Simultaneous Mass Estimation and Class Classification of Scrap Metals using Deep Learning

Dillam Jossue Díaz-Romero[1,2*], Simon Van den Eynde[1], Wouter Sterkens[1,2], Bart Engelen[1,3], Isiah Zaplana[1], Wim Dewulf[1], Toon Goedemé[2] , Jef Peeters[1]

[1]Department of Mechanical Engineering KU Leuven, Celestijnenlaan 300, Leuven 3000, Belgium

[2]PSI-EAVISE-KU Leuven, Jan Pieter de Nayerlaan 5, Sint-Katelijne-Waver 2860, Belgium

[3]Department of Mechanical Engineering KU Leuven, Wetenschapspark 27, Diepenbeek 3590, Belgium

Abstract —***While deep learning has helped improve the performance of classification, object detection, and segmentation in recycling, its potential for mass prediction has not yet been explored. Therefore, this study proposes a system for mass prediction with and without feature extraction and selection, including principal component analysis (PCA). These feature extraction methods are evaluated on a combined Cast (C), Wrought (W) and Stainless Steel (SS) image dataset using state-of-the-art machine learning and deep learning algorithms for mass prediction. After that, the best mass prediction framework is combined with a DenseNet classifier, resulting in multiple outputs that perform both object classification and object mass prediction. The proposed architecture consists of a DenseNet neural network for classification and a backpropagation neural network (BPNN) for mass prediction, which uses up to 24 features extracted from depth images. The proposed method obtained 0.82 R^2, 0.2 RMSE, and 0.28 MAE for the regression for mass prediction with a classification performance of 95% for the C&W test dataset using the DenseNet+BPNN+PCA model. The DenseNet+BPNN+None model without the selected feature (None) used for the CW&SS test data had a lower performance for both classification of 80% and the regression (0.71 R^2, 0.31 RMSE, and 0.32 MAE). The presented method has the potential to improve the monitoring of the mass composition of waste streams and to optimize robotic and pneumatic sorting systems by providing a better understanding of the physical properties of the objects being sorted.***

Keywords: *Artificial Intelligence; Automatic Sorting; Metal Recycling; Stainless Steel; Cast and Wrought Aluminium scrap; Deep Learning Computer Vision; Backpropagation Neural Network; Mass/weight Prediction; Object Detection and Recognition*

1. Introduction

Aluminum (Al) alloys are of great interest for various sustainable technologies due to their light-weight and mechanical properties, which explains its constantly increasing demand *(Cullen and Allwood, 2013)*. Along with the production volumes, also the amount of collected scrap metal is increasing every year. Today, the majority of this scrap is used as a secondary feed for producing Al Cast (C) alloys, which are mainly used for the production of combustion engine motor blocks *(Johnson et al., 2013)*. However, as a consequence of the electrification of the automotive sector, the demand for cast alloys is expected to stagnate and possibly even decline in the coming decade *(Modaresi and Müller, 2012)*. As a result, alternative destinations will have to be searched to avoid the generation of an aluminium scrap surplus. One of the solutions to prevent the emergence of a scrap surplus is to design recycling-friendly alloys that can function as alternative sinks for aluminum scrap due to less stringent tolerances on the concentrations of alloying elements in the alloy *(Modaresi, 2015)*. Another possibility, which is complementary with the development of recycling-friendly alloys, is the development of more advanced recycling technologies that allow sorting different qualities, e.g. to sort between C and different Wrought (W) Al alloy groups, as well as sorting Al from other metals, such as Stainless Steel (SS). Today, aluminum recycling is typically carried out by adopting magnetic (over belt) separators to distinguish between ferrous and non-ferrous metals and eddy current separators to differentiate plastics from non-ferrous metals *(Nijhof, 1994)*.

Further, Al can be separated from most other non-ferrous metals by adopting sink-float techniques due to the relatively low density of Al. However, those techniques cannot be adapted to separate C from W alloys *(Eggers et al., 2019)*. In addition, sink-float separators will likely never result in a perfect separation due to the surfing of mainly flat

*Corresponding author: dillam.diaz@kuleuven.be

objects, the floating of pieces due to the inclusion of air or attachment of other lighter materials and the inclusion of suspension material in hollow parts.

In this regard, X-Ray Fluorescence (XRF) or Laser-Induced Breakdown Spectrometry (LIBS), and/or machine vision technologies, using X-Ray Transmission, and/or Color and Depth cameras are considered to encompass substantial potential to sort Al based on the alloying elements, e.g. C from W alloys, and to obtain higher purity output fractions *(Díaz-Romero et al., 2021)*. Combining these technologies with a pneumatic valve block and/or a robotic gripping system opens the possibility of developing robust and cost-efficient systems for sorting scrap Al. However, in order to successfully plan and execute the physical sorting tasks, such as robotic picking or pneumatic object ejection, an optimally functioning sorting system requires a multimodal understanding of the objects to be sorted. This includes their semantic, geometric, and physical properties, preferably before any decision or contact with the object is made *(Standley et al., 2017)*. One of the most critical physical properties that influence both optimal grasping strategies and optimal control of a pneumatic ejection system is the object's mass *(Correll et al., 2016)*.

Object mass estimation is not only relevant for optimizing the actual sorting processes but also for reporting purposes as recycling rates are commonly quantified by a weight-based target *(Nelen et al., 2014)*, which may result in undesired behaviours, such as prioritizing the collection and sorting of the heaviest materials instead of the most environmentally relevant light-weight material. Therefore, object mass estimation technologies are also valuable when combined with object classification to monitor and report on the actual mass composition of waste streams in a more continuous and standardized manner *(Hotta et al., 2016)*. The object mass can be estimated by combining 3D information with the average material density of the waste stream. In contrast, such an approach is prone to significant errors when objects are of different materials, and/or irregular-shaped (hollow), and/or not making full contact with the surface on which they are positioned during analysis. Various applications of computer vision and convolutional neural networks (CNNs), which use imagery to gain a higher level of understanding of objects or classes, have been demonstrated in waste management applications *(He et al., 2016; Shao et al., 2017; Chu et al., 2018; Sterkens et al., 2021; Zhang et al., 2021)*.

1.1. *Deep Learning in Recycling*

The increasing use of automated sorting systems based on image recognition could help to reduce repetitive manual sorting tasks in the recycling field. Sterkens *et al*. investigated the use of the Yolo v2 Deep Learning network for object detection using X-Ray images of the internal structure of Waste Electric and Electronic Equipment (WEEE). The researchers collected a dataset of 532 X-Ray transmission images with two different X-Ray source configurations, obtaining a 91% true-positive rate and only a 6% false-positive rate for classifying battery-containing devices. (Sterkens et al., 2021). Mao *et al*. proposed to use DenseNet121 optimized by a genetic algorithm (GA) to enhance the classification accuracy on the TrashNet dataset, which has 2525 images grouped into six different object classes (glass, paper, cardboard, plastic, metal and trash), reaching up to 99.40 % of accuracy. Additionally, they proposed the gradient-weighted class activation mapping to help to highlight the waste image's rough features and validate the proposed method (Mao et al., 2021).

Zhang *et al.* used computer vision to classify household waste. They proposed a recognition-retrieval model to classify waste into four categories: Recyclable Waste, Residual Waste, Household Food Waste, and Hazardous waste. As a benchmark, a one-stage waste classification model was trained. Both systems were implemented in an automatic sorting machine, showing a sorting performance average accuracy of up to 94.71% ± 1.69 (Zhang et al., 2021). In an earlier study, the presented research built on the classification of C&W Al by evaluating five CNN Deep Learning models and two transfer learning methods (Díaz-Romero et al., 2021). This study showed that the fusion of RGB and 3D images at the last layer of the DenseNet network improves the classification of the evaluated dataset. Furthermore, it was concluded that DenseNet could classify C&W Al with up to 98% accuracy.

1.2. *Related Work for Mass Estimation*

Computer-aided mass estimation of irregularly-shaped metal waste is beneficial for developing recycling technologies. This is the first study investigating simultaneous classification and mass estimation of metal scrap to the authors' best knowledge. However, research has been performed on mass estimation in several domains such as medicine, agriculture and robotics. In 2017, Santley *et al*. proposed using colour images to predict the mass of various objects (image2mass). The study developed a dataset of web products on Amazon containing information on the image, object size, and mass. Then, using 14 features and 2 Xception networks, the authors predicted the object's mass. A human operator was asked to perform the same mass estimation to compare. Results showed that the system could predict mass with a coefficient of determination (R^2) of 0.691 and the minimum ratio error of 0.675 (Standley et al., 2017).

In 2019, Utai *et al*. investigated the input of feature extraction from the image into an artificial neural network (ANN) for the mass estimation of irregularly-shaped fruits, showing the highest success rates of 97% and 99% for R^2 using ANN input with area and thickness or length, width, and thickness parameters, respectively (Utai et al., 2019). Konovalov *et al*. used two instances of the LinkNet-34 segmentation CNN to segment the images and estimate the mass of harvested fish by using the weight-from-area model, which resulted in a mean absolute percentage error of 4.36% (Konovalov et al., 2019). However, both approaches cannot be used for irregularly-shaped objects because the area is calculated based on the homogeneous mask of the object.

In 2020, Zhang *et al*. proposed a more robust method for fish mass prediction using image analysis and neural networks. The authors proposed to calculate nine features extracted from the image; then, they evaluated a PCA to select the best features and, finally, they trained the BPNN network to predict their mass. Their results showed a mean absolute error (MAE) of 0.0104, a R^2 of 0.92, and a root means square error (RMSE) of 0.0134, demonstrating that the proposed method accurately estimates the mass (L. Zhang et al., 2020). An overview of related work and obtained performances are provided in Table I.

The classification method used in prior research could be used to define the average density of an object class but would not allow overcoming the difficulties of obtaining reasonable volumetric estimations for irregular shapes, which are typical for scrap metals. Therefore, this paper presents a novel approach to simultaneously estimate the mass of unknown metal scrap objects and the material class to which they belong. By combining two feature selection methods and seven machine learning models, the combination of a CNN and a backpropagation neural network (BPNN) was evaluated to outperform all other combinations. Therefore, the performance of the combined DenseNet+BPNN network is presented for the mass estimation and object classification for the combined datasets of Cast & Wrought (C&W) and Cast, Wrought and Stainless Steel (CW&SS).

The novel contributions of this paper are:

- Implementation of a multi-out network for simultaneous classification of non-homogeneous shaped metal scraps (C, W and SS) and prediction of their mass. To the best of our knowledge, this paper is the first to benchmark the performance of Deep Learning methods to classify scrap metals such as C, W, and SS and simultaneously predict their masses.
- The application and evaluation of handcrafted features for the mass estimation in recycling datasets using various machine-learning methods, which open the possibility of creating an online system for monitoring the material in the early or late stages of sorting.
- The use of the backpropagation neural network (BPNN) algorithm to obtain a more accurate mass estimation model for scrap metal compared to traditional machine learning methods.

The paper is organized as follows. Section 2 outlines the material and data pre-processing. Section 3 presents feature extraction methods and two types of feature selection algorithms. Section 4 describes the applied machine and Deep Learning methodology for metal classification and mass prediction using 3D images and evaluation metrics. Section 5 presents the results and discussions. Finally, Section 6 concludes the paper and discusses future work.

TABLE I
PERFORMANCES ACHIEVED IN PRIOR RESEARCH FOR DEEP LEARNING CLASSIFICATION IN RECYCLING APPLICATIONS AND FOR MASS ESTIMATION

	Author(s) Year	Objective	Algorithm	Type of Dataset	Result
Deep Learning in Recycling	Sterkens *et al.* 2021	The detection of batteries in waste electrical and electronic equipment (WEEE)	Yolo V2	532 X-ray transmission images for classifying battery-containing device	a 91% true-positive and a 6% false-positive rate
	Mao *et al.* 2021	The use of deep learning to optimize waste stream detection accuracy	DenseNet121 + a genetic algorithm (GA)	TrashNet: 2525 images grouped into six different object classes	an average accuracy of up to 99.40 %
	Zhang *et al. 2021*	The use of a recognition-retrieval model for the classification of waste	ResNet18 with a self-monitoring module (SMM)	TrashNet: 2525 images grouped into four different	an average accuracy of up to 94.71% ± 1.69
	Díaz-Romero *et al. 2021*	The fusion of RGB and 3D images for the classification of aluminum	DenseNet with early or late fusion	548 images of scrap aluminum scraps	an average accuracy of up to 98%
Mass Estimation	Santley *et al. 2017*	The use of the geometry module and the volume tower to predict the mass of the object	14 features + 2x Xception network	Amazon test set of 147k images The household test set of 479 images	R^2 of 0.691 and the RMSE of 0.675
	Utai *et al.* 2019	The use of feature extraction from the image to estimate the mass	Four features + ANN	Images of irregularly-shaped fruits	The highest success for R^2 with 97% and 99%
	Konovalov *et al.* 2019	The use of instance segmentation from image to estimate the mass	LinkNet-34 + weight-from-area model	1400 images of harvested fish and 300 segmented fish masks	MAE of 4.36%
	Zhang *et al.* 2020	The use of PCA and a calibration factor CF for mass estimation	Nine features +PCA+BPNN	455 images of the Crucian carp fish	R^2 of 0.92, MAE of 0.0104 and RMSE of 0.0134

2. Material

A dataset of 120 C, 428 W Al scrap samples and 134 SS samples of different shapes (e.g., compact, bar, sheet, pipe, and irregular) with a mass distribution between 5 to 200 grams (g) was collected from a Belgian recycling facility. The Wrought and Cast pieces were used in a previous study to classify Al scraps (Díaz-Romero et al., 2021). The 548 Al samples (C&W) were collected randomly from the Twitch fraction. The 134 SS pieces were extracted from the Zorba fraction, consisting of shredded non-ferrous metals. The ferrous metals and non-metals were separated from this non-ferrous fraction in earlier sorting steps.

The regression's ground truth was defined by weighing the metal pieces with a 1g resolution Sartorius Bp34 High-Capacity Basic Plus Balance and error of ±0.5 g. The classification's ground truth was defined by combining captured images on a conveyor belt using a Niton™ XL2 XRF analyzer, suitable for cross-analyzing all the metal scrap samples by linking each image with its mass and chemical composition.

The analysis of the collected 3D images, the detection of the Region of Interest (ROI), the extraction of 24 features from the ROI 3D images, the statistical analysis, and the implementation of machine-learning algorithms were carried out in Python. The images were captured on a conveyor belt with two LMI GOCATOR 2340 3D laser line profile sensors with a scan rate of 5 kHz synchronized by an LMI GOCATOR MASTER 810 with a resolution of 0.15 (mm) on the x and y axes and 0.0001 (mm) on the z-axis (Díaz-Romero et al., 2021).

For the experiments, the 682 scrap metal objects are randomly divided into 70% training, 10 % validation, and 20% testing for all the experiments. All the experiments were computed on a single GPU: NVIDIA RTX3070 8 GB. A CPU: Intel® i7 with 3.20 GHz with 32 GB DDR4 RDIMM memory was used for the training and testing.

2.1. *Data Pre-processing*

The 3D camera has a resolution of 16 bits and requires a pre-processing step to transform the images into 8 bits. Hence, the first step to detect the ROI is to calculate the Mean and the Standard Deviation (Std) of the image and then clip the images before using a scale factor, as seen in equations (1-3).

$X = \{x_{ij}\}_{i,\,j}$ represents a point cloud matrix used to calculate the mean (Img_{mean}) and standard deviation (Img_{Std}) of the 3D image using the number of rows n (equation 1). $v_{\max}$ and $v_{\min}$ represent the maximum and minimum intensities plus or minus three times Std, and are calculated to clip the image resolution (equation 2). The clip function ($clip(x, v_{\max}, v_{\min})$) limits the x array values that lie outside the specified interval at the edges of the interval. Finally, the *scale#* factor is used to get a better object mask (Img_{mask}) (equation 3); the scaling factor is directly proportional to the pixel size.

$$Img_{mean} = \frac{\sum x}{n} \qquad Img_{std} = \frac{\sum (x - Img_{mean})^2}{n - 1} \tag{1}$$

$$v_{max} = Img_{mean} + (3 \cdot Img_{std}) \qquad v_{min} = Img_{mean} - (3 \cdot Img_{std}) \tag{2}$$

$$Img_{mask} = \frac{clip(x, v_{max}, v_{min})}{scale\#} \tag{3}$$

Once the image mask is defined, the OpenCV library is used to identify the ROI, as shown in Fig. 1. The first step to detecting the ROI is applying the function *cv2.threshold* to transform the images from grayscale into a binary image (Mordvintsev and Abid, 2014). The following parameters were used for this step: THRESH_BINARY_INV as thresholding type, a threshold value (thresh) of 181, and a maximum value (maxval) of 150.

The last step in detecting the ROI requires applying the function *findContour,* using the mode RETR_EXTERNAL and the method CHAIN_APPROX_SIMPLE, as shown in Fig. 1b. The method returns the object contours used for cropping the object from the gathered image, as shown in Fig.1(b-c) (Mordvintsev and Abid, 2014; Suzuki, 1985).

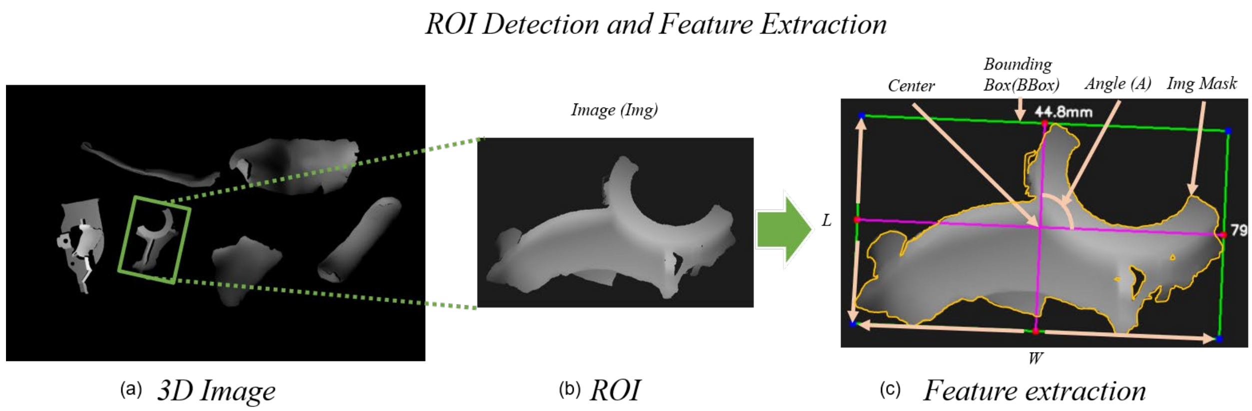

Fig. 1, (a) shows an illustrative example of a 3D image with multiple objects. ROI in Fig. 1b is calculated for scrap metal surrounded by a green bounding box (BBox), as shown in the middle image. Once the object is selected and cropped, seven features are shown in Fig. 1c image.

3. Methods

The aim is to calculate how accurately the mass of C, W, and SS can be estimated. The first step is to calculate 3D image features to estimate the mass of scrap parts. Second, feature selection methods are presented to identify the most relevant features for C, W, and SS. Third, machine learning and the BPNN method are presented to evaluate the feature selection-based mass estimation. Finally, CNN and the best mass estimation method are combined to simultaneously classify and predict the object's mass and investigate the performance.

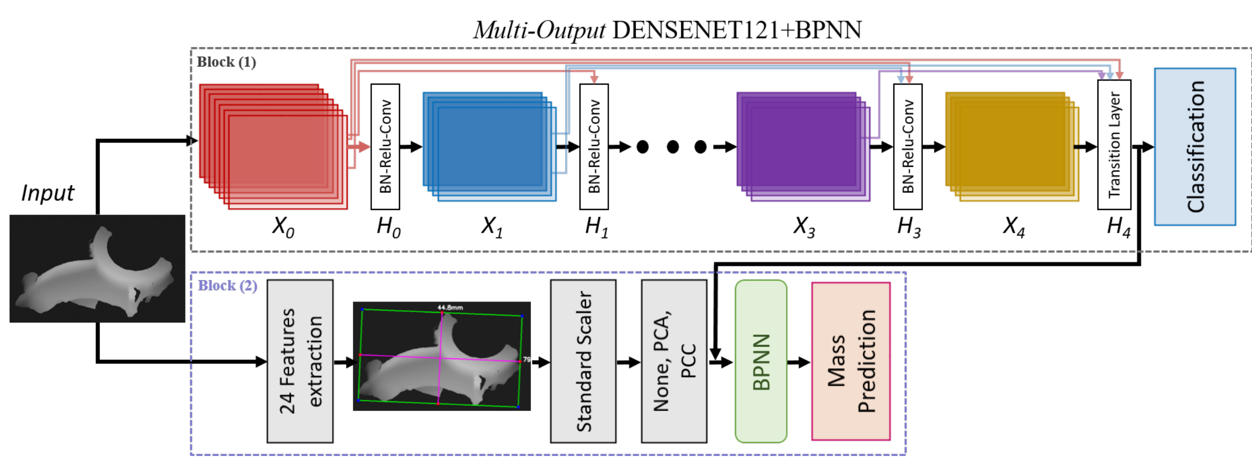

Fig.2 shows the proposed approach for mass prediction (Block 2) and classification (Block 1). Once a scrap piece of interest is cropped, it is fed into our pipe plan for classifying and predicting its mass.

Fig. 2 depicts the flowchart of the proposed combined architecture. It has two main building blocks: (1) the traditional DenseNet neural network for classification and (2) the BPNN for mass prediction powered with up to 24 extracted features. Combining both architectures is expected to positively impact metal scrap sorting by determining the average density of an object class and providing reasonable volumetric estimates for irregular shapes.

The mass estimation is correlated to the density and volume of the object, which strongly depends on the object class. Therefore, a better understanding of each objects' physical properties is achieved, and, thus, a robotic and/or pneumatic ejection system can effectively and accurately sort the metal scrap. Furthermore, the Pytorch library and Scikit-learn are used to modify the network architectures, train models, and evaluate the results.

3.1. *Feature Extraction*

Previous studies in the mass prediction's field for food processing (salmon, beef, pork, fruits) and household objects have used handcrafted features to determine an object's 3D properties and analyze its density and volume to aid in the understanding of the object's mass (Standley et al., 2017; Konovalov et al., 2019; L. Zhang et al., 2020; B. Zhang et al., 2020). The feature extraction is based on the bounding boxes (Bbox) extracted for all 3D images of the scrap metal objects, as depicted in Fig. 1c. We calculated the length (*L*), width (*W*) and center of the BBox, the maximum and minimum of the average image mask, and the object thickness's height correction (*Hc*). Then, we calculated the highest image mask point of the object to create a 3D BBox based on the *L* x *W* x *Hc*. The area features are calculated based on the 2D BBox defined by *L* x *W*, as shown in the right image in Fig. 1 based on the description in Table II. Finally, the mean, Std, and root in volumetric and area features were calculated. In total, we obtained 23 features from the cropped 3D image. Table I summarizes these features and states how they are derived. The 24th feature is the material type (*class*), which is only applied in the combined datasets to predict their mass and class simultaneously.

TABLE II
LIST OF EXTRACTED FEATURES AND EQUATIONS FROM 3D IMAGES

Extracted Features	**Feature Description**	**Symbols and formulas**
CenterX & CenterY	*X*, *Y* location of 2D BBox Center	Center(BBox)
CenterZ	height of the object in the 2D BBox Center	Ct = BBox[CenterX, CenterY]
Length & Width	ROI distances in the *X* & *Y* axes	*L* & *W*
HeightCorr (AvgHeight, MaxHeight and MinHeight)	Height correction is a subtraction between maximum and minimum height in the object mask (ImgMask)	*Hc* = Max(ImgMask)-Min(ImgMask) *Havg* = Avg(ImgMask)
Angle	The rotation angle of the bounding box	*A*
EstimatedArea FilledArea RootArea	Area estimation (*L*, *W*) is created based on the 2D BB The filled area is the correlation between the object's area and the empty space in the 2D BB Square root of the estimated area	EstArea = *prod*(*L*,*W*) FilledArea =*sum*(mask)/ EstArea rootEstArea = root(EstArea)
ImgIntensity ZeroIntensity NonZeroMask	Image intensity counts equal to 1 in ImgMask Zero intensity is the difference between image intensity and minimum height Image intensity counts equal to 0 in ImgMask	*ImgInt = Img[ImgMask==1]* *ZeroInt = ImgInt - Min*(ImgMask) *NonZero = sum*(*Img[ImgMask==0]*)
EstimatedVolume ObjVolume RootVolumen	Volume (L,W,Hc) is calculated based on the 3D BBox Object volume is the sum of *ZeroInt* or all the values in the thickness Square root of the volume estimation	EstVol = prod(*L*,*W*,*Hc*)) ObjVol = *sum*(ImgInt)/ EstArea rootEstVol = root(EstVol)
FilledVolume, EmptyVolume	Filled proportion is the correlation between the object's volume and the empty space in the 3D BB Empty space is the difference between the 3D BB minus the estimated volume.	FiPro = ObjVol /EstVol EmpVol = EstVol - ObjVol
SortedB/sortedM SortedB/sortedS SortedB*sortedM SortedM*sortedS	The sorted equations are proportional metrics to calculate the correction between *L*, *W*, *Hc*. The values are sorted between the bigger (B), middle (M), and smaller (S). The correlations between the sizes are calculated by dividing or multiplying them.	sB/sM = sortedB/sortedM sB/sS = sortedB/sortedS sB·sM = sortedB/sortedM sB·sS = sortedB/sortedS

3.2. *Feature Selection*

Feature selection is applied as a natural method to avoid redundant features and improve machine-learning model performance (Gharsalli et al., 2015). The first step consists of finding the correlation between the 24 features for the CW&SS datasets. For that, two different methods were used: (1) we compute the Pearson correlation coefficient (PCC) for each pair of features and, with them, the Pearson matrix (Sedgwick, 2012), as depicted in Fig. 3 and (2) we use the principal component analysis (PCA), as shown in Table III.

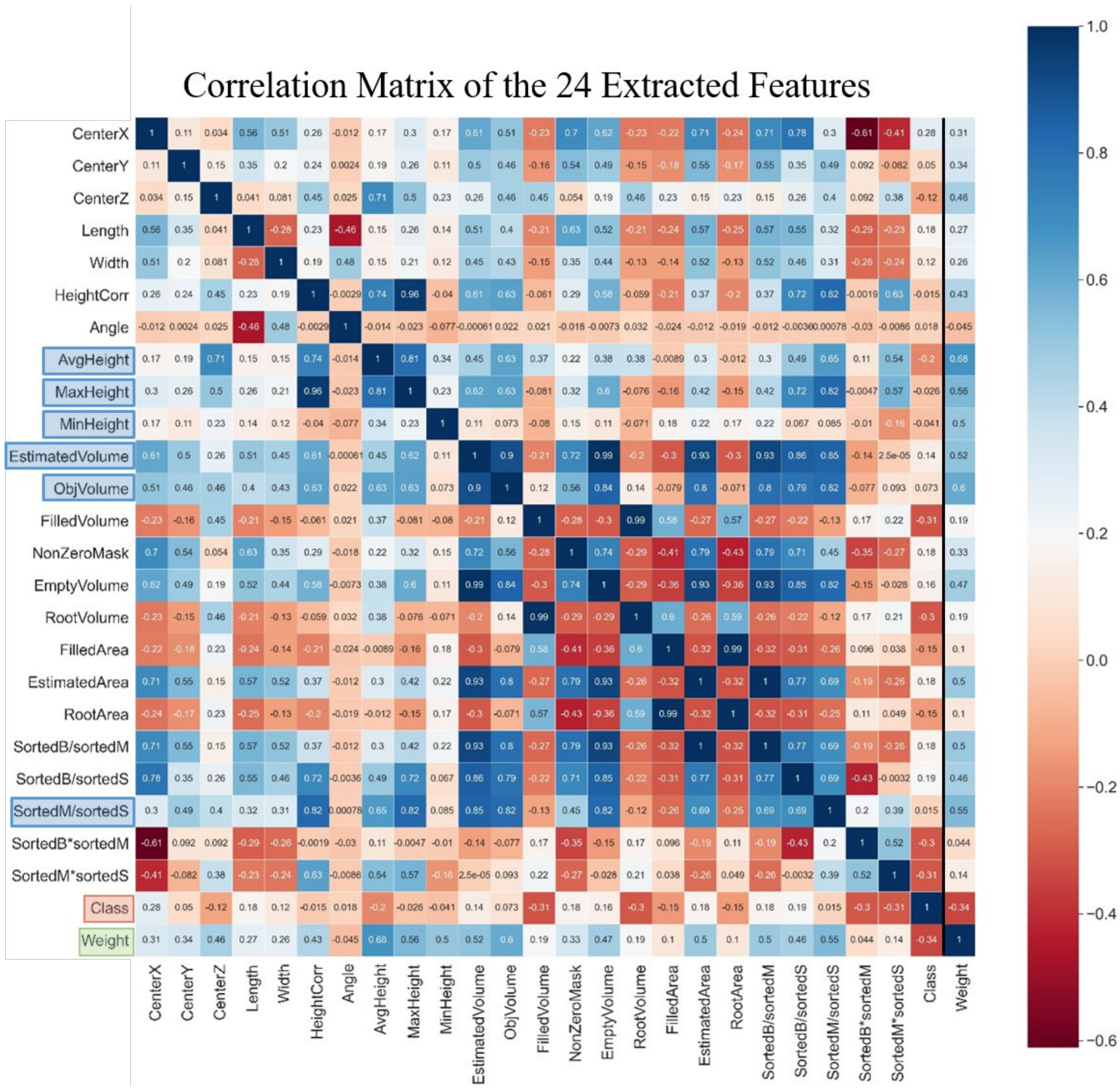

Fig. 3: Correlation heatmap based on Pearson's correlation coefficient between the 24 extracted features. The highlighted blue features are the only ones selected since they correlate greater than 0.5 regarding the object mass (weight). In contrast, the highlighted red feature (*class*) is selected as a substantial negative since its correlation value is smaller than -0.2. Finally, the highlighted green is the weight (mass) of the evaluation class.

The Pearson's correlation coefficient measures the linear association between the different features (seen as independent input variables) and the mass (weight) that acts as the output variable. The output coefficient ranges between -1, representing a stronger negative correlation, and 1, representing a stronger positive correlation (Sedgwick, 2012). Six features, namely, *AvgHeight, MaxHeight*, *MinHeight*, *EstimatedVolume, ObjVolume* and *SortedM/sortedS,* correlate higher than 0.5 concerning the weight, as a result, were selected. Additionally, features such as *SortedM*sortedS, Angle, and FilledVolume* do not affect the model's performance since their correlations with respect to the mass are close to zero.

Previous research demonstrated that using PCA as a feature selection method in the context of mass estimation can improve the performance of the regression model and the BPNN algorithms (L. Zhang et al., 2020). PCA is a multivariate statistical method for dataset dimension reduction that highlights those components with the most significant variance within the dataset (Wold et al., 1987). PCA identifies the relationships between characteristics and expresses them as a covariance matrix. Then, the existing data is converted into principal components using the eigenvalues of the covariance matrix. The most important features are selected, and the least relevant are eliminated (Wold et al., 1987). However, the data must be normalized before PCA is applied with a zero mean and variance equal to one.

The PCA method was adopted five times to calculate the feature selection, one per class – C, W, and SS – and two for their combinations – C&W and CW&SS (see Table II). The features depicted in each column of Table III correspond to those with a more significant influence on the components and the largest eigenvalues. The seven features highlighted in blue describe the common features between the five datasets used. Analogously, the feature highlighted in red (*class*) is relevant for the combined datasets C&W and CW&SS. The 12 features obtained have been used as input parameters for the metal scrap multi-output model. In particular, it is observed that both the PCA and PCC point out that the relevant features are *class*, *EstimatedVolume,* and *ObjVolume*. This is expected since the mass of an object can be defined as the multiplication of volume and density, where the volume is determined by the object's 3D geometry, while the density is determined by the object's material or class (Standley et al., 2017).

TABLE III
FEATURE EXTRACTION PER MATERIAL AND THEIR COMBINATIONS AND FEATURES SELECTED FOR MASS PREDICTION BASED ON PRINCIPAL COMPONENT ANALYSIS (PCA).
(THE STANDARD FEATURES BETWEEN METAL SCRAPS ARE HIGHLIGHTING IN BLUE AND RED FOR MULTICLASS REGRESSION)

The 11 Remaining Principal Features After the Dimension Reduction					
No	**Cast (C)**	**Wrought (W)**	**Stainless Steel (SS)**	**C&W**	**CW&SS**
1	EmptyVolume	EstimatedVolume	EstimatedVolume	EstimatedVolume	EstimatedVolume
2	SortedM*sortedS	RootVolume	RootVolume	RootVolume	AvgHeight
3	FilledArea	MinHeight	SortedM*sortedS	SortedM*sortedS	SortedM*sortedS
4	Angle	Width	Angle	MinHeight	Angle
5	SortedB*sortedM	SortedB*sortedM	SortedB*sortedM	SortedB*sortedM	SortedB*sortedM
6	MinHeight	FilledArea	MinHeight	MinHeight	MinHeight
7	CenterX	CenterY	RootArea	Class	Class
8	CenterY	Angle	CenterY	CenterY	CenterY
9	CenterZ	CenterZ	CenterZ	CenterZ	CenterZ
10	NonZeroMask	NonZeroMask	NonZeroMask	NonZeroMask	NonZeroMask
11	ObjVolume	CenterX	ObjVolume	ObjVolume	CenterX
12	SortedB/sortedM	ObjVolume	SortedB/sortedS	SortedM*sortedS	ObjVolume

3.3. *Mass Estimation Based on Machine learning*

Linear Regression (LR) (Pedregosa et al., 2011), Support Vector Regression (SVR) (Platt, 1999), K-Neighbors Regression (KNR) (Cover and Hart, 1967), Decision Trees Regression (DTR) (Quinlan, 1986), and Random Forests Regression (RFR) (Breiman, 2001) are the selected machine learning algorithms used to address how accurately the C, W, and SS mass can be estimated, based on their proven effectiveness *(B. Zhang et al., 2020; L. Zhang et al., 2020)*.

The library Scikit-learn 0.24 in Python was used to train and tune the model. The parameters adopted for the LR are the default parameters, while for SVR, they are kernel: 'RBF,' C: 300, gamma: 0.001 and degree: 3. For the KNR, the parameters are n_neighbors = 8, weights = uniform and algorithm 'auto,' while for the DTR, they are criteria: mse, min_samples_leaf = 2, and max_features = 4. The parameters adopted for the RFR are the number of trees in the forest: 192, criteria: mse, min_sample_split: 2, bootstrap: True, and oob_score: True. All the optimal parameters were found using Grid Search on the validation set, and the parameters not mentioned are set to their default values. Then, we compare the algorithm's performance for the machine-learning algorithms listed before and the BPNN for each metal scrap with and without feature selection.

4. Deep Learning Methodology and Evaluation Metrics

4.1. *Mass Estimation Based on The BPNN*

The BPNN was developed with the aim of solving the problems of training multi-layer perceptron, i.e., the problems derived from the use of hard-limit transfer functions, by adjusting each node of the network depending on the error rate obtained in the previous epoch (Rumelhart et al., 1986). The BPNN consists of one input, one hidden layer, and one output layer with activation functions after each layer. In addition, it has nonlinear noise assignment capabilities, and it exhibits excellent performance in various prediction domains (Utai et al., 2019; Liu et al., 2020; L. Zhang et al., 2020). As seen in the above-reviewed contributions, the BPNN is one of the most accurate and efficient ways of estimating the mass. Therefore, it has been selected as one of the building blocks for the mass estimation method proposed in this work. The designed BPNN (depicted in Fig. 4) has three layers: an input layer (with between 14 to 512 nodes), a hidden layer (with 45 to 150 nodes) with the rectified linear unit (ReLu) as an activation function (Glorot et al., 2011) and output layer (with one node) without a linear activation function.

Moreover, the number of nodes on the input and hidden layers can differ depending on the feature selection method applied in each experiment. Finally, the number of nodes was determined based on Kolmogorov's theorem, where the number of nodes in the hidden layer is determined by twice the number of nodes in the input layer plus one (i.e., $s = 2n + 1$, where s is the number of nodes in the hidden layer and n is the number of inputs) (Hecht-Nielsen, 1987).

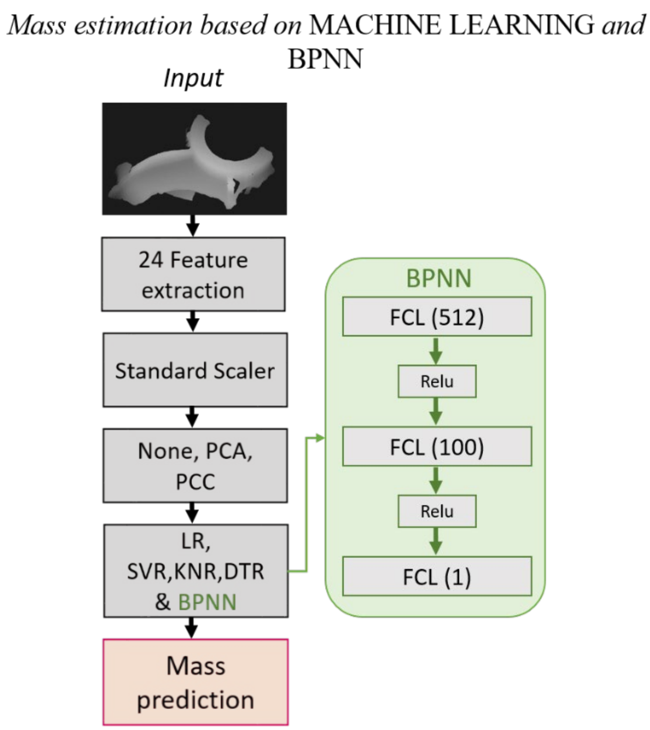

Fig. 4 The mass estimation pipeline consists of several machine learning algorithms and the BPNN. The general structure of the latter is further detailed in the green block.

4.2. *Multi-output DenseNet*

Deep Learning was inspired by the visual cortex system (Hubel and Wiesel, 1968), which provides a natural way for humans to communicate with digital devices (Sejnowski, 2020). One of the leading Deep Learning architectures for analyzing visual imagery is the convolutional neural network (CNN) (Valueva et al., 2020). With sufficient training, CNNs can determine a map of spatial and temporal dependencies, emphasizing the presence of a given characteristic in the image, such as the class, volume, colour, and/or shape information.

CNNs are typically constructed by combining convolutional, pooling and fully connected layers. Convolutional layers facilitate the extraction of different image characteristics by applying several filters and kernels. Pooling layers are used to select the most significant values in the feature maps and use them as input for subsequent layers. Finally, two or three fully connected layers are positioned at the end of the CNN to perform the classification, i.e., to estimate the probability of being in a given class.

Furthermore, a multi-output CNN can be built by adding layers to the end of its backbone. Typically, a CNN model, such as the Faster-RCNN, has two outputs for object detection: the bounding box (defined by a point, width and height), which is calculated with regression, and the object class, which is calculated in the classification layer. The multi-outputs added at the end of the network have a unique common loss function, formed by the weighted sum of the classification and regression loss functions (Ren et al., 2015). DenseNet was adapted to have multi-outputs to facilitate mass estimation and classification of scrap metals as part of this research.

DenseNet is a CNN architecture designed to mitigate the vanishing-gradient problem, reinforce feature propagation, reassure feature reuse, and substantially decrease the number of parameters (Huang et al., 2017). In conventional feed-forward neural networks, the layer's output constitutes the input of the subsequent layer after applying a function composition. In the case of DenseNet, each layer has direct access to the gradients from the loss function.

Whereas previously, only the feature map from the previous layer is fed to the next, the DenseBlocks strategy is implemented instead: the feature maps from all previous layers are concatenated and passed to all the subsequent layers, resulting in *deep supervision* as depicted in Fig. 2 block one and Fig. 5. The structure of DenseNet121 consists of four DenseBlocks, three transition layers, and an average pooling connected to a fully connected layer with a softmax activation. A DenseBlock comprises two separate convolutions of kernel sizes 1x1 and 3x3; the convolution operation is split into a depth, and channel-wise operation, respectively, which drastically speeds up the operation. Each transition layer halves the number of existing channels by using a 1x1-convolution layer and a 2x2 pooling layer between two consecutive DenseBlocks. Finally, a fully connected layer helps learn nonlinear combinations of the feature space for classification (Huang et al., 2017).

In general, the traditional DenseNet structure is used for scrap metal classification, which is simultaneously combined with a BPNN for mass prediction. This paper replaced the last layer of DenseNet121 with a linear regression output (DNR), which allows performing mass estimation without using additional features, as shown in Fig. 5. The DNR structure has the advantage that the extracted features of the CNN can be used to detect the mass estimation without handcrafted features. It also allowed a comparison between the extracted features in Table I and the features extracted from the network. Finally, two experiments were performed to evaluate the combination of the DenseNet+BPNN+PCA and DenseNet+BPNN+None.

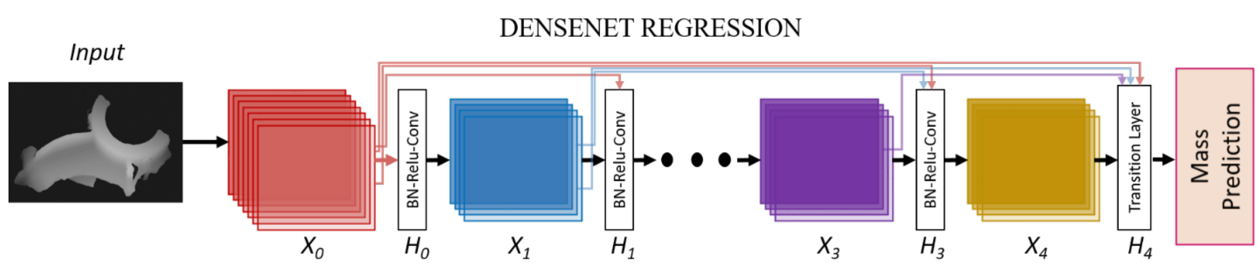

Fig. 5 shows the proposed approach for predicting the mass by replacing the last layer of DenseNet 121.

4.3. *Training Parameters and Loss Function*

Due to the absence of a large dataset, a fine-tuning transfer learning method was required for the training. For the fine-tuning, the model starts from a set of pre-trained parameters updated for the new task (to perform either regression or classification) by retraining the entire model.

In the performed experiment, we used a pre-trained DenseNet in Pytorch on the 100-class ImageNet dataset for fine-tuning, which has been successfully used in previous research (He et al., 2016; Schwarz et al., 2015). During the retraining, *Vertical* and *Horizontal Random Rotation* and *Color Jitter* are applied as data augmentation methods to enhance the image classification and the regression model (Perez and Wang, 2017; Wong et al., 2016).

The learning rate for the mass estimation and object classification is set to 0.01, while in the case of mass estimation with the BPNN, the learning rate is set to 0.001. In both cases, the stochastic gradient descent (SGD) (Sutskever et al., 2013) is used as an optimization method with a momentum of 0.92 and 0.95, respectively. In both experiments, 30 batches and over 120 epochs were trained. Moreover, the proposed architecture for metal scrap classification and mass estimation only needs a single input (as shown in Fig. 2).

DenseNet architectures use the Cross-Entropy loss function, which combines *LogSoftmax* and *Negative Log-likelihood Loss* (NLLloss) in one single function to improve the training of unbalanced datasets (Paszke et al., 2019). For the BPNN architecture, two different loss functions have been evaluated: the *Mean squared error* (MSELoss or L2-Squared norm) and the *Mean Absolute Error* (MAE or L1Loss). The addition of the cross-entropy loss function with one of the loss functions evaluated for the BPNN defines the loss function used in our proposed architecture.

4.4. *Evaluation Metrics*

The regression machine learning and Deep Learning algorithms were trained to find the best regression model. The performance of the regression was evaluated using three different metrics, namely R Square (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). They are defined in equations (4-6), where y_i represents the ground truth, $\hat{y}_i$ is the mass predicted value, $\bar{y}_i$ is known as vector f_i and N the number of elements.

R^2 is used to determine how well the model fits the dependent variables; RMSE measures how the residuals are distributed, showing how much the predicted mass deviates from the actual mass. Finally, MAE measures the average magnitude of the error in a prediction set without considering its direction.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{N}} \tag{5}$$

$$MAE = \frac{1}{N} \sum_i |y_i - \hat{y}_i| \tag{6}$$

The DenseNet+BPNN was trained only for the best models found in previous experiments, with and without feature selection for the C&W and CW&SS datasets. The performance of the classifiers was assessed using three quality indexes, namely Precision, Recall, and F1-score:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

Where TP is the number of true-positives, i.e., when the predicted class is "Cast" and the data is also labelled as "Cast," FP is the number of false-positives, i.e. when the data is labelled as "Wrought," but the predicted class is "Cast," and FN is the number of false-negatives, i.e. when the data is labelled as "Cast" but the predicted class is "Wrought" (Díaz-Romero et al., 2021). The F1-score is the harmonic mean of the Precision and Recall indices and is also used to evaluate the classification. It gives a better measure to evaluate the number of misclassifications in unbalanced datasets. Additionally, the area under the receiver operating characteristic curve (ROC) is used to evaluate several thresholds between Recall and the false-positive rate (FPR), defined as FP/ (FP+TN). The accuracy of the detection results (F1-score) is assessed using the Precision vs Recall curve, which focuses on evaluating the performance of a classifier for different probability thresholds on the minority class (He and Ma, 2013).

5. Results and Discussion

Section 5.1 examines how accurately the C, W, and SS mass is estimated. Furthermore, in Section 5.2, based on the obtained results, we investigate how accurately the C&W and the CW&SS are classified, and Deep Learning estimates their mass.

5.1. *Mass estimation Based on Machine Learning and Deep learning*

Table IV compares the test set's LR, SVR, KNR, DTR, RFR and BPNN repressors. Results show that the BPNN without feature selection generally performs best from the seven methods tested. For the proposed BPNN architecture, the C&W with PCA has an R^2 of 0.83, an RMSE of 0.17 and an MAE of 0.14. CW&SS without feature selection shows an R^2 of 0.76, an RMSE of 0.32 and an MAE of 0.24. Furthermore, in the C&W mass prediction cases, the BPNN with PCA enhances the regression by reducing the error by 0.11 and 0.04 for RMSE and MAE, respectively, while the R^2 score increased by 0.03. In general, the machine learning models across the entire test set perform with R^2 scores ranging between 47% and 77% for C, W and SS, concluding that RFR has the best and DRT has the worst performance.

Overall, the mass can be predicted based on 3D images through the features extracted from the images. Furthermore, the results show that feature selection does not provide a significant improvement. Since there is no standard established protocol for such studies, a direct comparison of the results is not possible. However, looking at the most closely related studies, we can see that our results are competitive (Konovalov et al., 2019; Agarwal et al., 2020; Liu et al., 2020; B. Zhang et al., 2020).

TABLE IV

COMPARISON OF REGRESSION PERFORMANCE OF THE TEST DATA SET FOR LINEAR REGRESSION (LR), SUPPORT VECTOR REGRESSION (SVR), K-NEIGHBORS REGRESSOR (KNR), DECISION TREE REGRESSION (DTR), RANDOM FOREST REGRESSION (RFR) AND BACKPROPAGATION NEURAL NETWORK (BPNN) WITH L2 LOSS FUNCTION

		RMSE (↓)					**MAE (↓)**					**R^2 (↑)**				
	Features	C	W	SS	CW	All	C	W	SS	CW	All	C	W	SS	CW	All
LR	None	0.60	0.60	0.85	0.49	0.56	0.40	0.41	0.70	0.36	0.44	0.68	0.65	0.43	0.77	0.74
	PCC	0.55	0.60	0.95	0.56	0.61	0.44	0.43	0.74	0.38	0.45	0.74	0.63	0.28	0.68	0.69
	PCA	0.58	0.60	0.76	0.50	0.56	0.42	0.41	0.67	0.36	0.44	0.70	0.65	0.55	0.76	0.73
SVR	None	0.71	0.70	0.75	0.65	0.63	0.53	0.49	0.63	0.42	0.43	0.55	0.52	0.56	0.60	0.66
	PCC	0.51	0.59	0.68	0.62	0.65	0.38	0.38	0.54	0.39	0.44	0.77	0.65	0.62	0.62	0.65
	PCA	0.73	0.71	0.75	0.64	0.63	0.56	0.49	0.63	0.42	0.43	0.52	0.51	0.56	0.60	0.67
KNR	None	0.82	0.56	0.82	0.59	0.63	0.60	0.38	0.74	0.38	0.44	0.40	0.69	0.46	0.66	0.67
	PCC	0.67	0.65	0.74	0.64	0.62	0.51	0.43	0.67	0.39	0.44	0.60	0.58	0.56	0.60	0.68
	PCA	0.82	0.56	0.83	0.59	0.62	0.60	0.38	0.75	0.38	0.43	0.39	0.69	0.46	0.67	0.68
DTR	None	0.77	0.72	1.04	0.61	0.69	0.56	0.44	0.82	0.43	0.48	0.47	0.49	0.14	0.64	0.59
	PCC	0.84	0.75	0.92	0.73	0.75	0.63	0.48	0.66	0.50	0.51	0.37	0.44	0.31	0.47	0.53
	PCA	0.97	0.79	1.04	0.80	0.72	0.72	0.49	0.83	0.54	0.52	0.15	0.39	0.14	0.39	0.56
RFR	None	0.53	0.60	0.88	0.49	0.53	0.41	0.41	0.70	0.33	0.39	0.75	0.64	0.39	0.77	0.76
	PCC	0.57	0.67	0.85	0.64	0.64	0.43	0.46	0.64	0.41	0.46	0.71	0.55	0.42	0.60	0.66
	PCA	0.64	0.61	0.84	0.57	0.60	0.49	0.43	0.74	0.40	0.44	0.63	0.63	0.44	0.69	0.69
BPNN	**None**	**0.46**	**0.54**	**0.67**	0.28	**0.32**	**0.34**	**0.37**	**0.48**	0.18	**0.24**	**0.78**	**0.82**	**0.71**	0.82	**0.76**
	PCC	0.51	0.73	0.73	0.75	0.44	0.35	0.52	0.52	0.48	0.24	0.72	0.78	0.53	0.81	0.74
	PCA	0.50	0.55	0.70	**0.17**	0.61	0.37	0.36	0.51	**0.14**	0.38	0.75	0.72	0.62	**0.83**	0.75

Standley *et al*. used RGB images to estimate the object's mass (Standley et al., 2017). In particular, they proposed using two Xception networks and 14 features to calculate the object's density and volume and then estimate its mass. The first Xception network was used to compute the bounding box and, thus, the 3D volume of the object. Then, the results obtained were fused to the second Xception network to estimate the object's density. In order to evaluate the system, two datasets were used: the household test set (56 items, 423 images) and the amazon test set (924 items). Overall, the household test set performed at 0.69 R^2, 0.67 RMSE and 0.68 MAE, while the Amazon test set performed at 0.77 R^2, 0.67 RMSE and 0.61 MAE. In addition, the study showed the mass prediction performance of 4 participants in the household dataset, achieving an R^2 score between 0.49 and 0.68 for the mass estimation.

Zhang *et al*. designed a dataset for fish mass estimation (455 images) using image analysis and neural networks (Zhang et al., 2020). The adopted approach aimed to use image segmentation, enhancement and pre-processing. A total of 14 features were extracted, filtering the best of them by using PCA. Finally, the fish mass was estimated by using the BPNN architecture. Overall, their system showed a performance of 0.90 R^2, 0.01 RMSE and 0.01 MAE; Although a direct comparison with previously performed mass-estimation research is not possible, the error obtained in our proposed method might be more significant since we are not using homogenous objects such as fish or fruits (Konovalov et al., 2019; Utai et al., 2019).

Before combining the BPNN with DenseNet121, the DenseNet-Regression (DNR) algorithm shown in Fig. 5 was evaluated, as shown in Table V. The results show the performance of the two-loss functions L1 and L2 for the mass prediction of scrap metals on the test set. Overall, the best performance was achieved with the DNR and L2. However, the RMSE error is 0.02 lower for the W mass prediction using L1. A DNR network without any additional features could predict the mass of metals scrap objects based on a 3D image and a pre-trained network, obtaining an R^2 score between 0.61 and 0.77. DNR models generally have a lower RMSE and MAE due to the gradient descent optimization applied during training and their multiplex iterations.

TABLE V
COMPARISON OF REGRESSION PERFORMANCE OF TEST DATA SET FOR DENSENET REGRESSION (DNR) BY JUST USING DEEP LEARNING WITHOUT FEATURE EXTRACTION

		RMSE (↓)					**MAE (↓)**					**R^2 (↑)**				
	Loss	C	W	SS	CW	All	C	W	SS	CW	All	C	W	SS	CW	All
DNR	L1	0.61	**0.16**	0.82	0.42	0.56	0.62	0.14	0.61	0.32	0.36	0.59	0.65	0.40	0.68	0.64
	L2	**0.54**	0.18	**0.68**	**0.27**	**0.28**	**0.35**	**0.14**	**0.50**	**0.23**	**0.20**	**0.70**	**0.61**	**0.59**	**0.77**	**0.72**

Nonetheless, the performance of the DNR is not better than the BPNN because of insufficient data and because the retraining of the networks was done with unbalanced classes. The authors believe DNR could outperform BPNN with a more comprehensive and balanced training set for each class. The study of Konovalov *et al.* for mass estimation in the research field of agriculture showed that by using the CNN, the mass of an object could be predicted with a high R^2 score and low error (Konovalov et al., 2019). In the presented case, the use of additional features improved the robustness of the model for unknown new metal scrap, resulting in a better overall performance. Although the mass prediction by computer vision is not as accurate as measuring mass with a scale, it still provides an essential approximation allowing monitoring of waste composition in an early stage.

5.2. *Mass Estimation And Classification of Metal Scrap based on Deep Learning*

The best regression performances were obtained for the BPNN+PCA and the BPNN+None without feature selection for C&W and CW&SS, respectively, as shown in Table IV.

The DenseNet+BPNN+PCA model results for the C&W test dataset are shown in Fig. 6, containing four subplots. Fig. 6a shows the classification results, indicating that C&W can be classified with a weighted average F1-score and Precision of 95%. However, the proposed classification is solely based on 3D images, leading to a slightly lower classification score than using fused RGB and 3D images, and the Recall for the C is around 76% due to the training with an imbalanced dataset. A marginally lower recall could produce a lower recovery volume, reducing the marginal benefit of scrap metals, resulting in moderately increased scrap metal recycling action costs. Fig. 6b depicts the regression results, performing at 0.82 for R^2, 0.2 for RMSE, and 0.28 for MAE for the DenseNet+BPNN+PCA model. The resulting regression lines with a 95% confidence interval for each regression are intended to show only the data trend, presenting a slightly higher slope for the Cast class. In general, the performance of DenseNet+BPNN+PCA (output: regression + classification) *vs* BPNN+PCA (output: regression) is not significantly divergent with a score difference of 0.06 for R^2, 0.02 for RMSE, and 0.14 for MAE. However, it should be noted that DenseNet+BPNN+PCA has the advantage of a multi-output pipeline compared to a single-output as in the case of BPNN+PCA, due to the possibility of classifying and estimating the mass of scrap metal pieces. Fig. 6c represents the Precision-Recall curve; the best result obtained on the test data was a Recall of 0.96 with a Precision of 0.94. The classification performance model at all classification thresholds is shown in Fig. 6d, where an area value of 0.81 and 0.94 have been achieved for C and W, respectively. The evaluation of the ROC curve is used to determine the most favourable operating point depending on the application function. A 0.82 TP at 0.18 FP rate is obtained in the presented results. The relatively high rate of FPs is expected to be a result of the absence of a color camera. Specifically, the red channel of the color image is relevant for differentiating materials with similar shapes and degrees of light absorption/reflection, such as C and W Al (Díaz-Romero et al., 2021).

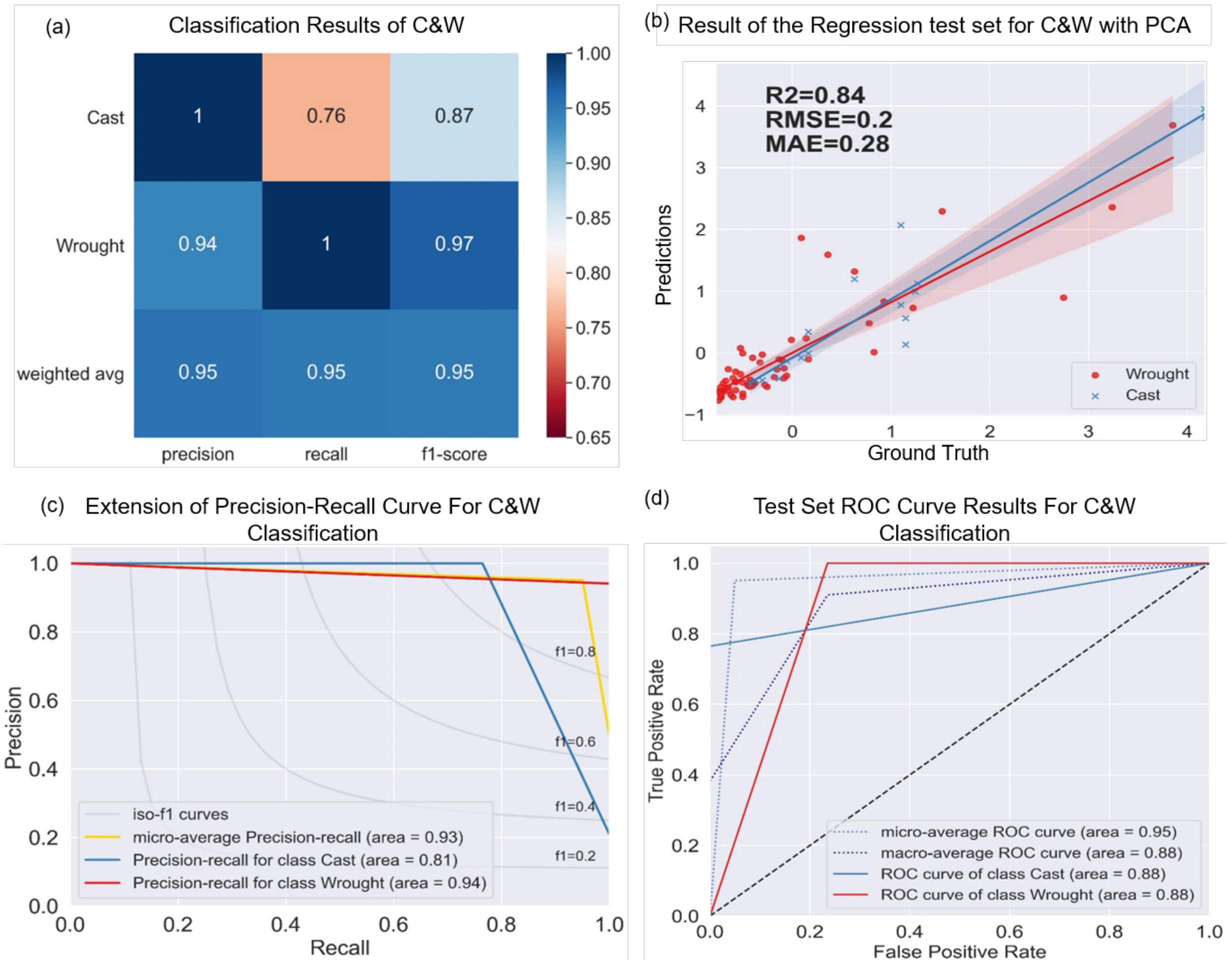

Fig. 6: Results of the C&W Al classification and mass estimation: (a) Classification results including the following evaluation metrics: weighted average, F1-Score, Recall, and Precision; (b) Regression results using the DenseNet+BPNN+PCA architecture and including the R^2, RMSE and MAE metrics, as well as the resulting regression lines with a 95% confidence interval for each regression (intended to show only the data trend); (c) The Precision-Recall curve, showing the balance between Precision and Recall for different thresholds and (d) The ROC curve, which represents the performance of the proposed classification model at all classification thresholds.

Overall, these experiments demonstrate that using DenseNet with the BPNN is a novel and promising alternative for mass estimation and C&W classification with high performance. The proposed method could be adapted to different materials and used as a first-step monitoring system to assess performance during (pre-) sorting. Furthermore, the system can be used at the end of the recycling line to enhance the understanding of the objects' physical characteristics, which, in turn, could enhance the control of a robotic and/or pneumatic sorting system.

The results for the CW&SS test dataset using the DenseNet+BPNN+None model are shown in Fig. 6. Compared to the C&W dataset, there is a significant reduction in the classification and regression performance because the characteristics of the SS class, such as shape and size, are similar to those of the W class, resulting in higher misclassification between the W and SS classes. However, this problem could be solved by using an RGB camera, which has a clear difference between the material colour and reflectance properties for the human eye.

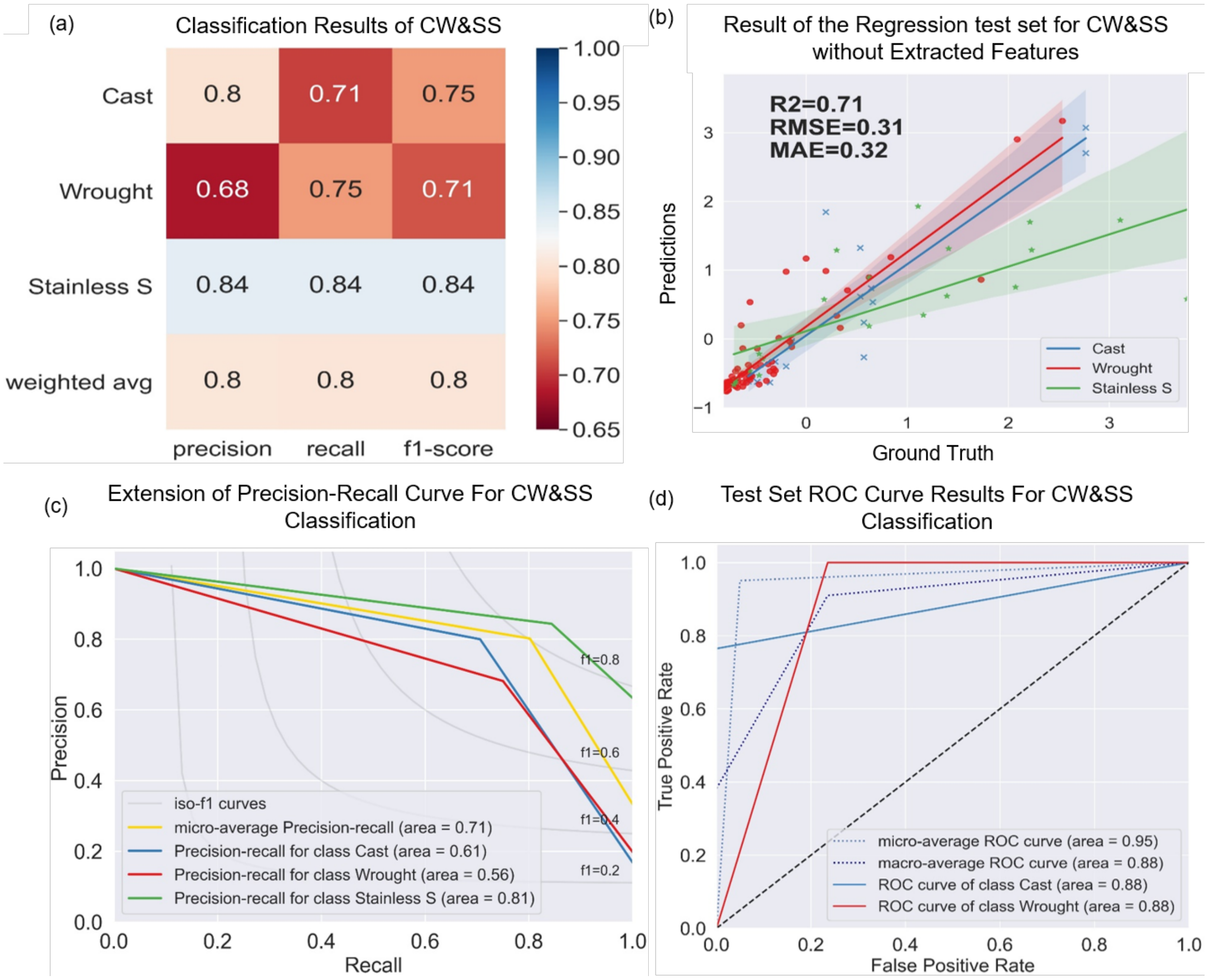

Fig. 7: Results of the CW&SS classification and mass estimation are shown.

The model's performance is evaluated using the Precision-Recall and ROC curves (Fig. 7c and d), achieving, in general, a micro-average area of 0.71 and 0.85 for the testing data, respectively. The best performance was obtained for a recall of 0.81 and a precision of 0.80 (see Fig. 7c). The best performance for the ROC curve can be seen in Fig. 7d at a 0.81 TRP with a 0.19 FPR.

Fig. 7a shows that stainless steel classification has a higher performance than the other two classes with an F1-score, Precision and Recall of 84%. In general, the classification has a weighted average performance of 80% for all the classification metrics, showing the possibility to use the intensity and 3D images for multi-object detection. The regression results are shown in Fig. 7b, representing the resulting regression lines for mass prediction and performance of the DenseNet+BPNN+None model with 0.71 for R^2, 0.31 for RMSE, and 0.32 for MAE. The trend of the SS class lines differs from that of the C&W class due to the density differences between SS and Al, which range from 7,500kg/m^3 to 8,000kg/m^3 and 2,640kg/m^3 to 2,810kg/m^3, respectively.

6. Envisaged Industrial Application

The first envisaged industrial application is the use of the developed method for assessing the composition and purity of mixed plastic and metal waste streams. To trade most of these waste fractions, minimal weight-based purity targets need to be reached, where higher purities typically result in a higher market value. The waste streams' composition, shape, and mass distribution can vary significantly depending on the process input mix. Therefore, a simple count of the detected objects per class does not accurately estimate a weight-based material composition. The classification and mass estimation techniques presented in this work offer opportunities to provide better insight into the actual purity achieved thresholds.

In addition, compositional information can be used for improved recycling process control. Al remelters producing secondary wrought Al alloys only buy scrap that meets the specific compositional constraints (Dispinar and Campbell, 2004). Therefore, recycling companies that operate a sorting process desire to maximize the amount of material that can be commercialized as a wrought fraction, which can be marketed at a higher value while still meeting the remelter's composition requirements. Since it is inherent of a sorting process that a trade-off needs to be made between a higher purity and a higher yield, the proposed method can provide helpful information on the actual weight composition achieved of the sorted fraction by considering the weight of all sorted objects. Therefore, in future research, the benefits of using the developed method to optimize the output purity of a sorting system with a laser-induced breakdown spectrometer, which can provide information on the alloy composition of every object, will be investigated.

Another envisaged application is using the proposed method to enhance the control of a pneumatic valve block and/or a robotic gripping system. Nowadays, the duration of the valve opening or the gripper to be used and the robot path are either fixed or solely based on the geometrical information extracted from (depth) images. Hence, integrating the developed class and weight prediction methods enable enhanced control of these sorting mechanisms. It allows the use of the semantic, geometric and physical properties calculated for every object.

7. Conclusion and Future Work

The presented results demonstrate the potential of state-of-the-art machine learning techniques and Deep Learning for simultaneous mass estimation and classification of scrap metal objects to enhance the control of either or both robotic and pneumatic sorting systems.

The study investigates the benefits and limitations of machine learning, BPNN and DenseNet for mass estimation. Furthermore, it identifies the best feature selection methods and the most suitable algorithms to work only with the data extracted from a 3D camera. The results obtained with the CNN DenseNet and the BPNN show that the developed method could monitor the proportion of metal classes based on their mass estimation. The best results for mass prediction were obtained with BPNN+PCA and BPNN+None, attaining an R^2 of 0.83, an RMSE of 0.17 and an MAE of 0.14, and an R^2 of 0.76, an RMSE of 0.32 and an MAE of 0.24, respectively. Therefore, the mass prediction method can be considered a follow-up or supplementary system in sorting C&W and CW&SS. In addition, it has a significant potential to develop a better understanding of the physical properties of an object which, in turn, will be helpful for its manipulation in automated systems.

Additionally, the experiments presented demonstrate that DenseNet+BPNN+PCA and DenseNet+BPNN+None can classify and predict the object's mass without losing performance in its classification. The results of the DenseNet+BPNN+PCA model for the C&W test data are 0.82 for the R^2, 0.20 for the RMSE, 0.28 for the MAE. The classification performance is 95%, computed as the weighted average of the F1-score, Recall and Precision indexes. The DenseNet+BPNN+None applied to the CW&SS test data has a weighted average performance of 80% for all the ranking metrics and 0.71 R^2, 0.31 RMSE, and 0.32 MAE for the regression metrics.

The data sets will be scaled up and balanced to reduce bias and increase the network's performance in future experiments. In addition, the dataset will be extended for light and heavy metals to explore whether density detection can improve the class detection of different metals. We will further develop an early end-to-end Deep Learning system to monitor the mass and classes of impurities and recycled materials and integrate our approach into the CNN mask. . Furthermore, the researchers will evaluate whether the combination of the regression and classification could enhance or help improve the classification prediction from different materials based on their density deviations. Finally, the system will be integrated into a real-time system for sorting aluminum

alloys, helping to reduce the threat of scrap surplus, and perhaps, more value could be recovered from post-consumer aluminum scrap.

Acknowledgments

Thanks to F. Arslan for constructive discussions, revision and support. This activity has received funding from the European Institute of Innovation and Technology (EIT), a body of the European Union, under the Horizon 2020, the EU Framework Programmed for Research and Innovation (project name: automatic sorting of mixed scrap metals, project number: 19294).

Reference

Agarwal, R., Díaz, O., Yap, M.H., Llado, X., Marti, R., 2020. Deep learning for mass detection in Full Field Digital Mammograms. Comput. Biol. Med. 121, 103774.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S., Xiong, X., 2018. Multilayer hybrid deep-learning method for waste classification and recycling. Comput. Intell. Neurosci. 2018.

Correll, N., Bekris, K.E., Berenson, D., Brock, O., Causo, A., Hauser, K., Okada, K., Rodriguez, A., Romano, J.M., Wurman, P.R., 2016. Lessons from the Amazon Picking Challenge. CoRR abs/1601.05484 (2016).

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27.

Cullen, J.M., Allwood, J.M., 2013. Mapping the global flow of aluminum: From liquid aluminum to end-use goods. Environ. Sci. Technol. 47, 3057–3064.

Díaz-Romero, D., Sterkens, W., Van den Eynde, S., Goedemé, T., Dewulf, W., Peeters, J., 2021. Deep learning computer vision for the separation of Cast-and Wrought-Aluminum scrap. Resour. Conserv. Recycl. 172, 105685.

Dispinar, D., Campbell, J., 2004. Metal quality studies in secondary remelting of aluminium. Foundry Trade J. 178, 78–81.

Eggers, A., Peeters, J.R., Waignein, L., Noppe, B., Dewulf, W., Vanierschot, M., 2019. Development of a computational fluid dynamics model of an industrial scale dense medium drum separator. Eng. Appl. Comput. Fluid Mech. 13, 1001–1012.

Gharsalli, S., Emile, B., Laurent, H., Desquesnes, X., Vivet, D., 2015. Random forest-based feature selection for emotion recognition, in: 2015 International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, pp. 268–272.

Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks, in: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, pp. 315–323.

He, H., Ma, Y., 2013. Imbalanced learning: foundations, algorithms, and applications.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hecht-Nielsen, R., 1987. Kolmogorov's mapping neural network existence theorem, in: Proceedings of the International Conference on Neural Networks. IEEE Press New York, pp. 11–14.

Hotta, Y., Visvanathan, C., Kojima, M., 2016. Recycling rate and target setting: challenges for standardized measurement. J. Mater. Cycles Waste Manag. 18, 14–21.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4700–4708.

Johnson, J.X., McMillan, C.A., Keoleian, G.A., 2013. Evaluation of life cycle assessment recycling allocation methods: The case study of aluminum. J. Ind. Ecol. 17, 700–711.

Konovalov, D.A., Saleh, A., Efremova, D.B., Domingos, J.A., Jerry, D.R., 2019. Automatic weight estimation of harvested fish from images, in: 2019 Digital Image Computing: Techniques and Applications (DICTA). IEEE, pp. 1–7.

Liu, B., Wang, R., Zhao, G., Guo, X., Wang, Y., Li, J., Wang, S., 2020. Prediction of rock mass parameters in the TBM tunnel based on BP neural network integrated simulated annealing algorithm. Tunn. Undergr. Space Technol. 95, 103103.

Mao, W.-L., Chen, W.-C., Wang, C.-T., Lin, Y.-H., 2021. Recycling waste classification using optimized convolutional neural network. Resour. Conserv. Recycl. 164, 105132.

Modaresi, R., 2015. Dynamics of aluminum use in the global passenger car system: challenges and solutions of recycling and material substitution.

Modaresi, R., Müller, D.B., 2012. The role of automobiles for the future of aluminum recycling. Environ. Sci. Technol. 46, 8587–8594.

Mordvintsev, A., Abid, K., 2014. Opencv-python tutorials documentation. Obtenido Httpsmedia Readthedocs Orgpdfopencv-Python-Tutroalslatestopencv-Python-Tutroals Pdf.

Nelen, D., Manshoven, S., Peeters, J.R., Vanegas, P., D'Haese, N., Vrancken, K., 2014. A multidimensional indicator set to assess the benefits of WEEE material recycling. J. Clean. Prod. 83, 305–316.

Nijhof, G.H., 1994. Aluminium separation out of household waste using the Eddy Current technique and reuse of the metal fraction. Resour. Conserv. Recycl. 10, 161–169.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., 2019. Pytorch: An imperative style, high-performance deep learning library. ArXiv Prepr. ArXiv191201703.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Perez, L., Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. ArXiv Prepr. ArXiv171204621.

Platt, J., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv. Large Margin Classif. 10, 61–74.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81–106.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 28, 91–99.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. nature 323, 533–536.

Schwarz, M., Schulz, H., Behnke, S., 2015. RGB-D object recognition and pose estimation based on pre-trained convolutional neural network features, in: 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 1329–1335.

Sedgwick, P., 2012. Pearson's correlation coefficient. Bmj 345.

Sejnowski, T.J., 2020. The unreasonable effectiveness of deep learning in artificial intelligence. Proc. Natl. Acad. Sci. 117, 30033–30038.

Shao, L., Cai, Z., Liu, L., Lu, K., 2017. Performance evaluation of deep feature learning for RGB-D image/video classification. Inf. Sci. 385, 266–283.

Standley, T., Sener, O., Chen, D., Savarese, S., 2017. image2mass: Estimating the Mass of an Object from Its Image, in: Conference on Robot Learning. PMLR, pp. 324–333.

Sterkens, W., Diaz-Romero, D., Goedemé, T., Dewulf, W., Peeters, J.R., 2021. Detection and recognition of batteries on X-Ray images of waste electrical and electronic equipment using deep learning. Resour. Conserv. Recycl. 168, 105246.

Sutskever, I., Martens, J., Dahl, G., Hinton, G., 2013. On the importance of initialization and momentum in deep learning, in: International Conference on Machine Learning. PMLR, pp. 1139–1147.

Suzuki, S., 1985. Topological structural analysis of digitized binary images by border following. Comput. Vis. Graph. Image Process. 30, 32–46.

Utai, K., Nagle, M., Hämmerle, S., Spreer, W., Mahayothee, B., Müller, J., 2019. Mass estimation of mango fruits (Mangifera indica L., cv.'Nam Dokmai') by linking image processing and artificial neural network. Eng. Agric. Environ. Food 12, 103–110.

Valueva, M.V., Nagornov, N.N., Lyakhov, P.A., Valuev, G.V., Chervyakov, N.I., 2020. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. Math. Comput. Simul. 177, 232–243.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2, 37–52.

Wong, S.C., Gatt, A., Stamatescu, V., McDonnell, M.D., 2016. Understanding data augmentation for classification: when to warp?, in: 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, pp. 1–6.

Zhang, B., Guo, N., Huang, J., Gu, B., Zhou, J., 2020. Computer Vision Estimation of the Volume and Weight of Apples by Using 3D Reconstruction and Noncontact Measuring Methods. J. Sens. 2020.

Zhang, L., Wang, J., Duan, Q., 2020. Estimation for fish mass using image analysis and neural network. Comput. Electron. Agric. 173, 105439.

Zhang, S., Chen, Y., Yang, Z., Gong, H., 2021. Computer vision based two-stage waste recognition-retrieval algorithm for waste classification. Resour. Conserv. Recycl. 169, 105543.