analytical chemistry

pubs.acs.org/ac Article

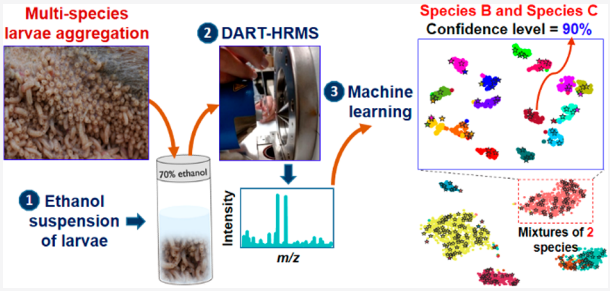
Identification of the Species Constituents of Maggot Populations Feeding on Decomposing Remains—Facilitation of the Determination of Post Mortem Interval and Time Since Tissue Infestation through Application of Machine Learning and Direct Analysis in Real Time-Mass Spectrometry

Samira Beyramysoltan, Mónica I. Ventura, Jennifer Y. Rosati, Justine E. Giffen-Lemieux, and Rabi A. Musah*

Cite This: *Anal. Chem.* 2020, 92, 5439−5446 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The utilization of entomological specimens such as larvae (maggots) for the estimation of time since oviposition (i.e., egg laying) for post mortem interval determination, or for estimation of time since tissue infestation (in investigations of elder or child care neglect and animal abuse cases), requires accurate determination of insect species identity. Because the larvae of multiple species are visually highly similar and difficult to distinguish, it is customary for species determination of maggots to be made by rearing them to maturity so that the gross morphological features of the adult can be used to accurately identify the species. This is a time-consuming and resource-intensive process which also requires that the sample be viable. The situation is further complicated when the maggot mass being sampled is comprised of multiple species. Therefore, a method for accurate species identification, particularly for mixtures, is needed. It is demonstrated here that direct analysis in real time−high resolution mass spectrometric (DART-HRMS) analysis of ethanol suspensions containing combinations of maggots representing *Calliphora vicina*, *Chrysomya rufifacies*, *Lucilia coeruleiviridis*, *L. sericata*, *Phormia regina*, and *Phoridae* exhibit highly reproducible chemical signatures. An aggregated hierarchical conformal predictor applied to a hierarchical classification tree that was trained against the DART-HRMS data enabled, for the first time, multispecies identification of maggots in mixtures of two, three, four, five, and six species. The conformal predictor provided label specific regions with confidence limits between 80 and 99% for species identification. The study demonstrates a novel, rapid, facile, and powerful approach for identification of maggot species in field-derived samples.



Larva represent one of the immature life stages of flies. Those that are associated with necrophagous insects (e.g., blow flies that colonize decomposing tissue), such as the members of the Calliphoridae family, utilize remains as a primary food source. Gravid females, which are able to detect carrion and corpses within a few minutes of death and from a distance of up to 2 miles away,[1,2] oviposit on the remains. When the eggs hatch, the emergent larvae consume the tissue as they progress through the first, second, and third instar stages, before subsequently pupating. While their importance in facilitating the decomposition of remains is well-appreciated, their presence can also reveal several types of information.[3] For example, maggots retrieved from tissue can be used for the determination of post-mortem interval in death investigations. In elder and child care abuse and animal neglect cases, maggots retrieved from infested wounds can provide estimates of the time frame over which maltreatment occurred. In illegal hunting/poaching investigations, larvae collected from the dead animal can be used to develop an accurate timeline of when the criminal activity occurred.[4,5] The reason that such information can be revealed is due to a combination of the following factors: (1) blow flies are able to detect the presence of a corpse or carrion within minutes of death and will lay eggs (oviposit) shortly after their arrival at the body, and (2) the insect development life cycle and timeline, which is highly species dependent and correlated to temperature, humidity, and other external factors, is well mapped and understood.[4,6−8] Thus, for a given species collected at a particular life stage, the

Received: January 15, 2020
Accepted: February 24, 2020
Published: February 24, 2020

time since the eggs were laid (and which would provide an estimate of when death occurred based on the premise that oviposition on the remains occurred within minutes to a few hours after death) can be calculated from knowledge of the species identity and the external factor conditions.[9−11] Traditionally, species identity determination is performed by an entomologist[5,12] and is accomplished most often by rearing immature life stages to maturity so that identification can be reliably made based on visual inspection of the phenotypic attributes of the adult.[13] This is often required because of the challenges imposed by the extent to which multiple species within the same life stage look alike, making species identification based on visual examination difficult if not impossible. The ability to use the results of DNA analysis would be ideal, but the paucity of mapped genomes for relevant fly species makes this approach impossible in many cases.[14−18] To address this issue, several studies have proposed alternative more rapid and less resource intensive approaches to achieve species identification.[19−24] Microscopy with and without staining has been used to magnify the features of eggs and larvae so that species attribution can be based on morphological distinctions that are normally microscopic.[22−24] A number of preliminary studies have indicated a correlation between the small-molecule profiles of a given life stage and species identity.[20,21] However, the proof-of-concept successes that have been achieved so far using such approaches for species identification were accomplished using single species samples. In practice, multispecies samples are often encountered. When entomological samples are procured, they are usually submerged in aqueous ethanol for storage, often in the form of multiples of eggs, larvae, or pupae, which can make it challenging to accurately determine the identities of the species that are present. Due to their small size and the tackiness of the egg samples, multiple eggs may remain adhered to one another, while some distinguishing larval morphological features (i.e., spiracular structures, protuberances, etc.) may become difficult to view. These features may be reduced or no longer visible in the pupal stage. When feeding, larvae engage in aggregation behavior that involves hundreds or thousands of individuals,[25−27] as illustrated in Figure S1, which shows typical feeding activity of tens of thousands of blow fly larvae. This aggregation affords a number of benefits, including better nutrient absorption and enhanced protection against predators and parasites, as well as more rapid maturation to adulthood as a consequence of the increases in local temperature that occur because of aggregation (up to 20 °C above ambient).[28]

Due to the occurrence of multispecies oviposition at the same site, maggot masses can be comprised of individuals of a single or multiple species.[26] Since knowledge of the species of insect is crucial to being able to glean useful information about infestation, an approach to the determination of the identities of the species present in mixtures is of high interest. In a machine learning context, problems of this type, that involve a combination of several "labels" (in this case, species) for instance, are known as multilabel problems, and two main approaches have been developed for the classification of such systems: problem transformation and algorithm adaptation methods.[29] The former, which, along with its derivatives, are binary relevance and label powerset approaches, transform the multilabel classification problem to one or more single-label classifications. In the study described herein, a novel hierarchical conformal predictor that uses a problem transformation concept-label powerset method at its core, was developed to achieve multilabel classification of entomological samples using DART-HRMS data. The strategy facilitated determination of the constituents of multispecies samples of larvae while resolving the complexity (i.e., a multiclass problem with many (∼44 classes)) and the class imbalance that occurs in instances where there is a preponderance of one or more classes over the others in the model. DART-HRMS data were acquired from analysis of 70% aqueous ethanol suspensions of individual and mixture combinations of two, three, four, five, and six species that were acquired in the Manhattan area of New York.

METHODS

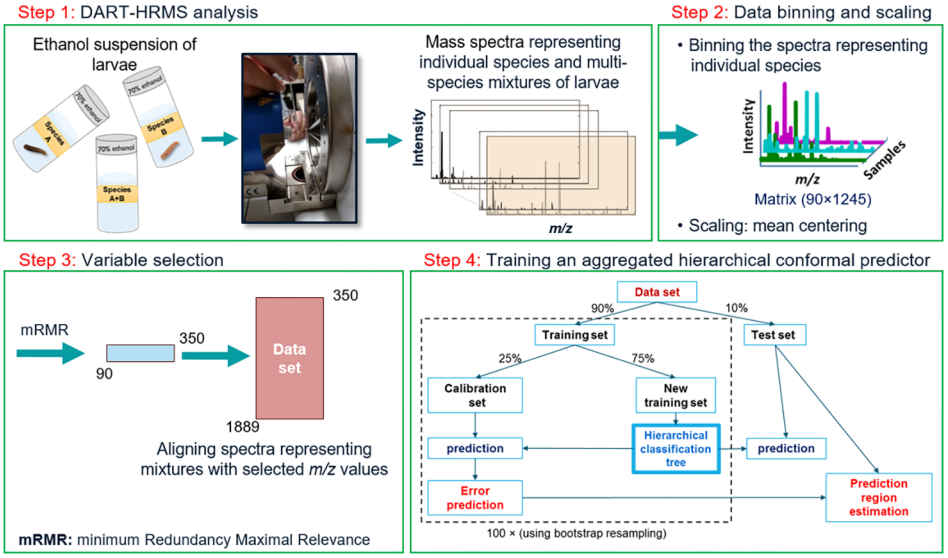
Collection and Preservation of Necrophagous Fly Samples. Blow fly stock colonies of *Calliphora vicina*, *Chrysomya rufifacies*, *Lucilia coeruleiviridis*, *L. sericata*, *Phormia regina*, and *Phoridae* species were maintained in the laboratory of Professor Jennifer Rosati at the John Jay College of Criminal Justice (New York, NY, USA). The details associated with their establishment and maintenance are as previously described.[20,21]

Sample Preparation for DART-HRMS Analysis. It has been shown previously that the aqueous ethanol suspensions in which blow fly eggs, larvae, pupae, and adults are stored exhibit species-specific chemical signatures that can be used to rapidly identify species.[20,21] To investigate whether this approach could be used to determine the identities of multiple species (both blow fly and Phorid, with the latter used as a control) contained within a mixture of larvae that were stored in 70% aqueous ethanol, a range of mixture suspensions representing combinations of two, three, four, five, and six species were prepared for chemical fingerprint analysis. Mixtures representing a range of proportions of contributing species were used in order to address the possibility of samples containing unequal distributions of species—in other words, to build into the created model the ability to accurately predict the species composition in a manner that was independent of whether the sample was dominated by an overabundance or under representation of a given species constituent. Therefore, samples were made by mixing together varying combinations of the 70% ethanol suspensions of single species (v/v) and diluting these as necessary with 70% aqueous ethanol to achieve a final volume of 25 μL. For example, for the two-species mixture series, where the two species are denoted "A" and "B," seven solutions were made in the following volumetric proportions of A/B/70% aqueous ethanol, respectively: 20:80:0, 45:45:10, 80:20:0, 45:10:45, 10:45:45, 5:10:85, and 10:5:85. The details of the combinations used for all of the mixture types are listed in Table S1.

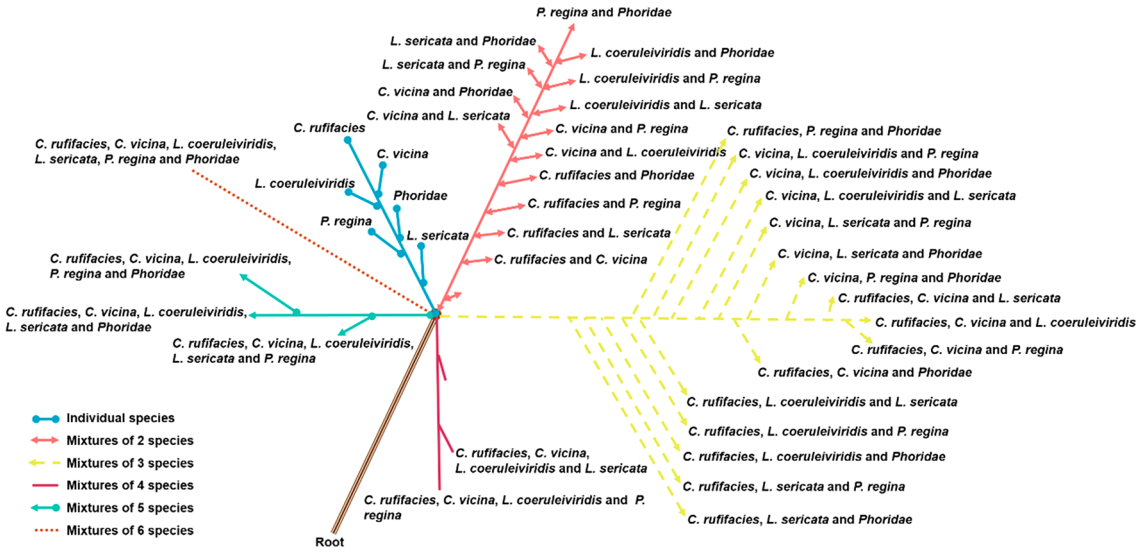
Instrument Parameters and Sample Analysis Conditions. A DART-SVP ion source (IonSense, Saugus, MA, USA) interfaced with a JEOL AccuTOF mass spectrometer (JEOL USA, Peabody, MA, USA) was used to collect mass spectral data in positive-ion mode. The optimized instrument parameter settings were as follows: helium gas flow rate, 2.0 L/min; gas temperature, 350 °C; DART ion source grid voltage, 50 V; ring lens voltage, 5 V; orifice 1 voltage, 20 V; orifice 2 voltage, 5 V; and peak voltage, 600 V (to detect m/z values ≥60). The samples were analyzed by dipping the closed end of a melting point capillary tube into the solution and presenting the coated surface to the open-air space between the mass spectrometer inlet and ion source. Spectra were collected at a rate of one spectrum per second over the mass range m/z 60−

https://dx.doi.org/10.1021/acs.analchem.0c00199
Anal. Chem. 2020, 92, 5439−5446

Scheme 1. An Overview of the Data Analysis Workflow for Multispecies Blow Fly Identification



Scheme 2. Structure of a Top-down Hierarchical Classification Tree for Multispecies Discrimination



800. For suspensions comprised of individual species samples, 15 of each were created. Each of these was analyzed by DART-HRMS five times, and the five spectra were averaged to generate a single representative spectrum. Each of the samples representative of a given mixture combination and the varying ratios of its contributing constituents (in a volume of 25 μL—see Table S1) was created in replicates of five. Each of these replicates was analyzed five times, and the resulting spectra were averaged to generate a representative spectrum. PEG 600 (PEG; Sigma-Aldrich, Louis, MO, USA) was used as a mass calibrant.

DART-HRMS Data Processing and Multivariate Statistical Analysis. Mass spectral peak centroiding and background subtraction were accomplished using TSSPro3 software (Schrader Analytical Laboratories, Detroit, MI, USA). The peaks corresponding to the ethanol dimer (nominal m/z 93) and a plasticizer derived from the capillary tubes used for sampling (m/z 371) were removed from the mass spectra prior to statistical analysis processing of the data.

The mass spectra were stored in text format and imported into MATLAB 9.3.0, R2019a software (The MathWorks, Inc., Natick, MA, USA) for further analysis. An overview of the data analysis workflow is presented in Scheme 1. It was comprised of DART-MS analysis (i.e., Scheme 1, step 1) and data treatment and statistical analysis processing (Scheme 1, steps 2 through 4). In step 2, the mass spectral data representing individual species (in replicates of 15 each) were aligned along common m/z values (i.e., binned). The optimal bin width and relative abundance threshold cutoff, determined by iterative application of principal component analysis-discriminant analysis (PCA-DA) and partial least squares-discriminant analysis (PLS-DA) algorithms to the mass spectral data matrices created by varying the bin widths and relative intensity thresholds from 5 to 10 mmu and 0.1−1%, respectively, were found to be 10 mmu and 0.2% respectively. This treatment resulted in a matrix with dimensions of 90 × 1245. It was mean centered and then subjected to minimum redundancy maximal relevance (mRMR) feature selection[30] (step 3). This step revealed 350 top ranked m/z values that

https://dx.doi.org/10.1021/acs.analchem.0c00199
Anal. Chem. 2020, 92, 5439−5446

were the most significant for species discrimination of larvae and individualization of mixture samples. The mass spectra representative of the larvae mixtures were aligned with the aforementioned selected m/z values and then scaled with individual species data to create a data matrix (named "**X**" (1889 × 350)) that included data representative of both individual species and species mixtures. The composite of samples was divided into 44 label sets defined as follows: each of the six species represented was classified as a single label (yielding six single species labels), 15 combinations of two species, 16 combinations of three species, three combinations of four species, three combinations of five species, and one containing all six species. Step 4 encompassed a sequence of operations that comprised the strategy designed to enable the prediction of the characteristics of the 44 class labels. Given that this circumstance can be categorized as a "multi-label" problem (with 44 labels), a multilabel classification approach known as the "Powerset method" was applied to convert each label combination to a single label, in one multiclass problem. While the samples themselves do not share taxonomic relationships, they can still be considered as related in hierarchy based on mixture type. This enabled the use of a top-down level-based hierarchical classification tree[31] in which several multiclasses could be hierarchically trained, to simplify the one 44-class problem to seven multiclass problems. Scheme 2 presents the structure of the resulting hierarchical classification tree exhibiting two levels of discrimination. In the first, the samples are differentiated according to mixture type (i.e., individual species and mixtures of two, three, four, five, and six species). In the second, five multispecies classes characterize the label sets of the samples corresponding to each mixture type, and there was a single class representing the group that contained the six individual species. In Scheme 2, the mixture types are color coded, and the samples within mixture types are illustrated as leaves within the nodes of the indicated tree branches.

Classification of the samples in the nodes of the classification tree was accomplished using a feed forward neural network with nonlinear *sigmoid* and *softmax* functions for the hidden and output layers, respectively. A regularization parameter and early stopping technique were applied to avoid the possibility of overfitting. The Bayesian optimization technique was applied to optimize the hyper parameters of the neural network (i.e., the number of hidden layer nodes, learning rate, and regularization parameters) to enable high accuracy prediction. The neural network was retrained 10 times, and the average of the trained weights was applied to making the predictions. To define the level of confidence of a predicted label set using the hierarchical classification tree, the tree was given an underlying classifier within a conformal prediction protocol,[32] termed the "aggregated inductive Mondrian conformal predictor." This approach is outlined in step 4 of Scheme 1. The data set matrix (1889 × 350) was partitioned into training (90% of the matrix) and test sets (10% of the matrix), each of which was composed of randomly chosen observations. The training set was subjected to 100× bootstrap resampling without a replacement strategy, to canvass 100 training and calibration sets for the development of a conformal predictor. In each bootstrap iteration, 25% of the samples were randomly selected for calibration in order to evaluate how similar the test samples were to the training samples.

The remaining samples (i.e., 75%) were used to train the hierarchical classification tree. Both calibration and test sets were then predicted by the trained tree. The error prediction rate was quantified using a nonconformity measure (termed "α"). Nonconformity measures of calibration sets and test samples were used to compute p-values for test samples aimed at estimating the prediction region for test samples. Three nonconformity measures[33] (shown in eqs S1−S3 in Scheme S1-Appendix 1, Supporting Information) were used in this study. To assess the class imbalance (i.e., the proportions of labels within a given multiclass that are significantly different), which is one of the major problems associated with the label powerset method, the Mondrian conformal prediction (label conditional) concept was used in computing the p-value. Therefore, in computing the p-value for a class-k assignment, the nonconformity measures of test samples for class k were compared with the nonconformity measures of the calibration samples with class k. The result from completion of a bootstrap run was 100 sets of calculated p-values for the test samples that were averaged to generate the p-values of the aggregated conformal predictor. Scheme S1 shows, in detail, the algorithm of the conformal predictor that was applied to the hierarchical classification tree. The performance of the conformal predictor[34,35] was assessed based on the test set prediction using the following merit criteria: multiple prediction rate (E-criterion), observed fuzziness (OF), error rate, and the "not assigned" (NA) rate, in the significance level (ε) range of 0−0.2. The merits are described in Scheme S1-Appendix 2 and presented in eqs S4−S6.

In order to trust the results of the developed classifier, it is essential to establish whether or not its high accuracy predictions occur purely by chance (as an artifact of a small data size, for example). One way to assess this is to test the fitting of the model to incorrect labels (also known as y-permuted labels).[36] The observation of a poor fit when this exercise is applied indicates that the developed classifier is not fitting to the random variations in the data and that the predictions are reliable and not occurring by chance. To apply the y-permutation procedure to the aggregated conformal predictor developed in this work, the labels of the training samples in each iteration of the 100× bootstrap resampling were randomly shuffled 10 times, while the training mass spectra were left unchanged. The 10 permuted labels and MS data were then used to train the 10 models that were applied for prediction of the calibration and test samples. The performance characteristics for the y-permuted model in each iteration of the bootstrapping were the average of the characteristics of the 10 created models. Following the bootstrapping, the performance of the y-permuted model was compared with the performance of the normal model (i.e., using the data with real labels).

An extended form of t-distributed stochastic neighbor embedding, termed global stochastic neighbor embedding (g-SNE), was used to provide a readily interpretable 2-D scatter plot by which to visualize the performance of the neural network. Therefore, the features in the hidden layer of the trained neural network (referred to as the hidden layer feature space) were projected to the 2-D space of a g-SNE plot.[37] The hidden data structure of the samples was computed by mapping the samples' spectra onto the trained network hidden layer space based on the weights and biases. Utilization of the neighbor-embedding technique preserves the pairwise sim-

https://dx.doi.org/10.1021/acs.analchem.0c00199
Anal. Chem. 2020, 92, 5439−5446

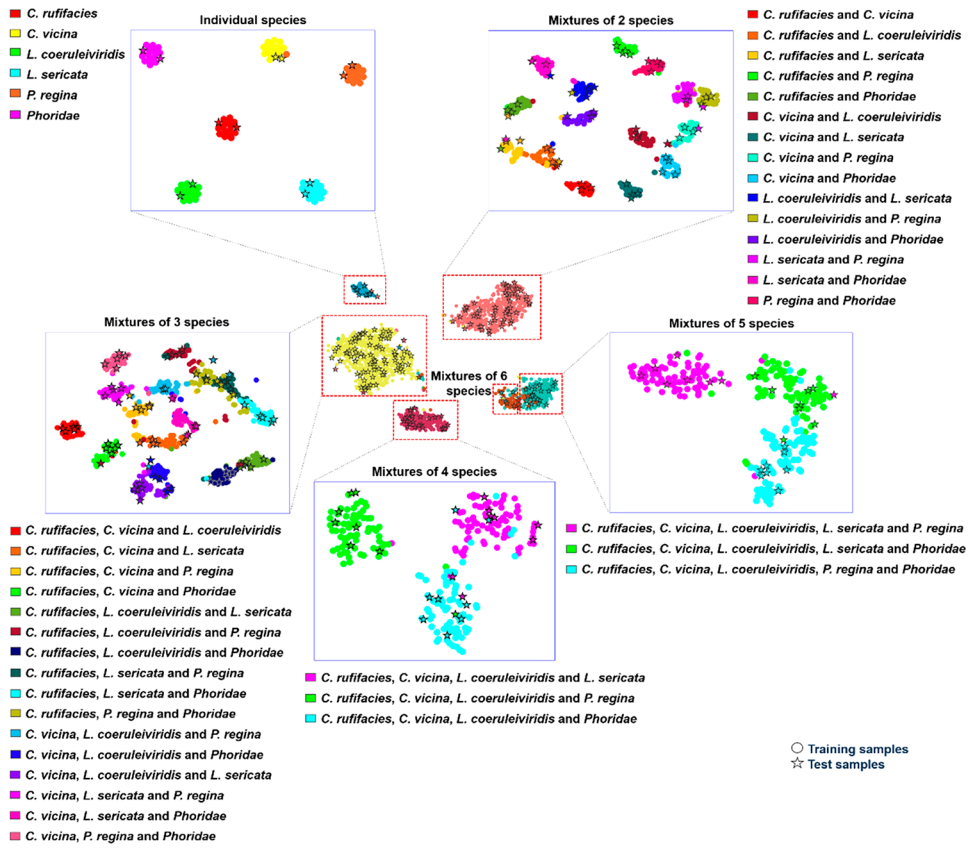


Figure 1. Neural network hidden feature space visualization using *g*-SNE. The circles and stars show the relative location of the training and test samples, respectively, in space. The data are color coded by class.

ilarities of the data points of neighbors by minimizing the divergence of similarity distributions.

RESULTS AND DISCUSSION

The multispecies aggregation of blow fly larvae that occurs in the context of insect colonization of tissue often results in the collection of entomological samples comprised of multiple species. We sought in this work to determine if species combinations furnished unique chemical signatures that could enable the identities of the contributing species to be deduced (for subsequent use in PMI estimation determinations for example). That this might be possible was inferred from the observation that individual species of larvae can be identified and distinguished based on the diagnostic chemical fingerprints that are exhibited by the aqueous ethanol solutions in which they are stored.[20]

Figure S2 displays the DART-HRMS spectra of aqueous ethanol suspensions of individual species. Following DART mass spectrum acquisitions, the data were compiled (according to the procedure described in the DART-HRMS Data Processing and Multivariate Statistical Analysis section) and subjected to the proposed hierarchical conformal predictor. The underlying classification structure is that of a hierarchical classification tree with six classification nodes and two layers of discrimination (as shown in Scheme 2). Table S2 shows the dimensions of the training, calibration, and the test data sets for the neural networks at each iteration of the boostrap resampling in the nodes of the classification tree. In addition, the optimized hyper parameters determined by the Bayesian method are displayed for each node (Table S2). In order to more readily visualize the performance of the classification tree, the hidden layer data structure of the trained neural network in the classification nodes of the tree was explored by projection of the training (90% of the original data) and test (the remaining 10%) data sets in the hidden feature space of the neural network, onto the 2-D space of a *g*-SNE plot. The dimensions of the components of this hidden data set (i.e., for the individual species, mixtures of two, mixtures of three, mixtures of four, mixtures of five, and mixtures of six) were 1683×15, 78×14, 467×44, 642×37, 209×30, and 216×36, respectively. The results, color-coded by species identity and mixture type, are displayed in Figure 1. The colored circles and black stars represent training and test samples respectively. Visual assessment of the results reveals that the model performs well as demonstrated by the apparent separation of most classes at both hierarchical levels. The first level of separation (i.e., mixture type) is illustrated by the samples that appear within red squares, of which there are six. At this level, it is apparent that the individual species cluster (turquoise symbols), and two-, three-, and four-species clusters (indicated by salmon, yellow, and pink colored symbols, respectively) are separated from one another. The clusters representing the five- and six-species mixtures (teal and brown symbols, respectively) overlap with one another to some extent, indicating that their mass spectral characteristics are highly similar and may be difficult to distinguish. Analysis of mixture-type clusters at the second level of the hierarchy reveals the cluster constituent species identities. For example, expansion of the two-species mixture cluster reveals 15 subclusters which are color-coded according to the identities of the pairs of species of which they are comprised. While most pairs are separated, there are two

https://dx.doi.org/10.1021/acs.analchem.0c00199
Anal. Chem. 2020, 92, 5439−5446

sets that abut one another: the *L. sericata*/*P. regina* and *L. coeruleiviridis*/*P. regina* pairs and the *C. rufifacies*/*L. sericata* and *C. rufifacies*/*L. coeruleiviridis* pairs. The merits of the neural network models in the absence of a conformal predictor for the calibration and test samples, determined by 100× bootstrap resampling, are detailed in Tables S3 and S4, respectively. Each table presents the performance metrics results for both the actual data learning and the average of the 10 times *y*-permuted data learning. Each row of the table lists the results associated with one of the classification nodes in the classification tree according to three criteria that were used to compute the accuracy of the neural network's class assignment predictions: the maximum output value of the neural network, output values of >0.5, and output values >0.4. The results presented in Table S3 show that acceptable prediction accuracies were achieved for the calibration samples. For example, the maximum accuracy of the mixture type model for prediction of the calibration samples was 94.4%, which dropped to 92.6% at the 0.5 threshold and rose to 94.3% at the 0.4 threshold. As anticipated, the accuracies for the two to five member species nodes were lower than for individual species because of the increased chemical complexity of these samples when compared to the single-species samples.

The result of the application of the *y*-permuted model was a dramatic reduction in prediction accuracy, which confirmed the suitability of the developed neural network models for discrimination and identification of species mixture compositions. For example, in Table S4, which shows the performance for the test samples' prediction, the accuracy of the model for the highest output value is 93.2% for discrimination of mixture types, and it remained high at the thresholds of 0.5 and 0.4 (91% and 92.7%, respectively). On the other hand, the accuracies drop and are extremely poor for the *y*-permuted data (38%, 0%, and 0% for the highest output value and at the thresholds of 0.5 and 0.4, respectively).

While the classification tree did well in performing two levels of discrimination for the six defined nodes, it would be of greater practical utility to also have a measure of the confidence interval associated with test sample predictions. To accomplish this, the algorithm outlined in Scheme S1 was applied to the classification tree. A key element of this algorithm is the estimation of the nonconformity measure (α), which gives an indication of the extent to which the results for the test samples deviate from those of the calibration/training samples. The equations that can be used to calculate the nonconformity measure are presented as eqs S1−S3. Their independent application to the data revealed that, while the results were similar, eq S3 gave a slightly better outcome. Therefore, eq S3 was used in the algorithm presented in Scheme S1. The general sequence of steps in Scheme S1 is more fully described in Scheme S1-Appendix 3.

The output results of bootstrapping in Scheme S1 can be conveyed in multiple ways, depending upon which attributes of the model are of interest. These include (a) whether the magnitude of the aggregate of p_{global} and p_{type} is less than a given significance level (i.e., p_{global} and $p_{type} < \varepsilon$) or less than the minimum and maximum of several thresholds (e.g., p_{global} and $p_{type} < \min(\varepsilon_1$ and $\varepsilon_2)$);[38] (b) assessment of p_{global} and p_{type} as independent entities for the first and second discrimination levels, respectively (i.e., $p_{global} < \varepsilon_{global}$ and $p_{type} < \varepsilon_{type}$); and (c) defining a label-specific significance threshold ($p^{k\_g}_{global} > \varepsilon^{k\_g}$ and $p^{k\_t}_{type} > \varepsilon^{k\_t}$) as proposed by Vovek et al.[39] for the label conditional conformal predictor. Given our specific interest in predicting, with knowledge of the confidence level, of the constituent species identities of complex mixtures, we focused on option c as shown in Scheme S1. If the $p(\alpha^{k\_g}_{global}) > \varepsilon^{k\_g}$ and $p(\alpha^{k\_g}_{type}) > \varepsilon^{k\_t}$, the test sample was categorized as class *k_g* (i.e., mixture type) and class *k_t* (i.e., within mixture type *k_g*), where ε^{k_g} and ε^{k_t} define the significance levels for assignment of classes *k_g* and *k_t*, respectively. The estimated prediction region τ for each sample included classes with *p*-values higher than the defined levels. Therefore, the prediction region can be empty, or involve single and/or multiple labels. The prediction is considered to be true if the region involves the true label.

The significance level, which is defined by the investigator, represents an acceptable error rate. A conformal predictor is valid if the error rate is less than the significance level (i.e., 1 − confidence level). In addition, the conformal predictor is optimal if the estimated region is as small as possible (i.e., small E-criterion). A visual rendering of the merits within the significance levels 0−0.2 is presented in Figure S3, while Table S5 displays the significance levels with their corresponding error, multiple prediction, and unassigned rates. The top row of plots shows the merits for the mixture-type discrimination levels (i.e., individual species, and mixtures of two, three, four, five, and six species), while the rows of plots beneath reveal the merits at the second level of discrimination (i.e., the level which reveals the specific identities of the constituents of a given mixture). The red vertical line in each plot demarcates the significance level that yields the maximum accuracy for assignment of the class. Thus, each plot in Figure S3 provides information on the *p*-values for the label assignment in the classification tree. For example, in the first plot on the top row (i.e., the turquois box representing the case of individual species), it can be seen that the multiple prediction rate is ∼0.25 at the 100% confidence level, which means that, in principle, at this level, other classes are possible along with the correct class. At the 0.06 significance level (i.e., the 94% confidence level), the multiple prediction rate is zero, which means that the probability of predicting other classes for the individual label is lower than 0.06. However, the unassigned rate starts to increase from the 93% confidence level, which means that the *p*-value of the correct class for a particular sample will be lower than 0.07. The error rate (blue circles) and observed fuzziness (green triangles) overlap at zero at the significance level 0.06, which is the best possible result (i.e., the model is 100% accurate for this class). The significance level 0.06 (indicated by the red vertical line) is considered to be the threshold level for the individual class assignment in the mixture type, since all merits are equal to zero. There was a small subset of classes for which the identification error was high, and in these instances, a theoretical significance level of 0.05 was considered as the threshold. These cases included *C. rufifacies*/*L. sericata* and *L. sericata*/*P. regina* in the mixtures of two species node and *C. vicina*/*P. regina*/*Phoridae* in the mixture of three species node.

Additionally, Table S5 presents the false positive (fp) and false negative (fn) rates for prediction of test sets for each label space, in order to confirm the suitability of the selected significance levels for label assignment. The samples with multiple and empty prediction regions were not involved in calculation of fp and fn rates. False positive and false negative rates reveal *type I* and *type II* errors, respectively; *type I* errors are generally considered to be more serious than *type II* errors. The probability of a *type I* error is revealed by the significance level and is set by the investigator. However, there is a balance

https://dx.doi.org/10.1021/acs.analchem.0c00199
Anal. Chem. 2020, 92, 5439−5446

that must be achieved between *type I* and *type II* errors. While the results can be protected against *type I* errors by choosing a low significance level, this can lead to an increase of the *type II* error.[40] The results show that the error is considerable for a subset of labels. These include: (1) six-species mixture determination at the first level of discrimination; (2) *C. rufifacies*/*L. sericata* and *L. sericata*/*P. regina* in the two-component mixture node (second level discrimination); (3) *C. vicina*/*P. regina*/*Phoridae* in the three-component mixture node (second level discrimination); (4) *C. rufifacies*/*C. vicina*/*L. coeruleiviridis*/*Phoridae* in the four-component mixture node (second level discrimination); and (5) *C. rufifacies*/*C. vicina*/*L. coeruleiviridis*/*L. sericata*/*P. regina* and *C. rufifacies*/*C. vicina*/*L. coeruleiviridis*/*L. sericata*/*Phoridae* in the five-component node. We evaluated the merits of the conformal predictor using *y*-permuted labels, correctly anticipating that the fits would yield poor results. Thus, the reported merits of the conformal predictor were considered in each classification node based on the average of 10 × *y*-permuted modeling. As displayed in Figure S4, the performance in each node dropped for the *y*-permuted data sets. For example, for determination of the mixture type for the samples comprised of multiple species, the error and multiple prediction rates were 0.25 and ∼0.74 at a significance level of 0−0.2, respectively. This shows that the samples were assigned to incorrect labels or to multiple labels (∼ four labels in total).

The results demonstrate proof-of-concept for application of the developed conformal predictor method and DART-HRMS for identification of the species constituents of complex blow fly larvae mixtures that are commonly encountered as maggot masses on decomposing remains or in tissue infestations.

CONCLUSIONS

Multispecies samples of blow fly larvae that are commonly collected in the context of medical, agricultural, forensic, or other types of investigations are difficult to identify because of the paucity of relevant taxonomic keys, and the absence of readily observable species-specific morphological features that can be used to tell them apart. Using aqueous ethanol insect storage suspensions of the type generated in the field by entomologists, an approach to species identification of multispecies samples that utilizes DART-HRMS measurements followed by application of a machine learning approach was developed to determine the species composition of mixtures of larvae. A conformal prediction protocol (aggregated hierarchical conformal predictor), with an underlying top-down hierarchical classification tree, was used for the first time for multispecies discrimination. Evaluation of the merits of the conformal predictor performance indicated confidence levels between 80 and 99% in prediction of test samples. The confidence level was observed to be lower for five- and six-species component mixtures. Further evaluation of the conformal predictor using *y*-permuted data yielded results that validated the suitability of DART-HRMS data for multispecies discrimination of blow flies at the larval life stage when they manifest as mixtures.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c00199.

Four additional figures, one scheme, and five tables referenced in the text: Aggregation of blow fly larvae on a pig carcass; representative DART-HRMS spectra representing various species of larvae; the merits plots of the aggregated hierarchical conformal predictor in analysis of the actual and *y*-permuted data for class assignment in the hypothesized significance level threshold range of 0−0.2; the algorithm of the aggregated hierarchical conformal predictor; makeup of the 70% aqueous ethanol suspensions representing mixture types; information on the input data and parameters for the learning of the neural networks; classification tree performance for all calibration and test samples using the common strategies for accuracy computation; and percentages of errors, multiple predictions, not assigned predictions, and false positive and false negative rates at the threshold significance levels (PDF)

AUTHOR INFORMATION

Corresponding Author

Rabi A. Musah − *Department of Chemistry, University at Albany, State University of New York, Albany, New York 12222, United States;* orcid.org/0000-0002-3135-4130; Email: rmusah@albany.edu

Authors

Samira Beyramysoltan − *Department of Chemistry, University at Albany, State University of New York, Albany, New York 12222, United States*

Mónica I. Ventura − *Department of Chemistry, University at Albany, State University of New York, Albany, New York 12222, United States*

Jennifer Y. Rosati − *Department of Sciences, John Jay College of Criminal Justice, New York, New York 10019, United States*

Justine E. Giffen-Lemieux − *Department of Chemistry, University at Albany, State University of New York, Albany, New York 12222, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c00199

Author Contributions

R.A.M. and J.Y.R. conceived of the project. M.I.V and J.E.G. performed experiments. S.B. performed statistical analysis. J.Y.R. collected eggs, reared and maintained fly colonies, and conducted fly species identification. R.A.M. and S.B. wrote the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Joey Fragale and Veena Mehta for insect care. The financial support of the John Jay College for the Advancement of Research Faculty Scholarship Program and the United States National Institute of Justice is gratefully acknowledged. The project was supported in part by Award Numbers 2015-DN-BX-K057 and 2017-R2-CX-0020 awarded by the United States National Institute of Justice. The opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect those of the Department of Justice.

REFERENCES

(1) Reibe, S.; Madea, B. *Forensic Sci. Int.* **2010**, *195* (1−3), 52−7.
(2) Vasconcelos, S. D.; Soares, T. F.; Costa, D. L. *Int. J. Legal. Med.* **2014**, *128* (1), 229−33.
(3) Krinsky, W. L. Chapter 5 - Forensic Entomology. In *Medical and Veterinary Entomology*, 3rd ed.; Mullen, G. R., Durden, L. A., Eds.; Academic Press: 2019; pp 51−60.
(4) Donovan, S. E.; Hall, M. J. R.; Turner, B. D.; Moncrieff, C. B. *Med. Vet. Entomol.* **2006**, *20* (1), 106−114.
(5) Sharma, R.; Kumar Garg, R.; Gaur, J.R. *Egypt J. Forensic Sci.* **2015**, *5* (1), 1−12.
(6) Anderson, G. S. *J. Forensic Sci.* **2000**, *45* (4), 824−832.
(7) Frątczak, K.; Matuszewski, S. *Forensic Sci. Int.* **2014**, *241*, 20−26.
(8) Pechal, J. L.; Moore, H.; Drijfhout, F.; Benbow, M. E. *Forensic Sci. Int.* **2014**, *245*, 65−71.
(9) Greenberg, B.; Singh, D. *J. Med. Entomol.* **1995**, *32* (1), 21−26.
(10) Greenberg, B.; Szyska, M. L. *Ann. Entomol. Soc. Am.* **1984**, *77* (5), 488−517.
(11) Liu, D.; Greenberg, B. *Ann. Entomol. Soc. Am.* **1989**, *82* (1), 80−93.
(12) Campobasso, C. P.; Introna, F. *Forensic Sci. Int.* **2001**, *120* (1), 132−139.
(13) Bourel, B.; Callet, B. t.; Hédouin, V.; Gosset, D. *Forensic Sci. Int.* **2003**, *135* (1), 27−34.
(14) Rolo, E. A.; Oliveira, A. R.; Dourado, C. G.; Farinha, A.; Rebelo, M. T.; Dias, D. *Forensic Sci. Int.* **2013**, *228* (1), 160−164.
(15) Mazzanti, M.; Alessandrini, F.; Tagliabracci, A.; Wells, J. D.; Campobasso, C. P. *Forensic Sci. Int.* **2010**, *195* (1), 99−102.
(16) Malgorn, Y.; Coquoz, R. *Forensic Sci. Int.* **1999**, *102* (2), 111−119.
(17) Whitworth, T. L.; Dawson, R. D.; Magalon, H.; Baudry, E., DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proc. R. Soc. London, Ser. B* **2007**, *274* (1619).1731
(18) Wells, J. D.; Williams, D. W., Validation of a DNA-based method for identifying Chrysomyinae (Diptera: Calliphoridae) used in a death investigation. *Int. J. Legal Med.* **2006**, *121*.1
(19) Musah, R. A.; Espinoza, E. O.; Cody, R. B.; Lesiak, A. D.; Christensen, E. D.; Moore, H. E.; Maleknia, S.; Drijfhout, F. P. *Sci. Rep.* **2015**, *5* (1), 11520.
(20) Beyramysoltan, S.; Giffen, J. E.; Rosati, J. Y.; Musah, R. A. *Anal. Chem.* **2018**, *90* (15), 9206−9217.
(21) Giffen, J. E.; Rosati, J. Y.; Longo, C. M.; Musah, R. A. *Anal. Chem.* **2017**, *89* (14), 7719−7726.
(22) Sanit, S.; Sukontason, K.; Kurahashi, H.; Tomberlin, J. K.; Wannasan, A.; Kraisittipanit, R.; Sukontason, K. L. *Acta Trop.* **2017**, *176*, 395−401.
(23) Sanit, S.; Limsopatham, K.; Klong-Klaew, T.; Samerjai, C.; Yasanga, T.; Sukontason, K.; Tomberlin, J. K.; Sukontason, K. L. *Acta Trop.* **2018**, *188*, 168−179.
(24) Bunchu, N.; Thaipakdee, C.; Vitta, A.; Sanit, S.; Sukontason, K.; Sukontason, K. L. *J. Parasitol. Res.* **2012**, *2012*, 10.
(25) Aubernon, C.; Hedouin, V.; Charabidze, D. *J. Adv. Res.* **2019**, *16*, 67−73.
(26) Boulay, J.; Aubernon, C.; Ruxton, G. D.; Hédouin, V.; Deneubourg, J.-L.; Charabidzé, D. *Insect Sci.* **2019**, *26* (1), 2−19.
(27) Dombrovski, M.; Poussard, L.; Moalem, K.; Kmecova, L.; Hogan, N.; Schott, E.; Vaccari, A.; Acton, S.; Condron, B. *Curr. Biol.* **2017**, *27* (18), 2821−2826 e2.
(28) Fouche, Q.; Hedouin, V.; Charabidze, D. *Sci. Rep.* **2018**, *8* (1), 2844.
(29) Tsoumakas, G.; Katakis, I. *Int. J. Data Warehousing Mining* **2007**, *3*, 1−3.
(30) Peng, H.; Long, F.; Ding, C. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27* (8), 1226−38.
(31) Beyramysoltan, S.; Abdul-Rahman, N.-H.; Musah, R. A. *Talanta* **2019**, *204*, 739−746.
(32) Classification with conformal predictors. In *Algorithmic Learning in a Random World*, Vovk, V., Gammerman, A., Shafer, G., Eds.; Springer US: Boston, MA, 2005; pp 53−96.
(33) Papadopoulos, H. Inductive conformal prediction: theory and application to neural networks. In *Tools in Artificial Intelligence*; Fritzsche, P., Ed.; InTech: Vienna, 2008; pp 315−330.
(34) Gauraha, N.; Spjuth, O. *arXiv:1804.05494* **2018**, 1−4.
(35) Vovk, V.; Fedorova, V.; Nouretdinov, I.; Gammerman, A. Criteria of efficiency for conformal prediction. In *Conformal and Probabilistic Prediction with Applications*; Gammerman, A., Luo, Z., Vega, J., Vovk, V., Eds.; Springer International Publishing: New York, 2016; pp 23−39.
(36) Golland, P.; Fischl, B. Permutation tests for classification: Towards statistical significance in image-based studies. In *Information Processing in Medical Imaging*; Taylor, C., Noble, J. A., Eds.; Springer: Berlin, 2003; pp 330−341.
(37) Rauber, P. E.; Fadel, S. G.; Falcão, A. X.; Telea, A. C. *IEEE Trans. Visual. Comput. Graphics* **2017**, *23* (1), 101−110.
(38) Smith, J.; Nouretdinov, I.; Craddock, R.; Offer, C.; Gammerman, A. Conformal anomaly detection of trajectories with a multi-class hierarchy. In *Statistical Learning and Data Sciences*; Gammerman, A., Vovk, V., Papadopoulos, H., Eds.; Springer International Publishing: Cham, 2015; pp 281−290.
(39) Vovk, V. *Mach. Learn.* **2013**, *92*, 349−376.
(40) Kaur, P.; Stoltzfus, J. *Int. J. Acad. Med.* **2017**, *3* (2), 268−270.

https://dx.doi.org/10.1021/acs.analchem.0c00199
Anal. Chem. 2020, 92, 5439−5446