

Learning-based Material Classification in X-ray Security Images

Benedykciuk Emil^a, Denkowski Marcin^b and Dmitruk Krzysztof^c

Institute of Computer Science, Maria Curie-Skłodowska University, Lublin, Poland

Keywords: X-ray Imaging, Material Classification, Machine Learning Algorithms.

Abstract: Although a large number of papers have been published on material classification in the X-ray images, relatively few of them study X-ray security raw images as regards of material classification. This paper takes into consideration the task of materials classification into four main types of organics and metals in images obtained from Dual-Energy X-ray (DEXA) security scanner. We adopt well-known methods of machine learning and conduct experiments to examine the effects of various combinations of data and algorithms for generalization of the material classification problem. The methods giving the best results (Random Forests and Support Vector Machine) were used to predict the materials at every pixel in the testing image. The results motivate a novel segmentation scheme based on the multi-scale patch classification. This paper also introduces a new, open dataset of X-ray images (MDD) of various materials. The database contains over one million samples, labelled and stored in its raw, original 16-bit depth form.

1 INTRODUCTION


Automatic baggage inspection systems using computer vision techniques have not been common for general threat detection in X-ray images yet. A significant obstruction is a difficulty of collecting large datasets of different materials with pixel-level and object labelling. To automate the X-ray scanners inspection a couple of problems such as material discrimination, image segmentation, object detection and finally threats identification should be resolved. Most of the security scanners use the dual-energy X-ray absorption techniques (DEXA). These devices apply radiation generated in two ranges classified as high energy (HE) and low energy (LE) X-rays. These two energies give two different absorption readings for the scanned objects that create one two-channel image with 16 bits per pixel. This makes the data unique if they are compared to the classic RGB images. Hence, this requires developing a different approach and considering some of the features of information obtained from the X-ray images.


A major contribution of our paper is to propose the best of the most-common machine learning methods to solve the problem of classification of the materials in the X-ray scans. We perform some experiments to


examine the effects of various combinations of data obtained from the DEXA scanners for generalization of the material classification problem. For this purpose different machine learning algorithms and training data sizes on the subregions (patches) of the full scene image are investigated. Further, we build classification results on our patch and demonstrate simultaneous material recognition as well as initial segmentation of a full resolution DEXA scan by the naive sliding window approach. In addition, we have built a large dataset, called the Materials in DEXA Scans Database (MDD) with about 1 million samples. The dataset has samples in five material classes. The entire dataset is discussed in more detail in section 3. MDD gives us the opportunity to train the above-mentioned machine learning algorithms and validate its effectiveness and accuracy.

In summary, we make two main contributions:

- conclusions regarding the data representation and selection of the machine learning methods, giving the best results for the problem of material classification in the X-ray images for simultaneous material recognition and initial segmentation;
- introduction of a new material dataset MDD for the DEXA images with a simple crowdsourcing pipeline for efficient acquisition of other millions of labelled patches.

^a  <https://orcid.org/0000-0002-1542-6747>

^b  <https://orcid.org/0000-0002-2491-091X>

^c  <https://orcid.org/0000-0003-1464-5822>

2 PRIOR WORK

2.1 X-ray Images Databases

The increase in the popularity of data-based methods has led to the creation of many datasets, including those with X-ray images. X-Ray scanners provide images based on a different level of radiation absorption for different materials. As a service to the X-ray testing and computer vision communities, Mery et al. (Mery et al., 2015) collected more than 19400 X-ray images for the development, testing and evaluation of image analysis and computer vision algorithms. Unfortunately, these images have been made available in the form of classic grayscale 8-bit images, thus preventing the use of DEXA methods for material recognition. In the domain of luggage inspection, many papers are focused on object detection. This leads to creating some datasets like ALERT Datasets (ALERT, 2019). ALERT has four active datasets of CT scans available to the research community that has developed out of transition tasks. One of the collections contains approximately 900 objects fully segmented from 62 luggage datasets to span the spectrum of packing, density, arrangement, orientation, and size difficulty. Another important dataset that should be mentioned in the context of materials recognition is MINC (Materials in the Context Database) (Bell et al., 2014) but this is a collection of samples of materials from the conventional images. There are also many public medical X-ray image databases but such images contain very limited and specific materials compared to the huge variety of materials and their combinations in security the X-ray images.

Unfortunately, there are no large X-ray images datasets of different materials with pixel or patch level labelling. Therefore, to fill this gap, we have created our own dataset. Our X-ray scans dataset MDD will allow us to verify different material recognition methods in the X-ray images which should help analyzing and interpreting X-ray scans and detecting threats.

2.2 Material Classification in the X-ray Images

X-Ray scanners provide images based on different levels of absorption of radiation of different materials. Such raw X-ray images are not always easy to analyze and interpret. Some image processing methods like object detection, frequency resolution increase or pseudocolouring are used (Dmitruk et al., 2015; Dmitruk et al., 2018). There are four main colors used widely in X-ray scanners to label material classes. In the newer generation of X-ray scanners, six classes

are often used to differentiate materials (see Figure 1). Nevertheless, the best solution would be to provide a specific type of material or an average value of the atomic number from which the material is made. The main problem in discriminating materials of a given object from its single projection image is to determine its thickness, density and composition itself. Two very different materials (e.g. steel and water) can give identical readings on the X-ray detectors if they have different densities and/or thicknesses. For this reason, multi-energy techniques are used that allow such a distinction for a single material. Dual-energy X-ray imaging is a such well-known technique. These X-ray scanners provide two images based on different levels of radiation absorption of different materials. In order to determine approximately the material class of the object being studied, the coefficient k (equation 1) is used (Watabiki et al., 2013; Rebuffel and Dinten, 2007):

$$k = \frac{\mu_{mL}}{\mu_{mH}} \quad (1)$$

where: μ_{mL} is the mass absorption coefficient for low energy scans and μ_{mH} is the mass absorption coefficient for high energy scans. However, the classic material discrimination methods used for the dual-energy X-ray scans do not cope well with determining the type of material when there are many layers of different types of material at a given point. As discussed in (Alvarez and Macovski, 1976; Lehmann et al., 1981; Chuang and Huang, 2000; Rebuffel and Dinten, 2007), the unambiguous definition of a material class for a composite of more than three substances is not obtainable in the case of a fixed X-ray tube system.

Material class	Examples	Threads
Background		
Light organic	Wood, Fabric	Explosives C4, Trotyl
Heavy organic	Plastics, Glass	Drugs, Heroin, Cocaine
Non-organic	Aluminium, Salt	Gunpowder
Metals	Iron, Steel, Brass	Guns, Bullets
Heavy metals	Silver, Gold	Jewellery contraband
Not-penetrable	Lead	Threat hiding

Figure 1: Material pseudo-colours and its classes used widely in the X-ray security scanners.

Dual-energy techniques allow simply to recognize a type of material only for a homogeneous object, regardless of its thickness. For an object composed of various materials, the obtained information is the averaged absorption value for all materials and their

thicknesses on the radiation line. But the biggest challenge is the case where the materials with a very different X-ray radiation absorption are assembled in a given place of the scan. Such a case can indicate a possible threat in the process of luggage control, i.e. when light objects (e.g. dry bulk goods like drugs or tobacco products) are placed on the background or inside the objects made of heavy materials (e.g. fire extinguisher). To overcome these difficulties, the community tries to use machine learning methods. In (Ko et al., 2011; Nedjar et al., 2015; Mehta and Sebro, 2019) the authors applied a random forest classifier or support vector machine (Elmasri et al., 2016; M.S. Kavitha et al., 2012; Sivakumar and Chandrasekar, 2013) to achieve fast and accurate classification task. Additionally, there are many works where machine learning methods have been used for material recognition in the conventional images (Caputo et al., 2005; Bhattacharjee et al., 2015). This allows to conclude that the methods also work for the problem presented in this paper giving satisfactory results. On the other hand, using machine learning in material discrimination in the X-ray scans is still in its initial form as for now and this problem has not thoroughly investigated by the computer vision community.

3 MATERIAL DEXA DATASET (MDD)

3.1 X-ray Samples

In our studies we use an ARIDA X-ray scanner (Arida, 2019), equipped with a Metrix SAX 1712A (Metrix, 2019) X-ray lamp and a strip of eight DT X-DAQ 0.8 DualEnergy detectors (DetectionTechnology, 2019) giving 1024 pixels of data in a single line of scan. All scans were made for the following X-ray beam setting: $130kV \times 0.86mA = 112W$.

The scans of all samples are raw, two-channel images where the pixel value in each channel is represented by a 16-bit integer with the values from 0 to 65536. The materials were classified into three main groups: organic materials, mixed materials, and heavy metals. The collection of organic materials for practical reasons has been divided into two sub-groups: light organic materials and heavy organic materials. For the first group of materials we selected the following liquids: water, ethanol, and slightly heavier vegetable oil. We have completed the first group with sugar, CD, and plexiglass from the set of solid and light substances. In addition, a set of samples was

supplemented with two types of wood. The approximate atomic number Z for organic light materials does not exceed 8 so the substances from this group have absorption values similar to those of explosives, e.g. C-4, TNT, Sentex. Paper and plasticine were selected from heavy organic materials. They give readings analogous to the specific type of dangers like powdery drugs, i.e. heroin or cocaine. Their approximate atomic number Z ranges from 8 to 10. Characteristic examples of materials with the average mass ratio values - non organic class, also called light metals, are aluminum and salt. In this group, there is also a kind of threats. The approximate atomic number of these substances is from 10 to 17. This is similar to that of gunpowder or heavy fuel. Whereas steel and brass belong to the group of substances called heavy metals, the approximate value of the atomic number Z is greater than 17. This material class represents those listed in section 2, metals and heavy metals. The reason for this simplification is the lack of a sufficient amount of substances with a very large atomic number Z . In addition, the impact on creation of one class of the two in our lamp settings will not affect readings on the detectors too much. Similarly to the previous classes of materials in this group there are many threats. The most common items in this group are the objects such as white weapons, firearms, cartridges, and high-value smuggling materials i.e. silver and gold. For all materials samples of different thicknesses were prepared.

To avoid overfitting and the 'leakage' of knowledge about the test data an additional test data set was created. It consists of selected common items and the specialized test case scans. Thus, we have expanded the MDD with an additional test set. Moreover, for the correct training process, we split our training samples and created a validation data set which contains 20% of all samples. The validation set helps us avoid overfitting and determines the accuracy of the classifier during the learning process.

3.2 Patches

The type of annotations or labels to collect for training is guided to a large extent by the tasks we wish to generate training data for. For some tasks such as scene recognition, whole-image labels can suffice (Xiao et al., 2014; Zhou et al., 2015). For object detection, the labelled bounding boxes as in PASCAL are often used (Everingham et al., 2010). For segmentation or scene parsing tasks, the per-pixel segmentations are required (Russell et al., 2008; Gould et al., 2009). Each style of annotation comes with a cost proportional to its complexity. For the materials

in the X-ray scans we decided to focus on one problem – patch material classification.

For training different types of classifiers, it is useful to have data in the form of fixed-sized patches. We use a patch center and a patch scale to define the image subregion that makes a patch. Further in Section 5, we justify our choice with experiments that vary the patch scale for the test suitcase scan. We generated about 1 million patches from all materials scans. The patches have the following sizes: 3x3, 5x5, 7x7, 9x9, 15x15. Table 1 shows the exact number of patches that make up each class of materials for learning and testing the dataset. It is worth mentioning that the samples are very noisy which is characteristic of X-ray images.

4 METHODS AND METHODOLOGY

Our goal is to create a system that will identify materials for each pixel of the scan. The result of the system will be an initial segmented image. For full-resolution segmentation of the image, the exact definition of the edge and noise removal will be missing. However, our initial segmentation is a solid input for the full segmentation method that simplifies and/or changes the representation of an image into something that is more meaningful and easier to analyze.

Among the algorithms and parameter variations we tested the best performing and the most common machine learning algorithms:

- 1) (LDA) – Linear Discriminant Analysis,
- 2) (LR) – Logistic Regression,
- 3) (NB) – Naive Bayesian,
- 4) (SVM) – Support Vector Machine,
- 5) (CART) – Decision Tree,
- 6) (RF) – Random Forests,
- 7) (KNN) – k-Nearest Neighbours,
- 8) (MLP) – Simple Multilayer Perceptron,

that produce a single prediction for a given input patch. Then the one that gives the most efficient predictions for all sizes of the input patches is chosen. After analyzing the performance in each stage our classifier is transformed into a sliding window and materials on a grid across the image are predicted. We do this with multiple scales of patches and then their simple arithmetic mean is computed. The above allows to predict the material class for every pixel of the scan.

Figure 2 gives the overview of our method for simultaneously initial segmenting and recognizing the material class. Given an estimator that can classify individual points in the image, based on the pixel dif-

ferent size regions, we use it for the naive sliding window approach. This way of initial segmentation of a full resolution DEXA scan allows to present transitional steps of full scan material classification. The probability of each of the material classes in a given pixel of the scan can be presented. The input image can be any size, depending on the size of the patch, the corresponding number of pixels on the edges is copied. It should be noted that increasing the patch size reduces the resolution of the results. Therefore for prediction of the pixel material class, we use the results of a given classifier for different sizes of the pixel region. Finally, depending on the probability of each material class, we select the colour of class with the greatest probability.

5 EXPERIMENTS AND RESULTS

5.1 Mean Classification Accuracy

Table 2 presents the patch material classification results on our dataset during the learning process and for the additional test scans. The input data images for all algorithms are three-channel images with the following composition: the first channel is high energy, the second channel is low energy and the last channel is filled by zeros.

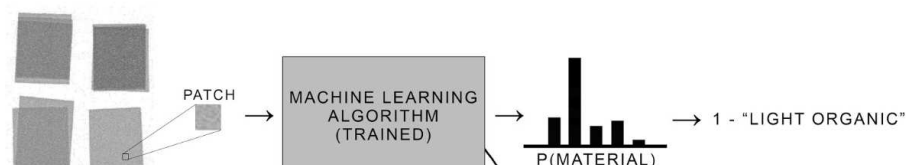
The algorithms that obtained the best results for the validation set (i.e. accuracy above 90 percent) are KNN, RF and SVM. We can try to choose an algorithm based on the results for a test set but we have to be careful because this may lead to matching the algorithm to the test set. This situation is called overfitting (see Section 4). The choice based only on the results for the test set would suggest that the LDA and LR algorithms are a good choice. However, a close look at the results for the validation set shows that this is not the case. This suggests that the algorithm does not generalize the problem sufficiently. The algorithms that obtain the most stable and best results are random forest (RF) and support vector machine (SVM).

For a more detailed comparison of these two algorithms in Tables 3 and 4 the accuracy of the algorithms for each class and patch size is presented. Comparing these two tables, we can conclude that the SVM algorithm gets better results particularly for larger patches where the decision is made for a larger number of image features. Due to the fact that the differences in accuracy of these two algorithms are not too big and because the scanning of the luggage has to be carried out very quickly (see Section 1) the RF algorithm is considered to be a better solution because RF is a much faster method in the learning process

Table 1: MDD training/testing data set - number of patches for all material classes.

Material Patch size	Background	Light organic	Heavy organic	Light metals	Heavy metals	Total
3 × 3	82526/29538	210028/15940	152082/3186	109376/7406	89374/11216	643386/67286
5 × 5	29436/10382	73966/5690	54176/1120	38716/2614	31432/3894	227726/23700
7 × 7	14852/5352	37000/2860	27400/550	19390/1326	15826/1982	114468/12070
9 × 9	8902/3198	21822/1664	16344/342	11574/778	9388/184	68030/7166
15 × 15	3108/1112	7258/578	5776/110	3862/270	3170/368	23174/2438
Total	138824/49582	350074/26732	255778/5308	182918/12394	149190/18644	1076784/112660

A) PATCH MATERIAL CLASSIFICATION



B) FULL SCENE MATERIAL CLASSIFICATION

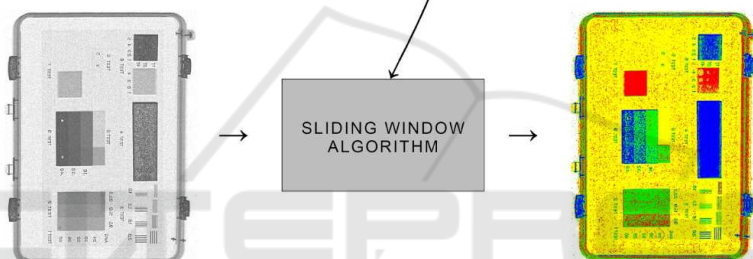


Figure 2: Pipeline for full scan material classification. WHITE - background, YELLOW - light organic, RED - heavy organic, GREEN - light metals, BLUE - heavy metals.

Table 2: Mean accuracy for all materials and patch sizes for different machine learning algorithms trained and tested on MDD. The performance of all algorithms was verified using a computer equipped with AMD Ryzen 7 2700x.

Algorithm	Validation Set	Test Set	Learning time [s]	Prediction time [s]
LDA	0.666 (0.027)	0.931 (0.044)	4.43	0.04
LR	0.743 (0.012)	0.931 (0.010)	30.77	0.04
NB	0.565 (0.005)	0.885 (0.005)	0.23	0.14
MLP	0.549 (0.230)	0.614 (0.347)	177.65	0.08
SVM	0.925 (0.030)	0.950 (0.049)	17776.42	182.82
KNN	0.965 (0.012)	0.882 (0.011)	2.50	27.65
CART	0.889 (0.018)	0.797 (0.027)	27.30	0.03
RF	0.965 (0.008)	0.935 (0.017)	177.72	0.88

as well as in the test process compared to SVM, as shown in Table 2.

As we approve the RF as the best solution for our problem, it is worth nothing that the efficiency of the classifier is the smallest for the class of heavy organic materials. This is probably due to the fact that this class mixes with that of light metals very much. Unfortunately, the random forest algorithm separates these two groups of materials poorly. In addition, it can be deduced that the accuracy for these

classes changes along with the change in the size of the patches. The effectiveness value for the light metal class increases with the increasing patch area because most of the patches from both groups are recognized as a heavier material. If time was not an important attribute of material prediction, the SVM algorithm would be a much better solution of this problem. Also noteworthy is the signal to noise ratio for all pixels in the patch which affects the prediction in these both algorithms. For the RF algorithm, this is

Table 3: Random forest accuracy for each class and patch size. Material classes: 0 – background, 1 – light organic, 2 – heavy organic, 3 – light metals, 4 – heavy metals.

Class	3x3	5x5	7x7	9x9	15x15
0	0.975	0.999	0.978	0.971	1.000
1	0.937	0.985	0.997	0.999	1.000
2	0.897	0.952	0.782	0.509	0.455
3	0.805	0.909	0.941	0.943	1.000
4	0.797	0.804	0.818	0.785	0.810

Table 4: Support vector machine accuracy for each class and patch size. Material classes: 0 – background, 1 – light organic, 2 – heavy organic, 3 – light metals, 4 – heavy metals.

Class	3x3	5x5	7x7	9x9	15x15
0	0.953	0.980	0.971	0.970	1.000
1	0.905	0.950	0.974	0.992	1.000
2	0.379	0.629	0.913	0.994	1.000
3	0.664	0.816	0.919	0.892	1.000
4	0.891	0.963	1.000	1.000	1.000

evident for heavy material classes. The heavier materials have more noisier samples which can be seen in the accuracy results. On the other hand, the ratio of noise to all pixels is larger for smaller patches which is clearly evident in the effectiveness of the SVM that achieves less accuracy for smaller patches. Our observations are consistent with what was discussed in (Ogorodnikov et al., 2002). The authors note that the discrimination error increases for lower mass thickness because there is no sufficient contrast between the low and high energy images, and for larger mass thickness due to the decreasing signal-to-noise ratio.

5.2 Training Dataset Size

We have also verified the impact of the size of the training database on the prediction accuracy. We checked the dataset containing 100%, 50%, 25% and 10% of all MDD training samples (see Table 1). The results of the mean accuracy of the algorithms for different amounts of training data are presented in Figure 3.

Evaluation of the classifiers results shows that a decrease in accuracy can be noted for only LDA. Prediction decreases as the training set decreases. The reverse situation is for SVM and RF because algorithms are overfitted. A much larger accuracy difference between 100% and 10% of the training set can be seen for SVM. This depends on the selected penalty parameter C and kernel parameters. Even though SVM is an approximate implementation of a bound on the generalization error, that depends on the margin, but is independent of the dimensionality of the feature

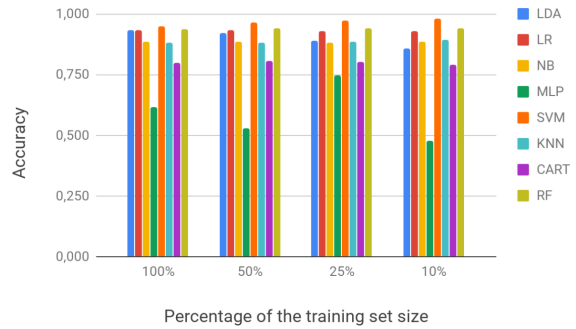


Figure 3: Impact of the training set size on the predictions accuracy for test set.

space. So in principle, SVMs should be highly resistant to over-fitting but in practice this depends on the careful choice of C and the kernel parameters (Cawley and Talbot, 2007; Cawley and Talbot, 2010).

5.3 Patch Material Classification

To classify a point in an image we must decide how large context to include around it. The context, expressed as an area of the pixel, is the patch size. A priori, it is not clear which scale is best since small patches have a better spatial resolution but larger patches have more contextual information. The visual impact of patch scales is shown in Figure 4. As you can see in this figure, the smaller size of the patch is a better reflection of the shape of objects appearing on the scan. Unfortunately, this is not an ideal solution as is largely susceptible to the previously mentioned noise appearing in the X-ray images. On the other hand, too large patches cause loss of information about details.

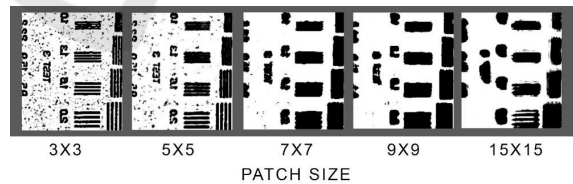


Figure 4: Visualization of the impact of patches size on the recognition of material classes (presented in the gray background). We present the probability of occurrence of the light organic class (using a gray scale) for the scan of special test suitcase fragment.

Good information can be the fact that with such patch sizes the accuracy of most of the tested classifiers is not very different (Figure 5). However, attention should be paid to the lack of stability of the MLP algorithm whose results come from its simple architecture, which made it impossible to generalize the classification process.

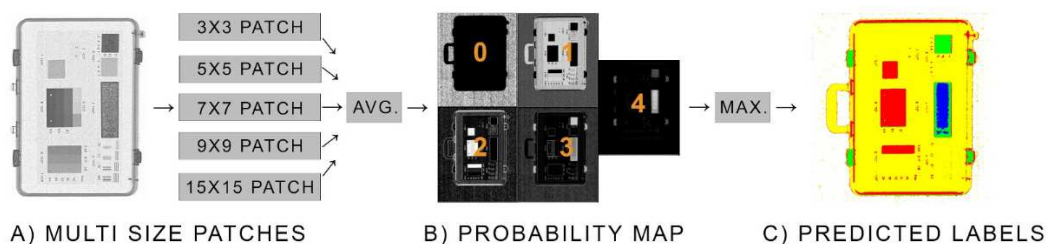


Figure 6: Pipeline for full scan material classification. Each colour indicates a specific type of material: WHITE (0) – background, YELLOW (1) – light organic, RED (2) – heavy organic, GREEN (3) – light metals, BLUE (4) – heavy metals.

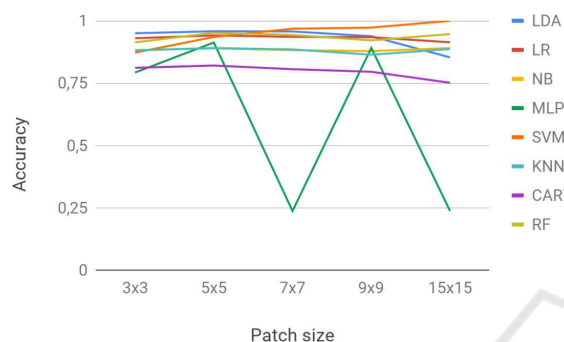


Figure 5: Accuracy as a function of a patch size for the tested machine learning algorithms on the test dataset and [LE, HE, zeros] input image.

5.4 Initial Full Scan Material Segmentation

The scheme of the image initial segmentation process is shown in Figure 6. Presentation of the probability of each material class is very important to determine the certainty of the receivables of each pixel. It also allows different presentations of the estimator results. We decided that the initial segmentation of the scans is based on selecting the class that is most likely for all patch sizes. As for now, the procedure is quite simple. We get all patches size for each pixel scan (Figure 6A). Then the average probability of the class for each pixel is calculated (Figure 6B). This generates probability maps for all classes. The last step is to choose the class with the greatest probability (Figure 6C). Figure 7 presents some results for the classifier which obtained the best results in the previous tests.

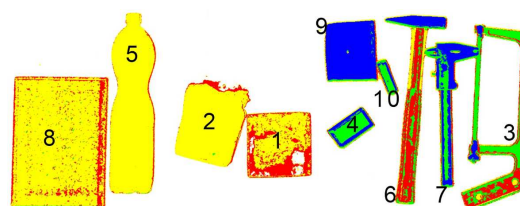


Figure 7: Full scan material classification examples: test set predictions by our method based on the random forest classifier. Each colour indicates a specific type of material as stated in Figure 1.

multaneous material classification and initial segmentation. We have proved that training on a dataset which includes the surrounding context is crucial for material classification in the DEXA scans. Many future avenues of work remain. Expanding the dataset to a broader range of categories will require new ways to mine images that have more variety, and new annotation tasks that are cost-effective. We also believe that further exploration of the material and object classification will be successful and lead to improvements in both tasks. Another issue that we want to deal with in the future study is to propose an original algorithm that will provide a faster learning process and better predictions. In addition, the improvement of deep convolutional neural networks suggests that the solutions will also work for the problem presented in this paper. The issue that can be solved by machine learning or deep learning algorithms is the unnatural separation of material classes. We believe that the clusters' algorithms would enforce the classes more naturally, based on the data set because, as we can see, the unnatural division of classes can cause classification errors.

6 CONCLUSIONS

Material recognition in the X-ray images is a long-standing, challenging problem. We introduce a new large, open, material database – MDD that includes a diverse range of materials, and this is the first such dataset. Using this large database we conduct an evaluation of recent machine learning algorithms for si-

REFERENCES

ALERT (2019). Alert datasets. <http://www.north-eastern.edu/alert/transitioning-technology/alert-datasets/>, (accessed: 2019-08).

Alvarez, R. and Macovski, A. (1976). Energy-selective reconstructions in x-ray computerized tomography. *Physics in Medicine and Biology*, 21:733–744.

- Arida (2019). Arida systems sp. z o.o. <http://arida.pl/index.php#products>, (accessed: 2019-10).
- Bell, S., Upchurch, P., Snavely, N., and Bala, K. (2014). Material recognition in the wild with the materials in context database. in *IEEE CVPR*.
- Bhattacharjee, T., Wade, J., and Kemp, C. C. (2015). Material recognition from heat transfer given varying initial conditions and short-duration contact. In *Robotics: Science and Systems*.
- Caputo, B., Hayman, E., and Mallikarjuna, P. (2005). Class-specific material categorisation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1597–1604 Vol. 2.
- Cawley, G. and Talbot, N. (2007). Preventing over-fitting during model selection via bayesian regularisation of the hyper-parameters. *Journal of Machine Learning Research*, 8:841–861.
- Cawley, G. and Talbot, N. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11:2079–2107.
- Chuang, K.-S. and Huang, H. (2000). Comparison of four dual energy image decomposition methods. *Physics in Medicine and Biology*, 33:455–466.
- DetectionTechnology (2019). Dt x-daq 0.8 dualenergy detector documentation. http://www.datvision.co.kr/upload/DT_20121019132944.pdf, (accessed: 2019-10).
- Dmitruk, K., Denkowski, M., Mazur, M., and Mikoajczak, P. (2018). Sharpening filter for false color imaging of dual-energy x-ray scans. *Signal, Image and Video Processing*, 11(4):613–620.
- Dmitruk, K., Mazur, M., Denkowski, M., and Mikolajczak, P. (2015). Method for filling and sharpening false colour layers of dual energy x-ray images. *IFAC-PapersOnLine*, 48:342–347.
- Elmasri, K., Hicks, Y., Yang, X., Sun, X., Pettit, R., and Evans, W. (2016). Automatic detection and quantification of abdominal aortic calcification in dual energy x-ray absorptiometry. *20th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems 5-7 September 2016*, 96:1011–1021.
- Everingham, M., Gool, L., Williams, C., and J. Winn, A. Z. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 1(2):303–338.
- Gould, S., Fulton, R., and Koller, D. (2009). Decomposing a scene into geometric and semantically consistent regions. pages 1 – 8.
- Ko, B., Kim, S., and Nam, J.-Y. (2011). X-ray image classification using random forests with local wavelet-based cs-local binary patterns. *Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology*, 24(6):1141–1151.
- Lehmann, L. A., Alvarez, R. E., Macovski, A., Brody, W. R., Pelc, N. J., Riederer, S. J., and Hall, A. L. (1981). Generalized image combinations in dual kvp digital radiography. *Medical Physics*, 8(5):659–667.
- Mehta, S. and Sebro, R. (2019). Random forest classifiers aid in the detection of incidental osteoblastic osseous metastases in dexa studies. *International Journal of Computer Assisted Radiology and Surgery*, 14(5):903–909.
- Mery, D., Rizzo, V., and Lobel, H. (2015). Gdxray: The database of x-ray images for nondestructive testing. *Journal of Nondestructive Evaluation*, pages 34–42.
- Metrix (2019). Metrix sax 1712a lamp documentation. <http://metrixndt.com/generators-brochures/SAXG%201712A%20v1%20LR.pdf>, (accessed: 2019-10).
- M.S. Kavitha, A. A., Taguchi, A., Kurita, T., and Sanada, M. (2012). Diagnosis of osteoporosis from dental panoramic radiographs using the support vector machine method in a computer-aided system. *BMC Medical Imaging*, 12:1011–1021.
- Nedjar, I., EL HABIB DAHO, M., Settouti, N., Sad, M., and Chikh, M. (2015). Random forest based classification of medical x-ray images using a genetic algorithm for feature selection. *Journal of Mechanics in Medicine and Biology*, 15(2):1540025.
- Ogorodnikov, S., Petrunin, V., and Vorogushin, M. (2002). Radioscopic discrimination of materials in 1-10 mev range for customs applications. pages 2807–2809.
- Rebuffel, V. and Dinten, J.-M. (2007). Dual-energy x-ray imaging: Benefits and limits. *Insight - Non-Destructive Testing and Condition Monitoring*, 49:589–594.
- Russell, B., Torralba, A., Murphy, K., and Freeman, W. (2008). Labelme: A database and web-based tool for image annotation. *International Journal of Computer Vision*, 77.
- Sivakumar, S. and Chandrasekar, C. (2013). Lung nodule detection using fuzzy clustering and support vector machines. *International Journal of Engineering and Technology (IJET)*.
- Watabiki, H., Takeda, T., and Mitani, S. (2013). Development of dual-energy x-ray inspection system. *Anritsu Technical Review*, (20):60–61.
- Xiao, J., Ehinger, K., Hays, J., Torralba, A., and Oliva, A. (2014). Sun database: Exploring a large collection of scene categories. *International Journal of Computer Vision*, 119(1):3–22.
- Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. (2015). Learning deep features for scene recognition using places database. *Advances in Neural Information Processing Systems*, 1:487–495.