

RESEARCH

Open Access



Utilization of a natural language processing-based approach to determine the composition of artifact residues

Tung Tho Nguyen^{1*} and Korey J. Brownstein^{2*}

*Correspondence:
nguyenthotung@gmail.com;
korey.brownstein@usda.gov

¹ Department of Mathematics,
The University of Chicago,
Chicago, IL 60637, USA

² Department of Molecular
Genetics and Cell Biology, The
University of Chicago, Chicago, IL
60637, USA

Abstract

Background: Determining the composition of artifact residues is a central problem in ancient residue metabolomics. This is done by comparing mass spectral features in common with an experimental artifact and an ancient artifact (standard method). While this method is simple and straightforward, we sought to increase the accuracy of predicting which plant species had been used in which artifacts.

Results: Here, we introduce an algorithm (new method) based on ideas from the field of natural language processing (NLP) to solve this problem. We tested our strategy on a set of modern clay pipes. To limit biases, we were not provided information on which plant species had been smoked in which clay pipes. The results indicate that our new method performed 12.5% better than the standard method in predicting the plant species smoked in each artifact.

Conclusions: Utilizing an NLP-based approach, we developed a robust algorithm for characterizing the composition of artifact residues. This work also discusses other general applications in which our algorithm could be used in the field of metabolomics, such as datasets where there are a limited number of replicates.

Keywords: Ancient residue metabolomics, Cosine similarity, Term frequency-inverse document frequency

Background

Metabolomics is the systematic quantitative and qualitative study of small molecules (or mass spectral features) in biological systems. Brownstein et al. [1] expanded upon this field with their ancient residue metabolomics-based method. Albeit the mass spectral features in this study were not derived from biological systems, they were residues left behind from biological processes, i.e., originating from plants including several *Nicotiana* species that were smoked by Indigenous Peoples. Before the Brownstein et al. [1] study, ancient residue analysis relied on the biomarker approach. However, the biomarker approach failed to distinguish between related species, leaving open questions about the relationship between plants and people. For ancient residue metabolomics, significant mass spectral features (i.e., singular peaks of small molecules above a



specified noise threshold) are of interest, which can improve the resolution of determining which plants species had been smoked in a particular artifact [1].

In ancient residue metabolomics, data from hyphenated chromatography instruments (such as gas chromatography- or liquid chromatography-mass spectrometer) are processed and aligned in MZmine 2 [2], Progenesis QI (Waters Corporation, Milford, MA, USA), or other “omics” software. Afterwards, these data are exported from the “omics” software and then processed manually (standard method) as described in Brownstein et al. [1] to determine which plant species may have been used in an ancient artifact. Because this step requires a manual process, it can introduce errors and is time consuming. Various metabolomics data analysis and interpretation platforms exist including MetaboAnalyst 5.0 [3] and XCMS Online [4]; however, these platforms are limited in their ability to process datasets from ancient residue metabolomics studies. Therefore, we introduce a novel, automated approach (new method) for determining the composition of organic residues in modern smoking artifacts utilizing techniques and ideas from the field of natural language processing (NLP).

Results and discussion

We used Python scripts because of the availability of data analysis, deep learning, and machine learning libraries. We also developed a script that automates the standard method described in Brownstein et al. [1], as well as utilized recent advances in NLP to better predict which plant species had been smoked in a particular artifact. All scripts and datasets are freely available on GitHub: https://github.com/tungprime/NLP_and_composition_of_artifact_residues. Term frequency-inverse document frequency (TF-IDF) has been previously used in imaging mass spectrometry for the co-localization of ions [5], as well as to score mass spectral feature outputs against theoretical spectrums [6]. These use cases exemplify that TF-IDF can be implemented as a method to determine similarities between mass spectrums, or in our study, samples. As shown in Table 1, the new method predicted that blind clay pipe 1 (BCP1) was most likely smoked with *Nta* (0.0370). Table 2 summarizes the model predictions of four separate methods, and the key provides the expected results. While the standard method only predicted four out of eight (50.0%) of the samples correctly, the new method performed slightly better, i.e., it classified five out of eight (62.5%) of the samples correctly (Table 2). This is a 12.5% improvement in accuracy. A second method, where *tf* was replaced with $1 + \log(\text{tf})$, classified 62.5% of the samples correctly; however, the similarity scores for this method (aside from BCP7) were lower than the new method (Table 2). Testing a third method, as well as a pointwise mutual information (PMI) method revealed that only three out of eight (37.5%) and four out of eight (50.0%) samples were correctly classified, respectively.

Table 1 Similarity scores for blind clay pipe 1 (BCP1) smoked with an unknown plant sample

	<i>Nta</i>	<i>Ngl</i>	<i>Nat</i>	<i>AmSp</i>	<i>Nob</i>
Similarity scores	0.0370	0.0288	0.0233	0.0186	0.0168

Clay pipes experimentally smoked with a single, known plant species or AmSp were individually compared to BCP1. Only the top five scores were included in the table. *Nta* is the most likely candidate smoked in BCP1. *Nat*, *Nicotiana attenuata*; *Ngl*, *Nicotiana glauca*; *Nob*, *Nicotiana obtusifolia*; *Nta*, *Nicotiana tabacum*; and AmSp, American Spirit

Table 2 Prediction of plant species in each blind clay pipe (BCP)

Blind clay pipe (BCP)	Standard method (number of mass spectral features shared with the experimental clay pipe)	Key	New method (similarity score)	Second method (similarity score)	Third method (similarity score)
BCP1	<i>Auv</i> (3)	<i>Nta</i>	<i>Nta</i> (0.0370)	<i>Nta</i> (0.0301)	<i>Nta</i> (0.0539)
BCP2	<i>Nat</i> (6) and <i>Nqu</i> (6)	<i>Nqu</i>	<i>Nqu</i> (0.1102)	<i>Nqu</i> (0.1066)	<i>Nqu</i> (0.1400)
BCP3	<i>Nob</i> (10)	<i>Nob</i>	<i>Nob</i> (0.1132)	<i>Nob</i> (0.1107)	<i>Ngl</i> (0.1499)
BCP4	<i>Nat</i> (4)	<i>Alu</i>	<i>Nta</i> (0.0928)	<i>Nta</i> (0.0875)	<i>Nta</i> (0.1336)
BCP5	<i>Ngl</i> (6)	<i>Ngl</i>	<i>Ngl</i> (0.0874)	<i>Ngl</i> (0.0799)	<i>Ngl</i> (0.1103)
BCP6	<i>Lin</i> (10)	<i>Lin</i>	<i>Ngl</i> (0.0844)	<i>Ngl</i> (0.0749)	<i>Ngl</i> (0.1452)
BCP7	<i>Auv</i> (16)	<i>Auv</i>	<i>Auv</i> (0.0792)	<i>Auv</i> (0.0800)	<i>Cse</i> (0.1128)
BCP8	<i>Auv</i> (10) and <i>Cse</i> (5)	<i>Auv</i> and <i>Nta</i> mixture	<i>Cse</i> (0.0853) and <i>Auv</i> (0.0662)	<i>Cse</i> (0.0778) and <i>Auv</i> (0.0643)	<i>Cse</i> (0.1336) and <i>Auv</i> (0.0951)

Clay pipes experimentally smoked with a single, known plant species or AmSp were compared individually to each blind clay pipe (BCP) to determine which plant species had been smoked in a BCP. The expected results are the plant species listed under the column labeled “Key”. The second method is where tf was replaced with $1 + \log(tf)$, and the third method is where $idf = 1$. *Alu*, *Artemisia ludoviciana*; *Auv*, *Arctostaphylos uva-ursi*; *Cse*, *Cornus sericea*; *Lin*, *Lobelia inflata*; *Nat*, *Nicotiana attenuata*; *Ngl*, *Nicotiana glauca*; *Nob*, *Nicotiana obtusifolia*; *Nqu*, *Nicotiana quadrivalvis*; *Nta*, *Nicotiana tabacum*; and AmSp, American Spirit

Contamination is a significant concern for ancient residue metabolomics [7–10]. For instance, residues from commercial tobacco smoke may contaminate the surface of artifacts at excavation sites or on display at a museum. Thus, we included AmSp in our study as a contaminant control. With the contaminant control, we were still able to accurately determine the composition of BCP1 (Table 1) and the other blind clay pipes. Utilizing contaminant controls improves confidence in determining if a particular artifact had been smoked with an endemic tobacco. Furthermore, our new method will enable researchers to confidently determine if the caffeine present in/on an artifact resulted from ancient cacao or holly brewing practices instead of modern contaminants from caffeinated beverages such as coffee [8, 9].

It was also revealed that none of the methods could accurately predict BCP8, which had a mixture of *Auv* and *Nta* (Table 2). The new and standard methods partially predicted the composition of BCP8. Though the standard method performed slightly better because it ranked *Auv* higher than *Cse* (Table 2). Nonetheless, the experimental clay pipes compared to BCP8 had all been smoked with only *one* plant species. It is possible that training the new method with experimental clay pipes smoked with complex mixtures may improve the likelihood of predicting if an artifact had been smoked with more than one plant species. A similarity score equal to one, or experimental clay pipes sharing all the mass spectra features found in an ancient artifact using the standard Venn diagram method is difficult to achieve [1, 10]. This is due to several factors including environmental contaminants, diagenesis, and differences between modern and ancient plant varieties. Like the blind clay pipes, compound degradation and smoking characteristics (i.e., duration of smoking, packing density) may also play a contributing factor in differences between experimental clay pipes and ancient artifacts. Ideally, researchers would need to “brew” or “smoke” pre-contact plant materials used by Indigenous Peoples. To achieve this, researchers would need to not only analyze the metabolite composition of

ancient herbarium specimens, but also review ethnobotanical literature. Ethnobotanical accounts often include information on which plant tissues were used, as well as Indigenous Peoples' harvesting, processing, and curing techniques. This knowledge has the potential to improve comparison values and reduce variability in datasets.

Conclusions

Deep and machine learning have been vital tools for solving problems in biology where traditional methods seem inadequate or are time-consuming. Utilizing NLP-based methods, such as our new method, will aid researchers in their quest to determine which plants had been used in ancient smoking pipes, brewing vessels, and other artifacts. Furthermore, we believe our method is robust enough to be implemented in other challenging problems in the field of metabolomics, particularly when distinguishing relationships between biological samples. Incorporating "biological" replicates is often not feasible in ancient residue metabolomics and other metabolomics studies; however, this method is robust in its ability to identify similarities between samples that have a limited number of replicates. Instead of solely relying on Venn diagrams (such as the standard method described in Brownstein et al. [1]), this new method can be used in concert with the standard method to improve confidence in characterizing unknown samples. Thus, this work opens new opportunities for interpreting atypical metabolomics datasets, as well as predicting the chemical composition and identity of samples with unknown histories.

Materials and methods

Mathematical model

A novel functionality of our approach is to introduce a new and automated method to compare the mass spectral feature similarities between experimental and ancient artifact samples. As with this study and other ancient residue metabolomics studies, datasets containing replicates is often not feasible [1]. This limits our ability to apply multivariate statistical methods. Thus, we developed an algorithm inspired by advances in NLP [11–14]. Here, we use the following analogy:

$$\begin{array}{l} \text{Words} \longleftrightarrow \text{Mass Spectral Feature Abundances} \\ \text{Documents} \longleftrightarrow \text{Samples.} \end{array}$$

That is, if words between documents can tell us which documents are similar, mass spectral features between samples can tell us which samples are more likely to be similar. The standard technique in NLP is to first transform the original data into the term frequency-inverse document frequency (TF-IDF) matrix [13]. This transformation helps to resolve the fact that some words (or mass spectral features) appear more often than others. More precisely, the importance of a term (or mass spectral feature) is not solely determined by its frequency (or abundance) in a text (or sample) but also how rare this term (or quantifiable intensity of a particular mass spectral feature) is in other texts in the corpus (or collection of all samples). We note that Brownstein et al. [1] previously used a method more qualitative in nature. As with comparing words between documents to identify commonalities, we can identify which samples are likely to be similar based on their shared mass spectral features. In other words, analyzing the common mass spectral features can allow for inferring

which experimental artifact (or experimental clay pipe) matches with which ancient artifact (or blind clay pipe). Let us recall these terminologies mathematically. Term frequency refers to the frequency (or abundance) of a word (or mass spectral feature) in a particular document (or sample).

$$tf(w, d) = \frac{\text{count of } w \text{ in } d}{\text{number of words in } d}$$

The inverse of the document frequency which measures the informativeness/prevalence/abundance of term, t (or mass spectral features).

$$idf(w) = \log\left(\frac{N}{df(w)}\right) + 1$$

N is the number of documents (or samples) and $df(w)$ is the number of documents (or samples) containing word, w (or mass spectral feature). We remark that the above formula for IDF is based on what Sklearn library (scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfTransformer) uses for its implementation; therefore, it is slightly different from the standard textbook definition. Specifically, some authors use the following convention for $idf(w)$.

$$idf(w) = \log\left(\frac{N}{df(w) + 1}\right)$$

We believe that due to the popularity and simplicity of Sklearn, its use, shown herein, can be applied to similar problems in the field. We also would like to remark that we applied other weighting methods available in Sklearn, but the method above performed the best. We refer to our GitHub repository for the performance comparison of these weighting schemes. Finally, once $tf(w, d)$ and $idf(w)$ are computed, the TF-IDF score is calculated by the following formula:

$$tf - idf(w, d) = tf(w, d) * idf(w).$$

In our context, the TF-IDF score describes the relevance of a mass spectral feature in a sample, as well as the relevance of that feature in different samples. Once the TF-IDF is computed, we can then use cosine similarity to compare two different documents (or samples). Recall that for two vectors v, w their cosine similarity is defined to be cosine of the angle θ between them.

$$\text{similarity} = \cos(\theta) = \frac{\langle v, w \rangle}{\|v\| \|w\|}$$

Here, $\langle v, w \rangle$ is the inner product of v, w and $\|v\|, \|w\|$ is the Euclidean norm of v, w . We note that the similarity score (or frequency/abundance of a mass spectral feature shared between samples) ranges from -1 meaning exactly opposite to 1 meaning the same, with 0 indicating orthogonality, while in-between values indicate intermediate similarity or dissimilarity.

Preparation of samples

Seeds of *Artemisia ludoviciana* (Strictly Medicinal, Williams, OR, USA), *Lobelia inflata* (Strictly Medicinal, Williams, OR, USA), *Nicotiana attenuata* (USDA Agricultural Research Services [ARS] National Plant Germplasm System; Accession Number: PI 555476), *Nicotiana glauca* (USDA ARS National Plant Germplasm System; Accession Number: PI 555686), *Nicotiana obtusifolia* (USDA ARS National Plant Germplasm System; Accession Number: PI 555573), *Nicotiana quadrivalvis* (USDA ARS National Plant Germplasm System; Accession Number: PI 555485), *Nicotiana rustica* (USDA ARS National Plant Germplasm System; Accession Number: PI 555554), *Nicotiana tabacum* (Strictly Medicinal, Williams, OR, USA), *Salvia sonomensis* (USDA ARS National Plant Germplasm System; Accession Number: PI 45388), and *Verbascum thapsus* (Companion Plants, Athens, OH, USA) were sown on Sunshine Mix LC1 soil (sphagnum peat moss and perlite; Sun Gro Horticulture Inc., Agawam, MA, USA). For 60 days, the plants were grown with the following greenhouse conditions—average temperatures of 24/17 °C (day/night), and a photoperiod of 16/8 h (day/night) under 1000 W metal-halide lights to supplement natural daylight. Lights were set to come on when the outside light intensity fell below 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$. During the day, the light intensity averaged 350–400 $\mu\text{mol m}^{-2} \text{s}^{-1}$ in the greenhouse. The plants were fertilized twice a week with Peters 20–20–20 (N–P–K; JR Peters Inc., Allentown, PA, USA) containing iron chelate, magnesium sulfate, and trace elements.

Arctostaphylos uva-ursi (collected: April 2015; voucher ID: 393408), *Cornus sericea* (collected: September 2016; voucher ID: 393409), and *Rhus glabra* (collected: September 2016; voucher ID: 393395) were collected on the Washington State University, Pullman campus. *Taxus brevifolia* (collected: October 2016; voucher ID: 393425) was collected in the Iller Creek Conservation Area, WA, USA.

After Korey Brownstein confirmed the identity of the fourteen (14) different plants, *A. ludoviciana* Nutt. (*Alu*) leaves, *A. uva-ursi* (L.) Spreng. (*Auv*) leaves, *C. sericea* L. (*Cse*) bark, *L. inflata* L. (*Lin*) leaves, *N. attenuata* Torr. ex S. Watson (*Nat*) leaves, *N. glauca* Graham (*Ngl*) leaves, *N. obtusifolia* M. Martens & Galeotti (*Nob*) leaves, *N. quadrivalvis* Pursh (*Nqu*) leaves, *N. rustica* L. (*Nru*) leaves, *N. tabacum* L. (*Nta*) leaves, *R. glabra* L. (*Rgl*) autumn leaves, *S. sonomensis* Greene (*Sso*) leaves, *T. brevifolia* Nutt. (*Tbr*) needles, and *V. thapsus* L. (*Vth*) leaves were collected, freeze-dried for 3 days, and crushed for experimental smoking. Voucher specimens from the same plants were also collected by Korey Brownstein and filed in the Marion Ownbey Herbarium, Washington State University, Pullman, WA, USA (herbaria.wsu.edu/web/default.aspx). These specimens can be found by performing a “Collector’s Name” search, i.e., Korey Brownstein, in the following database: intermountainbiota.org/portal/collections/harvestparams.php.

American Spirit (AmSp) tobacco (Santa Fe Natural Tobacco Company, Oxford, NC, USA) was purchased from a local grocery store in Pullman, Washington, USA. The plant materials ($n=5$ for each species) and AmSp ($n=5$) were smoked in clay pipes following the experimental conditions detailed in Brownstein et al. [1]. The experimentally smoked clay pipes were then completely submerged in acetonitrile:2-propanol:water [3:2:2] and sonicated for 10 min. Five non-smoked blank clay pipes were extracted as controls using the same extraction methods as the experimentally smoked clay pipes. To prepare the experimental clay pipes for liquid chromatography-mass spectrometry

(LC–MS) analysis, 3.0 mL from each of the five replicates were combined into a single tube. Only experimental clay pipes subjected to the same conditions or smoked with the same plant species were combined (i.e., non-smoked blank clay pipes were combined, experimental clay pipes smoked with AmSp were combined, experimental clay pipes smoked with *Alu* were combined, and so forth for the other plant species). The 15.0 mL pooled experimental clay pipe samples were freeze-dried for 3 days and resuspended with 5.0 mL of 0.10% formic acid/water:acetonitrile [1:1]. Afterwards, the resuspended samples were filtered into glass vials using a 0.20 μm filter.

The blind clay pipes ($n=8$) were smoked and broken in fragments with a mallet to emulate artifacts found in the field. These fragments were completely submerged in acetonitrile:2-propanol:water [3:2:2] and sonicated for 10 min. Afterwards, the extracts were freeze-dried for 3 days and resuspended with 1.0 mL of 0.10% formic acid/water:acetonitrile [1:1]. The resuspended samples were then filtered through a 0.20 μm filter into a glass vial. To limit biases, the authors did not know which plant species had been smoked in which blind clay pipe. After the experimental clay pipes, non-smoked blank clay pipes, and eight (8) blind clay pipes were analyzed by LC–MS and processed in MZmine 2 following the parameters described in Brownstein et al. [1], the data were exported into .csv files. Mass spectral features with peak heights less than 2.0×10^3 had their abundance values set to zero. The .csv files were arranged in the following format: mass spectral features were in rows; each mass spectral features' unique identifier (ID) number, m/z value, and retention time value (in min) were in the first, second, and third columns, respectively; and mass spectral feature abundance values were listed under each sample in the remaining seventeen (17) columns. To eliminate solvent contaminant noise, mass spectral features present in the blank clay pipes were removed from the experimental clay pipes and blind clay pipes before processing the datasets in our algorithm. Python libraries, such as Sklearn and Pandas, were then used to apply the TF-IDF computation scores to these datasets. The extracted experimental clay pipes and blind clay pipes were allowed to air-dry on the lab bench. All solvents used for extraction and analysis were of mass spectrometry grade.

Acknowledgements

We would like to thank Shannon Tushingam and Mario Zimmermann for experimentally smoking the blind clay pipes (BCPs) and their editorial assistance, as well as the USDA ARS National Plant Germplasm System for providing the *Nicotiana* and *Salvia sonomensis* seeds.

Author contributions

T.T.N. and K.J.B. designed the project. T.T.N. developed and wrote the scripts. T.T.N. and K.J.B. wrote the paper. All authors approved of the final version of the paper.

Funding

Korey Brownstein, Ph.D., holds a Postdoctoral Enrichment Award from the Burroughs Wellcome Fund. This work was also supported by National Science Foundation grant number 1906607 and National Science Foundation grant number 1419506.

Availability of data and materials

All scripts and datasets used in this study are freely available on GitHub: https://github.com/tungprime/NLP_and_composition_of_artifact_residues.

Declarations

Ethics approval and consent to participate

The collection of cultivated and wild plant materials in the article were carried out in accordance with guidelines set forth by Washington State University, as well as state, national, and international regulations. This article does not include research on plants without ethical approval. Voucher specimens of the study plants were collected by Korey Brownstein and filed in the Marion Ownbey Herbarium, Washington State University, Pullman, WA, USA (herbaria.wsu.edu/web/

default.aspx). These specimens can be found by performing a “Collector’s Name” search, i.e., Korey Brownstein, in the following database: intermountainbiota.org/portal/collections/harvestparams.php.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 15 March 2022 Accepted: 30 July 2024

Published online: 27 September 2024

References

1. Brownstein KJ, Tushingam S, Damitio WJ, Nguyen T, Gang DR. An ancient residue metabolomics-based method to distinguish use of closely related plant species in ancient pipes. *Front Mol Biosci.* 2020;7:133.
2. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* 2010;11:395.
3. Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques PÉ, Li S, Xia J. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* 2021;49:W388–96.
4. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem.* 2012;84:5035–9.
5. Ovchinnikova K, Stuart L, Rakhlin A, Nikolenko S, Alexandrov T. ColocML: machine learning quantifies co-localization between mass spectrometry images. *Bioinformatics.* 2020;36:3215–24.
6. Tang Y, Li R, Lin G, Li L. PEP search in MyCompoundID: detection and identification of dipeptides and tripeptides using dimethyl labeling and hydrophilic interaction liquid chromatography tandem mass spectrometry. *Anal Chem.* 2014;86:3568–74.
7. Damitio WJ, Tushingam S, Brownstein KJ, Matson RG, Gang DR. The evolution of smoking and intoxicant plant use in ancient Northwestern North America. *Am Antiq.* 2021;86:1–19.
8. King A, Powis TG, Cheong KF, Gaikwad NW. Cautionary tales on the identification of caffeinated beverages in North America. *J Archaeol Sci.* 2017;85:30–40.
9. Washburn DK, Washburn WN, Shipkova PA, Pellemounter MA. Chemical analysis of cacao residues in archaeological ceramics from North America: considerations of contamination, sample size and systematic controls. *J Archaeol Sci.* 2014;50:191–207.
10. Zimmermann M, Brownstein KJ, Díaz LP, Aragón IA, Hutson S, Kidder B, Tushingam S, Gang DR. Metabolomics-based analysis of miniature flask contents identifies tobacco mixture use among the ancient Maya. *Sci Rep.* 2021;11:1–11.
11. Cohen KB, Hunter L. Natural language processing and systems biology. In: Dubitzky W, Pereira F, editors. *Artificial intelligence and systems biology.* Dordrecht: Springer; 2004. p. 147–75.
12. Cong Y, Chan YB, Phillips CA, Langston MA, Ragan MA. Robust inference of genetic exchange communities from microbial genomes using TF-IDF. *Front Microbiol.* 2017;8:21.
13. Goodfellow I, Bengio Y, Courville A. *Deep learning.* Cambridge: MIT Press; 2016.
14. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. *IEEE Comput Intell Mag.* 2018;13:55–75.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.