Findings from machine learning in clinical medical imaging applications – Lessons for translation to the forensic setting

Carlos A. Peña-Solórzano[a,*], David W. Albrecht[b], Richard B. Bassed[c,d], Michael D. Burke[c,d], Matthew R. Dimmock[a]

[a] Department of Medical Imaging and Radiation Sciences, Monash University, Wellington Rd, Clayton, Melbourne, VIC 3800, Australia
[b] Clayton School of Information Technology, Monash University, Wellington Rd, Clayton, Melbourne, VIC 3800, Australia
[c] Victorian Institute of Forensic Medicine, 57-83 Kavanagh St., Southbank, Melbourne, VIC 3006, Australia
[d] Department of Forensic Medicine, Monash University, Wellington Rd, Clayton, Melbourne, VIC 3800, Australia

ABSTRACT

Machine learning (ML) techniques are increasingly being used in clinical medical imaging to automate distinct processing tasks. In post-mortem forensic radiology, the use of these algorithms presents significant challenges due to variability in organ position, structural changes from decomposition, inconsistent body placement in the scanner, and the presence of foreign bodies. Existing ML approaches in clinical imaging can likely be transferred to the forensic setting with careful consideration to account for the increased variability and temporal factors that affect the data used to train these algorithms. Additional steps are required to deal with these issues, by incorporating the possible variability into the training data through data augmentation, or by using atlases as a pre-processing step to account for death-related factors. A key application of ML would be then to highlight anatomical and gross pathological features of interest, or present information to help optimally determine the cause of death. In this review, we highlight results and limitations of applications in clinical medical imaging that use ML to determine key implications for their application in the forensic setting.

© 2020 Elsevier B.V. All rights reserved.

Contents

---

* Corresponding author.
  E-mail addresses: carlos.penasolorzano@monash.edu
(david.albrecht@monash.edu (D.W. Albrecht), richard.bassed@vifm.org (R.B. Bassed), michael.burke@vifm.org (M.D. Burke), matthew.dimmock@monash.edu
(M.R. Dimmock).

1. Introduction

Forensic radiology is not clinical radiology applied to a deceased person. In the forensic setting, findings that a clinical radiologist may not typically have encountered are commonplace [1], e.g. post-mortem gas formation [2]. Post-mortem computed tomography (PMCT) is widely used in forensic investigations, where acquisition protocols used during clinical CT are not applicable due to rigor mortis and aversion to repositioning the decedent to avoid tampering with evidence. However, CT scans can be acquired with higher doses and there is no patient motion, therefore improving image quality. Additionally, recent developments such as PMCT angiography (PMCTA) with specialized pumps allows the diagnosis of vascular lesions whilst maintaining the integrity of anatomic structures, thus preserving evidence integrity [3,4].

In order to overcome the limitations of soft tissue contrast and a lack of vascular visualization provided by PMCT [5], postmortem magnetic resonance imaging (PMMRI) is increasing in impact, albeit in a small way thus far. Whilst PMMRI offers improved soft tissue contrast, for vascular diagnoses it presents similar performance to PMCTA, with higher associated cost. However, applications to cardiac imaging are an exception, due to improved visualization of the coronary arteries and myocardium [5].

Recently, there has been much progress in the automation of image processing tasks to enhance medical imaging workflow [6,7] for the key modalities of plain film X-ray [8], CT [9,10], and MRI [11]. In the forensic setting, the completion of these tasks suffers from added complications such as decedent decomposition, trauma, incineration, variability in positioning of normal anatomical structures, and artefacts from foreign bodies. This review initially introduces the basic concepts of machine learning (ML) that relate to image processing, before discussing the limited literature on the use of ML in post-mortem forensic imaging. The review then synthesizes the existing literature on the relevant preclinical and clinical uses of ML, and contextualizes the information relative to future use in the forensic setting. Whilst the use of MRI is not yet widespread in forensic medicine, its growing popularity and extensive use with ML in clinical imaging yields important conclusions for long-term forensic implementation considerations. In addition, it should be noted that whilst there are extensive applications for the use of ML in both clinical and forensic histopathology, these are not considered.

2. Machine learning (ML) algorithms

Image processing typically involves segmentation, feature extraction, and classification. Image segmentation refers to the partitioning of a digital image into multiple segments that are sets of pixels (or voxels) which usually represent discrete structures. Approaches to image segmentation prior to ML included probabilistic atlases [12,13], statistical shape models (SSMs) [14,15], graph-cut (GC) algorithms [16,17], and multi-atlas segmentation (MAS) [18]. Feature extraction is a dimensionality reduction technique used to efficiently represent parts of an image as a compact feature vector. Feature extraction was traditionally performed through determining properties such as first order textures (e.g. mean or entropy) or correlations [19,20]. Image classification is the process of taking an image or volume and predicting whether it belongs to a list of predefined classes. Traditional approaches to classification included linear- and normal-discriminant analysis [21,22]. A variety of ML alternatives to each of these image processing tasks have now been proposed and pipelines that can automate many diagnostic and prognostic tasks have been introduced to reduce the burden on radiologists [23,24].

ML techniques can be categorized as supervised learning, unsupervised learning, and reinforcement learning. In supervised environments, data is composed of input-output patterns, and the task is to find a deterministic function that can predict the output from an observed input. Unsupervised techniques are a type of self-organized learning that extracts structures from the training samples directly, without pre-existing labels [25]. More recently, self-supervised techniques, a type of unsupervised learning where the training data is automatically labelled by exploiting the relations between different input signals, are being studied for better utilizing unlabeled data [26]. Reinforcement learning on the other hand is based on trial-and-error, where the algorithm evaluates a current situation, takes an action, and receives feedback from the environment; this feedback can be positive or negative [27]. The most common ML techniques used in medical applications are summarized below.

2.1. Random forests (RFs)

RFs operate by creating a multitude of decision trees (Fig. 1) that can be trained for classification and regression tasks [28,29], where the output is obtained by majority vote. Majority vote is a technique utilized to combine the outputs from multiple classifiers, with the voting rule following one of three forms: (i) unanimous voting, where all the individual votes must agree in one output class, (ii) simple majority, where the class with one more than 50 % of votes is selected, and (iii) plurality or majority voting, where the class with the highest number of votes is chosen [30].

2.2. k-nearest neighbors (k-NN)

In *k*-NN, the training samples are divided into classes, and the prediction of a new sample or test point is classified by a majority vote of its neighbors (Fig. 2). The algorithm uses a distance measurement function to search the *k* (defined by the user) closest training samples in the feature space, and assigns the case of the class that is the most common in the subset.

2.3. Support vector machines (SVMs)

SVMs originated from statistical learning theory [31] and are used for classification as they can model highly non-linear systems. SVMs project the data onto a high-dimensional space and apply a linear classifier on the projected data (Fig. 3) [25,32]. SVMs are intrinsically more suited to two-class problems as opposed to RFs which are best for those with multiple classes.

Fig. 3 SVM maps the input data (left) to a high-dimensional feature space (right) and calculates a hyperplane able to separate the different classes. The class assigned to new samples is decided according to the location of the data points on the high-dimensional space with respect to the generated hyperplane.
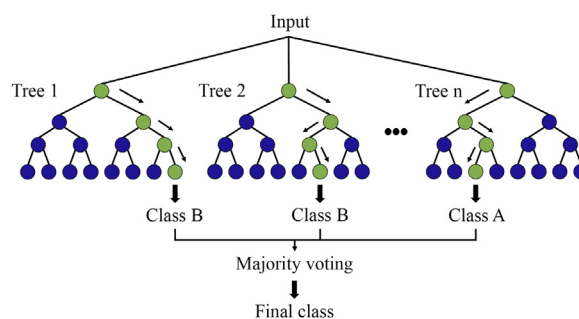
Fig. 1. RFs present a multitude of decision trees at training and outputs a class according to a defined majority vote technique (for classification tasks) or mean prediction (for regression tasks). In this example, the result would be the selection of Class B.
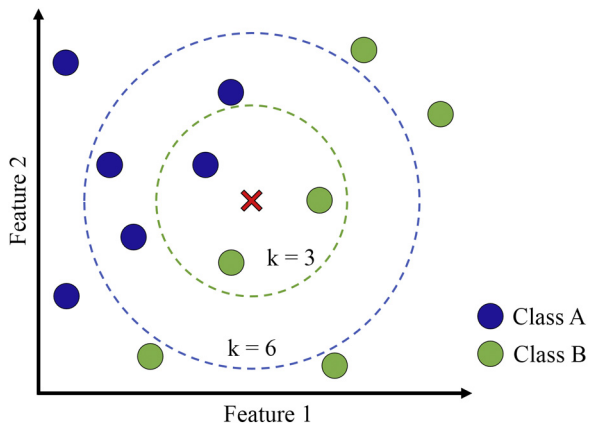
Fig. 2. *k*-NN assigns a class to the new data point, denoted here by X, according to the class of k neighbors. In this example, if k = 3 (inner circle), the new sample is assigned to Class B; if k = 6 (outer circle), the assignment changes to Class A.

2.4. Artificial neural networks (ANNs)

ANNs are inspired by the biological nervous system. ANNs contain a large number of highly interconnected nodes (called neurons) separated into layers (Fig. 4), enabling the network to process different pieces of information while considering constraints to coordinate internal processing, and to optimize its final output [25,33].

2.5. Convolutional neural networks (CNNs)

CNNs were inspired by the connectivity pattern of the animal visual cortex. Neurons respond to stimuli only in a restricted region (receptive field) of the previous layer, where receptive fields of different neurons partially overlap until they cover the entire visual field (Fig. 5). Unlike other ML techniques, the network learns the filters that are usually "hand crafted". Also, CNNs exploit the strong spatially local correlation found on images, allowing the features to be detected regardless of their position. In recent years, Deep Neural Networks (DNNs), which differ from ANNs by their depth (the number of neuron layers), have proven to be successful in solving diverse problems, mainly for their capacity to learn features from large datasets [34].

It should be noted that in the following discussion, algorithmic performance is assessed in terms of Dice's coefficient (DC), the modified Hausdorff distance (MHD) and the area under the receiver operating characteristic curve (AUC or AUROC), where possible. The quantification usually starts with calculation of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs). TP refers to cases correctly classified as pertaining to the class, opposite to FP, when the case is wrongly classified. Inversely, TN and FN refer to a case correctly and incorrectly classified as not belonging to the class, respectively. The DC quantifies overlap between the processed image from the technique with a defined ground truth, ranging from zero (no overlap) to unity (identical segmentation). MHD is a measure of similarity between two objects based on their shape attributes. AUC combines information of the true positive rate or sensitivity, and false positive rate or fall-out. Sensitivity measures the proportion of actual positives that are correctly identified, while fall-out indicates the proportion of cases wrongly classified as positives. Inversely, the specificity measures the proportion of negatives that are correctly identified. Recall is the ratio of TPs to the sum of TPs and FNs, indicating the proportion of actual positives that are correctly identified. Precision is defined as the ratio of TPs to the sum of TPs and FPs, indicating the proportion of identified positives that are correct [35].

3. Forensic applications

In terms of the currently reported use of ML in forensic post-mortem imaging, it is in its infancy. ML has only been trialed in a few specific forensic applications including automatic forensic dental identification [36]; sex determination [37–39]; the automation of bone age assessment [40,41]; prediction of bone fractures [42]; and the automatic detection of hemorrhagic pericardial effusion [43]. As far as we are aware, none of these studies has translated into daily forensic practice, despite the potential to streamline case-work.

The legally robust identification of a decedent is the first objective when their body is triaged for a post-mortem. Dental analysis and comparison of ante-mortem and post-mortem information is one of the recognized tools for determining a decedent's identity. This traditionally requires an odontologist to find the best match to an ante-mortem database, using features such as dental restorations, pathologies, and tooth and bone morphologies. Zhang et al. [36] proposed a new descriptor that encodes the local shape of a person's dental features. They subsequently used an RF classifier to match the features of the unknown person to those in the database (n = 200). The result yielded 100 % accuracy for complete (n = 20) and incomplete

Fig. 3. SVM maps the input data (left) to a high-dimensional feature space (right) and calculates a hyperplane able to separate the different classes. The class assigned to new samples is decided according to the location of the data points on the high-dimensional space with respect to the generated hyperplane.
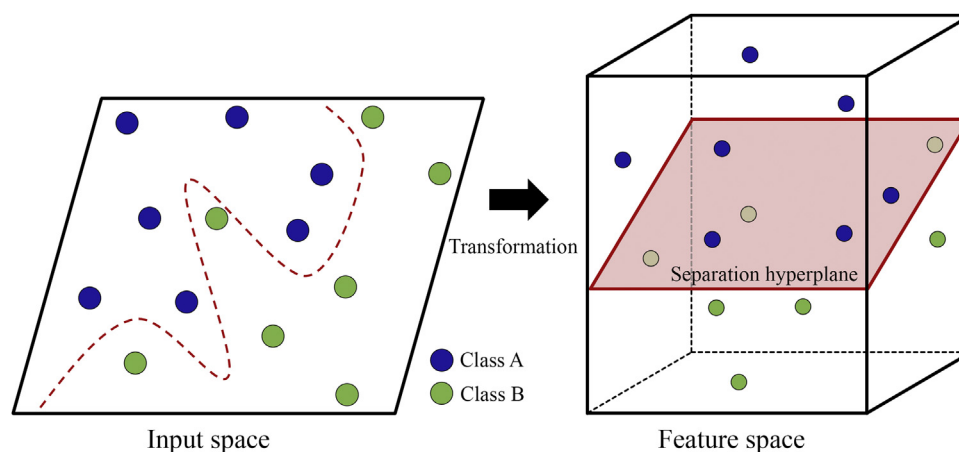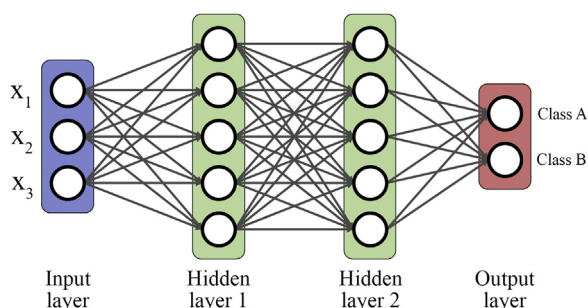
Fig. 4. Fully connected ANN with three inputs (X_1-X_3), two output classes (A and B), and two hidden layers of five neurons each. The weights that exist in the hidden layers are determined through the process of back-propagation which maximizes the classification success as the network is supplied with increased levels of training data.

(n = 20) feature datasets. Incomplete datasets were derived from cases involving trauma. The method presented was shown to be rotationally and translationally invariant, and was orders of magnitude faster than conventional 2D methods. It is important to note that the database was constructed using a surface laser scanner on plaster samples in contrast to PMCT scans.

Accurate determination of the sex of a decedent also aides in the identification process. Several different approaches have been used for sex estimation. Arigbabu et al. [37] utilized 100 head PMCT scans. They combined and evaluated six local feature representations, two feature learning, and three classification algorithms. This technique of combining multiple features and classifiers is often used in ML pipelines as it has been shown to improve accuracy and reliability. The best prediction rate was 86 %, which was within the reported sex prediction range for applications that use cranial features. The small number of cases obtained only from South East Asia limited the generalizability of the results. Anderson et al. [38] utilized morphological gray matter differences on MRIs to differentiate between male and female incarcerated offenders, with implications to cognitive neuroscience research. Preprocessing steps were described, including realignment and image registration, to obtain the volume and density of the gray matter on each case utilizing Statistical Parametric Mapping software (SPM12; http://www.fil.ion.ucl.ac.uk/spm). Source-based morphometry (SBM) was utilized to extract features from the gray matter spatial information, with SBM being able to identify distinct regions with common covariation between subjects. A number of ML classification approaches were trialed, however, only an SVM and logistic regression were described due to present the highest classification accuracy of 94 %. Limitations included the use of volumetric brain data only, without accounting for other

moderating variables and quantitative methods, such as age, functional activity, and structural and functional connectivity. Ortiz et al. [39] compared five different ML techniques in the assessment of panoramic radiographs. The ANN outperformed the rest of the models, including *k*-NNs and logistic regression, with an accuracy of 89 %. Only 100 panoramic radiographs were used, limiting the statistical significance of the results.

As with the identification of a decedent's sex, their estimated age is also an important parameter for streamlining the identification process. Štern, Payer and Urschler [40] compared two ML approaches, RFs and DCNNs to determine age (through regression) and distinguish minors from adults (classification) using bone ossification from MRI scans of the hand/wrist. As a general note, DCNNs are often compared with RFs as the DCNN can determine the most important features itself whereas the RF must be supplied with those deemed important by the user. To better study the impact of different input information on the decision process, three strategies were tested: the use of the whole hand, a cropped image with age relevant bones, or the hand-crafted filter-based enhanced epiphyseal gap. The best mean absolute error and standard deviation results with respect to the biological age (as estimated by radiologists) were 0.20 ± 0.42 and 0.23 ± 0.45 years for the DCNN using cropped structures and the RFs using enhanced images, respectively. The results were reported to achieve the new state-of-the-art accuracy compared with previous MRI-based methods and their earlier work. Furthermore, when the technique was adapted for 2D MRI, the method was in line with state-of-the-art methods using X-ray data. Limitations of this work included the requirement for age-relevant anatomical information, which implies a labor-intensive pre-processing step, and decreased accuracy for cases with biological ages greater than 18 years. In an alternative approach, Li et al. [41] utilized pelvic X-ray images and a DCNN to create a bone age assessment pipeline which yielded a mean error of 0.94 years, 0.36 years better than the existing reference standard. This work used transfer learning from a CNN pre-trained on the ImageNet database [44], achieving an appropriate accuracy for this type of input data. Transfer learning is widely used in ML applications and is particularly useful when small or unbalanced datasets are available. Limitations acknowledged by the authors included the lack of diversity in ethnicity of patients, and the exclusion of images with artefacts and diseases.

Many forensic institutions utilize PMCT to guide the pathologist in their approach to the autopsy. PMCT is particularly useful for identifying fractures due to the high attenuation of bone. Heimer et al. [42] used an undisclosed DCNN from a dedicated software (VIDI, Cognex, Natick, MA, USA) to predict the presence of skull fractures using 150 head PMCT scans (75 scans for each case: with and without fractures). The skulls were preprocessed through the
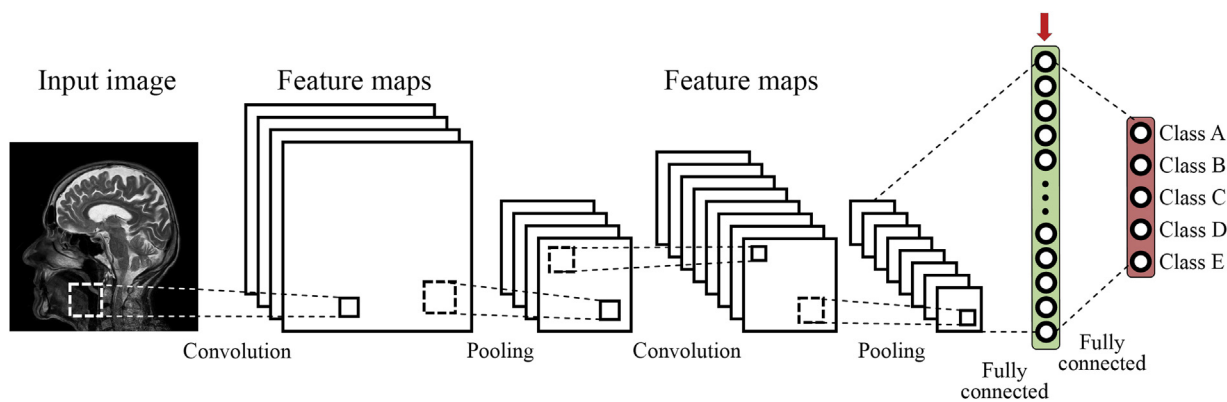

Fig. 5. A CNN for classification of the input image, e.g. MRI, into five categories. The CNN presents two convolutional layers, each followed by a pooling layer in charge of decreasing the size of the generated feature maps, and two fully connected layers, including the output layer, to increase the number of features closer to the network's output.

generation of curved maximum intensity projections, so that the skull's surface could be unfolded onto a single image. Deep learning was applied and the best-performing selected network yielded an AUROC of 0.965, a sensitivity of 91.4 % and a specificity of 87.5 %. An AUROC of 0.5 defines a model that classifies at random, while 1.0 is a completely accurate model.

PMCT is also useful for assessing many aspects of cardiac condition prior to autopsy, e.g. the appearance of discontinuities of the aortic wall can be a direct sign of injury in the aorta, whereas the appearance of a blood collection within the chest cavity (hemothorax or hemopericardium) can be an indirect sign [45,46]. These signs, observed on plain film X-ray or PMCT must be interpreted by radiologists and forensic pathologists. Ebert et al. [43] used two separated and undisclosed DCNNs from a dedicated software (VIDI) for the classification of images with or without hemopericardium and also the corresponding segmentation of the blood content in PMCT. The average DC, recall, and precision for the classification task were 77 %, 77 %, and 85 % respectively. For segmentation, the values obtained were 78 %, 78 %, and 79 %, respectively. Limitations of this study include the small number of training cases (n = 14 cases with hemopericardium), while the use of a dedicated software restricted the training data to individual slices, losing sometimes crucial volumetric information.

Due to the dearth of information relating to the application of ML to forensic imaging, it is important to review the state-of-the-art and establish lessons learned from the significant body of literature describing its application to clinical image analysis.

4. Current clinical applications

ML techniques have been used in the diagnosis and prognosis of diseases, as well as for segmentation, classification, and measurement of anatomical structures [24,47]. In this review, the ML applications have been grouped according to the tissue or organ studied, where brain, lungs, and skeleton were chosen to highlight results and limitations. Each anatomical section concludes with a summary evaluating the key implications determined from the clinical literature and their application in the forensic setting.

4.1. Brain tissue

Traditional atlas-based segmentations require registration to align the atlas images to the unseen image. Whereas, ML approaches can learn the variability between patients, making them especially useful in forensics, where variance is greater than for clinical imaging. ML can also be used in combination with atlas-based approaches or in its own right. As an example of the former, Srhoj-Egekher et al. [48] used atlas-based segmentation for pre-processing T2-weighted MRI neonatal brain images to obtain initial probabilities, subsequently refined using a *k*-NN approach. Whilst this approach achieved DCs and MHDs ranging from 77 % to 93 %, and 0.35 to 2.86 respectively, the assignment of a tissue classification to each voxel independently, post atlas registration, meant some voxels were attributed to more than one class, while background voxels were unclassified. Conversely, Zhang et al. [49] opted for purely ML approaches that analyzed image patches for segmentation into white matter (WM), gray matter (GM), and

cerebrospinal fluid (CSF) of infant brains (n = 10). Four network architectures were tested and, in most cases, the CNN method significantly outperformed SVMs and RFs with overall DC scores and MHDs of 85 % and 0.32, respectively. The CNN method also outperformed two other common image segmentation methods: coupled level sets (CLS) and majority voting (MV).

Three further publications were found where the authors segmented similar structures within adult brains. Van Opbroek et al. [50] applied an SVM for pixel-wise classification to registered volumes from a variety of MRI sequences for patients with diabetes and controls. The resulting segmentation of eight different tissue types demonstrated limited success (Table 1). The SVM showed poor performance in low contrast areas, while atlas misregistration caused voxels to be improperly classified. Moeskops et al. [51] used CNNs to process T1-weighted scans to segment the same eight tissue types. With CNNs, the use of different sized patches during training allowed for a smooth segmentation and analysis of local texture. In general, CNNs delivered better segmentation (Table 1), although this was a different patient cohort. A more recent application of 3D DCNNs [52] was used to identify 25 brain structures in T1-weighted MRI scans (n = 30). Again, image patches were utilized as input to the network. However, spectral and Cartesian coordinate information relating to the patches was added after the convolutional layers (e.g. see arrow in Fig. 5) in order to introduce spatial information, which substantially increased the segmentation accuracy.

ML can also be used for the assisted diagnosis of neurodegenerative diseases. Salvatore et al. [53] used a combination of principal component analysis (PCA) with an SVM to classify morphological MRI sequences as patients with Parkinson's disease (n = 28), progressive supranuclear palsy (PSP) (n = 28), or controls (n = 28). The large cohort sizes, inter-class cohort balance, and separation between PSP patients and other parkinsonian variants were identified as particular strengths, compared to other papers. The performance (accuracy, specificity and sensitivity were all > 80 %) of the model was shown to be limited by the number of principal components (16–26) utilized for classification. This dependence is an important consideration when using dimensionality reduction techniques and was also demonstrated for approaches that classified Alzheimer's disease [54].

Finally, ML techniques have also been used to segment and classify brain tumors. Zacharaki et al. [55] used conventional and perfusion MRI from patients with a diagnosis of intra-cranial neoplasm to classify them by type and grade of tumor (n = 98). Their approach consisted of region of interest (ROI) definition, feature extraction, feature selection, and classification by SVMs. For comparison, linear discrimination analysis (LDA) and *k*-NN were also implemented. The mean classification accuracy was 91 % for the SVM approach, compared with 81 % for LDA and 90 % for *k*-NN. Some of the limitations were related to the lack of features selected that described deformation of healthy structures due to the tumor, and the utilization of ROIs which yielded inter-observer variability.

Once the presence of tumors is verified, one possible subsequent step would be segmentation of the pathology, which is challenging even for experienced neuroradiologists [56]. To address this segmentation problem, a variant of CNNs named U-net

Table 1
Summary of selection of papers for non-infant brain tissue segmentation.

Authors	Type / No. images	Mean Dice's coefficient (%)
van Opbroek et al. [50]	MRI / 5 training, 12 testing	GM = 85, WM = 88, CSF = 78 (SVM)
Moeskops et al. [51]	MRI / 5 training, 10 testing	GM = 91, WM = 94, CSF = 85 (CNN)
Wachinger et al. [52]	MRI / 20 cases for training, 10 cases for testing (>256 images per case)	All structures = 91 (DCNN)

is often employed [57]. Beers et al. [58] utilized two 3D U-nets connected sequentially to perform whole tumor, enhancing tumor, and tumor core segmentation, achieving mean DCs for the test set (n = 95) of 84 %, 70 %, and 71 %, respectively. When the methodology was implemented on patients from ongoing clinical trials, the mean DCs decreased to 66 %, 54 %, and 45 %, respectively. The lower performances on the clinical trial patients were attributed to scans being post-operative, highlighting the importance of case selection for training.

Studies on brain tissues used mostly MRI data due to the multimodality information and a good soft-tissue contrast. Whilst the specific pathologies discussed are not all relevant to the forensic setting, the general conclusions deduced from the segmentation and localization of anatomical abnormalities are. Models that utilized dimensionality reduction techniques prior to classification were shown to yield performances dependent on the number of selected components. In addition, the identification of abnormalities in biological tissues required features capable of describing complicated deformations of the healthy structures. For CNNs, the performance of the pipeline depended significantly on the training set adequately representing expected cases. In general, CNNs outperformed algorithms such as SVMs, RFs, CLSs, and MV in segmentation and classification tasks. Note that some studies used small datasets, which limited statistical power. In addition, as will be demonstrated throughout this review, a combination of the variability in reporting of metrics, the lack of reporting of a diagnostic odds ratio [59], the unavailability of datasets and reference implementations, and the effect of imbalanced data in the classification accuracy, common in medical datasets [60,61], made it difficult to compare papers quantitatively.

In forensics, PMCT does not provide good resolution of internal cranial structures or brain metastases, and in general, the resolution is not sufficient to identify neurodegenerative issues, but degeneration can sometimes be observed in defined structures, e.g. in the caudate nucleus in Huntington's disease. On the other hand, PMCT is adequate in showing evolving brain infarcts and in displaying collections of blood, e.g. subdural hemorrhages (which are reasonably common). PMCT can also show intra-parenchymal hemorrhages and parenchymal hemorrhagic contusions. Intraparenchymal hemorrhages, e.g. hypertensive hemorrhage, classically involve distinct areas in the brain: basal ganglia, thalamus, pons, and cerebellar hemispheres. Parenchymal hemorrhagic contusions are classically seen with contra-coup basal frontal lobe contusions (bleed within brain tissue occurring on the opposite side of the head to the primary injury site) when someone falls onto the back of their head (often associated with a skull fracture – occipital).

4.2. Lungs

In ML, feature learning refers to the automatic discovery of meaningful representations from raw data, in contrast to manual feature engineering, where the features have to be chosen by a domain expert. Feature learning allows for end-to-end learning, where a complex system can be represented by a single model, bypassing the intermediate layers present in traditional workflow designs. Learning a representation of any tissue is a useful process if subsequent classification is required, or if the goal is to find differences between samples in the training data. The representation quality is highly dependent on the learned features.

A restricted Boltzmann machine (RBM) is a generative neural network that can be used to perform automatic feature learning. Li et al. [62] used a Gaussian RBM with a training dataset consisting of different sized patches obtained from high-resolution lung CT images (n = 92), with the purpose of classifying five tissue types using SVMs. The best accuracy obtained was 84 %, with a high rate of FPs caused by the similarity between tissues. Van Tulder and de Bruijne [63] utilized convolutional RBMs, adding learning objectives that helped the algorithm to extract features for description and training data classification. The training data consisted of CT scans (n = 73) with five types of tissues classified. Resulting accuracies were <75 % and 85–90 % for the classification of lung patches and airway centerlines, respectively. The low accuracies were attributed to small training sets and number of extracted filters due to computational restrictions.

Netto et al. [64] utilized examinations (n = 50) with 198 identified nodules and an SVM to classify the structure as nodule or non-nodule. The resulting accuracy was 91 %, with a sensitivity of 86 %. The largest errors were reported when the feature was very large or very small, where it could be mistaken for other structures or for being the continuation of one. Hua et al. [65] used images containing nodules from the Lung Image Database Consortium (LIDC) CT dataset to train both a CNN and a deep belief network (DBN) constructed by stacking RBMs. The performance of the two networks was then compared with two feature-based methods (Table 2). The major limitation reported was resizing of the input images, which discarded size cues that were important indicators of malignancy.

Kumar et al. [66] also classified the lung nodules in the LIDC images (Table 2) using an autoencoder (AE) and a binary decision tree classifier (BDT). An AE is an unsupervised deep learning technique utilized for feature extraction, while a binary decision tree is a specialized implementation for classification where every node has only two branches. The false positive rate of 39 % was attributed to the visual similarity between benign and malignant cases, which can be compared to a 27 % rate obtained on The National Lung Screening Trial (NLST) using low-dose CT (LDCT) [67].

A more recent study compared massive-training artificial neural networks (MTANNs) against CNNs [68] using a database of LDCT scans (n = 38), consisting of 1057 slices. MTANNs are an extension of ANNs, where a large number of overlapping subregions are created for each voxel of the original image and used as inputs to the network. The reported AUROC was 0.88 for the MTANN, and 0.78 for the best of the four CNN architectures. The MTANN required a smaller number of training samples than the CNNs for a better classification performance. This was attributed to the hierarchies of the learned features, where the MTANN learned to detect lesions utilizing low-level features, while the CNNs extracted low-, mid- and high-level features, increasing their reliance on irrelevant characteristics.

A recent focus of attention was related to the use of ML for early diagnosis, assessment of severity, and differentiation between the novel coronavirus (COVID-19) and community acquired pneumonia (CAP) from CT scans. Barstugan et al. [69] utilized n = 150 CT abdominal images from 53 infected patients, five feature

Table 2
Summary of selection of papers for lung nodule classification.

Authors	Type / dataset	Accuracy (%)	Sensitivity (%)
Hua et al. [65]	CT / 2545 nodules	–	73 (DBN), 73 (CNN), 76 (SIFT)
Kumar et al. [66]	CT / 4323 nodules	75	83 (AE + BDT)

extraction methods, and an SVM for the final classification, achieving a maximum accuracy of 99.7 %. The main limitation of their work was the manual selection of the patches obtained from the original images and used for the training, which restricts the usability and reproducibility of this approach. Tang et al. [70] assessed the severity (severe, non-severe) of the disease from chest CT images from 176 patients, utilizing quantitative measures, e.g. the ratio between the volume of the whole lung and the volume of ground-glass opaque regions, with several RF models. The best performing RF yielded results of 93 %, 75 %, 88 %, and 91 % for the sensitivity, specificity, accuracy, and AUC, respectively. To differentiate between COVID-19, CAP, or non-pneumonia, Li et al. [71] collected 4356 chest CT exams from 3322 patients. A DCNN was utilized, with an architecture denoted COVNet, able to classify the volumetric data with a sensitivity, specificity, and AUC of 90 %, 96 %, and 96 % for COVID-19 cases, 87 %, 92 %, and 95 % for CAP cases, and 94 %, 96 %, and 98 % for non-pneumonia cases, respectively. A limitation of this work included the lack of laboratory confirmation for each case, where COVID-19 could have similar imaging characteristics as other viral pneumonias.

Studies on lungs generally used CT scans for the segmentation of tissues and tumors, and classification of nodules for early cancer diagnosis. Due to the low contrast between different tissues in the lungs, the approaches reported were reliant on shape, texture, and feature size. The segmentation performance was poor for nodules at the size extremes. Major findings included lower performances due to image resizing, and the importance of reporting FP rates, which can yield high values in applications that intend to determine nodule malignancy.

Potential applications to the forensic setting include detection of emphysema, consolidation of lung parenchyma (pneumonia), and if appropriate windows are used, interstitial changes. Of crucial forensic interest is the presence of blood and fluid in the chest. Furthermore, establishing the presence of a lung lesion (and especially more than one) independently of the cause of death may indicate the presence of occult malignancy. In such cases, the deceased's next of kin can be alerted, and the family contact nurses can organize appropriate follow up for family members if a cancer is found. It is important to note that the appearance of the lungs in PMCTs can be affected by aspiration of gastric content that may occur in the process of dying, e.g. from a 'heart attack'.

4.3. Skeleton

Skeletal segmentation usually occurs before measurement and/ or diagnosis of bone or articular diseases. Koch et al. [72] segmented MRIs (n = 110) of the wrist using marginal space learning (MSL) and RFs, where MSL incrementally learned classifiers in marginal spaces of lower dimensions [73]. The segmented images were used to compute the 3D model of every carpal bone, with AUCs of 0.88 for both scan modalities. The approach was an order of magnitude faster than previous work using a semi-automatic method. Similar literature did not report segmentation errors and could not be used for comparison.

Bone age assessment from plain X-rays is used in pediatrics by comparing the results to chronological age for the evaluation of endocrine and metabolic disorders. A fully automated pipeline was presented by Lee et al. [74] using a pre-trained CNN (transfer learning). Both male and female test X-rays were assigned a bone age within 1 year of the correct value over 90 % of the time, and over 98 % within 2 years.

X-rays have also been widely used for fracture detection, e.g. of the tibia [75], where texture and shape features were fed into three different ML algorithms: an ANN, *k*-NN, and SVM, and the outputs fused using a majority vote scheme. The combination of the classifiers using both types of features presented a significant improvement over using just one classifier, or only one feature type. Reported accuracies, precisions, and sensitivities were above 97 %. Instead of fusing the results from the classifiers, multi-stage classifiers have also been used. Wels et al. [76] reported a fully automatic system using several RF stages, capable of detecting osteolytic spinal bone lesions from CT volumes, with an average sensitivity of 75 %. The performance was affected by differences in contrast and noise characteristics in the data used for training and testing, however, values for accuracy were not presented for further interrogation.

Sharma et al. [77] measured trabecular bone microarchitecture and used the information to discriminate between healthy cases (n = 10) and patients with Type 1 Gaucher disease (n = 20). SVMs were used to classify different genotypes of the disease, achieving an average 70 % classification accuracy, 74 % sensitivity, and 85 % precision. The structure of the trabecular bone obtained from MRI have also been used classify knees with osteoarthritis [78]. The characteristics found to relate to the disease were useful in classifying healthy from affected patients (n = 159) with an AUC of 0.92, as well as predicting the risk of cartilage loss. In a similar study, the fractal analysis of X-ray images with SVMs enabled the automatic classification of osteoporotic patients (n = 39) versus controls (n = 38) with accuracies of up to 95 % [79]. Reported limitations from the papers in this section include the small number of cases and the high percentages of patients at early stages of the disease.

Orthopedic ML applications include disease diagnosis, age assessment, and risk prediction e.g. osteoporosis, osteoarthritis. Plain film X-ray and CT were most common; however, MRI studies of joints are being increasingly reported. The performance of ML applications was shown to be affected by the number and selected features, which is significantly influenced by differences in contrast and noise characteristics in the datasets. Comparison or ranking of the results was limited by reported performance metrics and the use of databases that were not representative of the disease stages studied. Other limitations included small patient cohorts and the processing times.

The most common skeletal disorders that could be picked up on PMCT scans are osteoporosis and Paget's disease, while fracture diagnosis, and then pattern of fracture diagnosis, e.g. a "hangman's fracture", extension/tear-drop fractures of the cervical spine, and spiral fracture of a long bone in an infant are of significant forensic interest.

5. Discussion

Typical goals of ML techniques in medical imaging include the differentiation of healthy from diseased patients or tissues and the localization of pathologies in anatomic structures. Algorithmic performance can be significantly affected when trying to process a new sample that differs significantly from the training dataset. This characteristic is especially important when it comes to applications in forensic medicine, where there is a high variability in the structures and image acquisition protocols, and unclear definition of what normal implies, due to changes occurring because of circumstances of death, tissue decomposition, trauma, or incineration. However, some applications e.g. organ localization, can be immediately translated to the forensic setting by using the appropriate training data, or by using the clinical medical images for the initial training of CNNs and then fine-tuning using forensic information. This is usually referred to as transfer learning. On the other hand, due to the size and availability of forensic databases, the opposite is also possible, with applications being trained in forensic data and then fine-tuned to the clinical setting.

To improve the capabilities of ML techniques, the training data can be modified, or more informative features can be used as

inputs to the algorithms. The selection of features can be optimized using learning objectives [63] or by utilizing an unsupervised technique as a preprocessing step to the classification task [66,80]. The features selected can also be used to alleviate human labelling, by selecting more representative training data for the medical expert [81,82]. Another approach to the improvement of ML performance is the combination of several techniques using a majority vote scheme [75], or the use of multi-stage classifiers [58] for segmentation of different spatially related tissues.

A wide range of implemented algorithms were found during the review process, where SVMs outperformed techniques such as LDA and *k*-NN [55], however the trend in recent works has been the high performance of CNNs [49,51]. The main disadvantage of classic ML approaches compared to CNNs is the performance variability due to the quality of the features [53] that must be hand-crafted by an expert according to the goal and dataset. The selected feature pool is commonly processed to lower its dimensionality before training the classifier by using techniques such as PCA. It is important to note that the number of principal components or features selected at the end of this step plays a key role in the classification performance [53].

The performance of the algorithms can also be significantly affected if the labelling process (diagnosis) is prone to error [54]. Furthermore, for medical and forensic applications, the common practice of resizing input images can yield to a loss of information that could be essential for diagnostic purposes [65]. An additional consideration is that some authors use for example a radiologist to classify cases, then benchmark the performance of the algorithm against radiologists. Rajpurkar et al. [83], for instance, presented a CNN that achieved radiologist-level pneumonia detection on a database [84] for which no gold-standard label existed, and listed as limitation the lack of information in the database that affects the radiologists' accuracy. It is also important to note that the lack of reporting of a diagnostic odds ratio [59] and the variability in reporting of metrics makes it difficult to compare papers.

For the task of segmentation, both multi-atlas algorithms and DCNNs with multiple patch sizes showed comparable results [48,49], demonstrating CNNs were most successful. Patch-based techniques could be a good approach in forensic cases were organs or structures are not localized in the usual anatomic positions [63]. Furthermore, the use of different sized patches in segmentation tasks allows for both a smoother separation and the detailed analysis of local texture [51].

Three important results for the use of ML in clinically-related applications were found that can also be applied in the forensic setting: firstly, temporal efficiency through the use of transfer learning; secondly, improved accuracy through the combination of ML classifiers using majority voting techniques or multi-stage approaches; and finally, the addition of an active learning phase, where the human labor can be alleviated during labeling.

One of the main issues that affects both the clinical and forensic settings is the lack of interpretability of predictions by black-box approaches such as neural networks. This is active area of current research and a current approach to addressing this concern is the use of visual explanations for the class label under consideration, obtained from the convolutional layer feature maps [85,86], and attention mechanisms [87], able to determine the parts of the input images more relevant for a particular classification. Furthermore, depending on the application, it is not required and could be counter-productive to completely automate a task, for which a human-in-the-loop can be beneficial by reducing the complexity through human input and assistance [82].

Some applications of ML already found in clinical medicine, that could be repurposed for forensic medicine, include segmentation and classification of organs and structures, including arteries, tiny blood vessels, the liver, spleen, stomach, gallbladder, and pancreas

[88,89]; computation of organ 3D models [72] for virtual autopsies; detection of lesions and calcification on vascular cross-sections [90]; identification of bone and joint atrophies or disorders [81,77–79]; fluid volume and composition on body cavities (blood, pus, ascites) [91]; and organ volume estimation, e.g. heart size with respect to body size [92].

Tasks in forensic radiology that to our knowledge have not been tackled using ML include: segmentation and classification of foreign bodies, differentiation between ante-mortem and post-mortem gases, calculation of body mass index, and determination of skeletal completeness after accidents.

For the segmentation and classification of foreign bodies, e.g. bullets, metallic dental fillings, the main challenge becomes finding the object that does not belong inside the body. Furthermore, metallic components can create artefacts such as beam-hardening on CT scans or field distortions in MRI [93], which can also be addressed using deep learning [94].

Differentiation between ante-mortem and post-mortem gases can be difficult using the voxel values of CT scans or MRI, so emphasis should be placed on understanding the expected location and evolution of these gases at different points in time [95]; also, differentiation between acute and remote infarction on the brain, which on a CT scan can be characterized by voxel values and tissue volume changes, can be tackled utilizing existing tissue classification techniques [50,51,54], with the addition of new classes to differentiate the types of infarction.

In forensic anthropology, tasks that could be addressed using ML include: determination of skeletal completeness after accidents [96], e.g. plane crashes; 3D reconstruction of incomplete bones, that could be extrapolated from the work by Hermoza and Sipiran [97] on incomplete archaeological objects; and 3D reconstruction of fractured skulls [98–100], used to infer a cause of death, or to perform facial reconstruction.

In addition to the aforementioned applications traditionally related to medical imaging, there is the potential for the use of CT scans for facial identification [101,102]. As a final note, the release this year of the New Mexico Decedent Image Database (NMDID, https://nmdid.unm.edu/) [103] should be acknowledged as a significant step forward for the development of tools that can be used to enhance the post-mortem workflow.

6. Conclusions

ML techniques have been applied to a large number of tasks that can be used in clinical medicine, where the algorithms most widely utilized in applications with medical images include RFs, SVMs, and CNNs. CNNs have shown better performance in the literature.

Techniques to improve the ML performance in radiology include data augmentation, improved feature selection and algorithmic combination, e.g. majority voting. Performance was shown to be affected by resizing of the input images and the accuracy of the labels provided with the training data. In addition, benchmarking was found to be difficult due to the lack of gold-standard labels, as well as the variability in reporting of metrics, and lack of reporting of a diagnostic odds ratio.

ML applications investigated for clinical medicine could be repurposed to the forensic domain with careful consideration to account for the increased variability and temporal factors, e.g. decomposition, that affect the data used to train the ML techniques. Due to the complexity of the autopsy process, a key application of ML to forensic radiology would be to streamline decedent identification and highlight and annotate areas of forensic interest. ML pipelines could be used to present information to optimally determine the cause of death, including differentiation between body cavity fluid accumulations (blood, pus, ascites) and their corresponding volumes, calculation of organ

volumes and weights, percentage of coronary artery calcification, identification of subtle fractures especially in critical areas such as the cervical spine, and determination of skeletal completeness and skeletal commingling after mass fatality incidents.

Funding sources

References

[1] T. Sutherland, C. O'Donnell, The artefacts of death: CT post-mortem findings, J Med Imag Radiat On. 62 (2) (2018) 203–210, doi:http://dx.doi.org/10.1111/1754-9485.12691.

[2] S.A. Bolliger, M.J. Thali, Imaging and virtual autopsy: looking back and forward, Philos. Trans. Biol. Sci. 370 (1674) (2015), doi:http://dx.doi.org/10.1098/rstb.2014.0253.

[3] S.G. Ross, S.A. Bolliger, G. Ampanozi, L. Oesterhelweg, M.J. Thali, P.M. Flach, Postmortem CT angiography: capabilities and limitations in traumatic and natural causes of death, Radiographics 34 (3) (2014) 830–846, doi:http://dx.doi.org/10.1148/rg.343115169.

[4] S.L. Saunders, B. Morgan, V. Raj, G.N. Rutty, Post-mortem computed tomography angiography: past, present and future, Forensic Sci. Med. Pat. 7 (3) (2011) 271–277, doi:http://dx.doi.org/10.1007/s12024-010-9208-3.

[5] J.M. Grimm, S. Grabherr, Future prospects of forensic imaging, Atlas of Postmortem Angiography, Springer, 2016, pp. 583–588, doi:http://dx.doi.org/10.1007/978-3-319-28537-5_36.

[6] E. Gibson, W. Li, C. Sudre, L. Fidon, D.I. Shakir, G. Wang, et al., NiftyNet: a deep-learning platform for medical imaging, Comput. Meth. Prog. Bio. 158 (2018) 113–122, doi:http://dx.doi.org/10.1016/j.cmpb.2018.01.025.

[7] M.R. Arbabshirani, B.K. Fornwalt, G.J. Mongelluzzo, J.D. Suever, B.D. Geise, A.A. Patel, et al., Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration, NPJ Digit. Med. 1 (1) (2018) 9.

[8] J. Johnston, T.L. Fauber, Essentials of Radiographic Physics and Imaging, Elsevier Health Sciences, 2015.

[9] E. Seeram, Computed Tomography: Physical Principles, Clinical Applications, and Quality Control, Elsevier Health Sciences, 2015.

[10] N. Norman, M.R. Dimmock, K. Lee, J. Graham, R. Bassed, The applicability of Dual-Energy Computed Tomography (DECT) in forensic odontology–A review, J. Forensic Radiol. Imaging 10 (2017) 15–22, doi:http://dx.doi.org/10.1016/j.jofri.2017.07.002.

[11] C. Constantinides, Magnetic Resonance Imaging: The Basics, CRC Press, 2016.

[12] E. Dura, J. Domingo, G. Ayala, L. Marti-Bonmati, E. Goceri, Probabilistic liver atlas construction, Biomed. Eng. Online 16 (1) (2017) 15, doi:http://dx.doi.org/10.1186/s12938-016-0305-0308.

[13] H. Park, P.H. Bland, C.R. Meyer, Construction of an abdominal probabilistic atlas and its application in segmentation, IEEE T Med. Imaging. 22 (April 4) (2003) 483–492, doi:http://dx.doi.org/10.1109/TMI.2003.809139.

[14] T. Okada, R. Shimada, M. Hori, M. Nakamoto, Y.W. Chen, H. Nakamura, et al., Automated segmentation of the liver from 3D CT images using probabilistic atlas and multilevel statistical shape model, Acad. Radiol. 15 (11) (2008) 1390–1403, doi:http://dx.doi.org/10.1016/j.acra.2008.07.008.

[15] H. Seim, D. Kainmueller, M. Heller, H. Lamecker, S. Zachow, H.C. Hege, Automatic segmentation of the pelvic bones from CT data based on a statistical shape model, Proceedings of the First Eurographics Conference on Visual Computing for Biomedicine. EG VCBM'08. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association (2008) 93–100.

[16] Y.Y. Boykov, M.P. Jolly, Interactive graph cuts for optimal boundary & region segmentation of objects in ND images, IEEE, Proceedings Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001, vol. 12001, pp. 105–112, doi:http://dx.doi.org/10.1109/ICCV.2001.937505.

[17] U. Bagci, X. Chen, J.K. Udupa, Hierarchical scale-based multiobject recognition of 3-D anatomical structures, IEEE T Med. Imaging 31 (March 3) (2012) 777–789, doi:http://dx.doi.org/10.1109/TMI.2011.2180920.

[18] Z. Xu, R.P. Burke, C.P. Lee, R.B. Baucom, B.K. Poulose, R.G. Abramson, et al., Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning, Med. Image Anal. 24 (1) (2015) 18–27, doi:http://dx.doi.org/10.1016/j.media.2015.05.009.

[19] L. Nanni, S. Ghidoni, S. Brahnam, Handcrafted vs. Non-handcrafted features for computer vision classification, Pattern Recogn. 71 (2017) 158–172.

[20] A. O'Neil, M. Shepherd, E. Beveridge, K. Goatman, A comparison of texture features versus deep learning for image classification in interstitial lung disease, Communications in Computer and Information Science, Springer International Publishing, 2017, pp. 743–753, doi:http://dx.doi.org/10.1007/978-3-319-60964-5_65.

[21] A. Phinyomark, H. Hu, P. Phukpattaranont, C. Limsakul, Application of linear discriminant analysis in dimensionality reduction for hand motion classification, Meas. Sci. Rev. 12 (3) (2012), doi:http://dx.doi.org/10.2478/v10048-012-0015-0018.

[22] Y. Liang, L. Sun, W. Ser, F. Lin, S.T.G. Thng, Q. Chen, et al., Classification of non-tumorous skin pigmentation disorders using voting based probabilistic linear discriminant analysis, Comput. Biol. Med. 99 (2018) 123–132, doi:http://dx.doi.org/10.1016/j.compbiomed.2018.05.026.

[23] M. Bhargavan, A.H. Kaye, H.P. Forman, J.H. Sunshine, Workload of radiologists in United States in 2006–2007 and trends since 1991–1992, Radiology 252 (2) (2009) 458–467, doi:http://dx.doi.org/10.1148/radiol.2522081895.

[24] S. Wang, R.M. Summers, Machine learning and radiology, Med. Image Anal. 16 (5) (2012) 933–951, doi:http://dx.doi.org/10.1016/j.media.2012.02.005.

[25] F. Camastra, A. Vinciarelli, Machine Learning for Audio, Image and Video Analysis, Springer, 2008.

[26] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, D. Rueckert, Self-supervised learning for medical image analysis using image context restoration, Med. Image Anal. 58 (2019)101539.

[27] K. Arulkumaran, M.P. Deisenroth, M. Brundage, A.A. Bharath, Deep reinforcement learning: a brief survey, IEEE Signal Proc. Mag. 34 (6) (2017) 26–38, doi:http://dx.doi.org/10.1109/msp.2017.2743240.

[28] S.R. Kesler, A. Rao, D.W. Blayney, I.A. Oakley-Girvan, O. Palesh, Predicting long-term cognitive outcome following breast Cancer with pre-treatment resting state fMRI and random forest machine learning, Front. Hum. Neurosci. 11 (2017), doi:http://dx.doi.org/10.3389/fnhum.2017.00555.

[29] S. Pereira, R. Meier, R. McKinley, R. Wiest, V. Alves, C.A. Silva, et al., Enhancing interpretability of automatically extracted machine learning features: application to a RBM-Random Forest system on brain lesion segmentation, Med. Image Anal. 44 (2018) 228–244, doi:http://dx.doi.org/10.1016/j.media.2018.04.008.

[30] K. Raza, Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule, U-Healthcare Monitoring Systems, Elsevier, 2019, pp. 179–196, doi:http://dx.doi.org/10.1016/b978-0-12-815370-3.00008-6.

[31] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995 Inc..

[32] S. Chaplot, L. Patnaik, N. Jagannathan, Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network, Biomed. Signal Process. Control 1 (1) (2006) 86–92, doi:http://dx.doi.org/10.1016/j.bspc.2006.05.002.

[33] J. Jiang, P. Trundle, J. Ren, Medical image analysis with artificial neural networks, Comput. Med. Imaging Graph. 34 (8) (2010) 617–631, doi:http://dx.doi.org/10.1016/j.compmedimag.2010.07.003.

[34] D. Štern, C. Payer, V. Lepetit, M. Urschler, Automated age estimation from hand MRI volumes using deep learning, International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (2016) 194–202, doi:http://dx.doi.org/10.1007/978-3-319-46723-8_23.

[35] A. Tharwat, Classification assessment methods, Appl. Comput. Inf. (2018), doi:http://dx.doi.org/10.1016/j.aci.2018.08.003.

[36] Z. Zhang, S.H. Ong, X. Zhong, K.W. Foong, Efficient 3D dental identification via signed feature histogram and learning keypoint detection, Pattern Recogn. 60 (2016) 189–204, doi:http://dx.doi.org/10.1016/j.patcog.2016.05.007.

[37] O.A. Arigbabu, I.Y. Liao, N. Abdullah, M.H. Mohamad Noor, Computer vision methods for cranial sex estimation, IPSJ Trans. Comput. Vis. Appl. 9 (1) (2017) 19, doi:http://dx.doi.org/10.1186/s41074-017-0031-0036.

[38] N.E. Anderson, K.A. Harenski, C.L. Harenski, M.R. Koenigs, J. Decety, V.D. Calhoun, et al., Machine learning of brain gray matter differentiates sex in a large forensic sample, Hum. Brain Mapp. 40 (5) (2019) 1496–1506, doi:http://dx.doi.org/10.1002/hbm.24462.

[39] A. Ortiz, C. Costa, R. Silva, M. Biazevic, E. Michel-Crosato, Sex estimation: anatomical references on panoramic radiographs using Machine Learning, Forensic Imaging 20 (2020)200356, doi:http://dx.doi.org/10.1007/s00330-018-5791-5796.

[40] D. Štern, C. Payer, M. Urschler, Automated age estimation from MRI volumes of the hand, Med. Image Anal. 58 (2019)101538.

[41] Y. Li, Z. Huang, X. Dong, W. Liang, H. Xue, L. Zhang, et al., Forensic age estimation for pelvic X-ray images using deep learning, Eur. Radiol. 29 (5) (2019) 2322–2329, doi:http://dx.doi.org/10.1007/s00330-018-5791-5796.

[42] J. Heimer, M.J. Thali, L. Ebert, Classification based on the presence of skull fractures on curved maximum intensity skull projections by means of deep learning, J. Forensic Radiol. Imaging 14 (2018) 16–20, doi:http://dx.doi.org/10.1016/j.jofri.2018.08.001.

[43] L.C. Ebert, J. Heimer, W. Schweitzer, T. Sieberth, A. Leipner, M. Thali, et al., Automatic detection of hemorrhagic pericardial effusion on PMCT using deep learning-a feasibility study, Forensic Sci. Med. Pat. 13 (4) (2017) 426–431, doi:http://dx.doi.org/10.1007/s12024-017-9906-1.

[44] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90, doi:http://dx.doi.org/10.1145/3065386.

[45] C. Basso, B. Aguilera, J. Banner, S. Cohle, G. d'Amati, et al., Guidelines for autopsy investigation of sudden cardiac death: 2017 update from the Association for European Cardiovascular Pathology, Virchows Arch. 471 (6) (2017) 691–705.

[46] I. Sifaoui, C. Nedelcu, G. Beltran, V. Dupont, J. Lebigot, A. Gaudin, et al., Evaluation of unenhanced post-mortem computed tomography to detect chest injuries in violent death, Diagn. Interv. Imag. 98 (May 5) (2017) 393–400.

[47] K. Suzuki, Pixel-based machine learning in medical imaging, Int. J. Biomed. Imaging 2012 (2012) 1, doi:http://dx.doi.org/10.1155/2012/792079.

[48] V. Srhoj-Egekher, M. Benders, K.J. Kersbergen, M.A. Viergever, I. Isgum, Automatic segmentation of neonatal brain MRI using atlas based segmenta-

tion and machine learning approach, MICCAI Grand Challenge: Neonatal Brain Segmentation, (2012) .

[49] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, et al., Deep convolutional neural networks for multi-modality isointense infant brain image segmentation, Neuroimage 108 (2015) 214–224, doi:http://dx.doi.org/10.1016/j.neuro-image.2014.12.061.

[50] A. van Opbroek, F. van der Lijn, M. de Bruijne, Automated brain-tissue segmentation by multi-feature SVM classification, The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS13), (2013) OA.

[51] P. Moeskops, M.A. Viergever, A.M. Mendrik, L.S. de Vries, M.J. Benders, I. Išgum, Automatic segmentation of MR brain images with a convolutional neural network, IEEE T Med. Imaging 35 (5) (2016) 1252–1261, doi:http://dx.doi.org/10.1109/TMI.2016.2548501.

[52] C. Wachinger, M. Reuter, T. Klein, DeepNAT: Deep convolutional neural network for segmenting neuroanatomy, NeuroImage 170 (2018) 434–445, doi:http://dx.doi.org/10.1016/j.neuroimage.2017.02.035 Segmenting the Brain.

[53] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, et al., Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy, J Neurosci Meth. 222 (2014) 230–237, doi:http://dx.doi.org/10.1016/j.jneumeth.2013.11.016.

[54] L. Khedher, J. Ramirez, J.M. Górriz, A. Brahim, F. Segovia, Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images, Neurocomputing 151 (2015) 139–150, doi:http://dx.doi.org/10.1016/j.neu-com.2014.09.072.

[55] E.I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E.R. Melhem, et al., Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme, Magn. Reson. Med. 62 (6) (2009) 1609–1618, doi:http://dx.doi.org/10.1002/mrm.22147.

[56] S. Bonte, I. Goethals, R. Van Holen, Machine learning based brain tumour segmentation on limited data using local texture and abnormality, Comput. Biol. Med. 98 (2018) 39–47, doi:http://dx.doi.org/10.1016/j.comp-biomed.2018.05.005.

[57] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer Assisted Intervention (2015) 234–241, doi:http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[58] A. Beers, K. Chang, J. Brown, E. Gerstner, B. Rosen, J. Kalpathy-Cramer, Sequential neural networks for biologically-informed glioma segmentation, Medical Imaging 2018: Image Processing, vol. 10574, International Society for Optics and Photonics, 2018, pp. 1057433, doi:http://dx.doi.org/10.1117/12.2293941.

[59] A.S. Glas, J.G. Lijmer, M.H. Prins, G.J. Bonsel, P.M. Bossuyt, The diagnostic odds ratio: a single indicator of test performance, J. Clin. Epidemiol. 56 (11) (2003) 1129–1135, doi:http://dx.doi.org/10.1016/S0895-4356(03)00177-X.

[60] T. Liu, W. Fan, C. Wu, A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset, Artif. Intell. Med. 101 (2019) 101723, doi:http://dx.doi.org/10.1016/j.artmed.2019.101723.

[61] F. Thabtah, S. Hammoud, F. Kamalov, A. Gonsalves, Data imbalance in classification: experimental evaluation, Inform Sci. 513 (2020) 429–441, doi:http://dx.doi.org/10.1016/j.ins.2019.11.004.

[62] Q. Li, W. Cai, D.D. Feng, Lung image patch classification with automatic feature learning, 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2013, pp. 6079–6082, doi:http://dx.doi.org/10.1109/EMBC.2013.6610939.

[63] G. van Tulder, M. de Bruijne, Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted boltzmann machines, IEEE T Med. Imaging 35 (5) (2016) 1262–1272, doi:http://dx.doi.org/10.1109/TMI.2016.2526687.

[64] S.M.B. Netto, A.C. Silva, R.A. Nunes, M. Gattass, Automatic segmentation of lung nodules with growing neural gas and support vector machine, Comput. Biol. Med. 42 (11) (2012) 1110–1121, doi:http://dx.doi.org/10.1016/j.comp-biomed.2012.09.003.

[65] K.L. Hua, C.H. Hsu, S.C. Hidayati, W.H. Cheng, Y.J. Chen, Computer-aided classification of lung nodules on computed tomography images via deep learning technique, Oncotargets Ther. 8 (2015), doi:http://dx.doi.org/10.2147/OTT.S80733.

[66] D. Kumar, A. Wong, D.A. Clausi, Lung nodule classification using deep features in CT images, 2015 12th Conference on Computer and Robot Vision (CRV), IEEE, 2015, pp. 133–138, doi:http://dx.doi.org/10.1109/CRV.2015.25.

[67] H.J. Mehta, T.L. Mohammed, M.A. Jantz, The American college of radiology lung imaging reporting and data system: potential drawbacks and need for revision, Chest 151 (3) (2017) 539–543, doi:http://dx.doi.org/10.1016/j.chest.2016.07.028.

[68] N. Tajbakhsh, K. Suzuki, Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs, Pattern Recogn. 63 (2017) 476–486, doi:http://dx.doi.org/10.1016/j.patcog.2016.09.029.

[69] M. Barstugan, U. Ozkaya, S. Ozturk, Coronavirus (COVID-19) classification using CT images by machine learning methods, arXiv e-prints (2020) arXiv:2003 09424.

[70] Z. Tang, W. Zhao, X. Xie, Z. Zhong, F. Shi, J. Liu, et al., Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images, arXiv e-prints. (2020) arXiv:2003 11988.

[71] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, Radiology (2020) 200905.

[72] M. Koch, A.G. Schwing, D. Comaniciu, M. Pollefeys, Fully automatic segmentation of wrist bones for arthritis patients, 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, 2011, pp. 636–640, doi:http://dx.doi.org/10.1109/ISBI.2011.5872487.

[73] Y. Zheng, D. Comaniciu, Marginal space learning, Marginal Space Learning for Medical Image Analysis: Efficient Detection and Segmentation of Anatomical Structures, Springer New York, New York, NY, 2014, pp. 25–65, doi:http://dx.doi.org/10.1007/978-1-4939-0600-0_2.

[74] H. Lee, S. Tajmir, J. Lee, M. Zissen, B.A. Yeshiwas, T.K. Alkasab, et al., Fully automated deep learning system for bone age assessment, J. Digit. Imaging 30 (4) (2017) 427–441, doi:http://dx.doi.org/10.1007/s10278-017-9955-9958.

[75] N. Umadevi, S. Geethalakshmi, Multiple classification system for fracture detection in human bone X-ray images, 2012 Third International Conference on Computing Communication & Networking Technologies (ICCCNT), IEEE, 2012, pp. 1–8, doi:http://dx.doi.org/10.1109/ICCCNT.2012.6395889.

[76] M. Wels, B.M. Kelm, A. Tsymbal, M. Hammon, G. Soza, M. Sühling, et al., Multi-stage osteolytic spinal bone lesion detection from CT data with internal sensitivity control, Medical Imaging 2012: Computer-Aided Diagnosis, vol. 8315, International Society for Optics and Photonics, 2012, pp. 1–8, doi:http://dx.doi.org/10.1117/12.911169.

[77] G.B. Sharma, D.D. Robertson, D.A. Laney, M.J. Gambello, M. Terk, Machine learning based analytics of micro-MRI trabecular bone microarchitecture and texture in Type 1 Gaucher disease, J. Biomech. 49 (9) (2016) 1961–1968, doi:http://dx.doi.org/10.1016/j.jbiomech.2016.04.010.

[78] J. Marques, H.K. Genant, M. Lillholm, E.B. Dam, Diagnosis of osteoarthritis and prognosis of tibial cartilage loss by quantification of tibia trabecular bone from MRI, Magn. Reson. Med. 70 (2) (2013) 568–575, doi:http://dx.doi.org/10.1002/mrm.24477.

[79] A. Tafraouti, M. El Hassouni, H. Toumi, E. Lespessailles, R. Jennane, Osteoporosis diagnosis using fractal analysis and support vector machine, 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS), IEEE, 2014, pp. 73–77, doi:http://dx.doi.org/10.1109/SITIS.2014.49.

[80] M. Chen, X. Shi, Y. Zhang, D. Wu, M. Guizani, Deep features learning for medical image analysis with convolutional autoencoder neural network, IEEE T Big Data (1) (2017) 1, doi:http://dx.doi.org/10.1109/TBDATA.2017.2717439.

[81] S. Hao, J. Jiang, Y. Guo, H. Li, Active learning based intervertebral disk classification combining shape and texture similarities, Neurocomputing 101 (2013) 252–257, doi:http://dx.doi.org/10.1016/j.neucom.2012.08.008.

[82] A. Holzinger, Interactive machine learning for health informatics: when do we need the human-in-the-loop? Lect. Notes Artif. Int. 3 (2) (2016) 119–131, doi:http://dx.doi.org/10.1007/s40708-016-0042-0046.

[83] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, et al., CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning, arXiv preprint (2017) arXiv:171105225.

[84] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 2097–2106.

[85] A. Chattopadhay, A. Sarkar, P. Howlader, V.N. Balasubramanian, Grad-CAM: generalized gradient-based visual explanations for deep convolutional networks, IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018.

[86] Q. shi Zhang, S. chun Zhu, Visual interpretability for deep learning: a survey, Front. Inf. Technol. Electron. Eng. 19 (1) (2018) 27–39, doi:http://dx.doi.org/10.1631/fitee.1700808.

[87] J. Rebane, I. Karlsson, P. Papapetrou, An investigation of interpretable deep learning for adverse drug event prediction, IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2019, doi:http://dx.doi.org/10.1109/cbms.2019.00075.

[88] H.R. Roth, H. Oda, Y. Hayashi, M. Oda, N. Shimizu, M. Fujiwara, et al., Hierarchical 3D fully convolutional networks for multi-organ segmentation, arXiv preprint (2017) arXiv:170406382.

[89] C. Wang, M. Oda, Y. Yoshino, T. Yamamoto, K. Mori, Fine segmentation of tiny blood vessel based on fully-connected conditional random field, Medical Imaging 2018: Image Processing, vol. 10574, International Society for Optics and Photonics, 2018, pp. 105740K, doi:http://dx.doi.org/10.1117/12.2293486.

[90] M.A. Zuluaga, I.E. Magnin, M.H. Hoyos, E.J.D. Leyton, F. Lozano, M. Orkisz, Automatic detection of abnormal vascular cross-sections based on density level detection and support vector machines, Int. J. Comput. Assist. Radiol. Surg. 6 (2) (2011) 163–174, doi:http://dx.doi.org/10.1007/s11548-010-0494-0498.

[91] M. Cicero, A. Bilbily, E. Colak, T. Dowdell, B. Gray, K. Perampaladas, et al., Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs, Invest. Radiol. 52 (5) (2017) 281–287, doi:http://dx.doi.org/10.1097/RLI.0000000000000341.

[92] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, S. Li, Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation, Med. Image Anal. 30 (2016) 120–129, doi:http://dx.doi.org/10.1016/j.media.2015.07.003.

[93] M.J. Thali, M. Braun, J. Wirth, P. Vock, R. Dirnhofer, 3D surface and body documentation in forensic medicine: 3-D/CAD Photogrammetry merged with 3D radiological scanning, J. Forensic Sci. 48 (November 6) (2003) 1356–1365, doi:http://dx.doi.org/10.1520/JFS2003118.

[94] L. Gjesteby, Q. Yang, Y. Xi, Y. Zhou, J. Zhang, G. Wang, Deep learning methods to guide CT image reconstruction and reduce metal artifacts, SPIE Medical Imaging. International Society for Optics and Photonics, (2017), pp. 101322W, doi:http://dx.doi.org/10.1117/12.2254091.

[95] C.E. Offiah, J. Dean, Post-mortem CT and MRI: appropriate post-mortem imaging appearances and changes related to cardiopulmonary resuscitation, Br. J. Radiol. 89 (1058) (2015), doi:http://dx.doi.org/10.1259/bjr.20150851 20150851.

[96] M.Y. Iscan, M. Steyn, The Human Skeleton in Forensic Medicine, Charles C Thomas Publisher, 2013.

[97] R. Hermoza, I. Sipiran, 3D reconstruction of incomplete archaeological objects using a generative adversarial network, Proceedings of Computer Graphics International (CGI), CGI 2018. New York, NY, USA: ACM, 2018, pp. 5–11, doi:http://dx.doi.org/10.1145/3208159.3208173.

[98] M. Kettner, P. Schmidt, S. Potente, F. Ramsthaler, M. Schrodt, Reverse engineering-rapid prototyping of the skull in forensic trauma analysis, J.

Forensic Sci. 56 (4) (2011) 1015–1017, doi:http://dx.doi.org/10.1111/j.1556-4029.2011.01764.x.

[99] W. Yu, M. Li, X. Li, Fragmented skull modeling using heat kernels, Graph Models 74 (4) (2012) 140–151, doi:http://dx.doi.org/10.1016/j.gmod.2012.03.011.

[100] P. Gach, L. Tuchtan-Torrents, C. Delteil, P. Adalian, M.D. Piercecchi, L.C. Ebert, et al., Virtual reconstruction of paranasal sinuses from CT data: A feasibility study for forensic application, Diagn. Interv. Imag. 100 (3) (2019) 163–168, doi:http://dx.doi.org/10.1016/j.diii.2018.11.011.

[101] P.J. Galzi, J. Mullins, Case study: 3d application of the anatomical method of forensic facial reconstruction, J. Forensic Res. 07 (05) (2016), doi:http://dx.doi.org/10.4172/2157-7145.1000350.

[102] K. Imaizumi, K. Taniguchi, Y. Ogawa, K. Matsuzaki, H. Maekawa, T. Nagata, et al., Development of three-dimensional facial approximation system using head CT scans of Japanese living individuals, J. Forensic Radiol. Imaging 17 (2019) 36–45, doi:http://dx.doi.org/10.1016/j.jofri.2019.05.004.

[103] B.S. Daneshvari, H.J.H. Edgar, Standardizing data from the dead, St Heal T. 264 (2019) 1427–1428, doi:http://dx.doi.org/10.3233/SHTI190467 (MEDINFO 2019: Health and Wellbeing e-Networks for All).