Machine learning algorithms for categorization of agricultural dust emissions using image processing of wheat combine harvester[†]

Utpal Ekka[1], Himadri Shekhar Roy[2,*], Adarsh Kumar[1], S. P. Singh[1], Apratim Kumar Pandey[1] and Kamalika Nath[2]

[1]ICAR-Indian Agricultural Research Institute, Pusa, New Delhi 110 012, India
[2]ICAR-Indian Agricultural Statistics Research Institute, Pusa, New Delhi 110 012, India

India is the second largest wheat producer in the world after Russia. Wheat harvesting in the country was traditionally done using a sickle, a hand tool. However, in the last two decades, combined harvesters have been extensively used. The rapid development of mechanization has resulted in the production of dust and straw particles during the harvesting operation of wheat. These particles have severe health hazards for the machine operator. Exposure to various types of particulate matter has a variety of effects on human health. Such an effect can be minimized if the concentration of the generated particle is maintained within a permissible limit. Hence, the present study has been conducted to evaluate and categorize dust and straw particles in the workspace of a combine harvester operator during wheat harvesting. An image-processing technique was used to study a field data sample collected on sticky paper. It describes a novel method of collecting dust and straw particles while harvesting wheat. Few studies have been conducted in developing countries to analyse the characteristics of dust and wheat straw exposure of combined harvester operators. The number of dust and straw particles deposited per square millimetre was 9–12, with sizes ranging from 10 to 1400 μm. The extracted data were divided into three groups, viz. thoracic, inhalable and straw and modelled using machine learning algorithms, including support vector machine (SVM) and k-nearest neighbor. With an accuracy of 96%, SVM outperformed the other methods for categorising dust and straw particles, whereas linear discriminant analysis performed poorly with an accuracy of 88%.

Keywords: Agriculture, combine harvester, dust and straw particles, image processing, machine learning.

DUST consists of solid particles ranging from less than 1 to at least 100 μm that can be or become airborne depending on their source, physical features and ambient circumstances[1]. India is the second largest wheat-producing country in the world after Russia. As of 2019, India was indeed the second-largest producer of wheat in the world, with a production of 103.6 million tonnes (ref. 2). In India, wheat harvesting was traditionally performed using a sickle, a hand tool. However, in the last two decades, combined harvester have been used for harvesting wheat. Presently, the total number of combine harvesters used on Indian farms is about 40,000 (ref. 3), with a potential of 4500–5000 yearly additions[4]. Combine harvesting of wheat produces an enormous amount of particulate matter due to the interaction of machine tools with soil and crop. The unit operations involving cutting, conveying, threshing and blowing the threshed straw also produce dust and straw particles. Moreover, harvesting wheat becomes difficult due to peak summer temperatures ranging between 40°C and 46°C with different constraints, i.e. minimum relative humidity, high wind velocity and high solar radiation. A cabin in the combine harvester helps prevent dust exposure by reducing the concentration of dust particles by filtering and restricting their aerial movement inside the cabin[5]. Kirkhorn and Garry[6] reported dust reduction from 2–20 $\mu g\ m^{-3}$ to 0.1–1 $\mu g\ m^{-3}$, but in countries like India, combined harvesters are not equipped with environmental control cabins. Thus combined operators are directly exposed to dust and other environmental conditions (Figures 1 and 2). Farmers operate the combine harvester at high temperatures and low relative humidity (RH) for long periods each day, in contrast to the suggested comfort zones of 18°–24°C temperature and 30–70% RH[7]. Dust and straw exposure, in addition to extreme weather conditions, causes operational issues. Many studies have demonstrated high dust concentration in farm operations, particularly during harvesting. During rotary tilling, wheat harvesting and haymaking, Nieuwenhuijsen *et al.*[8] found that PM_{10}, $PM_{2.5}$, and PM_1 concentrations were higher than the human exposure limits. The increased workload in harsh conditions increases the respiration rates as well as the risk of respiratory diseases in farmers[9–11]. Ekka *et al.*[12] evaluated the particulate matter exposure of combine harvester operators and found that PM_{10} and $PM_{2.5}$ were 37 and 8 times (daily) and 62 and 11 times (annually) above the permissible limits respectively. A machine learning (ML)-based aerosol

[†]Data will be available upon request due to privacy/ethical restrictions. Code will be available from the corresponding author upon request.
*For correspondence. (e-mail: himadriiasri@gmail.com)

categorization method was presented by Siomos *et al.*[13] using observations from a double monochromator Brewer spectrophotometer.

Particulate matter exposure has been linked to respiratory and cardiovascular diseases[14]. A non-immunological process has been linked to the release of histamine and leukotriene from lung tissues after inhaling dust, particularly grain dust. Acute bronchial constriction among farmers is considered to be caused due to this process[15]. Grain dust containing mould spores can result in 'farmer's lung', a potentially fatal and disabling disease caused by a hypersensitivity reaction to moulds or the compounds they produce[16]. Inflammation of the eyes, lungs and skin can also result from dust contact[17]. Behera *et al.*[15] found that 22% of farmers in India suffered from respiratory problems. Sticking dust particles and straw to the body surface, inhalation in the respiratory system, dust particles in the eyes and throat causing irritation, headaches, etc. are all issues confronted by the combine harvester operators.

Figure 1. Working environment of combine harvester operator.

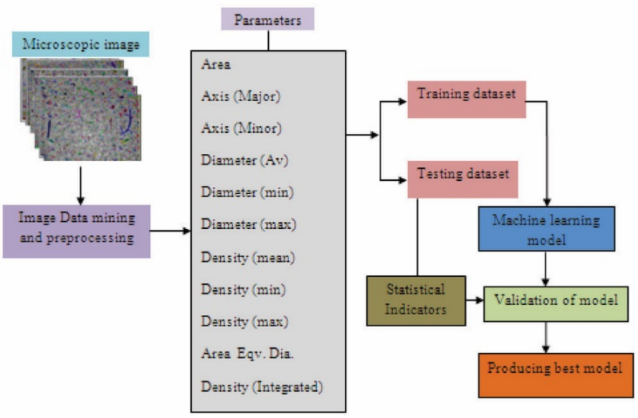

Figure 2. Diagrammatic representation of the classification of straw and dust particles.

ML algorithms can classify specific problems by examining the pattern of change in large datasets and predicting these data[18,19]. ML models, in essence, learn from previous exercises and calculations to provide more consistent and repeatable decisions and results[20]. In recent years, the use of ML for predicting real situations has gained popularity, particularly in soil science[21–23] and environmental science[24]. The most widely used supervised learning algorithms in studies related to the prediction of air quality[25], land susceptibility to erosion[26], dust sources[27] and dust storm index[28] include support vector machines (SVM), random forest (RF), naive Bayes (NB), decision tree, *k*-nearest neighbors (*k*-NN), extreme gradient boosting and Cubist.

ML techniques such as RF and neural networks are used to forecast particulate matter and identify the primary meteorological covariates. Czernecki *et al.*[29] estimated the concentration of PM_{10} and $PM_{2.5}$ at 11 urban air quality monitoring stations in Poland, including background, traffic and industrial sites, and used various ML techniques for analysis and cross-validation.

For forecasting ambient air pollutant trends, Lu and Wang[30] used SVM. Osowski and Garanty[31] used SVM and a wavelet decomposition procedure to forecast daily meteorological pollutants. Lee *et al.*[32] used satellite images to employ various ML-based algorithms for dust particle detection. To detect dust emissions on construction sites, Xiong and Tang[33] used several ML procedures and synthetic images. Friedl and Brodley[34] classified land-cover data using a decision tree algorithm for remote sensing data. With the proliferation of portable cameras and smartphones, images are playing an increasingly important role in information representation and description. If air-quality metrics like $PM_{2.5}$ and the air quality index (AQI) can be estimated from photographic images, they will provide an efficient and cost-effective method to monitor air quality using computer vision and ML techniques[35].

Few studies have assessed the characteristics of dust and wheat straw exposure of combine harvester operators in developing countries. The classification of particulate matter generated from a wheat combine harvester using ML has several advantages. First, wheat production is a crucial part of the agricultural sector, providing food and raw materials for a large part of the global population. However, wheat harvesting generates a significant amount of particulate matter, which can cause harm to human health and the environment.

Using ML to classify particulate matter generated from wheat combine harvesters, researchers can identify the different types of particles and their potential health and environmental impacts. This information can help policymakers and farmers develop and implement more effective strategies to reduce emissions and mitigate the negative effects of wheat harvesting on air quality.

ML can also help improve our understanding of the factors that influence the generation of particulate matter during wheat harvesting. Additionally, using ML to classify

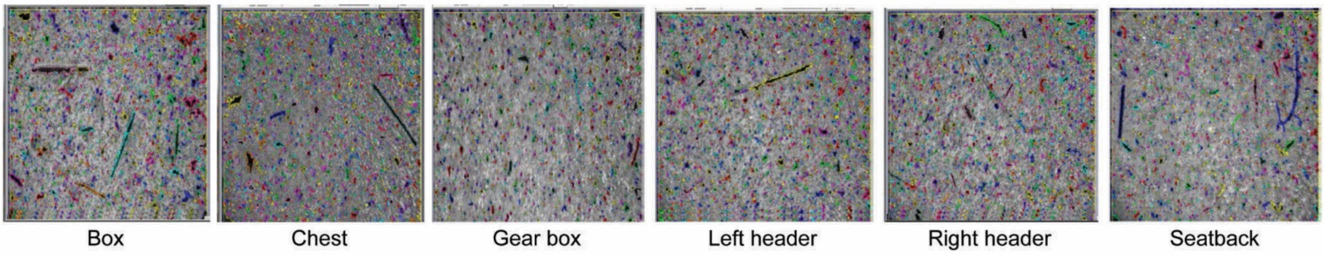

Figure 3. Microscopic image of particulate matter at different locations of the workspace.

particulate matter from wheat harvesting can help improve the accuracy and efficiency of current monitoring and measurement methods. Traditional methods for measuring particulate matter can be time-consuming and require specialized equipment and expertise. In contrast, ML algorithms can automate the process and provide more accurate and consistent results.

Overall, the classification of particulate matter generated from a wheat combine harvester using ML can help mitigate the negative impacts of wheat harvesting on air quality and public health while also providing valuable insights into the factors that influence particulate matter emissions. As a result, the present study was carried out to evaluate dust and straw particles in the workspace of a combine harvester operator during wheat harvesting.

Methodology

There are few reports on straw particles larger than 100 μm produced during wheat harvesting and their flow dynamics in combined harvester operation areas. The main objective of the current research was to determine the quantity of straw produced during the harvest and its distribution around the operator's workstation. For collecting wheat dust and straw samples, sticky paper measuring 29.7 cm × 21 cm was placed at six distinct locations in the workspace of a combine harvester operator (thoracic region of the operator, toolbox, gearbox, left header (LH) and right header (RH) and seatback). Due to the presence of adhesive in the sticky paper, the dust and straw particles stick to its surface. The experiments were conducted at the agricultural farm in the ICAR-Indian Agricultural Research Institute, New Delhi, that lasted for 15 min. During the experiment, the ambient temperature and RH were 41°C and 29% respectively. After the experiment was completed, the sticky paper was removed from the designated location and image-processing operations were performed to analyse the characteristics, type, size and shape of the particles. The sticky paper samples were also processed to determine particle size distribution and the number of particles per unit surface area. With the use of a microscopic camera, image vision technology was used to capture the image of the sticky paper. The total number of particles, the equivalent diameter of each particle, the major and minor axis of the particles and particle density were calculated using the Biovis particle size analyzer software.

The edge detection method was used to separate the particles. By calibrating the unit pixel length in the sample image and applying the measured pixel length to each particle, the dimensions of the particles were calculated. Due to the limitations of image processing technologies for identifying particles smaller than 10 μm, particles having more than 10 μm equivalent diameter were selected for the study. Digital picture acquisition, pre-processing, dimension calibration, image processing, analysis and evaluation were all part of the procedure. The edge-detection approach was used to separate the dust and straw particles and every adjacent edge considered as an individual particle. Several particles that were deposited over the sticky paper were counted using machine vision technology. Area, axis (major), axis (minor), diameter (minimum), diameter (average), diameter (maximum), density (mean), density (minimum), density (maximum), density (integrated) and area equivalent diameter of each particle were also measured. The determination of particle size distribution and its contribution in different size ranges, such as thoracic (10–20 μm), inhalable (20–100 μm) and straw particles (>100 μm), is important in understanding the potential health hazards associated with agricultural field operations. However, the high cost of sensors and their application in agricultural operations can be a significant challenge. As a result, computer vision was used to capture and distinguish each particle from the sample image, and ML was used to estimate the dust and straw particles and their share in different groups, such as thoracic, inhalable and straw particles (Figure 3).

Now, we summarise all the ML algorithms used to classify the dust and straw particles in this study.

Support vector machine

SVM is a supervised ML technique used to solve classification and regression problems. It is one of the most accurate ML algorithms because it is highly sophisticated and mathematically sound. The goal of the SVM algorithm is to identify a hyperplane in an n-dimensional space that categorizes data points. Each data point is represented as a point in an n-dimensional space, where each feature is one of the coordinates. The features include area, major axis, minor axis, minimum diameter, average diameter, maximum diameter, mean density, minimum density, maximum density, integrated density and area equivalent diameter. For two inputs and three features, the hyperplane is simply a line in the two-dimensional plane. This hyperplane with the greatest margin between classes is taken into account. These margins are calculated using support vectors.

The Support Vector Machine (SVM) algorithm is powerful machine learning technique that utilizes kernels to transform data into a higher-dimensional feature space. The use of kernels allows the algorithm to capture complex relationships between the data points that may not be apparent in the original feature space. It is known that the SVM algorithm produces a linear hyperplane. However, if the problem is nonlinear, the linear classifier sometimes fails. Therefore, the concept of kernel transformation is useful. By performing kernel transformation, a low-dimensional space is converted into a high-dimensional space where a linear hyperplane can easily classify the data points, thereby making SVM a de facto nonlinear classifier. Different types of kernels aid in the solution of various linear and nonlinear problems. Choosing these kernels is another hyper-parameter to deal with and tune appropriately.

KNN algorithm

k-NN is a straightforward, supervised ML algorithm that can be used to solve classification and regression problems[36]. The steps for solving the *k*-NN algorithm are given below.

(i) Set *k* to the desired number of neighbours.
(ii) Calculate the Euclidean distance of *k* number of neighbours.
(iii) Take the *k* closest neighbours based on the Euclidean distance.
(iv) Among these *k* neighbours, count the number of the data points in each category.
(v) Assign the new data points to that category for which the number of neighbours is maximum.
(vi) Pick the first *k* entries from the sorted collection.
(vii) For classification, return the mode of the *k* labels, and the model is ready to use.

Linear discriminant analysis

Linear discriminant analysis (LDA) is a technique for reducing dimensionality that is commonly used in supervised classification problems. It also models group differences, such as separating two or more classes. It is used to project features from higher to lower-dimension spaces. The steps of LDA are given below.

(i) Compute the *d*-dimensional mean vectors.
(ii) Compute the scatter matrices, both within-class and between-class.
(iii) Sort the eigenvectors by decreasing eigenvalues and chose eigenvectors with the largest eigenvalues to form a matrix (each column denotes an eigenvector).
(iv) Transform the matrix onto the new subspace.

Naive Bayes algorithm

NB algorithm is a probabilistic ML technique that can be applied to a wide range of classification applications. The NB classifier is a simple and effective classification method based on Bayes' theorem that aids in the development of fast ML models capable of making quick predictions[37]. Bayes' theorem can be defined as follows:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)},$$

where $P(A|B)$ is the posterior probability, $P(B|A)$ the likelihood probability, $P(A)$ the prior probability and $P(B)$ is the marginal probability.

Decision tree

Decision tree is a supervised learning technique that can be used to solve both classification and regression problems; however, it is most commonly employed to solve classification problems. Internal nodes represent dataset attributes, branches represent decision rules, and each leaf node provides the conclusion in this tree-structured classifier. A decision tree has two nodes: the decision node and the leaf node. Decision nodes are used to make a decision and have multiple branches, whereas leaf nodes are the output of those decisions and have no additional branches. It is a graphical representation of all possible solutions to a problem/decision given certain conditions.

Results and discussion

The boundary edge arrangement of the sticky paper was used to detect particle distribution. As shown in Figure 2, the dust particles were differentiated from each other by their irregular and unique size, represented by different random colours. A microscopic camera captured the average sample field area of the sticky paper, which was 270 mm^2. Experiments on particles in this area were considered. The total number of particles was calculated based on the closing of the boundary edge. It was found that the number of particles varied across all six working locations. In 15-min time frame, the load was 10, 12, 10, 10, 12 and 9/mm^2 with the total number of dust and straw particles of 2948, 3310, 2527, 2752, 3017 and 2430 in the box, chest, gearbox, LH, RH and seatback respectively.

There is a wide range of particles with equivalent diameters ranging from 10 to 1369.7 μm, 1325.1 μm, 815 μm,

Table 1. Particulate matter distribution

Tool box	Thoracic area	Gear box	Left header	Right header	Seatback
PMPUSA: 10 no./mm^2, AED: 10.9 (minimum), 1400 (maximum), Mean: 85.8, SD: 92.5	PMPUSA: 12 no./mm^2, AED: 10.6 (minimum), 1300 (maximum), Mean: 76.2, SD: 75.4	PMPUSA: 9 no./mm^2, AED: 10.5 (minimum), 800 (maximum), Mean: 78.8, SD: 68	PMPUSA: 10 no./mm^2, AED: 10.7 (minimum), 1300 (maximum), Mean: 87.9, SD: 77.2	PMPUSA: 12 no./mm^2, AED: 10.2 (minimum), 930 (maximum), Mean: 79.8, SD: 72.2	PMPUSA: 9 no./mm^2, AED: 10.8 (minimum), 1300 (maximum), Mean: 76.5, SD: 79.6

AED, Area equivalent diameter; PMPUSA, Particulate matter per unit surface area; SD, Standard deviation.

1312.2 μm, 936.4 μm and 1311.9 μm at the locations of the box, chest, gearbox, LR, RH and seatback respectively. These diameters contribute to varying percentages of thoracic, inhalable and straw particles. Due to the irregularity of particle size, large variation in size, and wide range of particle distribution, it is difficult to classify the particles. Therefore, parameters obtained from image processing methods were considered. Due to the varying speeds and directions of the wind and tractor, predicting a specific wind direction is difficult. As a result, a wind gust with dust and straw particles occurs inside the working area of the combine harvester. Due to high temperature and sweating, these particles adhere to the body surface of the operator and enter his eyes, nose and mouth, resulting in swelling and burning of the eyes and face. In the workspace, the number of particles per unit surface area was 12, 10, 9, 10, 12 and 9/mm^2. This is related to the number of particles that could adhere to an exposed body of the operator. According to the literature, the entire surface area of an adult male's body is 1.9 m^2. Given that the surface area of the hands, arms, and face is around 10% of this (0.2 m^2), the number of particles adhered to the skin is estimated to be 12×10^6. When the temperature increases above 40°C in summer, the straw particles act as irritants in the presence of sweat. The concentration of straw was found to be exceedingly high, irritating the skin and eyes of the operator as well as swelling his face.

Percentage of particle distribution in the workspace

Dust and straw particles are produced during the harvesting period as a result of soil–crop and machine interaction. The reciprocating cutting mechanism produces straw particles, which, combined with dirt and dust, migrate into the operator's workspace. The operator's eyes, throat, skin, as well as other exposed body parts, are seriously harmed by the sharp edges of the straw. In India, tractors and combine harvesters are not equipped with a closed chamber or cabin, exposing the operator to dust and straw. In this study, particles size greater than 10 μm were considered for analysis and separated using edge detection technology to identify each particle and straw particle distribution in the workspace of the combine harvester. The collected samples were divided into three groups: 10–20 μm (thoracic), 20–100 μm (inhalable) and >100 μm (straw). The percentage of inhalable dust was highest (65–71) at all workspace locations, according to the distribution. The percentage of straw with particle size greater than 100 μm was 21–29%. Thoracic particles accounted for roughly 6–8% of total particulate matter. At several places, large variations in particle equivalent diameter were recorded, ranging from 10 to 1400 μm. Particle size standard deviation in terms of equivalent diameter was recorded at different places as 92.5, 75.4, 68, 77.2, 77.2, 79.6 μm, with an average equivalent diameter of 85.8, 76.2, 78.8, 87.9, 79.8, 76.5 μm respectively (Table 1). The direction and speed of wind have a major impact on dust movement in the workspace. Furthermore, when the combine harvester moves across the farm, it comes into contact with the wind, contributing to particulate matter's movement. Pollen, fungal spores, fungal hyphae, mycotoxins, germs and endotoxins are transported by these particles to the workspace of the operator and are inhaled by him[38]. Grain dust exposure can cause a variety of acute and chronic respiratory symptoms, as well as a reduction in lung function[10,39]. Coughing, chest tightness and phlegm were among the most common impacts of dust particles, with 60% of smokers and non-smokers reporting coughing and 83% reporting phlegm, chest tightness and dyspnea. Non-smokers had a lower rate of coughing (47%) and a lower rate of chest discomfort (13%) than smokers[10]. During harvesting of wheat, both organic and inorganic dust is produced, causing allergic and non-allergic reactions[40]. The produced organic dust and other identified substances could cause health problems for the operator[16]. A small percentage of workers who are exposed may develop asthma as a result of the bacteria and dust mites or other components in the grain dust[41]. They may also have a non-allergic acute asthma-like response, which could be linked to endotoxin exposure[42]. Grain dust exposure has been demonstrated in studies to cause immediate conjunctival, nasal, respiratory and systemic symptoms[43].

Figure 4 shows the scatter plot for different variables used in this study. Figure 5 shows the observations of different groups, i.e. frequency of observations belonging to thoracic, inhalable and wheat straw groups. The collected dataset was pre-processed, scaled and divided into training and validation sets. Five ML approaches were used to evaluate the performance and validation of the models. The models were trained with 70% of the observations, and the accuracies were validated using a test sample. Table 2 shows the accuracy of different ML models, and the accuracies were 0.96, 0.93, 0.88, 0.89 and 0.94 for SVM, *k*-NN, LDA, NB and decision tree respectively. Under SVM, the

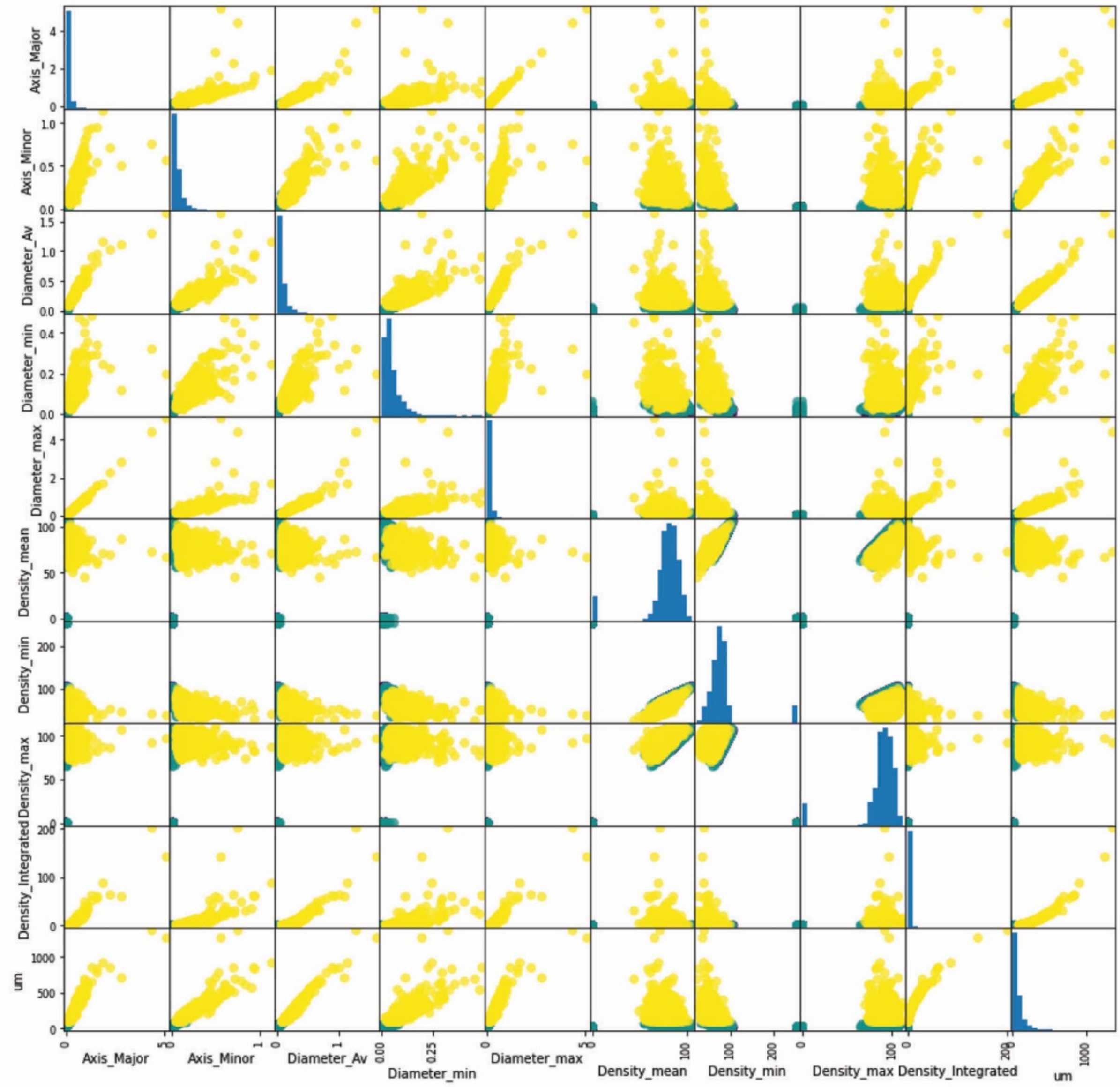

Figure 4. Scatter matrix plot for each variable.

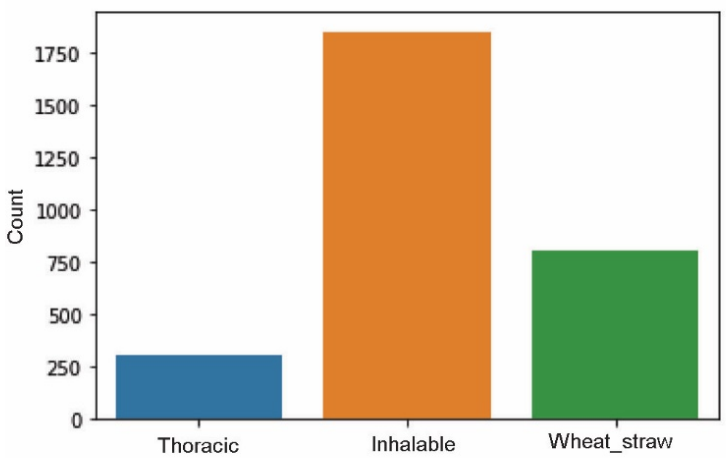

Figure 5. Plot of group variables.

Table 2. Comparison of different machine learning models

Experiment	Precision	Recall	f1-Score	Accuracy	
1	0.98	0.96	0.97	0.96	SVM
2	0.88	0.93	0.90		
3	0.96	0.98	0.97		
1	0.93	0.97	0.95	0.93	*k*-NN
2	0.96	0.59	0.73		
3	0.94	0.97	0.96		
1	0.85	0.99	0.92	0.88	Linear discriminant analysis
2	0.87	0.45	0.59		
3	1.00	0.79	0.88		
1	0.86	0.98	0.91	0.89	Gaussian naive Bayes
2	0.00	0.00	0.00		
3	0.95	0.99	0.97		
1	0.92	1.00	0.96	0.94	Decision tree
2	1.00	0.48	0.65		
3	0.99	1.00	1.00		

radial basis function is the most popular kernel function, which shows better fit and accuracy. Along with accuracy, precision, recall and f1-score were also calculated to validate the models properly. The higher the precision, the better the capacity to discriminate, and the recall rate is the ability to identify dust particles. SVM performed better than the other models in identifying dust particles.

Conclusion

Collecting samples during wheat harvesting in developing nations has been a challenge due to extreme weather conditions and the dynamic nature of the operation. However, a novel approach for collecting dust and straw particles during wheat harvesting has been developed here using sticky paper. The collected samples were examined using an image processing technique to assess the level of dust and straw exposure of the combine harvester operators.

The results of this study show that the number of dust and straw particles deposited per square millimetre ranges from 9 to 12, with particle sizes ranging from 10 to 1400. The extracted data were categorized into three groups and modelled using several ML approaches. Precision, recall, f1-score and accuracy were estimated for each method, and the results showed that SVM outperformed the other methods for categorization of dust and straw particles with an accuracy of 96%.

These findings are particularly relevant to developing nations where few studies have been conducted on dust and straw exposure during wheat harvesting. By utilizing this novel approach of collecting dust and straw particles, policymakers and researchers can make more informed decisions about the health and safety of combined harvester operators.

1. Berglund, B., Lindvall, T. and Schwela, D. H., *World Health Organization Occupational and Environmental Health Team. Guidelines for Community Noise*, World Health Organization, Geneva, Switzerland, 1999.
2. Devi, M., Kumar, J., Malik, D. P. and Mishra, P., Forecasting of wheat production in Haryana using hybrid time series model. *J. Agric. Food Res.*, 2021, **5**, 100175.
3. Singh, S. P. and Singh, S., Farm power availability and its perspective in Indian agriculture. *RASSA J. Sci. Soc.*, 2021, **3**(2), 114–126.
4. Mehta, C. R., Chandel, N. S., Jena, P. C. and Jha, A., Indian agriculture counting on farm mechanization. *Agric. Mech. Asia, Afr. Latin Am.*, 2019, **50**(1), 84–89.
5. Zander, J., Ergonomics in machine design: a case-study of the self-propelled combine harvester, Doctoral dissertation, Wageningen University of Research, Wageningen, 1972.
6. Kirkhorn, S. R. and Garry, V. F., Agricultural lung diseases. *Environ. Health Perspect.*, 2000, **108**(4), 705–712.
7. Suggs, C. W., *Thermal Environment of Agricultural Workers: Environmental Stress*, Module 10, ASAE, Michigan, USA, 1991.
8. Nieuwenhuijsen, M. J., Noderer, K. S., Schenker, M. B., Vallyathan, V. and Olenchock, S., Personal exposure to dust endotoxin and crystalline silica in California agriculture. *Ann. Occup. Hyg.*, 1999, **43**(1), 35–42.
9. Atiemo, M. A., Yoshida, K. and Zoerb, G. C., Dust measurements in tractor and combine cabs. *Trans. ASAE*, 1980, **23**(3), 0571–0576.
10. Becklake, M. R., Grain dust and lung health: not just a nuisance dust. *Can. Respir. J.*, 2007, **14**(7), 423–425.
11. Zock, J. P., Heederik, D. and Hans, K., Exposure to dust, endotoxin and micro-organisms in the potato processing industry. *Ann. Occup. Hyg.*, 1995, **39**(6), 841–854.
12. Ekka, U., Kumar, A. and Roy, H. S., Particulate matter exposure of combine harvester operator during wheat harvesting in northern India. *Indian J. Agric. Sci.*, 2021, **91**(5), 678–682.
13. Siomos, N., Fountoulakis, I., Natsis, A., Drosoglou, T. and Bais, A., Automated aerosol classification from spectral UV measurements using machine learning clustering. *Remote Sensing*, 2020, **12**(6), 965.
14. Du, Y., Xu, X., Chu, M., Guo, Y. and Wang, J., Air particulate matter and cardiovascular disease: the epidemiological, biomedical and clinical evidence. *J. Thorac. Dis.*, 2016, **8**(1), E8.
15. Behera, D., Pal, D. and Gupta, D., Respiratory symptoms among farmers in the vicinity of a north Indian city. *Lung India*, 2005, **22**, 45–49.
16. Swan, J. R. M., Blainey, D. and Crook, B., The HSE grain dust study – workers' exposure to grain dust contaminants, immunological and clinical response. Health and Safety Executive Research Report, 2007, p. 540.

17. Matthews, J. and Knight, A. A., *Ergonomics in Agricultural Equipment Design*, National Institute of Agricultural Engineering, Silsoe, UK, 1973.
18. L'heureux, A., Grolinger, K., Elyamany, H. F. and Capretz, M. A., Machine learning with big data: challenges and approaches. *IEEE Access*, 2017, **5**, 7776–7797.
19. Qiu, J., Wu, Q., Ding, G., Xu, Y. and Feng, S., A survey of machine learning for big data processing. *EURASIP J. Adv. Signal Process.*, 2016, **2016**(1), 1–16.
20. Shalev-Shwartz, S. and Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, New York, USA, 2014.
21. Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. and Fernandes Filho, E. I., Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 2019, **340**, 337–350.
22. Keskin, H., Grunwald, S. and Harris, W. G., Digital mapping of soil carbon fractions with machine learning. *Geoderma*, 2019, **339**, 40–58.
23. Khaledian, Y. and Miller, B. A., Selecting appropriate machine learning methods for digital soil mapping. *Appl. Math. Model.*, 2020, **81**, 401–418.
24. Tong, W., Machine learning for spatiotemporal big data in air pollution. In *Spatiotemporal Analysis of Air Pollution and Its Application in Public Health*, Elsevier, Amsterdam, The Netherlands, 2020, pp. 107–134.
25. Masood, A. and Ahmad, K., A model for particulate matter (PM2.5) prediction for Delhi based on machine learning approaches. *Procedia Comput. Sci.*, 2020, **167**, 2101–2110.
26. Pham, B. T., Prakash, I., Singh, S. K., Shirzadi, A., Shahabi, H. and Bui, D. T., Landslide susceptibility modeling using reduced error pruning trees and different ensemble techniques: hybrid machine learning approaches. *Catena*, 2019, **175**, 203–218.
27. Rahmati, O. *et al.*, Hybridized neural fuzzy ensembles for dust source modeling and prediction. *Atmos. Environ.*, 2020, **224**, 117320.
28. Ebrahimi, K. Z., Roustaei, F., Khusfi, M. E. and Naghavi, S., Investigation of the relationship between dust storm index, climatic parameters, and normalized difference vegetation index using the ridge regression method in arid regions of Central Iran. *Arid Land Res. Manage*., 2020, **34**, 239–263.
29. Czernecki, B., Marosz, M. and Jędruszkiewicz, J., Assessment of machine learning algorithms in short-term forecasting of PM10 and PM2.5 concentrations in selected polish agglomerations. *Aerosol Air Qual. Res.*, 2021, **21**, 200586.
30. Lu, W. Z. and Wang, W. J., Potential assessment of the 'support vector machine' method in forecasting ambient air pollutant trends. *Chemosphere*, 2005, **59**(5), 693–701.
31. Osowski, S. and Garanty, K., Forecasting of the daily meteorological pollution using wavelets and support vector machine. *Eng. Appl. Artif. Intell.*, 2007, **20**(6), 745–755.
32. Lee, J., Shi, Y. R., Cai, C., Ciren, P., Wang, J., Gangopadhyay, A. and Zhang, Z., Machine learning based algorithms for global dust aerosol detection from satellite images: inter-comparisons and evaluation. *Remote Sensing*, 2021, **13**(3), 456.
33. Xiong, R. and Tang, P., Machine learning using synthetic images for detecting dust emissions on construction sites. *Smart Sustain. Built Environ.*, 2021, **10**(3), 487–503.
34. Friedl, M. A. and Brodley, C. E., Decision tree classification of land cover from remotely sensed data. *Remote Sensing Environ.*, 1997, **61**(3), 399–409.
35. Chakma, A., Vizena, B., Cao, T., Lin, J. and Zhang, J., *Image-based Air Quality Analysis using Deep Convolutional Neural Network*, International Conference on image processing, Institute of Electrical and Electronics Engineers, Beijing, China, 2017, pp. 3949–3952.
36. Guo, G., Wang, H., Bell, D., Bi, Y. and Greer, K., KNN model-based approach in classification. In OTM Confederated International Conferences: On the Move to Meaningful Internet Systems, Springer, Berlin, Germany, 2003, pp. 986–996.
37. Zhang, M. L., Peña, J. M. and Robles, V., Feature selection for multi-label naive Bayes classification. *Inf. Sci.*, 2009, **179**(19), 3218–3229.
38. Spankie, S. and Cherrie, J. W., Exposure to grain dust in Great Britain. *Ann. Occup. Hyg.*, 2012, **56**(1), 25–36.
39. Huy, T., De Schipper, K., Chan-Yeung, M. and Kennedy, S. M., Grain dust and lung function. Dose–response relationships. *Am. Rev. Resp. Dis.*, 1991, **144**(6), 1314–1321.
40. Baker, J. B., Southard, R. J. and Mitchell, J. P., Agricultural dust production in standard and conservation tillage systems in the San Joaquin Valley. *J. Environ. Qual.*, 2005, **34**, 1260–1269.
41. Armentia, A., Martinez, A. and Castrodeza, R., Occupational allergic disease in cereal workers by stored grain pests. *J. Asthma*, 1997, **34**, 369–378.
42. Chan, Y. M., Dimich, W. H. and Enarson, D. A., Five cross-sectional studies of grain elevator workers. *Am. J. Ind. Med.*, 1992, **136**, 1269–1279.
43. Chan-Yeung, M., Schulzer, M., MacLean, L., Dorken, E. and Grzybowski, S., Epidemiologic health survey of grain elevator workers in British Columbia. *Am. Rev. Resp. Dis.*, 1980, **121**(2), 329–338.

Received 12 July 2022; revised accepted 16 February 2023

doi: 10.18520/cs/v124/i9/1074-1081