



OPEN

## Rapid identification of wood species using XRF and neural network machine learning

Aaron N. Shugar<sup>1✉</sup>, B. Lee Drake<sup>2</sup> & Greg Kelley<sup>3</sup>

**An innovative approach for the rapid identification of wood species is presented. By combining X-ray fluorescence spectrometry with convolutional neural network machine learning, 48 different wood specimens were clearly differentiated and identified with a 99% accuracy. Wood species identification is imperative to assess illegally logged and transported lumber. Alternative options for identification can be time consuming and require some level of sampling. This non-invasive technique offers a viable, cost-effective alternative to rapidly and accurately identify timber in efforts to support environmental protection laws and regulations.**

The identification of wood species is difficult at best. Typically, analytical techniques that have been used for identification are invasive, require some level of sampling, and can have expenses associated with them<sup>1</sup>. In some cases, the analysis can be undertaken using microscopic techniques<sup>2</sup>, while more advanced techniques using visual learning, chemical analysis and genetics are making rapid strides<sup>3–6</sup>. The importance of being able to identify tree species is essential in the continuing fight to control the illegal trade of lumber<sup>1,7</sup>.

Illegal logging and the trade in illegal timbers are a largely recognized threat to global biodiversity<sup>8,9</sup>. Resolution 23/1 was adopted by the United Nations in May 2014 promoting the development of tools and technologies that could be used in the fight against illicit trafficking of timbers<sup>10,11</sup>. Treaties such as the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) along with current laws and certification schemes provide a framework to protect over-exploited species and control illegal trade<sup>8</sup>. The use of a rapid, non-destructive forensic verification technique to ensure the compliance of wood products as they travel through the entire chain of providers to users would enhance governmental and certification bodies' efforts<sup>10</sup>. Without species validation, the economic incentive to provide false information is substantial, and timber is easily misattributed to circumvent trade regulations<sup>4</sup>.

Environmental protection laws exemplified by the US Lacey Act require the declaration of full scientific names with genus and species<sup>10</sup>. Since logs and sawn timber lack the morphological characters of leaves, fruits, seeds and flowers traditionally used in species determination<sup>12</sup>, accurate anatomical identification down to the species level both rapidly in the field and through more invasive and time-consuming laboratory investigations can be challenging<sup>1</sup>. This is especially true in tropical forests with high species diversity where the anatomy of a specimen is often not well known<sup>13</sup>.

Traditional microscopic examination of wood sections has been a staple technique to narrow down genus, and in some cases species, but is time consuming and requires specialist knowledge<sup>10</sup>. Front-line officers who perform initial inspections are trained and knowledgeable about related issues, but may have limited or insufficient taxonomical expertise to verify species and make quick decisions as to whether or not the wood in question needs additional scientific testing<sup>14</sup>. While it is recognized that rapid-field identification is important, the cost and effort has often been cited as a hindrance to the process<sup>1</sup>. To overcome the lack of trained anatomists and the analytical limitations for determining species, several separate scientific disciplines have focused their attention on this problem<sup>1,6</sup>.

No one scientific methodology can address all timber identification questions related to genus, species, geographic region of origin, age, and individual species, so multiple approaches are often needed<sup>8</sup>. For the reliable identification of select species, visual learning, chemical analysis, and genetics have been employed<sup>8</sup>. Examples include machine vision<sup>15</sup>, near infrared spectroscopy (NIRS)<sup>12</sup>, direct analysis in real time time-of-flight mass spectrometry (DART TOFMS)<sup>3</sup>, and DNA barcoding<sup>16</sup>. Most of these techniques are in their early stages of development with limited application, but with recognized potential to develop<sup>1</sup>. These methodologies are mainly

<sup>1</sup>Garman Art Conservation Department, Buffalo State College – SUNY, Buffalo, NY, USA. <sup>2</sup>Department of Anthropology, University of New Mexico, Albuquerque, NM, USA. <sup>3</sup>Collections, Curatorial and Conservation Branch, Indigenous Affairs and Cultural Heritage Directorate, Parks Canada, Government of Canada, Ottawa, ON, Canada. ✉email: shugaran@buffalostate.edu

lab-based systems that are currently expensive and not readily available. They require specific training both to use and interpret resulting data. DNA barcoding relies on extensive databases, and all but NIRS require small samples to be taken for analysis.

The development of a non-destructive, easy to employ, cost effective identifier that can be used by front-line workers would help to improve rapid field-level inspections as well as forensic wood identification. In addition, effective scientific verification at the beginning of the supply chain could have great impact tracing and verifying legal timber along the global supply network in efforts to curb illegal logging and prevent lumber substitution. Such technology could be used as an early screening process or a 'triage' that could further separate out tree species and select key trees for additional confirmation testing<sup>1,13,14</sup>. These requirements may well be met by the use of X-ray fluorescence spectrometers which could become a vital tool in the fight against illegally traded lumber and as well as assistance for species identification for other areas of study.

X-ray fluorescence (XRF) is a rapid elemental identifier that has been employed extensively in various fields for the characterization of materials. It has had limited use for the identification of wood mainly because of the slight variations in wood chemistry making it difficult to create accurate calibrations which could allow for the clear separation of genus or species. XRF has successfully been used to identify heavy metals in woods<sup>17</sup> and has shown some limited use to characterize variable density in wood<sup>18</sup>. This method has recently been taken one step further through the application of X-ray computed tomography, which is being used for wood identification but still requires samples to be taken and is time consuming<sup>19</sup>. The real value in XRF is that it can be portable (portable or handheld XRF), analysis is rapid, and testing can be done on site (scans can often be done in less than one minute per sample). Instrumentation is relatively cheap, has a small footprint, and is easy to learn how to use.

With the advent of more powerful computers and the development of machine learning and artificial intelligence, we are now able to reassess the potential for XRF to perform rapid wood-species identification. XRF analysis combined with machine learning and deep learning neural networks algorithms allow for more detailed investigations of materials that in the past were not seen as possible. For example, by using random forest regression on XRF data, moisture concentrations in manure have been determined<sup>20</sup>. In addition, using neural networks models in conjunction with XRF analysis, the total organic carbon content of geological shale has been analyzed<sup>21</sup>. The application of neural networks, which allows for the continual building of model design, has been fundamental to the development of wood-species identification using XRF as presented here.

The field of artificial intelligence has seen extraordinary progress since the groundbreaking AlexNet model made 2D convolutional neural networks (CNNs) mainstream<sup>22</sup>. CNN models are designed to mirror how vision systems work in animals through the use of shifting positions<sup>23</sup>. These models tend to diverge from other statistical methods in the following ways:

- They can interpret data in context-specific situations (e.g. facial features or the ability to classify and separate paintings by different artists<sup>24</sup>).
- When completed, they can be retrained on new data.

The output of these models is a set of interconnected neural weights that can be retrained as new data is made available. These ongoing 'learning models' develop and improve over time. The danger to these models is that they learn too much from datasets, a problem known as overfitting. One potential resolution to this problem is to force neurons to be dropped randomly<sup>25,26</sup>.

CNNs can be 1 dimensional and suited for strings of numbers, not unlike the way XRF spectra are presented. Because of lower dimensionality, these networks require exponentially fewer computer resources to train, making them practical on most modern computer systems. They are particularly well suited for time-series data-classification tasks<sup>27</sup>. These models have been used successfully with spectral data to detect atrial fibrillation<sup>28</sup> and identify acids in IR spectra<sup>29</sup>, among other applications.

The application of CNN for the identification of wood species has been developing over the last few years. The implementation of CNN combined with visual assessment (imaging) of gross features has been used for the identification of veneers, fine-grained wood specimens, and several tropical wood species<sup>14,30–32</sup>. Images of microscopic sections have also been classified using CNN<sup>33</sup>, and the sorting of several *Pinus* species has been successfully undertaken using an artificial neural network<sup>34</sup>. To date, there has been no use of XRF in combination with CNN for the identification of wood species.

This paper will present the first use of XRF spectral collection in combination with convolutional neural network machine learning to accurately identify various wood species. As a pilot study, it provides strong evidence for the validity of this rapid, non-destructive identification technique and allows future researchers a path forward for its development and implementation.

## Methodology

**Selection of samples.** Wood blanks used for this study are housed in the Garman Art Conservation Department, SUNY—Buffalo State. They are part of various educational wood identification sets that were purchased over 20 years ago and consist mostly of heartwood. No samples violated CITES regulations, nor crossed state or international borders during this study (Fig. 1, Table 1). If multiple samples were available of one species, they were used to develop a more robust model. Forty-eight wood specimens were investigated with 17 of those having more than one sample run for comparison giving a total of 66 datasets. Samples were all run on either tangential or radial sections, or slight rotations between these two orientations.

**Collection of spectra.** Blank wood samples were analyzed using a Bruker Tracer 5 g XRF coupled to a DeWitt MSS-150E scanning stage. The XRF was run at 50 kV 35uA with an 8 mm collimator and no filter for



**Figure 1.** Example of some of the wood samples used in this study (all housed in the Garman Art Conservation Department, SUNY – Buffalo State).

20 s per scan. Samples were placed so that the head of the instrument was no more than 1 mm from the surface of the wood. The scanning stage was set to move at 3 mm in the X axis per scan and then shift up 3 mm in the Y axis to create a matrix that produced approximately 260 scans per wood sample. Proprietary software outputted data as individual Bruker PDZ files. The spectra were then converted to csv files for processing.

**Application of machine learning algorithm.** Spectra data were imported using the open-source Cloud-Cal software<sup>35</sup> with spectra data normalized to total counts (e.g., each detector channel representing a % of the total, rather than counts). The spectrum was then compressed to 0.7 and 37 keV in 100 eV increments.

A one-dimensional convolutional neural network was generated using 3 1D CNN layers with filter sizes of 32, 64, and 128 with intermediate max pooling and batch normalization layers. Each 1D CNN layer had descending kernel sizes of 5, 3, and 2, which were fed into an interconnected network with 512, 256, 128, and 64 neurons with 60% dropout rates (Fig. 2).

This model was trained over 2000 epochs (e.g., forward and backward proportions of the network through the training dataset) using the Nadam optimizer, categorical crossentropy as a loss function, and a learning rate of 0.00001 with a batch size of 4. Fifteen percent of the data was randomly masked from training to be used for model validation.

A second smaller dataset including only wood samples with repeated representatives for each species was also used with the same model architecture and training method, though for only 1500 epochs. A third data set of random words with random spectra associations was made as a control to verify that the wood identification models were the product of meaningful signals, and not memorization by the network<sup>36</sup>.

The model is available in the open source SheetCrunch software<sup>37</sup> using the Keras framework in Tensorflow 2.2.

## Results and discussion

Tree species identification was found to be 99% accurate in the cross validation set for all 48 species investigated.

The 1D CNN model (Fig. 3) converged after 500 epochs for the full data set and after only 250 for the partial data set (Fig. 4). Accuracy for the full model saw considerably higher variation in cross-validation results (Fig. 3, Supplemental Table 1) compared to the repeated model (Fig. 4, Supplemental Table 2), though both models have nearly identical feature importance plots (Figs. 5, 6). Training and validation accuracies were 98.5% and above (Table 2). Randomized data had much poorer performance with training and validation accuracies near 7% (Table 2, Supplemental Table 3).

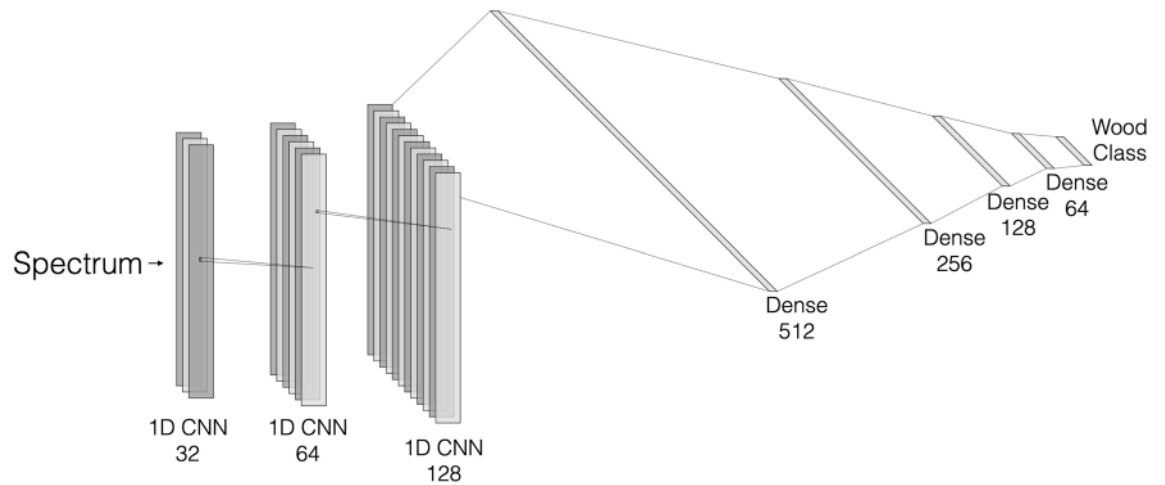
The model performed well on wood identification for both the full and partial data sets (Table 2, see supplementary data for more detail). The accuracy was high enough to warrant concern that wood ID was not being addressed as much as the memorization of the data<sup>33</sup>. While randomized cross-validation in the manner presented in this manuscript is the traditional way of evaluating a model, an additional random word dataset with randomized associations to the spectra were generated as an additional test. If the same neural architecture could learn random associations as strongly as it learned wood taxa, then this could suggest either malfunctioning random cross-validation or data leakage (e.g., the validation data was included in the model training). The randomized data showed poor classification with training and validation accuracies near 7% (Table 2). For comparison, a fully random process would be approximately 3.5% accurate in identifying 29 classes. The higher performance in the random model was the result of uneven randomness in variables—it simply predicted one word (she) that had the most instances in the dataset. This is why the specificity of the random model is higher

Scientific name	Common names	Samples run
<i>Abies balsamea</i>	Balsam fir	1
<i>Acer saccharum</i>	Sugar maple	2
<i>Castanea dentata</i>	American chestnut	1
<i>Cedrela odorata</i>	Spanish cedar	1
<i>Cedrela</i> spp.	Cedro	1
<i>Dyera costulata</i>	Jelutong	1
<i>Fagus grandifolia</i>	America beech	2
<i>Gleditsia triacanthos</i>	Honey locust	2
<i>Juglans cinerea</i>	Butternut	2
<i>Juglans nigra</i>	Black walnut	2
<i>Juniperus virginiana</i>	Red cedar	1
<i>Lagarostrobos franklinii</i>	Huon pine	1
<i>Liquidambar styraciflua</i>	Sweetgum	2
<i>Liriodendron tulipifera</i>	Yellow poplar, tulip poplar	2
<i>Maclura pomifera</i>	Osage orange	1
<i>Morus alba</i>	White mulberry	1
<i>Morus rubra</i>	Red mulberry	1
<i>Ostrya virginiana</i>	American hophornbeam, ironwood	1
<i>Picea rubens</i>	Eastern spruce	1
<i>Picea rubens</i>	Red spruce	1
<i>Picea</i> spp.	Spruce	1
<i>Pinus echinata</i>	Shortleaf pine	1
<i>Pinus monticola</i>	Western white pine	1
<i>Pinus palustris</i>	Longleaf pine	1
<i>Pinus ponderosa</i>	Ponderosa pine, western yellow pine	1
<i>Pinus</i> spp.	Southern yellow pine	2
<i>Pinus strobus</i>	Eastern white pine	2
<i>Pinus virginiana</i>	Virginia pine	1
<i>Platanus occidentalis</i>	American sycamore	2
<i>Populus balsamifera</i>	Balsam poplar	1
<i>Populus</i> spp.	Aspen	1
<i>Prosopis</i> spp.	Mesquite	1
<i>Prunus serotina</i>	Black cherry	2
<i>Prunus serrula</i>	Birch bark cherry	1
<i>Pyrus communis</i>	Pearwood	1
<i>Quercus alba</i>	White oak	2
<i>Quercus falcata</i>	Southern red oak	1
<i>Quercus rubra</i>	Northern red oak	1
<i>Salix nigra</i>	Black willow	2
<i>Sassafras</i> spp.	Sassafras	1
<i>Shorea</i> spp.	Philippine mahogany	2
<i>Swietenia macrophylla</i>	Caoba	1
<i>Swietenia macrophylla</i>	Honduran mahogany	2
<i>Taxodium</i> spp.	Cypress	2
<i>Thespesia grandiflora</i>	Maga	1
<i>Thuja plicata</i>	Western red cedar	1
<i>Tilia americana</i>	Basswood	3
<i>Ulmus americana</i>	American elm	1

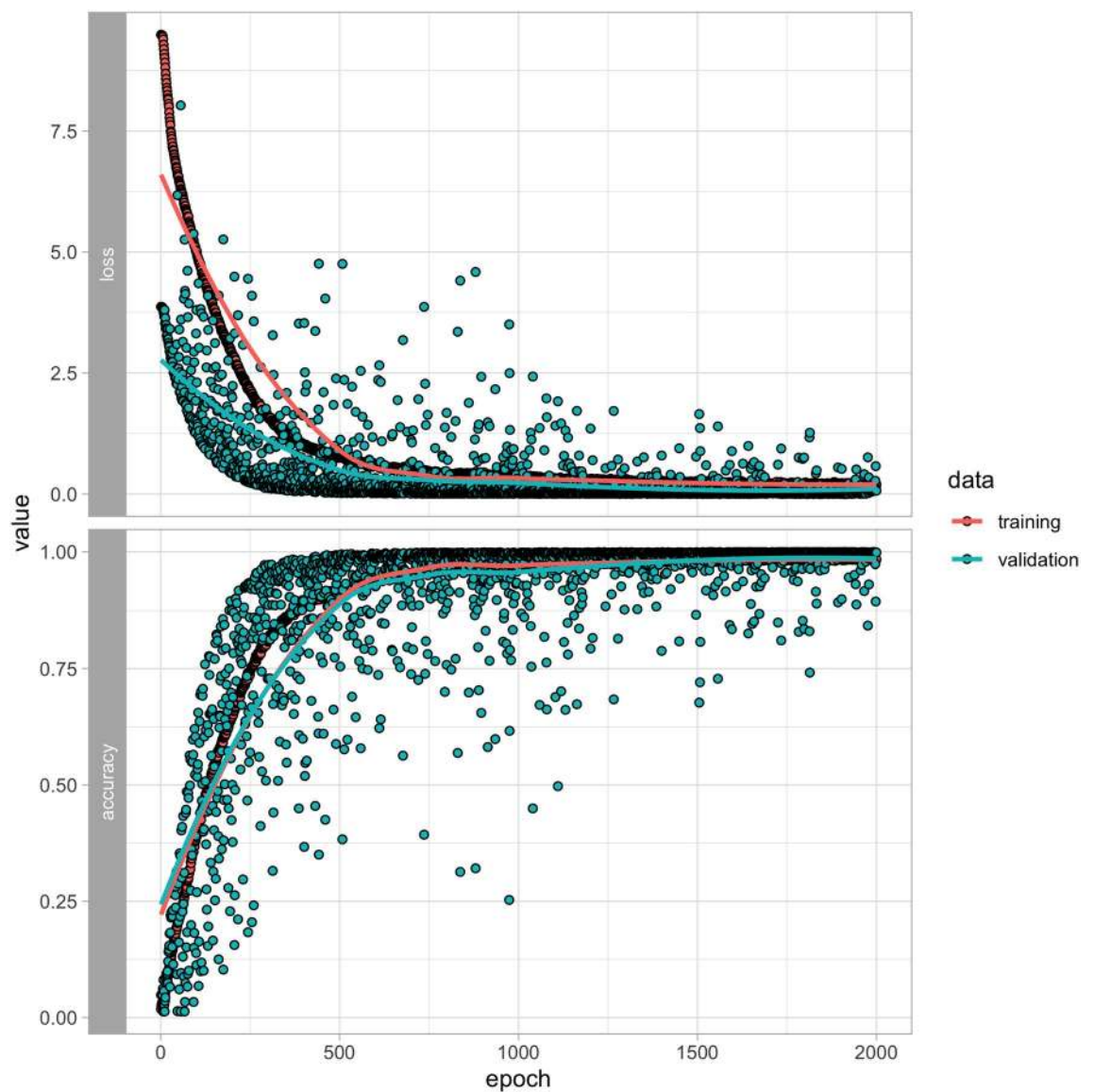
**Table 1.** List of wood species analyzed and the number of samples run for each.

than would be expected (Table 2). The random model test illustrates that neither data memorization nor data leakage can explain the high accuracy in the wood taxon classification models.

Other machine learning approaches have inferred unusual properties from XRF spectra including water<sup>20</sup> and carbon/nitrogen<sup>38</sup> from organic fertilizers, organic carbon from sedimentary cores<sup>21</sup>, and meteorite classification<sup>39</sup>. While these approaches do not use 1D CNN models, they do highlight functions that are not

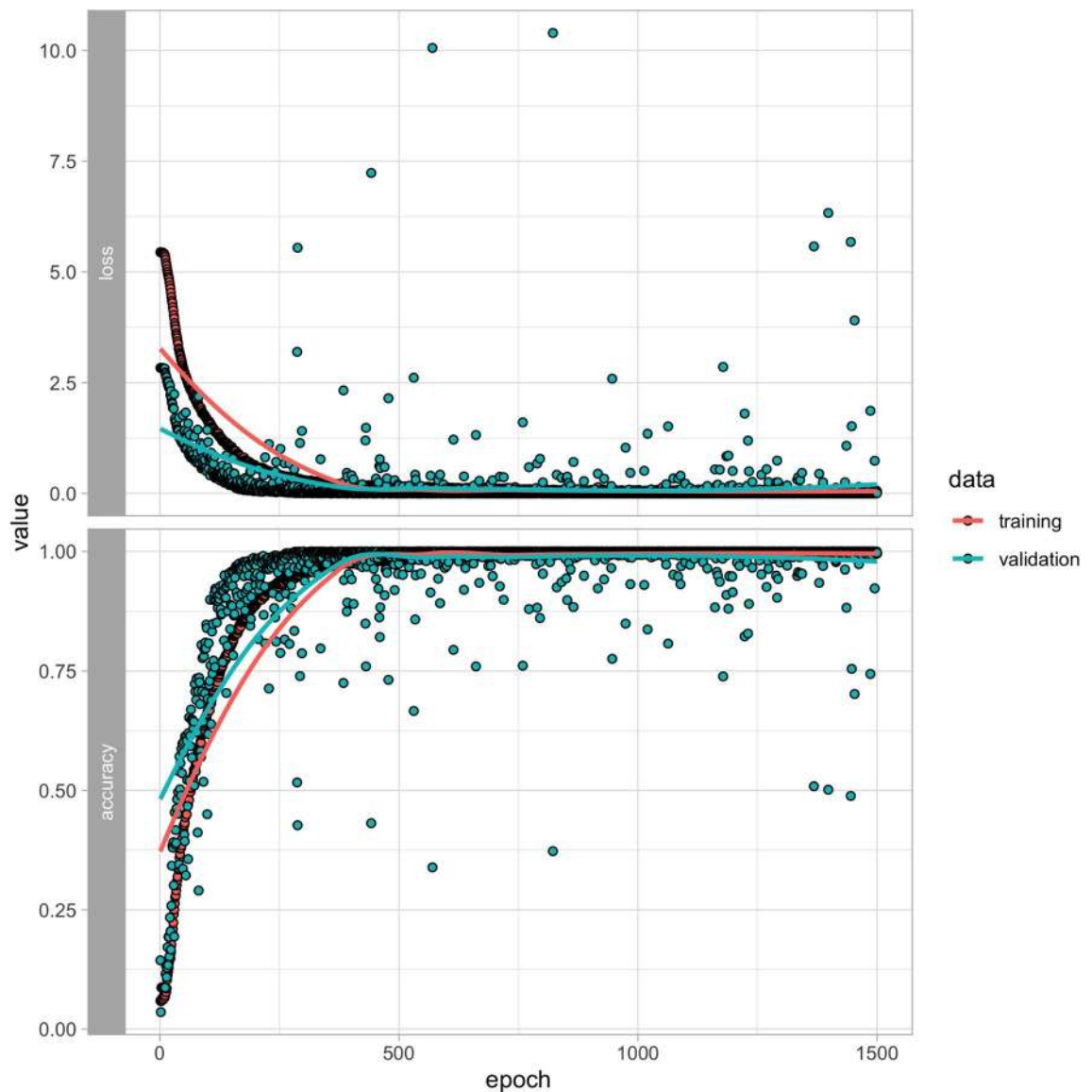


**Figure 2.** 1D CNN Architecture for classification.



**Figure 3.** Training and validation loss (top) and accuracy (bottom) for the training ( $n = 14,672$ ) and validation ( $n = 2590$ ) portions of the full dataset. Models converged most of the time after 500 epochs, though considerable variation in validation accuracy is seen throughout training.





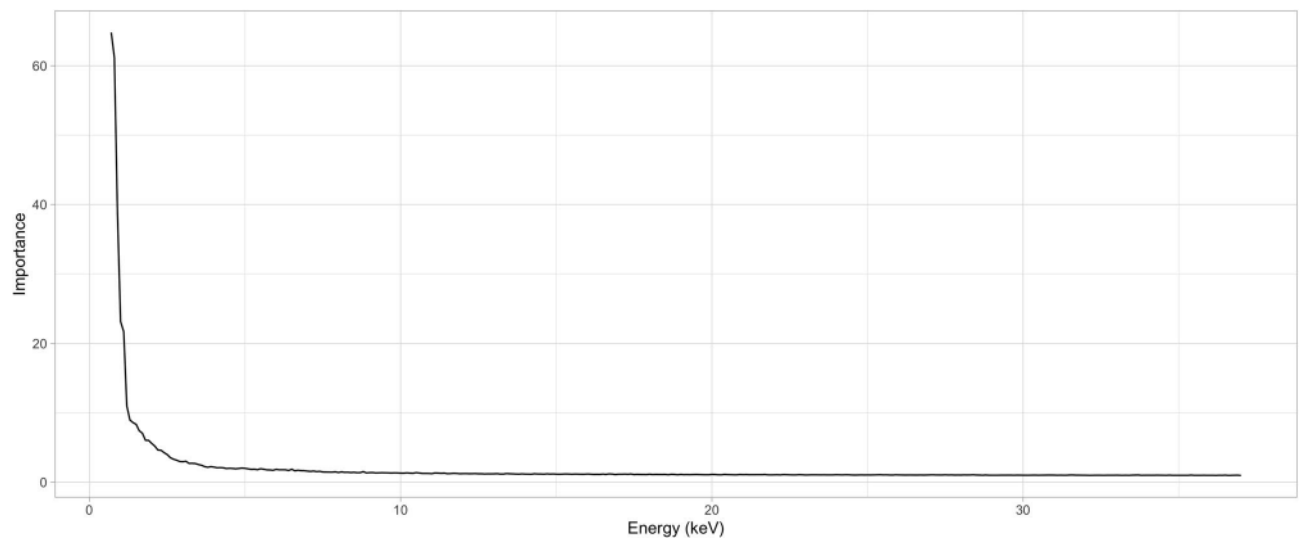
**Figure 4.** Training and validation loss (top) and accuracy (bottom) for the training ( $n=7825$ ) and validation ( $n=1382$ ) portions of the repeated dataset. Models converged most of the time after 250 epochs with a smaller range of validation accuracies compared to the full training set.

typically considered for XRF. Machine learning expands the applicability of spectrometer-based technologies considerably beyond their traditional uses.

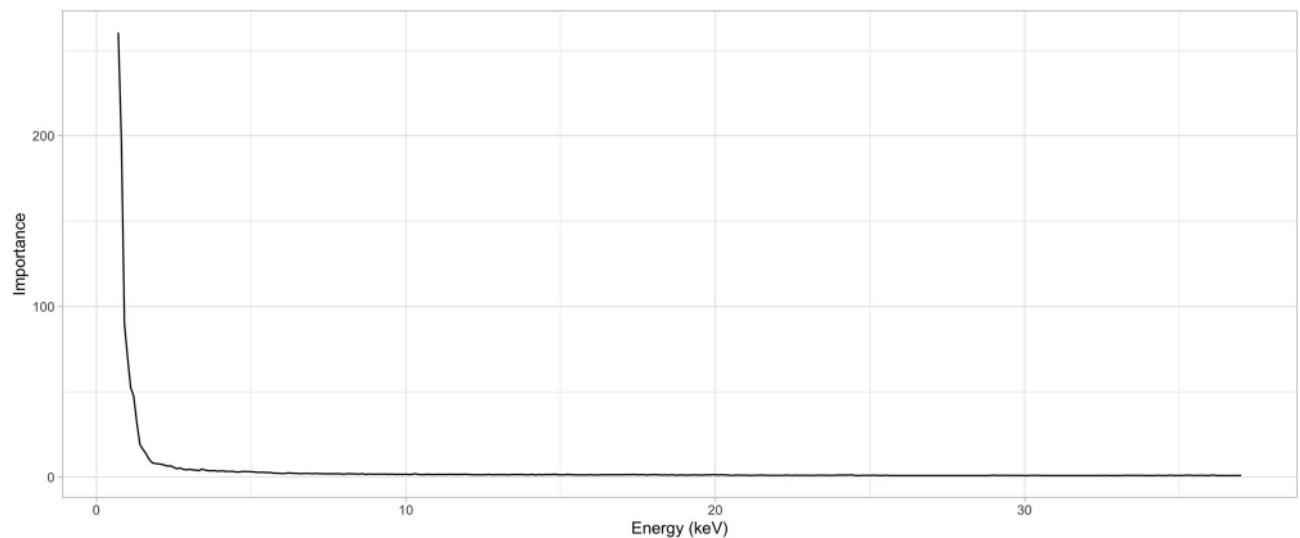
The full model saw accuracies as high as 100% and as low as 45% in the last 200 epochs (Fig. 3), while the range of validation accuracies is considerably narrower (Fig. 4). This is likely a product of having a single sample of a species represented for some groups. The multi-sample dataset provides a more representative example of how multiple observations of multiple objects is better suited to training networks.

Despite the limitation of single-sample beams to represent groups in the full dataset, the feature importance plots are nearly identical (Figs. 5, 6), with 0.7–1.7 keV being the most important portion of the spectra to classify wood. This is a surprising result as the elements that fluoresce in this region most prominently are sodium (Na), magnesium (Mg), aluminum (Al), and silicon (Si). Other elements found in woods include calcium (Ca), potassium (K), iron (Fe), sulfur (S), and phosphorous (P); however, all are found in rather small concentrations<sup>40</sup>. The majority of wood is comprised of carbon, oxygen, hydrogen, and nitrogen, mostly formed as cellulose and lignin. It is therefore unclear why this area of the spectrum is so important, but some hypotheses for the reason include:

1. The network infers most of what it needs to know from the first few keV of the spectra, making the rest of the spectrum redundant.
2. Variation in these elements may be more distinct (even in smaller concentrations) than previously thought, or components such as  $\text{SiO}_2$ , which is the main component of phytoliths, can be diagnostic for plant taxa.



**Figure 5.** Importance features for the full dataset using categorical crossentropy as a loss function. Note that the most relevant information is included from 0.7 to 1.7 keV.



**Figure 6.** Importance features for the repeated dataset using categorical crossentropy as a loss function. Feature importance is largely identical to the full dataset.

	N		Accuracy		Validation Mean Sensitivity (%)	Validation Mean Specificity (%)	Validation Mean Optimization Precision (%)
	Train	Validation	Train (%)	Validation (%)			
Full	14,672	2590	99.9	99.9	99.8	99.9	99.8
Repeats	7825	1382	99.9	99.7	99.7	99.8	99.6
Random	14,672	2590	7.2	6.9	3.4	96.7	−93.1

**Table 2.** Final training and validation accuracy, sensitivity, specificity, and optimization precision for full, partial, and randomized data sets. Note that the high specificity for the random model is a product of it repeating the same prediction.

3. The bremsstrahlung radiation reflected from 0.7 to 2 keV is reflective of density changes between wood taxa, which has been shown to be informative<sup>18</sup>.

A combination of these factors may well explain the nature of the model. Of the three hypotheses, number 1 is the most easily tested. By experimenting with different XRF spectra acquisition techniques (e.g., use an attenuating filter to minimize bremsstrahlung radiation) feature importance can be re-evaluated.

While this pilot study shows a high potential for success, there are some areas that need to be developed to corroborate the findings and create a more substantial working database. The samples investigated here derived from a study collection and may not meet the required taxonomic integrity of reference materials. Vouchered specimens should be analyzed to ensure that proper identification has been met<sup>8</sup>. Limited quantities of reference materials in existing collections and a lack of authenticated tree specimens worldwide<sup>1,7</sup> make this challenging. The development of a comprehensive database is also required to advance XRF (or any other analytical technique) for the identification of tree species.

The samples investigated in this study were mostly heartwood, and the scans were mainly done on radial and tangential sections, or slight rotations between these two orientations. It would be necessary to test this system on sapwood and on transverse sections of wood to see if the same results followed suit, or if alternative calibrations would be required for these orientations. In addition, variation in moisture content was not considered for this study. Future work should consider investigating the effects of moisture content of the photons scattering.

Ideally with deep learning, the more data analyzed the better both across and within sample types. In this study, 48 tree species were tested with only 17 of those having multiple samples (mostly 2 samples and 3 samples for *Tilia americana*). This needs to be expanded to ensure that the findings for each of the tree species is indeed an identifier for that species and not more related to associated soil chemistry or geographical defining features (e.g., weather alteration may cause variation in density). In addition, looking at the same species with different known geographic locations would be necessary to assess the model's validity for geographic variation in wood growth.

While deep learning traditionally requires significant hardware (typically GPUs), the 1D CNN model presented here can run on most CPU systems due to its lower dimensionality. Because the network is re-trainable (e.g., the neurons can be reloaded into a model), the data could be adapted for any XRF instrument provided its spectra are transformed in the same manner (0.7–37 keV in 100 eV increments), and there is some light retraining with representatives of the classes.

## Conclusion

This study presents the first use of XRF for the identification of tree species using a convolutional neural network machine learning model. This rapid, non-invasive, non-destructive analytical technique provides highly accurate identification of wood species. The findings show great promise for XRF's use to help limit the illegal trade and transport of lumber. The ability to rapidly use XRF with limited training offers border guards the ability to make more accurate assessments of lumber and better triage which samples might require additional cross validation using more invasive techniques. Although this pilot study shows highly accurate identification, additional research should be undertaken to consider the effect and results of various wood orientations, moisture content, addition of biocides, and variations of species and geographic locations. A clear benefit to using non-invasive analytical techniques such as NIRS or XRF for wood analysis is that they can be used on finished products or timbers that are deemed too valuable from which to remove a sample. Rapid XRF species identification would also benefit other fields of study including the art market and art conservation, where it could aid in authentication and provenance studies as well as ensure the use of proper replacement components in conservation treatments.

Received: 23 April 2021; Accepted: 13 August 2021

Published online: 02 September 2021

## References

- Wiedenhoeft, A. Best practice guide for forensic timber identification. In *United Nations Office on Drugs and Crime: International Consortium on Combating Wildlife Crime*. (Laboratory and Scientific Section, Global Programme for Combating Wildlife and Forest Crime Vienna, Austria, 2016).
- Gasson, P. How precise can wood identification be? Wood anatomy's role in support of the legal timber trade, especially CITES. *IAWA J.* **2011**, 137–154 (2011).
- Deklerck, V. *et al.* Chemical fingerprinting of wood sampled along a pith-to-bark gradient for individual comparison and provenance identification. *Forests* **11**, 107 (2020).
- Espinoza, E. O., Wiemann, M. C., Barajas-Morales, J., Chavarria, G. D. & McClure, P. J. Forensic analysis of CITES protected Dalbergia timber from the Americas. *IAWA J.* **36**, 311–325 (2015).
- McClure, P. J., Chavarria, G. D. & Espinoza, E. Metabolic chemotypes of CITES protected Dalbergia timbers from Africa, Madagascar, and Asia. *Rapid Commun. Mass Spectrom.* **29**, 783–788 (2015).
- Schmitz, N. *et al.* Overview of current practices in data analysis for wood identification. A guide for the different timber tracking methods. (2020).
- Gasson, P. E. *et al.* WorldForestID: Addressing the need for standardized wood reference collections to support authentication analysis technologies; a way forward for checking the origin and identity of traded timber. *Plants People Planet* **3**, 130–141 (2020).
- Dormontt, E. E. *et al.* Forensic timber identification: It's time to integrate disciplines to combat illegal logging. *Biol. Conserv.* **191**, 790–798 (2015).
- Union, E. Regulation (EU) No 995/2010 of the European Parliament and of the Council of 20 October 2010 laying down the obligations of operators who place timber and timber products on the market. *Off. J. L* **295**, 23–34 (2010).
- Lowe, A. J. *et al.* Opportunities for improved transparency in the timber trade through scientific verification. *Bioscience* **66**, 990–998 (2016).
- United Nations. Strengthening a targeted crime prevention and criminal justice response to combat illicit trafficking in forest products, including timber. United Nations office on drugs and crime: Commission on Crime Prevention and Criminal Justice (CCPCJ), 23rd Session (2014).
- Bergo, M. C. J., Pastore, T. C. M., Coradin, V. T. R., Wiedenhoeft, A. C. & Braga, J. W. B. NIRS identification of Swietenia macrophylla is robust across specimens from 27 countries. *IAWA J.* **37**(3), 420–430 (2016).



13. Wheeler, E. & Baas, P. Wood identification—A review. *IAWA J.* **19**, 241–264 (1998).
14. Olschofsky, K. & Köhl, M. Rapid field identification of CITES timber species by deep learning. *Trees For. People* **2**, 100016 (2020).
15. de Andrade, B. G., Basso, V. M. & de Figueiredo Latorraca, J. V. Machine vision for field-level wood identification. *IAWA J.* **41**, 681–698 (2020).
16. Jiao, L., Lu, Y., He, T., Guo, J. & Yin, Y. DNA barcoding for wood identification: Global review of the last decade and future perspective. *IAWA J.* **41**, 620–643 (2020).
17. Block, C. N., Shibata, T., Solo-Gabriele, H. M. & Townsend, T. G. Use of handheld X-ray fluorescence spectrometry units for identification of arsenic in treated wood. *Environ. Pollut.* **148**, 627–633 (2007).
18. Zielenkiewicz, T., Radomski, A. & Zawadzki, J. XRF examination of matrix uniqueness in chosen deciduous wood species. *Eur. J. Wood Prod.* **70**, 845–849 (2012).
19. Kobayashi, K., Hwang, S.-W., Okochi, T., Lee, W.-H. & Sugiyama, J. Non-destructive method for wood identification using conventional X-ray computed tomography data. *J. Cult. Heritage* **38**, 88–93 (2019).
20. Sapkota, Y., Drake, B. L., McDonald, L. M., Griggs, T. C. & Basden, T. J. Elemental composition and moisture prediction in manure by portable X-ray fluorescence spectroscopy using random forest regression. *J. Environ. Qual.* **49**, 472–482 (2020).
21. Lawal, L. O., Mahmoud, M., Alade, O. S. & Abdulraheem, A. Total organic carbon characterization using neural-network analysis of XRF data. *Petrophysics* **60**, 480–493 (2019).
22. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012).
23. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**, 193–202 (1980).
24. Lecoutre, A., Negrevergne, B. & Yger, F. Recognizing art style automatically in painting with deep learning. In *Asian Conference on Machine Learning, Proceedings of Machine Learning Research* Vol.77, 327–342 (2017).
25. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. Preprint at <https://arxiv.org/abs/1207.0580> (2012).
26. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
27. Hatami, N., Gavet, Y. & Debayle, J. Classification of time-series images using deep convolutional neural networks. In *Tenth International Conference on Machine Vision (ICMV 2017)* vol. 10696, 106960Y (International Society for Optics and Photonics, 2018).
28. Hsieh, C.-H., Li, Y.-S., Hwang, B.-J. & Hsiao, C.-H. Detection of atrial fibrillation using 1D convolutional neural network. *Sensors* **20**, 2136 (2020).
29. Chen, X., Chai, Q., Lin, N., Li, X. & Wang, W. 1D convolutional neural network for the discrimination of aristolochic acids and their analogues based on near-infrared spectroscopy. *Anal. Methods* **11**, 5118–5125 (2019).
30. Urbonas, A., Raudonis, V., Maskeliūnas, R. & Damaševičius, R. Automated identification of wood veneer surface defects using faster region-based convolutional neural network with data augmentation and transfer learning. *Appl. Sci.* **9**, 4898 (2019).
31. Shustrov, D., Eerola, T., Lensu, L., Kälviäinen, H. & Haario, H. Fine-grained wood species identification using convolutional neural networks. In *Image Analysis* (eds Felsberg, M. et al.) 67–77 (Springer, Berlin, 2019).
32. Kwon, O. et al. Automatic wood species identification of Korean softwood based on convolutional neural networks. *J. Korean Wood Sci. Technol.* **45**, 797–808 (2017).
33. Lens, F. et al. Computer-assisted timber identification based on features extracted from microscopic wood sections. *IAWA J.* **41**, 660–680 (2020).
34. Esteban, L. G. et al. Application of artificial neural networks as a predictive method to differentiate the wood of *Pinus sylvestris* L. and *Pinus nigra* Arn subsp. *salzmannii* (Dunal) Franco. *Wood Sci. Technol.* **51**(5), 1249–1258 (2017).
35. Drake, B. L. CloudCal v3.0. GitHub. <https://github.com/leedrake5/CloudCal>. Accessed 16 July 2021 (2018).
36. Gu, J. & Tresp, V. Neural network memorization dissection. Preprint at <https://arxiv.org/abs/1911.09537> (2019).
37. Drake, B. L. SheetCrunch. <https://github.com/leedrake5/sheetCrunch>. Accessed 16 July 2021 (2020).
38. Towett, E. K. et al. Comprehensive nutrient analysis in agricultural organic amendments through non-destructive assays using machine learning. *PLoS ONE* **15**, e0242821 (2020).
39. Allegretta, I. et al. Macro-classification of meteorites by portable energy dispersive X-ray fluorescence spectroscopy (pED-XRF), principal component analysis (PCA) and machine learning algorithms. *Talanta* **212**, 120785 (2020).
40. Esch, M. et al. Elemental composition of different types of wood. *Nucl. Instrum. Methods Phys. Res. Sect. B: Beam Interact. Mater. Atoms* **109–110**, 328–331 (1996).

## Acknowledgement

The authors would like to thank the Andrew W. Mellon Foundation for supporting this research.

## Author contributions

A.S. helped to conceptualize the experiments, perform the XRF analysis, review the data, and write the manuscript. L.D. helped to conceptualize the experiments, perform the machine learning analysis, review the data, and write the manuscript. G.K. helped to research the background information, provide expert advice on samples analyzed, and write and edit the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-96850-2>.

**Correspondence** and requests for materials should be addressed to A.N.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021, corrected publication 2022