Classification of municipal solid waste using deep convolutional neural network model applied to multispectral images

Harald Ian D.I. Muri[a] and Dag Roar Hjelme[a]

[a]Norwegian University of Science and Technology, O.S. Bragstads plass 2b, Trondheim, Norway

ABSTRACT

Minimization of the environmental impact of the incineration process and to produce energy efficiently are the most important considerations in obtaining efficient operation of waste-to-energy (WtE) plants. WtE operation can obtain significant improvements by predicting combustion properties of municipal solid waste (MSW) prior to incineration. Combustion properties of MSW can be assessed by estimating the weighted waste fractions such as paper and cardboard, plastic or inert and fines. Waste materials and fractions can be recognized using imaging techniques and image classification methods based on deep convolution neural networks (CNN). We have tested a new sensor system for image classification based on using multispectral (MS) images and deep CNN pretrained on the ImageNet database to recognize MSW categories. MS camera was used for sampling images above the walking floor of a WtE plant (StatKraft Varme Tiller, Trondheim). The waste load was automatically registered as industrial or household waste at the time of delivery. The MS images from 49 waste loads were used to perform transfer-learning on EfficientNetB0 model weighted with ImageNet-NoisyStudent parameters. Using the predefined classes, a test and training set were generated from the 49 waste loads delivered between June and September (2020). The training set consisted of 35 waste loads while the remaining 14 waste loads were used as test. The weights for the image feature extraction was constant during training while the fully connected layer (top-layer) was updated for each epoch. The model performance on the test set was assessed by making predictions on the household or industrial waste images. With a fixed threshold value at 0.5, the model showed 85% accuracy, 92% precision, 89% recall and 90% F-measure for industrial class, while for household class the model showed 80% accuracy, 94% precision, 81% recall and 87% F-measure. For all threshold values, the area under curve estimated from the receiver operating characteristic plot showed that the model has 87% confidence in distinguishing household waste images from industrial waste images and 90% confidence in distinguishing industrial waste images from household waste images.

Keywords: material characterization, municipal solid waste, waste to energy, machine vision, multispectral imaging, deep convolution neural network, image classification, field testing

1. INTRODUCTION

Municipal solid waste (MSW) materials that are highly crude and decrepit are difficult to be reused, refurbished and recycled. Last option for handling such type of materials are landfilling or incineration on a waste-to-energy plant (WtE). The efficiency of WtE plant operation is paramount to both produce energy efficiently and to minimize the environmental impact of the incineration process. The efficiency of the operation is highly dependent on moisture in MSW materials as well as waste fraction categories such as plastic, paper and cardboard or inserts and fines.[1–3] The waste mixtures and moisture content are difficult to characterize and are changing continuously. For off-line characterization bomb calorimeters or waste fraction quantification are used but are not usable for on-line monitoring.[1,2,4,5] New on-line monitoring methods are required to provide real time information on combustion properties of MSW for improving the WtE performance. Imaging methods are often used for recognizing material components in various classification tasks and should be evaluated for on-line waste monitoring considering that the operation of the WtE plant is highly dependent on waste fraction

Further author information: (Send correspondence to Harald Ian Muri)
Harald Ian Muri: E-mail: harald.muri@ntnu.no, Telephone: +4773412688
Dag Roar Hjelme: E-mail: dag.hjelme@ntnu.no, Telephone: +4773559604

categories.[6–8] Of particular interest are deep convolution neural network (CNN) methods for image classification that have proved to outperform traditional machine-learning techniques such as logistic regression or support vector machines.[8–11]

Image classification with deep CNN is a supervised machine learning method and is based on applying image convolutions with learnable filters to extract features relevant for image categories.[12–15] The feature maps obtained from using learnable filters represents the convolution layer with weight parameters to be updated during the training. The convolution layers are applied in series combined with rectified linear unit (ReLu) activation functions to contain non-linearity and with pooling functions to reduce the dimensionality. The image features extracted using architectures of convolution layers, ReLu, and pooling functions are further passed to the fully connected layers for mapping features to corresponding image categories. During training, the weights of the image feature extraction is updated as well as the weights for mapping extracted features to categories. The training of feature extraction and classification weights requires large amounts of categorized data and significant amounts of computational power to obtain acceptable training times with acceptable model performance.[12–15] Model training can be achieved with less training data, less computational power and smaller training times by performing transfer-learning on a model that have been trained for a similar classification task.[16,17] For transfer-learning, the image feature extraction can be performed with fixed weights on a new image data-set while the weights used for mapping features to classes are updated. A vast amount of pretrained models for image classification tasks can be found on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) that is arranged annually.[18] Deep CNN architectures are submitted to ILSVRC to compete on classifying images on the ImageNet database correctly. The ImageNet database contain more than 14 million images and 20000 categories and the 20 most accurate models submitted have achieved more than 85% accuracy.[15,19]

In this paper, we have developed a sensor system for characterizing MSW based on image classification using multispectral (MS) camera and EfficientNetB0 CNN pretrained on the ImageNet database to recognize MSW categories.[20,21] The images were sampled using MS camera above the walking floor of a WtE plant (StatKraft Varme Tiller, Norway). At the time of delivery, waste loads are automatically registered as household or industrial waste. MS images with 16 wavelength channels of 49 waste loads were used to perform transfer-learning on a pretrained EfficientNetB0 model weighted with ImageNet-NoisyStudent parameters.[20,22,23] A test and training-set were generated from the 49 waste loads delivered between June and September (2020). Images from 35 waste loads were used for training and the remaining 14 waste loads (7 loads of each industrial and household waste) were used for testing. Using the training set, the image features were extracted by the model and the outputs were used for training the fully connected layer (top-layer) to map features to either household or industrial class. Next, the model performance was assessed by making predictions on the predefined household or industrial categorized MS images of the test-set. The threshold value (TV) can be tuned to optimize the decision making such as for controlling the rate of true or false alarms as well as the sensitivity for activating alarms relevant for improving the MSW incineration process. By estimating the area under curve (AUC) for the receiver operating characteristic (ROC) plot the model performance was assessed for all TVs. From AUC estimations the model showed 87% confidence in distinguishing household waste images from industrial waste images and 90% confidence in distinguishing industrial waste images from household waste images for all TVs.

2. MATERIALS AND METHODS

2.1 Data sampling

The MS imaging system and light sources were installed on a WtE plant (Statkraft Varme in Tiller, Trondheim) to sample images above the walking floor in the waste bunker. The imaging of waste materials were performed as shown in the top of Figure 1. The waste is delivered by a truck that can tilt the load sideways onto the walking floor and the walking floor transports the waste into the waste bunker. The waste delivered are categorized as either household or industrial waste. The time-log for waste loads corresponding to either household or industrial category are used for labeling images and to generate training and test-set for the CNN image recognition. For optimal monitoring the camera and light sources were placed on a support bridge that crosses the walking floor shown in the bottom of Figure 1.

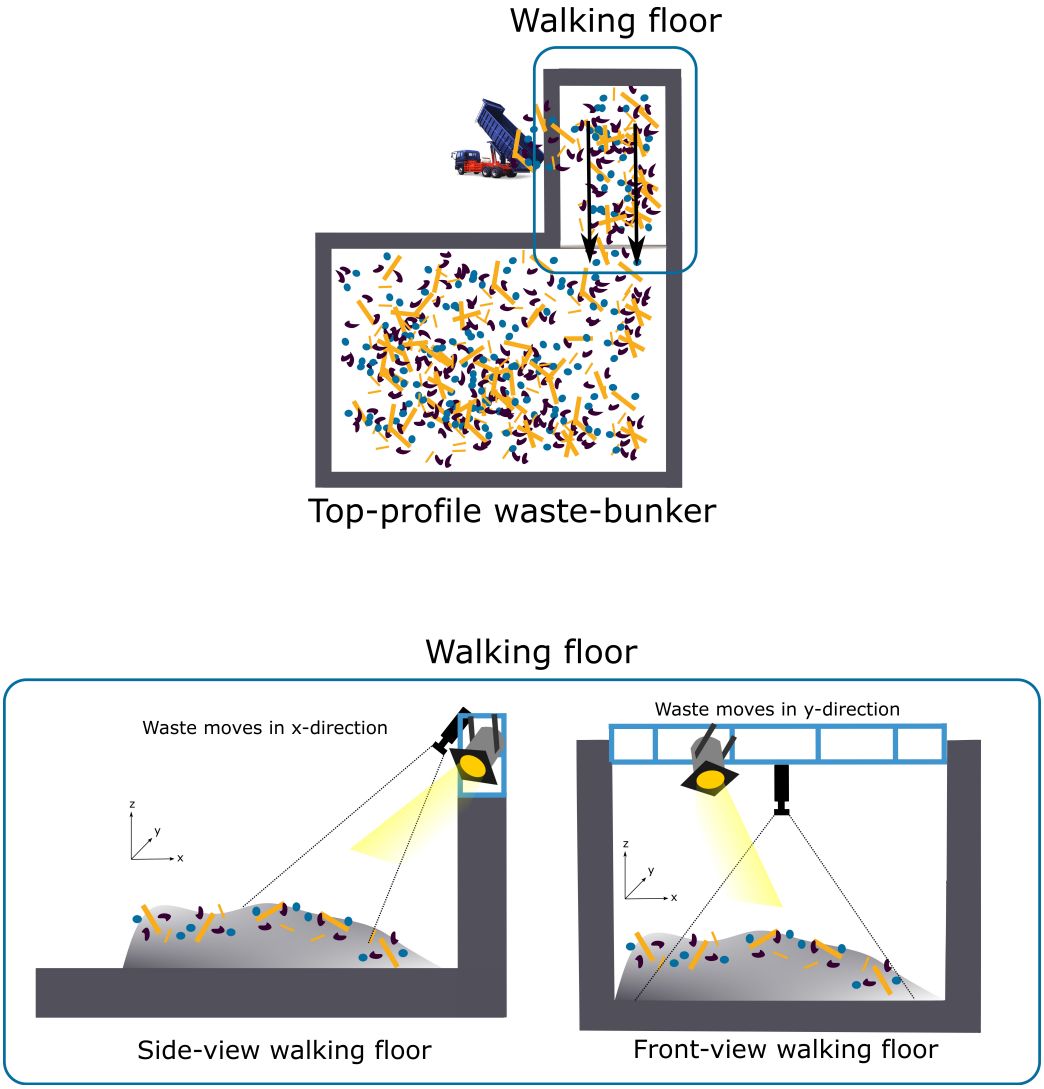

Figure 1. Top figure: Illustration of waste load delivered from a side-tipping truck onto the walking floor of waste bunker. Bottom figure: Illustration of imaging system and light source installed on support bridge used for sampling MS images above the walking floor.

To protect the MS camera it was placed inside a water and dust-proof housing (TVAC70200, IP68, Abus). The camera (Snapshot Mosaic Imager, IMEC) used is based on CMOSIS CMV2000 image sensor (Ximea xiQ camera) with a color filter of 16 wavelength channels in range 590-860 nm.[8,24] The total image resolution is 2048x1088 pixels and the 16 wavelength channels are encoded as mosaic patterns in the pixels similar to a Bayer masks with an image resolution of 512x272 pixels per band. A 16 mm lens (16mm VIS-NIR, TECHSPEC, Edmund Optics) was used with VIS-NIR anti-reflective coating and aperture at F/2.8 to satisfy optimal spectral imaging conditions as well as to obtain imaging with a field of view equal to the width of walking floor. The light sources used consist of a water and dust-proof xenon lamp (S XENON, IP68, 4500K, 35W, Vignal Group) and metal-halide lamp (ART. 8001-RAY/SM, IP65, 4200K,150W, TEC-MAR). As illustrated in Figure 2 the MS imaging was controlled with Raspberry Pi 4 (RPi 4, Okdo, RS Components AS) and the images were transferred to PC NTNU Database using a 4G router (RUT240, Teltonika). The RPi 4 and 4G router were protected inside a dust and water proof housing (Bocube, B 273612 PC-V0-G 7035, IP66, BOPLA).

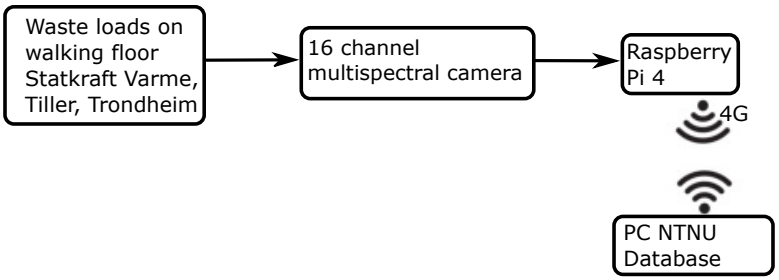

Figure 2. Schematic of image sampling above the walking floor.

With RPi 4, open source packages in python (ximea, xiapi, cv2) were used for handling the camera settings and the image sampling. The gain used for the imaging was constant while exposure time was automatically optimized internally by the camera. The imaging acquisition was set to free run while the interval for image sampling was set to 20 seconds. The folder destination with sampled images on the RPi 4 was synchronized with a folder destination on PC NTNU Database using the open source software Syncthing to optimize the usage of the varying network speeds provided by the 4G router. From the data sampling, 3173 images were further used to perform transfer-learning and model performance assessment of the deep CNN. From the 3173 images, 1307 images was categorized as household waste and 1866 images were categorized as industrial waste.

2.2 Transfer-learning on deep CNN

The time registered for each waste load delivered was used for mapping multispectral images to household or industrial waste categories. The label encoded MS images were further used to perform transfer-learning on an image recognition model EfficientNetB0. The EfficientNetB0 model is pretrained on the ImageNet database consisting of more than 14 million images and 20000 categories. The classification task of the pretrained model on ImageNet is assumed to be similar to the task of classifying images of waste materials. The transfer-learning require less training data, less computational power and smaller training times to obtain an image classifier with sufficient performance.[16]

First, demosaicing was performed on the MS images to retrieve the 16 wavelength channels as 16 monochrome images (information on mosaic structure received from IMEC upon request). The 16 images containing one wavelength channel each had to be adapted for the EfficientNetB0 model pretrained on images with 3 channels. To obtain 3 channel images, MS image were reconstructed to several RGB images, where 16 channels were combined into 3 channels as (1,2,3), (2,3,4), (3,4,5), ...(14,15,16). The reconstruction of a single MS image resulted in 14 sub-images (RGB channels). Python was used for the demosaicing and the sub-image reconstruction. Next, a training and test-set were generated from the MS images of the 49 waste loads delivered. 71% (35 waste loads) were used for training and 29% (14 waste loads) were used for test. During the model training, a validation-set composed of 50% of the test-set (7 waste loads) was also used. The training of the EfficientNetB0 model was performed as illustrated in Figure 3 based on using Keras and TensorFlow packages in python.[25,26]

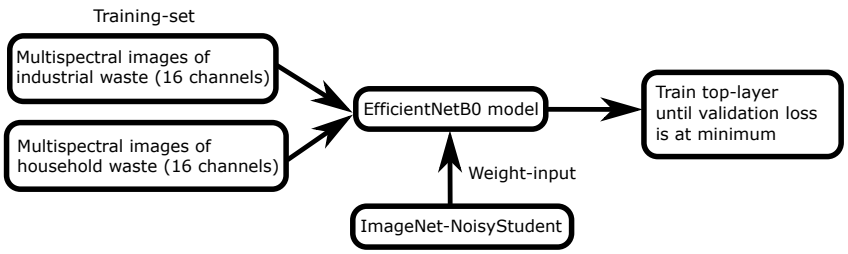

Figure 3. Schematic of the CNN training procedure where training is stopped when validation loss it at minimum (early stopping).

MS images and sub-images encoded in the industrial or household class were passed to EfficientNetB0 model weighted with ImageNet-NoisyStudent (available on the official repository of the EfficientNet-GitHub) produced from the pre-training on the ImageNet database.[21–23] For our procedure the weights used for the image feature extraction were not changed and only the fully connected layer (top-layer) were trained to map image features to either a household or industrial waste category. For each epoch the validation loss was estimated during model training. The validation loss was used for early stopping to optimally balance the under and over-fitting of the model.[27,28] In early stopping the training continues up to the next epoch where the validation loss is at minimum.[29] Before the validation loss is at minimum the model is considered to be under-fitting whereas with validation loss reaching minimum and after increasing the model is considered to be over-fitting. The following parameters were used for the early stop training; image size: 512x512 (sub-images resized using bilinear interpolation method), batch size: 64, data augmentation: (random rotation ±15%, translation ±10%, flip and contrast ±10%), drop-out: 0.4, activation function: softmax, adam learning rate: 0.01, loss function: categorical cross-entropy. Hence, data augmentation is performed for the training set but not for the validation or test-set. After the early-stopping the trained model exhibited a training loss of 0.49 and accuracy of 0.77 with validation loss of 0.42 and accuracy of 0.81.

For the model performance evaluation the prediction scores on the test-set were collected as illustrated in Figure 4.

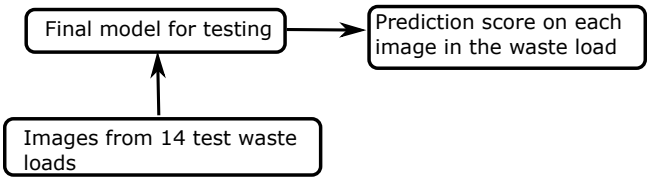

Figure 4. Schematic of the collection of prediction scores of the MS images and sub-images in the test-set.

The MS images and sub-images of the 14 waste loads were passed to the trained model. The model produces a prediction score on each sub-image as a vector $[h, i]$ where, h, is the fraction of image predicted to be household and, i, is the fraction of image predicted to be industrial waste. Next, the distribution of predictions scores, the confusion matrix, and the ROC curve were estimated.

2.2.1 Model performance assessment

For the model performance assessment, the confusion matrix and the ROC curve were used. The confusion matrix can be illustrated as in Figure 5 containing different amounts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN).[30]

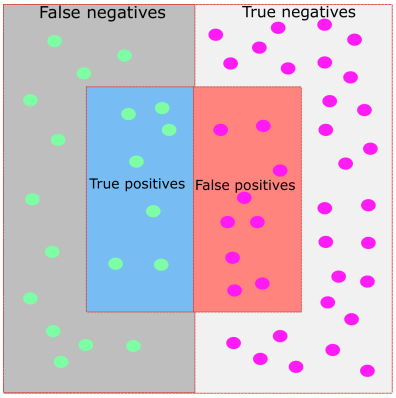

Figure 5. Illustration of the confusion matrix with different amounts of true positives, true negatives, false positives, and false negatives.

TP, TN, FP and FN can further be used for producing model performance parameters as accuracy, precision, recall and F-measure for a fixed TV.[31]

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F-measure} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

To assess the model performance for all TVs, T, the receiver operating characteristic (ROC) curve can be estimated where true positive rate (TPR) is plotted as a function of false positive rate (FPR). The TPR and FPR can be expressed as,

$$TPR(T) = \frac{TP(T)}{TP(T) + FN(T)} \tag{5}$$

$$FPR(T) = \frac{FP(T)}{FP(T) + TN(T)} \tag{6}$$

The ROC curve can be used for observing the trade-of between TPR and FPR for all TVs between zero and one. The ability of a model to distinguish between classes for all TVs can be performed by estimating the AUC of the ROC curve.[32] The AUC can be expressed using the trapezoid rule as,

$$AUC = \sum_{i=1}^{n} \frac{TPR(T_{i-1}) + TPR(T_i)}{2} \Delta FPR(T_i) \tag{7}$$

where T_i is a TV in a vector, T, with length n. For the model performance assessment, ROC curve and AUC were estimated with TVs between zero and one and with incremental steps of 0.001.

3. RESULTS AND DISCUSSION

The predicting accuracy of the model can be shown by observing the distribution of prediction scores and the confusion matrix as a function of TV. The distribution of positive prediction scores for industrial waste (PPI) and negative predictions scores for household waste (NPH) were estimated of test-set as shown in left side of Figure 6. Using histogram with bin-width at 0.05 the counts of prediction scores were normalized between zero and one. The distribution for PPI is rapidly increasing and NPH is rapidly decreasing with increasing predictions. The exponential-like distribution of prediction scores are expected since a softmax activation function is used for normalizing the neural network output between zero and one. A TV can be used to visualize the confusion matrix as shown in right side of Figure 6.

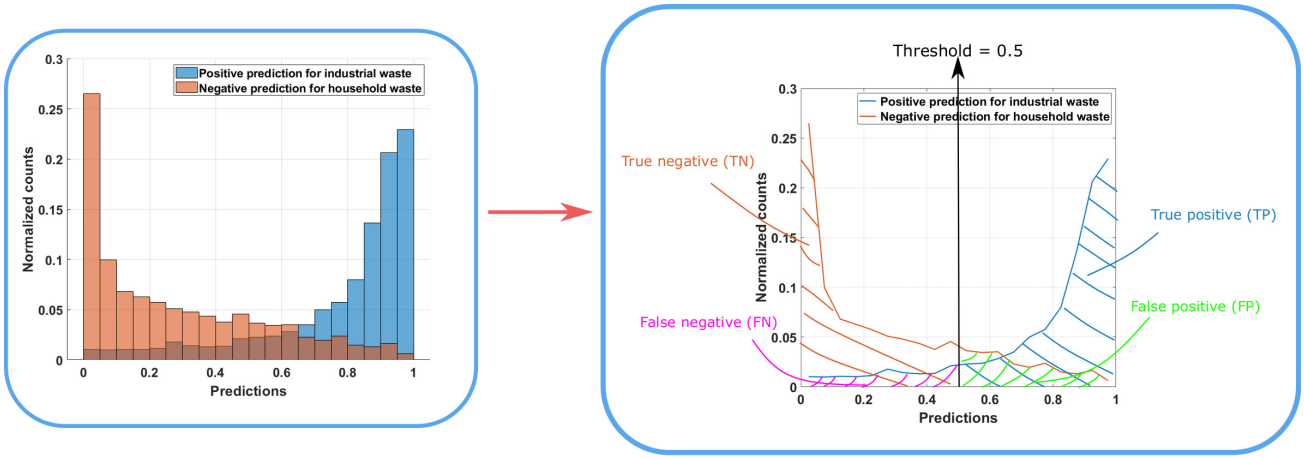

Figure 6. Left side: Distribution of positive and negative predicted score of industrial and household waste, respectively. Right side: Threshold value applied discriminating the distribution of positive and negative prediction scores of industrial and household waste, respectively. True positive colored as blue, false negative colored as violet, true negative colored as red, and false positives colored as green.

Using a TV of 0.5 the PPI and NPH prediction scores can be discriminated where TP, FP, TN and FN are shaded as blue, green, red, and violet, respectively. By observing the color shaded areas, the model accuracy can be represented as the red and blue ratio to the total color shaded area, precision can be represented as the blue ratio to the blue and green shaded area, and recall can be represented as the blue ratio to the blue and violet shaded area.

To assess model performance for each waste load, mean predicted can be estimated for the positive scores for industrial and household waste load as shown in Figure 7. The mean predicted score is estimated for each waste load and the load number is chronologically arranged (test-set from 49 waste loads). The mean predicted for industrial waste load 2, 9, 15, 20, 42 and 47 varies between 0.88 and 0.71 while waste load 25 has a mean predicted at 0.48. The mean predicted for household waste load 26, 31, 32, 35, 36, 40 are varying between 0.82 and 0.67 while waste load 18 has a mean predicted at 0.52. The drop in mean predicted score for waste load 25 and 18 could be due to image features shared between the industrial and household class.

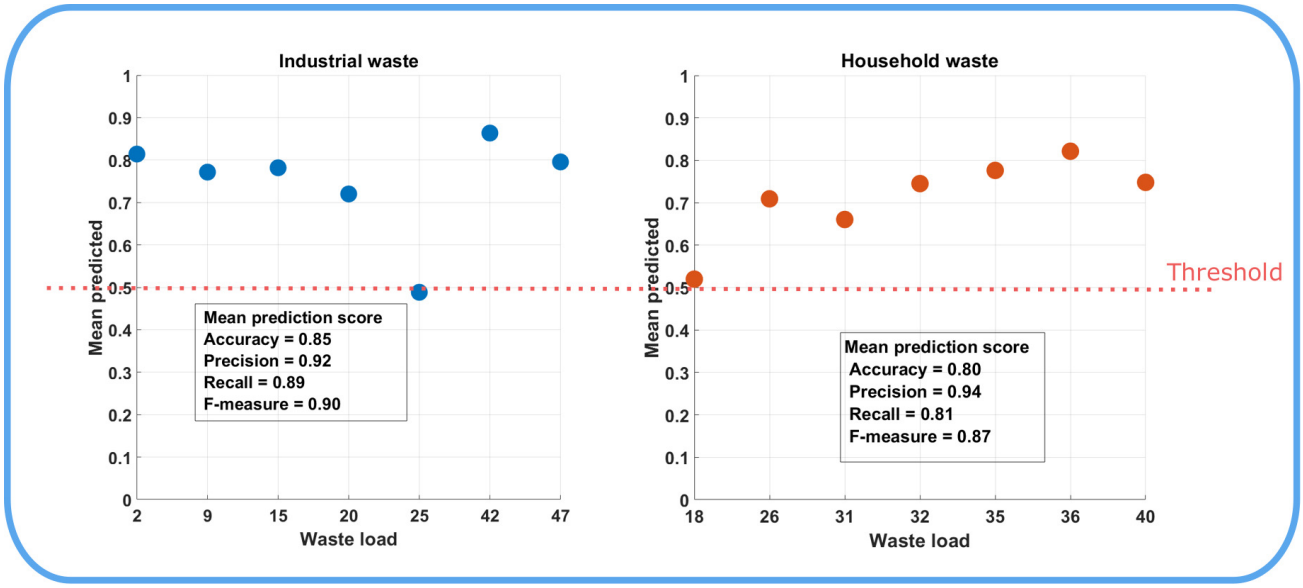

Figure 7. Mean predicted is estimated for positive scores for industrial and household waste loads with chronologically arrangement (test-set from 49 waste loads). Accuracy, precision, recall, and F-measure are estimated for the MS images and sub-images with threshold value at 0.5.

With a TV at 0.5 the model outputs one false negative for the industrial class and zero false negative for household class. Considering that there are only 14 waste loads available for the assessment the use of confusion matrix will not be sufficient for assessing the model performance in classifying waste loads. Using confusion matrix for assessing model performance in classifying images will be sufficient since there are 920 MS images in the test-set (29% of total 3173 MS images). As noted in Figure 7, for industrial waste class with TV at 0.5, the accuracy, precision, recall and F-measure were found to be 0.85, 0.92, 0.89 and 0.90, respectively. For household waste class with TV at 0.5, the accuracy, precision, recall and F-measure were found to be 0.8, 0.94, 0.81 and 0.87, respectively. The accuracy, precision, recall and F-measure values can be interpreted as follows. Accuracy; true positives and negatives are correctly predicted with a higher certainty for the industrial than for the household class. Precision; for positive predictions of industrial or household class, true positives are correctly predicted with a higher certainty for the household than for the industrial class. Recall: all classes are correctly predicted with a higher certainty for the industrial than for the household class. F-measure; the model has a higher confidence in distinguishing industrial waste images from household waste images than distinguishing household waste images from industrial waste images.

The threshold parameter set for the classification task is dependent on the decision making performed. To evaluate the model performance for different TVs it is possible to plot the ROC curve to observe the trade-off between TPR and FPR. The ROC curve for the industrial and household class are presented in Figure 8 and shows the TPR as a function of FPR for TVs between 0 and 1 with incremental steps of 0.001.

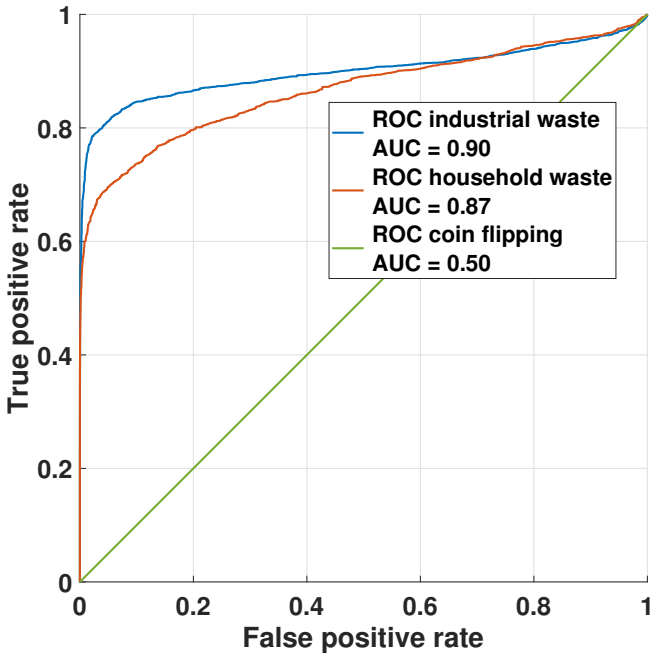

Figure 8. True positive rate as a function of false positive rate for threshold values between 0 and 1 with incremental steps of 0.001

For comparison, the ROC curve for a model without any discrimination capacity (equal to coin-flipping) is included as a linear plot. For the industrial class, the increase in TV more than 0.5 (TPR=0.86, FPR=0.21) reduces FPR with small cost of a decrease in TPR up to the point where the slope of the curve is high (TPR=0.81, FPR=0.05). For the household class, the change in TV lower or higher than 0.5 (TPR=0.7, FPR=0.04) will result in substantial changes both in FPR and TPR. The tuning of TV to optimize the trade-off between TPR and FPR depends on the decision making that is performed. If a high FPR (with many false alarms) have substantial negative impact on the waste combustion process performance, the increase in TV could be used to reduce false alarms. In view of the ROC curve results, with a low FPR preferred over small TPR to improve the process, the increase in TV with respect to the industrial class is feasible.

Similarly, if a low TPR (with low sensitivity for producing an alarm) have substantial negative impact on the waste combustion performance, the decrease in TV could be used. Concerning the ROC curve results, with a high TPR preferred over a small FPR to improve the process, the decrease in TV with respect to the household class is feasible, however with the cost of producing more false alarms due to the substantial increase in FPR.

The AUC of the ROC curve for industrial and household class are presented in Figure 8 and are compared with the AUC for a model without any discrimination capacity (coin-flipping). The AUC for the model is 0.90 for industrial class and 0.87 for the household class. The AUC values are equal to the F-measure values but are used differently for assessing the model. F-measure is estimated for a fixed TV whereas the AUC estimation is a result of using all TVs. Hence, the AUC shows that for all TVs the model has 90% confidence in distinguishing between industrial wast images from household images and 87% confidence in distinguishing household waste images from industrial waste images.

4. CONCLUSION

The WtE operation efficiency is dependent on the MSW weighted fractions of plastic, paper and cardboard or inerts and fines. Imaging methods can be used for recognizing material components in MSW and provide waste fractions estimations to improve the WtE performance. The waste material recognition was performed by using

a sensor system based on multispectral imaging and a pretrained (ImageNet-NoisyStudent weights) CNN model (EfficientNetB0) to discriminate between household and industrial waste categories. The prelimenary results on image classification for MSW characterization were assessed from observing the prediction scores for fixed and variable TVs. With fixed TV at 0.5, the prediction performance on the test-set for MS images of industrial and household waste were found as follows. Industrial class: accuracy at 85%, precision at 92%, recall at 89%, and F-measure at 90%. Household class: accuracy at 80%, precision at 94%, recall at 81%, and F-measure at 87%. The model has a higher accuracy, recall, and F-measure for the industrial class than for the household class, while the household class has a higher precision than the industrial class. In addition, the F-measure obtained using fixed TV shows that the model has a higher confidence in distinguishing industrial waste images from household images than from distinguishing household waste images from industrial waste images. For variable TVs the model performance was assessed based on ROC estimations. From the ROC curve, the AUC was estimated to be 0.90 and 0.87 for the industrial class and the household class, respectively. Thus, for all TVs the model shows to have 90% confidence in distinguish industrial waste images from household waste images and 87% confidence in distinguishing household waste images from industrial waste images. The main advantage of using this model for image classification is that the feature representation learned from waste images is not limited by the human visual perception due to the spectral imaging performed in the visible and near-infrared wavelengths. The higher order image features are extracted for the 16 spectral domains (in the range 590-860 nm) by the model and may contain a holistic information on the waste images that is not recognizable by human perception due the spectral information used. Thus, using MS images to train a deep learning model may be used for predicting waste classes related to MSW combustion properties even if the waste appears highly heterogeneous by the human vision.

Future work will consist of performing waste categorization and multi-class labeling of the images from the waste loads to recognize MSW fractions relevant for the WtE performance. The labeling strategy will be based on human vision of images of the waste loads assembled as videos to efficiently categorize images into waste fraction classes. Concerning model development, the tensor containing the (pretrained) weights will be extended so that the 16 channel spectral image can be passed to the deep neural network instead using sub-images with 3 channels. This will increase efficiency of the model training and a higher model regularization can be obtained on the spectral information in the images.

ACKNOWLEDGMENTS

This work was supported by the WtE 2030 funding (project number 280949) from the research council of Norway

REFERENCES

[1] Kumar, A. and Samadder, S. R., "A review on technological options of waste to energy for effective management of municipal solid waste," *Waste Management* **69**, 407–422 (2017).

[2] Komilis, D., Kissas, K., and Symeonidis, A., "Effect of organic matter and moisture on the calorific value of solid wastes: An update of the Tanner diagram," *Waste Management* **34**(2), 249–255 (2014).

[3] KIMURA, S., TAKAGI, Y., TONE, S., and OTAKE, T., "A RATE EQUATION FOR GAS-SOLID REACTIONS ACCOUNTING FOR THE EFFECT OF SOLID STRUCTURE AND ITS APPLICATION," *Journal of Chemical Engineering of Japan* **14**(6), 456–461 (1981).

[4] Abu-Qudais, M. and Abu-Qdais, H. A., "Energy content of municipal solid waste in Jordan and its potential utilization," *Energy Conversion and Management* **41**(9), 983–991 (2000).

[5] Kathiravale, S., Muhd Yunus, M. N., Sopian, K., Samsuddin, A. H., and Rahman, R. A., "Modeling the heating value of Municipal Solid Waste," *Fuel* **82**(9), 1119–1125 (2003).

[6] Makantasis, K., Karantzalos, K., Doulamis, A., and Doulamis, N., "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in [*2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*], 4959–4962 (2015).

[7] Gewali, U. B., Monteiro, S. T., and Saber, E., "Machine learning based hyperspectral image analysis: a survey," *arXiv preprint arXiv:1802.08701* (2018).

[8] Fotiadou, K., Tsagkatakis, G., and Tsakalides, P., "Deep convolutional neural networks for the classification of snapshot mosaic hyperspectral imagery."

[9] Caroppo, A., Leone, A., and Siciliano, P., "Comparison between deep learning models and traditional machine learning approaches for facial expression recognition in ageing adults," *Journal of Computer Science and Technology* **35**(5), 1127–1146 (2020).

[10] Wang, P., Fan, E., and Wang, P., "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognition Letters* **141**, 61 – 67 (2021).

[11] Lai, Y., "A comparison of traditional machine learning and deep learning in image recognition," *Journal of Physics: Conference Series* **1314**, 012148 (oct 2019).

[12] Li, Y., Zhang, H., Xue, X., Jiang, Y., and Shen, Q., "Deep learning for remote sensing image classification: A survey," *WIREs Data Mining and Knowledge Discovery* **8**(6), e1264 (2018).

[13] Cai, L., Gao, J., and Zhao, D., "A review of the application of deep learning in medical image classification and segmentation," *Annals of translational medicine* **8**, 713–713 (June 2020).

[14] Rawat, W. and Wang, Z., "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Computation* **29**(9), 2352–2449 (2017). PMID: 28599112.

[15] Khan, A., Sohail, A., Zahoora, U., and Qureshi, A. S., "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review* **53**(8), 5455–5516 (2020).

[16] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H., "How transferable are features in deep neural networks?," (2014).

[17] Morid, M. A., Borjali, A., and Del Fiol, G., "A scoping review of transfer learning research on medical image analysis using imagenet," **128**, 104115.

[18] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L., "Imagenet large scale visual recognition challenge," **115**(3), 211–252.

[19] Deng, J., Dong, W., Socher, R., Li, L., Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in [*2009 IEEE Conference on Computer Vision and Pattern Recognition*], 248–255.

[20] Chollet, F. et al., "Keras." https://keras.io (2015).

[21] Tan, M. and Le, Q. V., "Efficientnet: Rethinking model scaling for convolutional neural networks," **abs/1905.11946**.

[22] Xie, Q., Hovy, E. H., Luong, M., and Le, Q. V., "Self-training with noisy student improves imagenet classification," **abs/1911.04252**.

[23] EfficientNet, "Weights." https://github.com/tensorflow/tpu/tree/master/models/official/efficientnet (2021).

[24] Geelen, B., Blanch, C., Gonzalez, P., Tack, N., and Lambrechts, A., "A tiny VIS-NIR snapshot multispectral camera," in [*Advanced Fabrication Technologies for Micro/Nano Optics and Photonics VIII*], von Freymann, G., Schoenfeld, W. V., Rumpf, R. C., and Helvajian, H., eds., **9374**, 194 – 201, International Society for Optics and Photonics, SPIE.

[25] Chollet, F. et al., "Keras: Image classification via fine-tuning with efficientnet." https://keras.io/examples/vision/image_classification_efficientnet_fine_tuning/ (2015).

[26] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X., "TensorFlow: Large-scale machine learning on heterogeneous systems," (2015). Software available from tensorflow.org.

[27] Prechelt, L., "Automatic early stopping using cross validation: quantifying the criteria," **11**(4), 761 – 767.

[28] Finnoff, W., Hergert, F., and Zimmermann, H. G., "Improving model selection by nonconvergent methods," **6**(6), 771 – 783.

[29] Chollet, F. et al., "Keras: Early stopping." https://keras.io/api/callbacks/early_stopping/#earlystopping.

[30] Tharwat, A., "Classification assessment methods,"

[31] Boyd, K., Eng, K. H., and Page, C. D., "Area under the precision-recall curve: Point estimates and confidence intervals," in [*Machine Learning and Knowledge Discovery in Databases*], Blockeel, H., Kersting, K., Nijssen, S., and Železný, F., eds., 451–466, Springer Berlin Heidelberg.

[32] Fawcett, T., "An introduction to roc analysis," **27**(8), 861 – 874. ROC Analysis in Pattern Recognition.