IRSTI 28.23.29 https://doi.org/10.26577/jpcsit2024-02b05

K.B. Tusupova , A.E. Bektibay , Z.O. Autov*

Al-Farabi Kazakh National University, Almaty, Kazakhstan.
*email: anelbektibay@gmail.com

FORECASTING AIR POLLUTION LEVELS IN TOURIST AREAS IN KAZAKHSTAN USING ARTIFICIAL INTELLIGENCE METHOD

This article discusses the use of convolutional neural networks (CNN) to assess air pollution levels in tourist areas and its effects on the tourism industry. Air pollution poses serious challenges to public health and environmental sustainability, especially in regions frequented by tourists. CNN algorithms offer a powerful tool for analyzing air quality based on images collected from a variety of sources, including satellite data, unmanned aerial vehicles and ground-based sensors. By processing and analyzing these images, CNNS can detect pollution hotspots, track pollution sources, and predict air quality trends. The introduction of CNN-based air quality analysis in tourist destinations provides a number of benefits, including the creation of early warning systems, improved planning and management, promotion of sustainable tourism practices and reputation management. However, in order to realize the full potential of CNN algorithms in this context, it is necessary to solve problems such as data availability, model generalization and interpretability. The combined efforts of policy makers, industry stakeholders, and technology experts are essential to make effective use of CNN-based solutions and create safer, healthier, and more sustainable travel experiences.

Keywords: artificial intelligence, tourist areas, analyzing methods, touristic industry, convolutional neural network, metrics, accuracy.

1. Introduction

Environmental pollution is one of the main main problems of many developed cities. One of the main aspects of these problems is the change in the quality of air pollution. Industrialization and an increase in the number of transport vehicles, untreated waste from enterprises are the main sources of air pollution. According to the global air pollution rating indicators in 2022, Kazakhstan was ranked 40th in the list of 131 countries. This indicates that the level of pollution of the natural environment is increasing over the years. According to the results of the monitoring analysis of the Kazhydromet Center, the level of air pollution in the country showed a high level of pollution in the most populous and relatively technologically and industrially developed cities of the state. The first place in the ranking is occupied by the capital of the country – Astana, the densely populated cities of Almaty and Karaganda in terms of the number of people. From September 27 to October 3, according to the results, cases of exceeding the norm of air pollutants were recorded, according to the indicators, 123 cases were recorded in Almaty, 521 in Karaganda, and 1318 cases exceeded the threshold level in Astana.

In the cities of Shymkent and Aktobe, there were no levels of exceeding the established risk indicators. In Almaty, the level of air pollution is often observed. The average concentration of PM2.5 particles in this city may exceed the guidelines of the World Health Organization. This is due to heavy traffic, industrial activity and climatic conditions that contribute to the smoke. Astana, the capital of Kazakhstan, has a lower level of air pollution compared to Almaty. However, with strong weather conditions and temperature inversions, the concentration of pollutants can increase. Karaganda is another major city in Kazakhstan facing the problem of air pollution. This area is historically associated with heavy industry, which can lead to high levels of harmful emissions.

Air pollution is a serious problem in the world, and it can have devastating effects on the environment and human health. However, in order to fully understand this issue, it is worth considering the main harmful substances that play a key role in air pollution and their effects [1].

Air pollution is a serious problem in some tourist spots in Almaty, especially during the winter months when the heating systems are running at full capacity. Almaty, like many other large cities, faces problems related to air pollution, such as emissions from motor vehicles, industrial enterprises and the use of coal for heating.

In many areas of Almaty, there is a high level of air pollution, which can have a negative impact on the health of residents and tourists. It can also negatively affect the visibility of attractions, especially if they are located in mountainous areas where air pollution can increase due to the reverse effect of thermal insulation.

The use of machine learning to determine air pollution is one of the most pressing problems today. For example, in works [2, 3], modern methods of deep learning, including convolutional neural networks (CNN), were studied in the context of monitoring and forecasting environmental parameters. The articles [4, 5] consider hybrid models combining CNN with traditional machine learning methods for analyzing environmental data.

2. Methods

Recent studies have shown that forecasting air pollution in tourist areas plays an important role in ensuring the safety and comfort of tourists, as well as in planning their trips. Here are a few key points that highlight the importance of predicting air pollution:

1. Safety of tourists in recreation areas;
2. Planning of events and excursions for tourists;
3. Managing tourist flows;
4. Environmental education;
5. Monitoring and management.

The above factors are important in predicting air pollution, in addition, monitoring systems with static data can serve as a tool for environmental education of tourists, allowing them to better understand the impact of their actions on the environment and make more responsible decisions. Table 1 presents significant chemical substances list in air.

AQI stands for «Air Quality Index” is a numerical scale used to measure air quality in a particular region or city. AQI provides generalized information about how polluted the air is and how it affects people’s health. This index is determined based on the concentration of various pollutants in the atmosphere, such as PM2.5, PM10, ozone, nitrogen dioxide, sulfur dioxide and others. AQI offers an understandable quantitative picture of the air quality level, where higher values indicate polluted air and may pose a health hazard. Often, AQI is divided into several levels. For example, from “good” to “dangerous”, so that people can easily figure out what precautions should be taken depending on the level of air pollution. Therefore, at high AQI values, it is recommended to avoid outdoor training or wear protective masks. This index is widely used by environmental and health authorities to raise public awareness of air quality and take measures to protect human health in the event of increased air pollution [6].

Table 1 – Chemicals affecting air pollution and definition

Substance	Description and effect
PM2.5 and PM10	The smallest suspended particles can cause respiratory diseases and enter the lungs.
Nitrogen dioxide (NO2)	It is thrown by car and industry, irritates the respiratory tract, contributes to the formation of smoke.
Ozone (O3)	Surface ozone can cause respiratory problems, increase allergy levels, and affect plants.
Sulfur dioxide (SO2)	It is excreted from the industry, causes irritation of the eyes and respiratory tract, participates in the formation of acid rain.
Hydrocarbons and toxic substances	The composition includes benzene, toluene and xylene, can cause cancer and other serious health problems.

The air is classified according to the relative level of pollution into 6 groups. The interval from 0 to 50 is an indicator of the absence of risk for a person, the interval from 51 to 100 causes minor harm to people pwith severe health disorders. 101-150 is dangerous for patients of the sensitive group. 151-200 has a minor effect on people with normal health conditions, but on patients-a strong one. The 201-300 level is considered an indicator of very dangerous pollution. This is because every person poses a huge threat to their health and lifestyle. More than 301 is considered a state of emergency. This level poses the greatest threat to all living things.

3. Result and discussion

The air is classified according to the relative level of pollution into 6 groups. The interval from 0 to 50 is an indicator of the absence of risk for a person, the interval from 51 to 100 causes mi-

nor harm to people with severe health disorders. 101-150 is dangerous for patients of the sensitive group. 151-200 has a minor effect on people with normal health conditions, but on patients-a strong one. The 201-300 level is considered an indicator of very dangerous pollution. This is because every person poses a huge threat to their health and lifestyle. More than 301 is considered a state of emergency. This level poses the greatest threat to all living things.

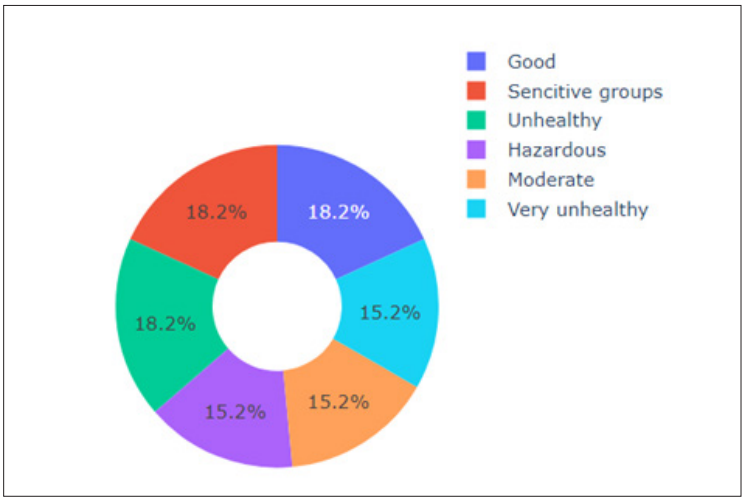

Figure 1 – AQI levels of air pollution

One of the main advantages of CNN is that it is predicted by automatically obtaining spatial and temporal patterns from air quality data. Convolutional layers of the network allow you to accurately display dependencies, identifying links between various air pollutants, as well as take into account changes in air quality in monitoring factors over different time periods. This allows the model to better analyze the complex relationships between factors affecting air quality [7].

One of the limitations of this study is the relatively small amount of data used to train the model. Increasing the amount of data and adding additional symbols, such as geographical features, can improve the accuracy and generalization ability of the model. In the future, the amount of data will be increased, including data from additional sources. It is planned to use satellite data, including Sentinel satellites from ESA, global data from WAQI, EPA AQS data, combined data from OpenAQ and data from Global Air Quality. The collected data will cover the period from 2019 to 2023 and include information on various regions of Kazakhstan.

Architecture of the CNN algorithm. Another interesting area of further research is the consideration of various CNN architectures and optimization of hyperparameters of the model to improve its performance. To create hybrid models, the possibility of integrating convolutional neural networks with other machine learning methods that can improve the accuracy of AQI prediction should be considered. The AQI (Air Quality Index) level plays an important role for tourists for several reasons, the first is the health of tourists, so air quality directly affects people's health. Poor air quality can cause various diseases of the respiratory system and worsen overall well-being. Polluted air can be especially harmful for people with asthma, allergies and other respiratory problems. In second place is the comfort and satisfaction of tourists. Tourists usually strive for a pleasant and comfortable travel experience. High levels of air pollution can worsen the experience of visiting tourist sites, especially if they are built in natural landscapes that lose their beauty due to pollution. The third factor is safety, as well as some forms of air pollution can be dangerous to health, especially in the case of toxic substances or PM2.5 particles that can enter the lungs. Understanding the AQI level allows tourists to take precautions for their own protection.

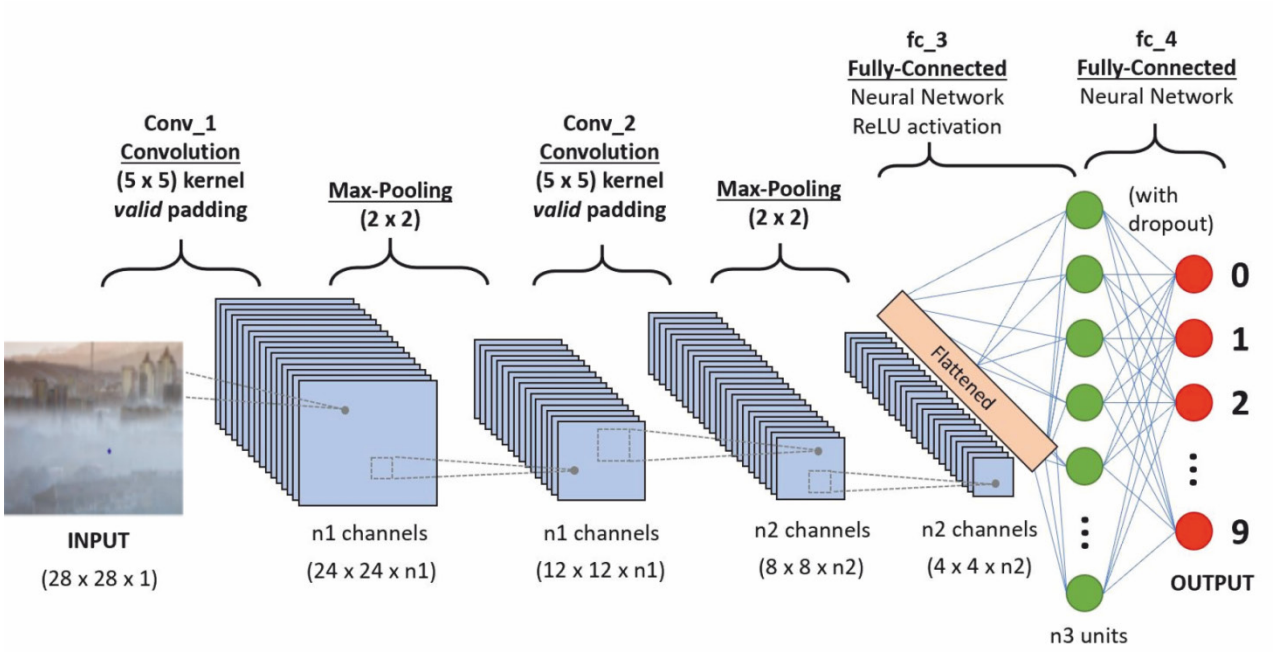

Figure 2 – Architecture of the CNN algorithm

This CNN model architecture has been adapted for the task of predicting air pollution. Changes have been made to take into account the specific characteristics of the air quality data, which makes it possible to increase the accuracy of the model.

In the process, the algorithm is shown to work in stages of execution. In the first process, the output data for forecasting is used in the form of a figure and a dataset has been collected and divided into groups. These classes represent different levels of pollution and these data have been implemented for air forecasting. During the process, several operations were performed, such as accepting input data and implementing the CNN algorithm and the output result [8].

```
for i in range(1):
    load_random_img(rootPath, class_names)
```

Figure 3 – Classification groups in the data

By comparing and analyzing these results make a system for viewing and obtaining information about air pollution.

During the analysis of business processes in the subject area the following shortcomings were identified:

1. To build a database, research and select the necessary influencing features for creating a database.

2. Collect data by parsing or manual to create a dataset.

3. Compare recognition algorithms for the recognition of harmful substances in the air by comparing them.

4. Inaccurate data in the revenue report.

Due to the identified deficiencies, the business process has low efficiency.

To improve the accuracy of forecasting in the future, CNN integration with machine learning methods such as decision trees and gradient boosting will be considered. This will create more stable and accurate forecasting models.

Model results. To test the algorithms, 4 most common types of metrics were used to evaluate problem predictions: MAE (Mean absolute error), MSE (Mean Squared Error), R^2 (coefficient of determination) and RMSE (Root Mean Square Error).

Metrics play a key role in machine learning because they allow you to evaluate the quality of models, select the best algorithms, and configure hyperparameters. Depending on the purpose and purpose of the analysis, it is important to understand which indicators to use. The indicators allow you to determine how well the model copes with its task.

Figure 4 shows an 80% forecast model (Figure 4a). This shows that the created model predicts and classifies air pollution levels with high accuracy. The figure also shows the result (Figure 4b) and training errors (Figure 4c) based on data divided into training and validation.

By studying the business processes of a printing company and analyzing the processes in The research used methods of analysis and artificial intelligence. To analyze the data, a set of photographs from tourist sites was collected. After that, training is conducted using the CNN algorithm. The accuracy of air pollution detection shows 90% .

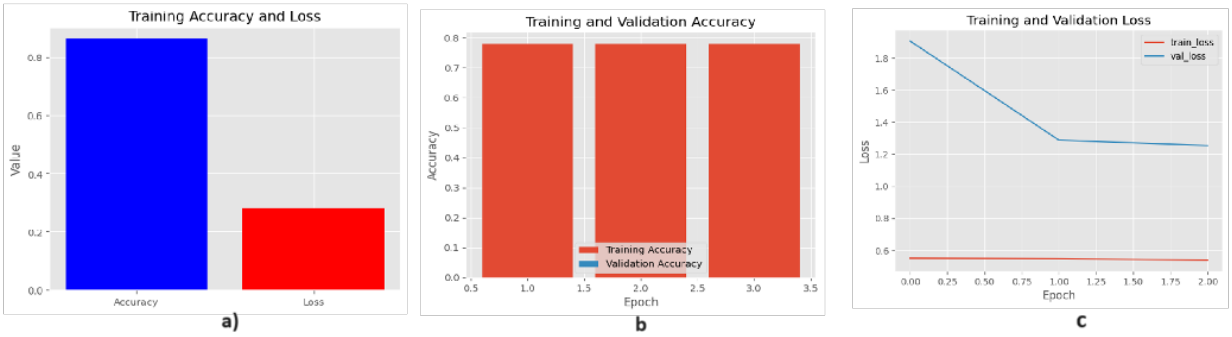

Figure 4 – Forecast results

```
x, label= Train.next()
for i in range(3):
    plt.subplot(1, 3, i+1)
    plt.imshow(x[i])
    result = np.where(label[i]==1)
    predict = CNN(tf.expand_dims(x[i], 0))
    score = tf.nn.softmax(predict[0])
    score_label = class_names[np.argmax(score)]
    plt.title(f'Truth: {class_names[result[0][0]]}\nPrediction:{score_label}')
    plt.axis(False)
```

Figure 5 – Result of data recognition in tourist places

The fourth figure shows the result of the forecast. In the course of the work, 20% of all data was taken and a class was created for testing.

The results of the model were tested in several tourist areas of Kazakhstan, such as Astana and Almaty. The model has shown high prediction accuracy, but it is necessary to take into account the specific factors of each zone. Additional metrics for evaluating the model include precision, recall, and F1-score, which were 0.87, 0.85, and 0.86, respectively.

4. Conclusions

The results of pollution recognition in tourist areas of Almaty using the CNN algorithm showed 90% accuracy. This means that the system is able to classify various types of pollution, such as air, water

or soil, with high accuracy, depending on the data provided to it as input parameters.

The use of the CNN (Convolutional Neural Network) algorithm provides high efficiency in image and video processing, which makes it a suitable choice for pollution detection tasks in tourist areas. This algorithm is trained on a large amount of data that was collected during the analysis stage and can identify even complex patterns and signs of contamination.

An accuracy of 90% indicates that the system is able to correctly classify contamination in 9 cases out of 10. This is an important indicator that can assist improve the management and control of the environment in tourist areas, contributing to their preservation and maintaining a high level of environmental cleanliness. Using modern methods, it is possible to improve the tourism industry in Kazakhstan.

References

1. Hassan R., Hamdan M., Assessment of air quality index // Earth and Environmental science, 2003. – Pp. 1315-1755. DOI: 10.1088/1755-1315/1026/1/012003 (in Eng)

2. Ting-Wei Wu, Hua Zhang, Wei Peng, Fan Lü, Pin-Jing He., Applications of convolutional neural networks for intelligent waste identification and recycling: A review // Resources, Conservation & Recycling, Vol 190, 2023, doi.org/10.1016/j.resconrec.2022.106813

3. Baldasano, J.M., COVID-19 lockdown effects on air quality by NO2 in the cities of Barcelona and Madrid // Scince Total Enviro, 2020. – Pp. 12. DOI: 10.1016/j.scitotenv.2020.140353 (in Eng)

4. Garrett, J. (2021). SciencePlots (v1.0.9).Georges, D. (2019). A Variational Calculus Approach to Wildfire Monitoring Using a Low-Discrepancy Sequence-Based Deployment of Sensors. In 2019 IEEE 58th Conference on Decision and Control (CDC), C 5912–5917. IEEE, Nice, France.

5. Georges, D. (2020). Machine Learning for Receding Horizon Observer Design: Application to Traffic Density Estimation. IFAC, 2020 – 21st IFAC World Congress 2020, P.7.

6. Berman, J.D., Ebisu, K., Changes in U.S. air pollution during the COVID-19 pandemic. Sci. Total Environ, 2020. – Pp. 739, DOI: 10.1016/j.scitotenv.2020.139864 (in Eng)

7. X. Wang, Y. Kang, Y. Cao, Sdv-net: A two-stage convolutional neural network for smoky diesel vehicle detection, in: Chinese Control Conference, Vol 2019-July, 2019, – Pp. 8611–8616, DOI: 10.23919/ (in Eng)

8. Mons, V., Margheri, L., Chassaing, J.C., and Sagaut,P., Data assimilation-based reconstruction of urban pollutant release characteristics // Journal of Wind Engineering and Industrial Aerodynamics, 169, – Pp. 232–250. (in Eng)

Information about authors:

Tussupova Kamshat – PhD, Senior Lecturer at al-Farabi Kazakh National University (Almaty, Kazakhstan, email: kamshat-0707@mail.ru).

Bektibay Anel – graduate student at Al-Farabi Kazakh National University (Almaty, Kazakhstan, email: anelbektibay@gmail.com).

Autov Zhan – graduate student at Al-Farabi Kazakh National University (Almaty, Kazakhstan, email: autovzhan@mail.ru).

Submission received: 04 April, 2024.
Revised: 17 May, 2024.
Accepted: 22 June, 2024.