis V, et al. 1082: The impact of ambient air pollution on the presence of high-risk cervicovaginal microbial communities. American Journal of Obstetrics and Gynecology, 2020, 222(1):S667-S668. https://doi.org/10.1016/j.ajog.2019.11.1094

[2] Masyagutova L M, Bakirov A B, Simonova N I, et al. Laboratory confirmation of staging and volume of preventive measures at work under the conditions of microbial pollution of work environment air. Klinicheskaia Laboratornaia Diagnostika, 2018, 63(9):584-587.

[3] Rooij M D, Hoek G, Schmitt H, et al. Insights into Livestock-Related Microbial Concentrations in Air at Residential Level in a Livestock Dense Area. Environmental Science and Technology, 2019, 53(13):7746-7758. https://doi.org/10.1021/acs.est.8b07029

[4] Landry K G, Ascension N M, Armelle D, et al. Assessment of Indoor Microbial Quality of Library's Premise: Case of Central Library of the University of Yaoundé I. Open Journal of Preventive Medicine, 2018, 8(4):109-120.

[5] Odonkor S T, Mahami T. Microbial Air Quality in Neighborhoods near Landfill Sites: Implications for Public Health. Journal of Environmental and Public Health, 2020, 2020(3):1-10. https://doi.org/10.1155/2020/4609164

[6] Tanikkul P, Boonyawanich S, Pisutpaisal N. Bioelectricity recovery and pollution reduction of distillery wastewater in air-cathode SCMFC. International Journal of Hydrogen Energy, 2019, 44(11):5481-5487.

[7] Osterwalder S, Huang J H, Shetaya W H, et al. Mercury emission from industrially contaminated soils in relation to chemical, microbial, and meteorological factors. Environmental Pollution, 2019, 250(JUL.):944-952.

[8] Mushi D, Kebede G, Linke R B, et al. Microbial faecal pollution of river water in a watershed

of tropical Ethiopian highlands is driven by diffuse pollution sources. Journal of water and health, 2021, 19(4):575-591. https://doi.org/10.2166/wh.2021.269

[9] Celik F, Camas M, Kyeremeh K, et al. Microbial Sorption of Uranium Using Amycolatopsis sp. K47 Isolated from Uranium Deposits. Water Air & Soil Pollution, 2018, 229(4):112.1-112.14.

[10] Abas A, Awang A, Aiyub K. Lichen As Bio-Indicator For Air Pollution In Klang Selangor. Pollution Research, 2018, 37(4):35-39.

[11] Mehmedinovi N I, Kesi A, Estan A, et al. Biosensor Chip for Monitoring of Water Pollution as Novel Test Tool in Public Health: Proof of Principle. International Journal of Engineering and Advanced Technology, 2021, 10(5):386-392.

[12] Fernandez-Baca C P, Omar A, Pollard J T, et al. Microbial communities controlling methane and nutrient cycling in leach field soils. Water Research, 2019, 151(MAR.15):456-467.

[13] Dengming, Yan, Xinshan, et al. Bioelectricity generation from air-cathode microbial fuel cell connected to constructed wetland.. Water science and technology : a journal of the International Association on Water Pollution Research, 2018, 78(9):1990-1996. https://doi.org/10.2166/wst.2018.471

[14] Landrigan P J, Stegeman J J, Fleming L E, et al. Human Health and Ocean Pollution. Annals of Global Health, 2020, 86(1):1-64. https://doi.org/10.5334/aogh.2831

[15] Kedare V J, Kharat S S, Tawde S. Effectiveness of Constructed Indoor Air Purifier to Enhance Indoor Air Quality. European Journal of Engineering Research and Science, 2020, 5(12):61-66.

[16] Kic P, Ruzek L, Popelarova E. Concentration of air-borne microorganisms in sport facilities. Agronomy Research, 2018, 16(4):1720-1727.

[17] Pompilio A, Bonaventura G D. Ambient air pollution and respiratory bacterial infections, a troubling association: epidemiology, underlying mechanisms, and future challenges. Critical Reviews in Microbiology, 2020, 46(5):600-630.

[18] Pivato A, Raga R, Marzorati S, et al. Mitigating long-term emissions of landfill aftercare: Preliminary results from experiments combining microbial electrochemical technologies andin situaeration:. Waste Management & Research, 2021, 40(5):596-606. https://doi.org/10.1177/0734242X20983895