Inter-laboratory workflow for forensic applications: Classification of car glass fragments

Omer Kaspi[a], Osnat Israelsohn-Azulay[b], Yigal Zidon[b], Hila Rosengarten[b], Matea Krmpotić[c], Sabrina Gouasmia[c], Iva Bogdanović Radović[c], Pasi Jalkanen[d], Anna Liski[d], Kenichiro Mizohata[d], Jyrki Räisänen[d], Olga Girshevitz[e,*], Hanoch Senderowitz[a,*]

[a] *Department of Chemistry, Bar-Ilan University, Ramat-Gan 5290002, Israel*

[b] *Israel Police HQ, Toolmarks and Materials Lab, Israel*

[c] *Laboratory for Ion Beam Interactions, Division of Experimental Physics, Ruđer Bošković Institute, Bijenička cesta 54, HR-10000 Zagreb, Croatia*

[d] *Department of Physics, University of Helsinki, P.O. Box 43, FI-00014 Helsinki, Finland*

[e] *Bar Ilan Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat-Gan 5290002, Israel*

* Corresponding authors.

E-mail addresses: Olga.Girshevitz@biu.ac.il (O. Girshevitz), Hanoch.Senderowitz@biu.ac.il (H. Senderowitz).

ABSTRACT

The International Atomic Energy Agency (IAEA) has coordinated a research project titled ”Enhancing Nuclear Analytical Techniques to Meet the Needs of Forensics Sciences” (CRP F11021) with the aim of empowering accelerator and reactor based techniques for applications in forensic sciences. One of the key topics of this project was the analysis and classification of forensic glass specimens using Ion Beam Analysis (IBA) techniques and in particular, Particle Induced X-ray Emission (PIXE).

To this end, glass fragments from car windows from different car models and manufacturers provided by the Israeli police force were subjected to PIXE measurements at three laboratories to determine their elemental compositions and possible glass corrosion. Major and trace elements were measured and given as an input to machine learning (ML) algorithms in order to develop classification models to determine the origin of the glass samples.

First, we have developed ML models based on the results obtained at each lab. These models successfully classified glass fragments into different car models with an accuracy > 80% on external test sets. Next, we demonstrated that following an appropriate pre-processing step, results from different labs could be combined into a single unified database for the derivation of a classification model. This model demonstrates good performances that matches or surpasses the performances of models derived from the individual labs. This finding paves the way towards establishing an international database that is composed of measurements from various PIXE labs. We believe that using this methodology of combining various sources of measurements will improve models' performances and generality and will make the models accessible to law enforcement agencies around the world.

Keywords: PIXE, car window glass fragments**,** Machine Learning, Forensics

Introduction

Detailed understanding of crime scenes is very important in a criminal investigation in particular if leading to a trial. The burden of providing insightful conclusions from accurate measurements performed on acquired forensic specimens becomes increasingly heavier with societies' aspiration to improve conviction rates while reducing wrongful imprisonment. To cope with this task, large amounts of accessible, standardized, and reliable data could be used, or in other words, a forensic database is required.

Harnessing the benefits of a large database to forensic applications is not a novel idea. In 2013 Rossy et al. [1] presented a statistical outlook on a retrospective database composed of data from six police forces covering the western part of Switzerland from 2009 to 2011. The study showed that forensic information links incidents using situational information, shoe marks pattern, DNA and images 29.7% of all cases, thus demonstrating the benefits a large database can provide. Erlich et al. [2] showed how a large DNA database (containing samples from 3% of US citizens) can identify at least a single third-cousin match with high probability (>99%). Using this information, the Golden State Killer Joseph James DeAngelo Jr., a former police officer who committed at least 13 murders, 50 rapes, and 120 burglaries across California between 1973 and 1986, was apprehended in 2018.

Glass fragments may form a significant piece of evidence. Glass fragments attach to clothes, shoes, hair, skin, or other objects [3]–[5]. A forensic investigation may link a fragment to its potential origin be it specific [6]–[9] or broader and related to the glass's original function (e.g., bottles, mugs, windows, etc.). Properly analyzing this important evidence may help forensic investigators get a better, more complete understanding of crime scenes and in favorable cases, provide an association between a suspect and a victim. However, as of now, glass evidence does not fulfill its full forensic potential. For example, every year, the Israeli Police Force reports on dozens of crimes that involved glass fragments as evidence that did not significantly influence the investigation and judicial process.

The fact that most glass fragments found in crime scenes are smaller than 0.5 mm coupled with recent improvements in glass manufacturer's quality control processes, renders classical forensic methods such as morphological analysis (e.g., thickness, color, etc.) or Refractive Index (RI) measurements, the gold standard for glass evidence comparison [10]–[12], insufficiently accurate. Thus, alternative approaches are required to fully extract the forensic potential of such fragments.

One such approach is based on the usage of Machine Learning (ML) algorithms. Several models derived by such algorithms were reported in the literature and used to estimate the similarity between a pair of glass fragments in order to determine whether they originate from the same source [13], [14] as well as to classify glass fragments into one of pre-defined classes [15]. These and other studies were summarized in several review papers [16], [17].

Recently Kaspi et al. [18] proposed a workflow for constructing ML-based classification models for glass fragments. This workflow is presented in Figure 1 and is composed of three steps, namely specimen acquisition, specimen measurement and model training, validation and application.

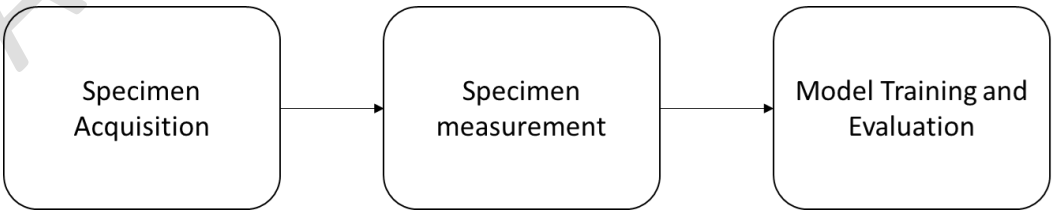


Figure 1: Workflow to derive a classification model

This workflow was initially implemented using Particle-induced X-ray emission (PIXE) measurements performed in a single laboratory on 48 glass specimens collected from junkyards by the Israeli DIFS (Division of Identification and Forensic Sciences) and covering 13 car models from ten car manufacturers. Classification models were derived using Random Forest (RF) and *k*-Nearest Neighbors algorithms with an overall accuracy of > 80%. Of note, these models were constructed based on available information pertaining to glass fragmets, namely, their car manufacturer origin. Thus the models are agnostic to additional information related to the glass origin and supply chain.

In this work we aim to consolidate data measured on the same set of fragments by three different laboratories into a unified database and demonstrate that source-agnostic, predictive ML models could be developed. This database will allow different groups to share their measurements and benefit from models derived from a large collection of diverse (i.e., in terms of car manufacturers) data.

Materials and methods

A recent work [18] presented a workflow that accepts glass specimens collected from car windows and determines the manufacturer of that car with good performances (> 80%). The current research demonstrates that with a small modification, this workflow could be made to accept data from laboratories using different experimental setups and combine them into a large, standardized database. Potential benefits of such database are its increased size (clearly, more labs can measure more specimens) and specimens' diversity (e.g., French cars are likely to be more abundant in France whereas German cars are likely to be more abundant in Germany). Moreover, such a global database may well give rise to more global models in comparison with local databases.

With this in mind, in this work we collected elemental analysis data on glass fragments from three laboratories all participating in the IAEA-initiated CRP-F11021 consortium, namely, Bar-Ilan Institute of Nanotechnology (BINA), Ruđer Bošković Institute (RBI), and the Department of Physics, Accelerator Laboratory, University of Helsinki (HU). All participating laboratories use the PIXE technique, albeit with different instrumentations and setups. Thus, to test for and ensure data compatibility across the three laboratories, the glass Standard Reference Materials NIST-620

and NIST-610 were used as external standards. The experimental setups were required to reproduce the known concentrations of Na, Mg, Al, Si, K, Ca, S, Cl, Ti, Fe, Cr, Sn, and Mn in these standards., For this purpose each lab prepared the sampels, tested their homogeneity, optimized beam energy, solid angle and sample's collection charge, and defined quantification strategies, and elemental menu, (see

Table 1).

Next, 48 glass specimens originally used by Kaspi et al. [18], were broken into three smaller specimens and distributed to the three labs for PIXE measurements. BINA and RBI further broke all specimens into smaller fragments. Fragments that had a smooth surface likely correspond to the surface of the glass (it is impossible to tell whether these fragments came from the inside surface or the outside surface), while those with ruptured surfaces likely come from the bulk of the glass. In contrast, HU measured both sides of the specimens (Side A, Side B). Thus, measurements were made on 96 specimens per laboratory. As a further assurance for homogeneity, each fragment was measured multiple times and then averaged. Due to the differences in the experimental setups, the elements identified by the groups varied, e.g., 16, 10 and 5 elements were identified by BINA, RBI and HU, respectively (see

Table 2). Of note, some elements are represented by multiple peaks, thus the total number of features may be larger than the number of elements.

Having verified the ability of the individual labs to reproduce the elements composition of the NIST-620 and NIST-610 standards, we first developed laboratory-specific classification models. In developing these models, we only used the five features (i.e., elements) measured by all three laboratories (Al, Si, K, Ca, Fe; see Table 2). To this end we have used the Random Forest (RF) algorithm which was previously shown by us to outperform *k*NN in what relates to the classification of glass fragments [18].

Table 1: Setups used by the different participating laboratories.

Lab ID	**BINA IBA**	**UH IBA**	**RBI IBA**
Method for Analysis	PIXE	PIXE	PIXE
Type of Accelerator	1.7 MV Tandem Pelletron	5 MV Tandem	1 MV Tandetron
Type of Ion, Energy (MeV)	2 MeV H+	3 MeV H+	2 MeV H+
Beam size, mm	1.5	1	1 and 3
Charge, µC	3	1	1
Detector Model	Amptek X-123FastSDD, 70mm C2window	KETEK AXAS-D SDD VITUS H20	SDD KETEK, model Vitus H20 Si(Li) Canberra, model SSL80155
Detector resolution, (eV)	122	133	140/155 for SDD and Si(Li) respectively
Detector filter	FF, 100 mm Kapton, 1.5% effective area hole	None	SDD N/A Si(Li) 335 µm Mylar
Coating of the sample	Carbon, 30 nm	None*	None**
Analysis Software	GUPIX v3.0.3	PyMCA v5.5.5	GUPIX v2.2.4

*HU used external PIXE and current normalization was done using 2.95 keV ArK peak.

**To avoid an increase in the background of the X-ray spectra due to charge buildup on the insulating glass surface, a small piece of double-adhesive carbon tape was glued on the sample surface to ensure electrical connection with the metal sample holder. Measurements were performed by positioning the ion beam as close as possible to the carbon tape which was sufficiently effective to prevent charging of the glass.

Table 2: Databases measured by the different laboratories.

Group	**# of registries**	**# of elements**	**# of features**	**elements**
BINA	92	16	18	Na, Mg, Al, Si, S, Cl, K, Ca, Ti, Cr, Mn, Fe, Co, Cu, Zn, Sr
RBI	104	10	10	Na, Mg, Al, Si, S, K, Ca, Cr, Mn, Fe
HU	96	5	5	Al, Si, K, Ca, Fe

Briefly, RF constructs multiple decision trees where each tree is a set of sequential rules that pair objects' attributes with their activities (e.g., class membership). For model robustness considerations, each tree is constructed using a randomly selected set of features (see Figure 2 for an example of a single tree). The "forest" in RF is the final decision made by a majority vote of all of the trees [19]. In this work, we have used the sklearn's implementation of the RF algorithm in Python using 500 trees.

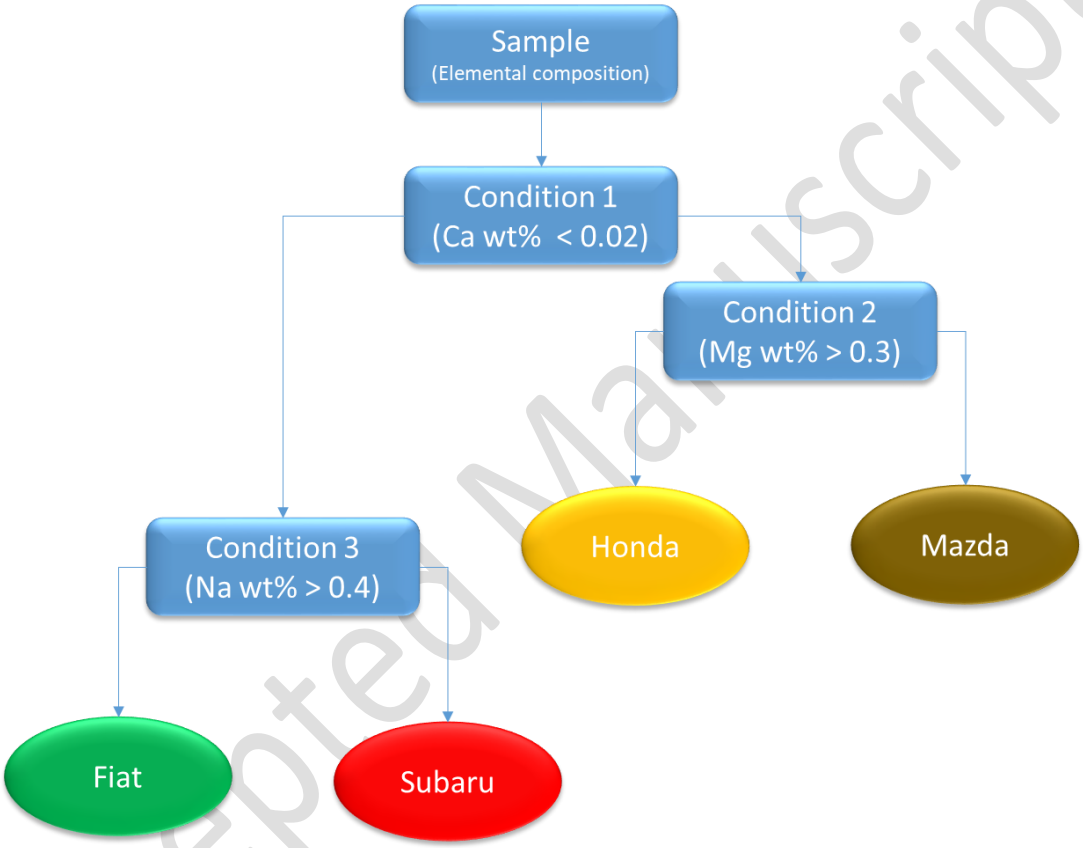


Figure 2: An example of a decision tree which classifies glass fragments into one of four classes based on the content of calcium (Ca), magnesium (Mg), and sodium (Na).

The resulting models were evaluated on external test sets by means of a confusion matrix (see **Error! Reference source not found.** in reference [18]). The rows in the matrix represent the known truth (i.e., actual classes of samples) while the columns represent the model's predictions (i.e., model's estimation for the samples' classes). Several evaluation metrics were derived from the confusion matrix including recall, precision, and F1-score. Precision is calculated as the number of objects correctly classified (True Positive, TP) divided by the total number of samples classified as active (True Positive (TP) + False Positive (FP)). Recall is calculated as the number

of objects correctly classified (TP) divided by the total number of objects in the class (TP + False Negative (FN)). F1-Score is the harmonic mean of the Precision and Recall $(\frac{precision*recall}{precision+recall}) * 2$.

To further validate the models, all models underwent a Y-scrambling procedure whereby the class memberships of the training group samples were randomly shuffled and used to create a faulty model that is expected to lead to poor predictions. Y-scrambling was repeated twenty times.

Having demonstrated the ability of lab-specific measurements to produce predictive classification models, we turned to develop a model based on all of the data. In order to consolidate data from the three laboratories into a unified database, the original workflow [18] that was also used here to develop lab-specific models was modified to include an additional Exploration and Preprocessing stage. In this stage the data is evaluated using various mathematical and ML tools, processed for the development of future classification models and unified to create a complete standardized database. Figure displays the adapted workflow.

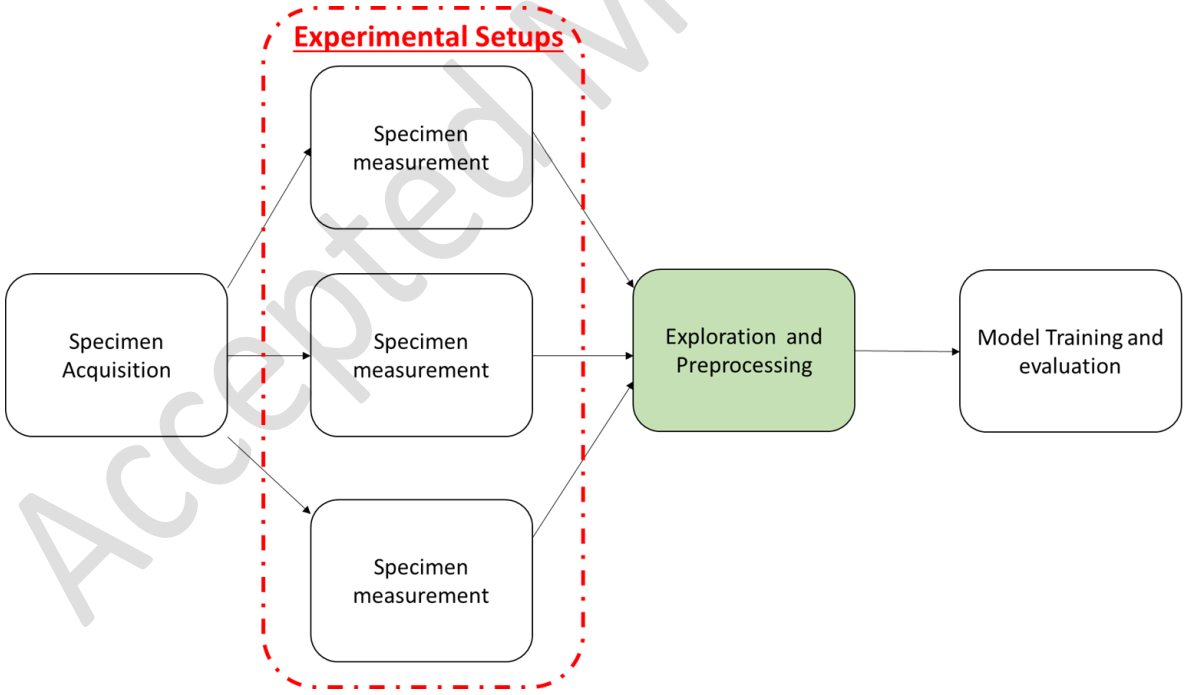


Figure 3: Workflow for glass fragments classification based on data collected from different laboratories.

The exploration and processing stage has two main objectives: (1) To provide a comparative overview of the data structure, and (2) To transform the data so that they could be readily and reliably combined into a unified database. The first objective was met through the usage of dimensionality reduction method whereas the second objective required Z-score normalization of the features.

Dimensionality reduction methods are used primarily for data visualization. They project a set of samples, originally defined in a high dimensional space into a lower space (for visualization purposes, the lower space would be 2-dimensional or 3-dimensional space). In this work we used two methods, namely, Principal Component Analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE). PCA creates a transformation matrix that allows to transform data from high dimensionality to lower dimensionality and vice versa while retaining the variance in the original, high dimensional space as much as possible [20]. t-SNE, in comparison, generates a probability distribution in the high dimensional space such that similar objects are assigned high probabilities, and minimizes its distance from a similar distribution generated in the low dimensional space. [21].

Normalizing each feature (i.e., element composition) from each individual laboratory was found to be required prior to the unification of measurements. This is because it became evident that different laboratories with different equipment and methodologies produce different nominal measured values for each element in the same specimen. Yet the overall "ordering" of the specimen within the dataset should be identical across all laboratories. To put all measured values from all labs on a common ground for comparison, Z-Score normalization was used. Z-Score normalizes each feature according to equation (1) where μ is the mean value of the feature and σ is its standard deviation.

$$z_{val} = \frac{x-\mu}{\sigma} \qquad (1)$$

Finally, the normalized databases from the individual labs were concatenated, registries from all labs were added and only common features were kept. The unified database contained 292 entries, each characterized by five features (Al, Si, K, Ca, Fe).

The processed and unified database was subjected to the same model derivation procedure used for the individual databases. Thus, models were derived from a training set and evaluated on a test

set using the Sensitivity, Recall, and F1-Score metrics and tested for chance correlation by Y-scrambling.

Last, to demonstrate that this unified database has merit in real forensic applications, the final model was challenged with the same test case used in Kaspi et al. [18] with satisfactory results.

RESULTS

At the start of the project, each participating laboratory calibrated its own experimental setup so as to reproduce the element composition of the NIST-610 and NIST-620 standards. The results are presented in Figure 4 and show an overall good agreement with the certified concentrations (the error is less than 10%). Additionally, it was observed that The HU detector was not sensitive to Na.

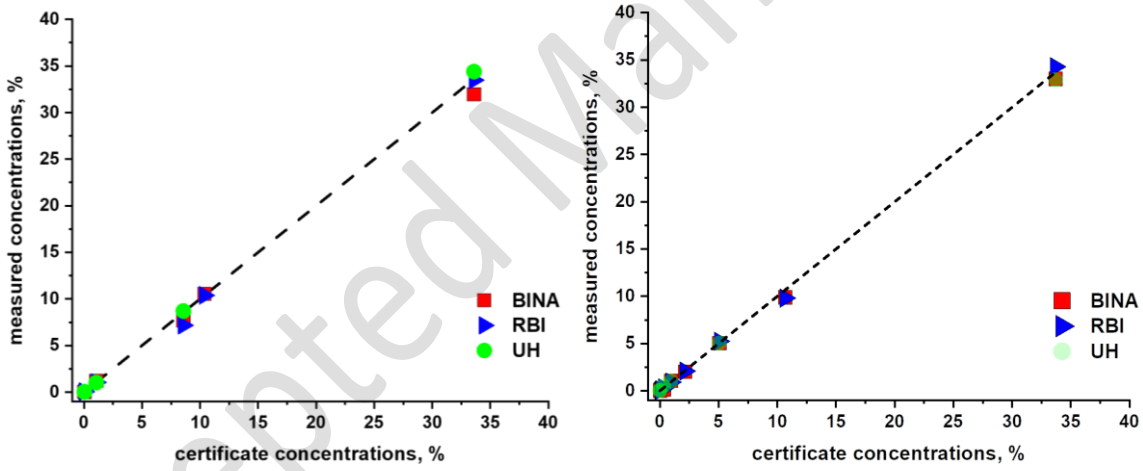
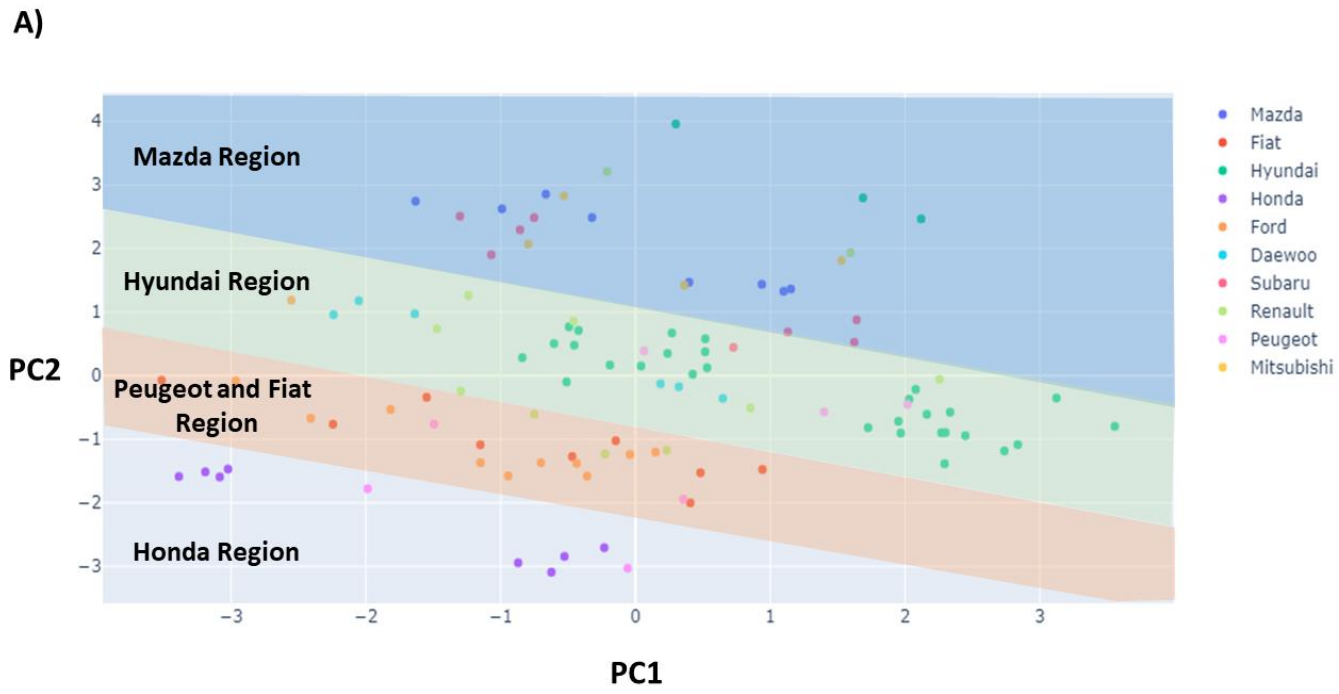


Figure 4: Comparison between the certificate and measured concentrations of the different elements in the NIST-610 (A) and NIST-620 (B) Standard Reference Materials as produced by the three participating laboratories.

Next, to better understand the data from the individual labs, PCA-based dimensionality reduction was applied. As is evident from Figure 5A&B (RBI and BINA, respectively), there are regions that contain most of the glass samples from a particular car manufacturer (A region defines the area where the likelihood for a particular manufacturer to be present is highest. Regions for several manufactures may overlap). Interestingly, the order in which these regions are located remains the same (albeit in the opposite direction) in the two projections. This observation demonstrates the

"inherent" order within the database by which it is possible to classify car manufacturers. Thus, a predictive ML model may be derived and more importantly, this model may be lab agnostic. Figure 5C (HU) also demonstrates the partition of the space into regions albeit in a slightly less precise manner. This suggests that a classification model may also be derived from the data measured at HU, however, with a smaller success rate than models derived based on BINA and RBI's data.



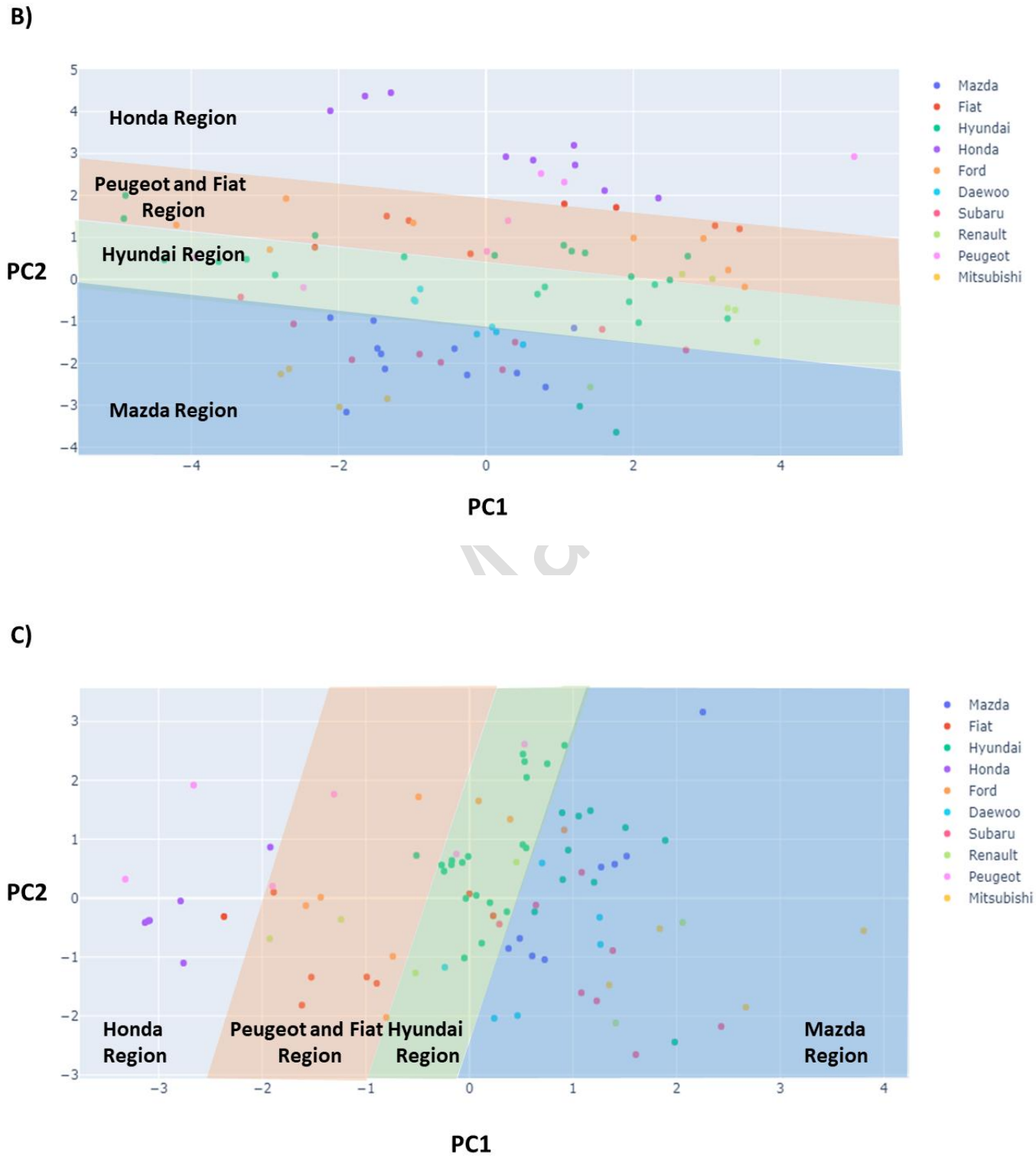


Figure 5:PCA for RBI (A), BINA (B), and HU (C). All labs present regions that contain most of the samples from a particular car manufacturer.

Next, we turned our attention to the derivation of lab-specific models. It is well established that ML models should be derived from one set of data (training set) and validated on another set (test set). Therefore, all groups constructed RF-based ML models with three types of training/test set partitions, namely, (1) using data obtained from the glass surface as a training set and the data obtained from the bulk of the glass as a test set, (2) reversing the roles of the surface and bulk data and (3) combining surface and bulk data and randomly splitting them into training and test set. The process of random splitting was repeated 20 times (For HU the training/test partition was based on side A and side B and not on bulk and surface). The performances of the derived models are presented in Table 3. All labs managed to construct reliable models that perform significantly better than a random model (a random model would produce ~0.1 for Precision, Recall and F1-Score based on a random draw of one out of ten classes). Y- scrambling provided F-Score values of 0.1 ± 0.06, 0.13 ± 0.05, 0.15 ± 0.07 for BINA, RBI and HU respectively as can be expected from a purely random model clearly indicating that the original models are not chance-correlated.

Table 3: Performances of models from the various laboratories on the respective test sets.

Lab group	Training	Test	Precision	Recall	F1 - Score
BINA	Bulk	Surface	0.88 ± 0.01	0.85 ± 0.04	0.84 ± 0.02
	Surface	Bulk	0.88 ± 0.02	0.84 ± 0.04	0.85 ± 0.02
	Random (67%)	Random (33%)	0.82 ± 0.07	0.83 ± 0.08	0.81 ± 0.07
RBI	Bulk	Surface	0.78 ± 0.03	0.72 ± 0.02	0.74 ± 0.03
	Surface	Bulk	0.69 ± 0.03	0.68 ± 0.04	0.63 ± 0.04
	Random (67%)	Random (33%)	0.85 ± 0.08	0.84 ± 0.06	0.82 ± 0.08
HU	Side A	Side B	0.36 ± 0.01	0.26 ± 0.02	0.23 ± 0.01
	Side B	Side A	0.48 ± 0.05	0.48 ± 0.02	0.44 ± 0.05
	Random (67%)	Random (33%)	0.62 ± 0.07	0.62 ± 0.08	0.59 ± 0.07

Next, in order to develop a global, lab-agnostic model, the data from the individual labs was z-scored normalized (see the discussion section for an explanation about the importance of normalizing the individual databases before combining them). Then, data from all sources was combined into a unified database with 292 registries and five features (common to all labs). Then the combined data was visualized in 2D using t-SNE for dimensionality reduction (t-SNE produces better clusters than PCA and hence was used for this purpose here [21]). The results are presented

in Figure 6 that shows that some manufacturers are easily distinguishable (i.e., Honda) while others are interwoven with one another (e.g., Ford, Renault and Fiat, as seen in the zoomed-in region in Figure 7).

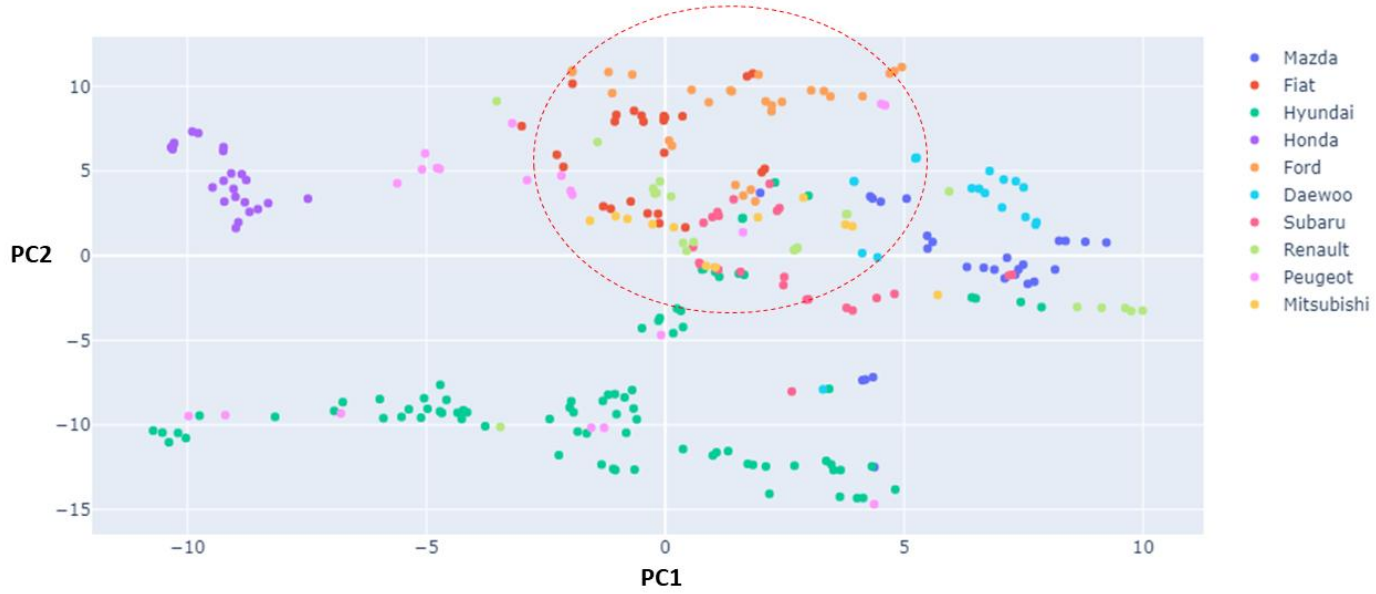


Figure 6: Dimensionality reduction of the combined database. A zoom-in on the circled region is provided in Figure 8.

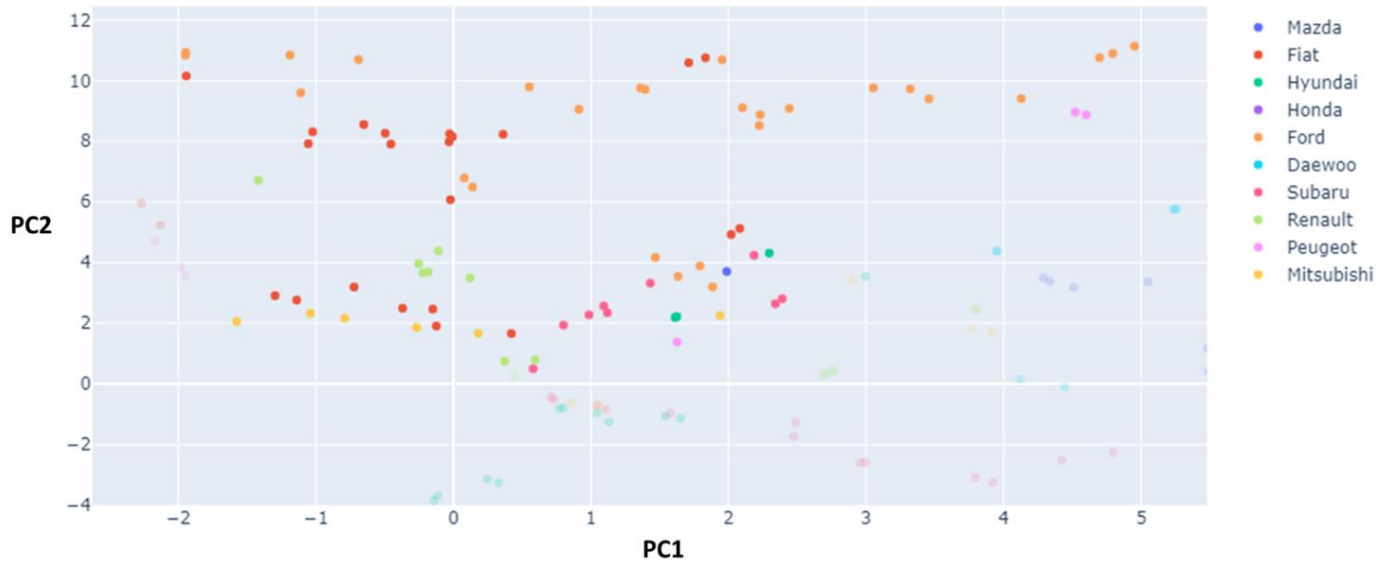


Figure 7: Zoom in on the Ford –Renault–Fiat region of the combined database.

The unified database was then used to train and validate a new RF model. However, since not all laboratories have "surface" and "bulk" specimens, data partition into training and test set was only performed randomly and repeated 20 times. The mean Recall, Precision and F1 Score of the resulting models were found to be 0.84 ± 0.03, 0.85 ± 0.04, and 0.84 ± 0.04, respectively. Y-scrambling provided a model with F1-score value of 0.14 ± 0.06 clearly indicating that the original models are not chance-correlated.

Finally, to demonstrate the applicability of the unified database, it was applied to a previously described test case [18] describing a hit-and-run accident. Subsequent to the accident, glass fragments were taken from the suspect, the hitting vehicle and the location the suspect claimed to be, at the time of the accident. Due to technical issues, the samples were only measured by BINA and RBI laboratories and the results were fed into 20 classification models derived from the unified database.

clearly shows that both laboratories link the suspect and the hitting car, as was indeed the case.

Table 4: Test case results

Test case outcome	**BINA**	**RBI**
Suspect matched with the hitting vehicle	65%	50%
Suspect's alibi was confirmed	0%	0%
Inconclusive (Suspect matched with both vehicle and potential alibi)	35%	30%
Inconclusive (Suspect matched with neither vehicle nor potential alibi)	0%	20%

DISCUSSION

The purpose of this inter-laboratory project was to evaluate the PIXE technique for determining major, minor and trace elements in glass fragments, to examine and compare the results obtained by the PIXE experimental setups of the CRP-F11021 consortium participants, and to test their usefulness in addressing forensic-related problems. We reasoned that from a practical standpoint, the PIXE technique could be evaluated based on its ability to afford data that could be used for developing predictive classification models for the aforementioned glass fragments. In addition, we wanted to determine whether PIXE results obtained from different laboratories could be

favorably combined into a unified database that could afford the development of a reliable glass fragments classifier.

In principle, the ever-increasing rigidness in industry's standards controls the composition of major elements [22]–[24] within manufacturer vehicle glass while trace elements are generally less considered. This fact led to the initial assumption that the key to differentiating between manufacturers lies in trace elements. However, a model based on the unified database which was composed of major elements only (Al, Si, K, Ca and Fe – All major elements) gave satisfactory results.

Data visualization via PCA (Figure) revealed that measurements performed at RBI and BINA afforded a reasonable separation between different car manufacturers suggesting the possibility of developing reliable classification models. The results presented in Table 3 clearly support this assumption. In contrast, poorer separation in the space of the PCs and less accurate models could be developed using the HU data. This might be attributed to the smaller number of elements measurable by the HU setup. Importantly, since all models were evaluated on external test sets, we do not attribute the better performances of the RBI and BINA models to overfitting.

Figure 5 also provides the first hint towards the feasibility of combining the data obtained by the three labs into a unified database by demonstrating a similar "inherent order" of the specimen regardless of the measuring lab. This ordering reflects the fact that the differences between the specimens are consistent enough to be independent of the measuring protocol. However, this is only true when considering the ratios between the elements rather than their nominal concentrations. In support of this observation, Figure 8 shows differences in measurements between labs using absolute counts (left) and differences in measurements between labs using normalized counts (right). Clearly, normalization of the data reduces variability that is caused by different measurement protocols and instruments.

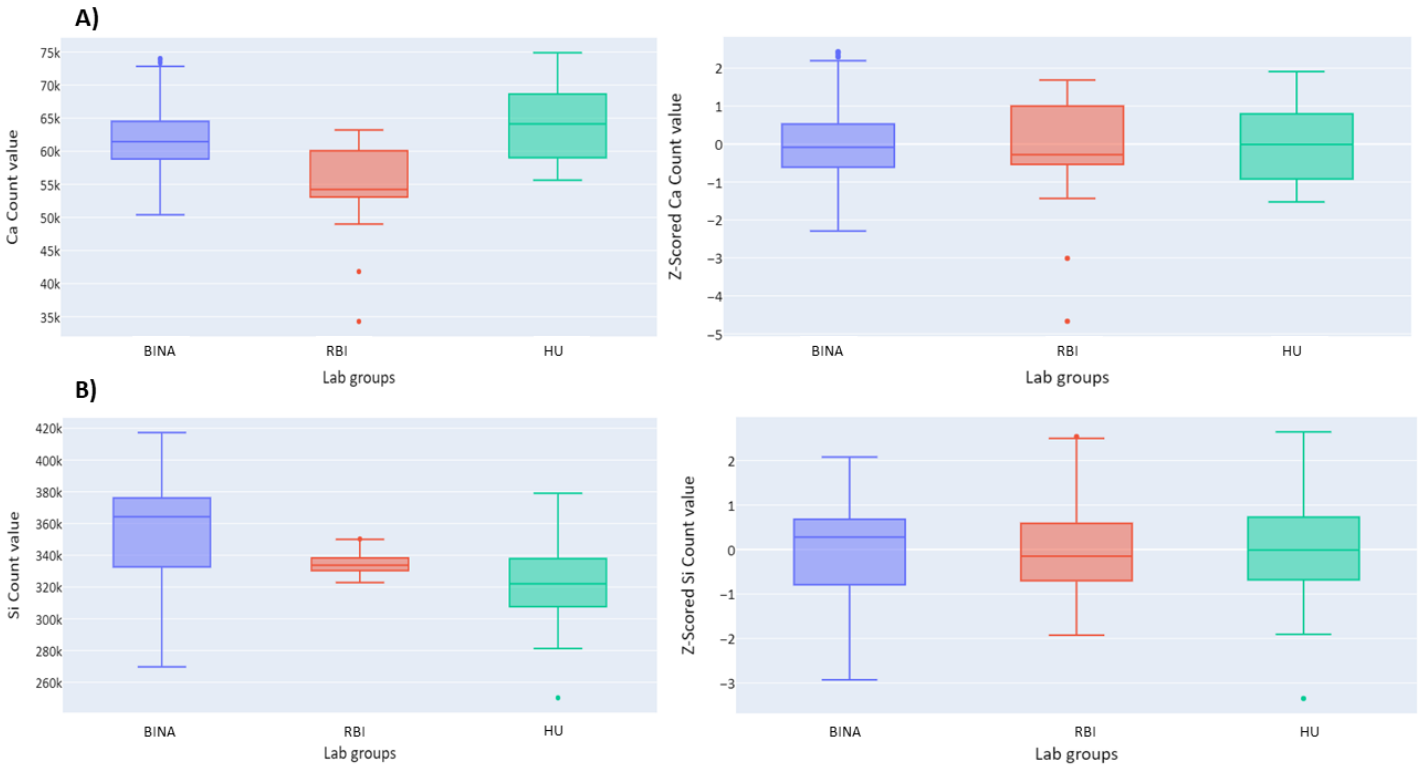


Figure 8: Comparison of measurements using different setups before (left) and after (right) normalization. A) Comparison of Ca counts, B) Comparison of Si counts.

As hinted by the data in Figure 5, consolidating PIXE measurements from three laboratories into a unified database afforded reliable classification models for glass fragments. Figure 9shows that the model produced by the unified database has performances metrics values (Precision, Recall, F1-Score) similar to the metrics of the best-performing, lab-specific model.

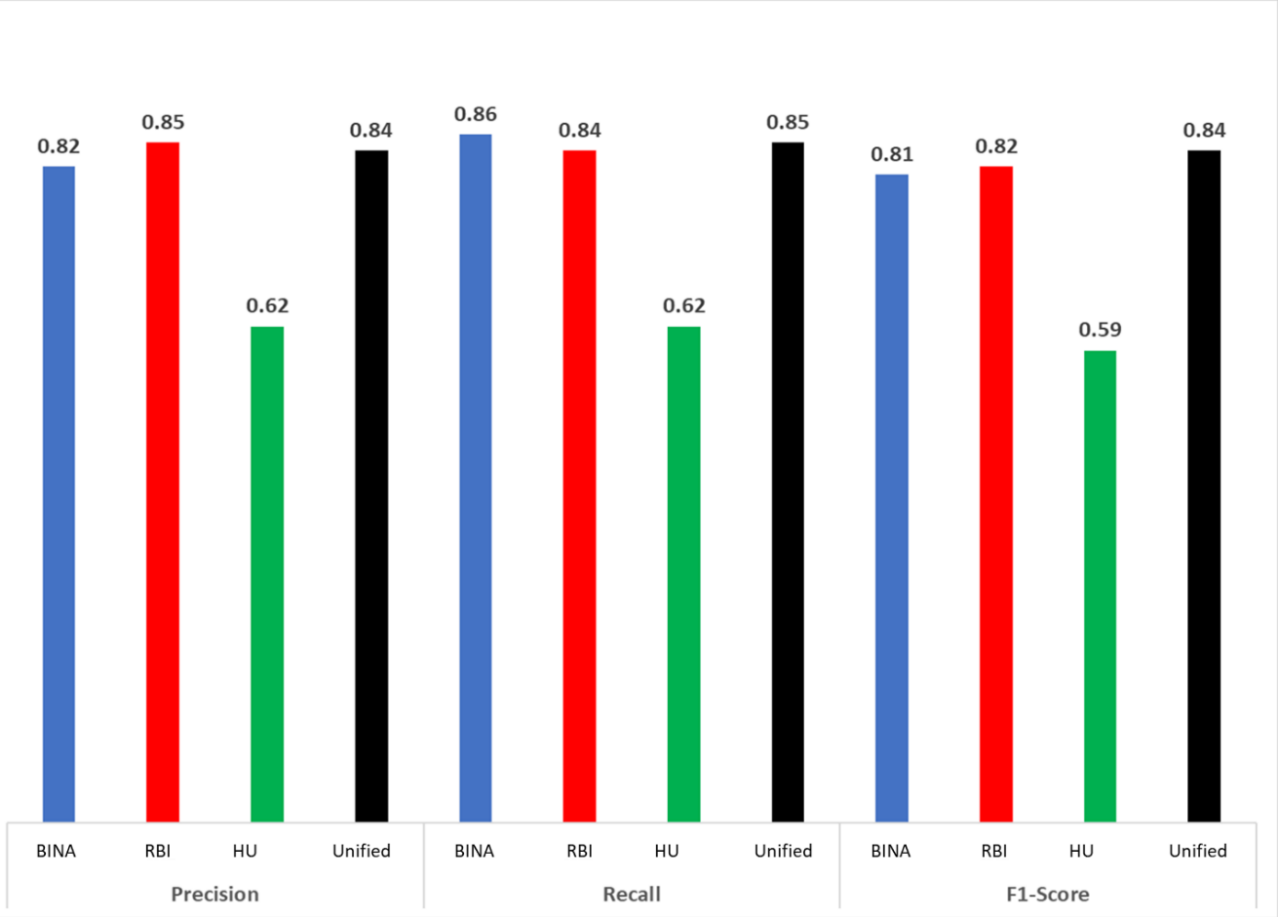


Figure 9: Performance metrics of models derived from the individual labs as well as those derived from the unified database.

Finally, the model derived from the unified database was successfully implemented on the test case. Thus, both participating laboratories (RBI and BINA) were able to correctly link the suspect with the vehicle (Table 4).

The high success rates of the models developed in this work result from differences in the elemental composition of the glass fragments which in turn likely reflect their manufacturing process and supply chain. This is apparent from Figure 6. Yet this figure also reveals that three manufacturers, Renault, Ford, and Fiat may possible share glass manufacturer, manufacturing processes and supply chain. However, since this type of information is generally not disclosed, we cannot verify this conjecture.

CONCLUSIONS AND FUTURE DIRECTIONS

In this work we have demonstrated that combining PIXE-based measurements of glass fragments from different laboratories into a unified database can produce a reliable classification model that can be potentially used by different laboratories around the world. The resulting model provided results which were equal to or better than any lab-specific model. Moreover, models derived from this database will likely increase in accuracy and generality as more specimens are measured and recorded. Finally, the workflow presented in this work can be extended to many other domains of forensics (e.g., gunshot residues, flammable liquids, substances of abuse, etc.) and thus increase performances of and cooperation between different law enforcement agencies.

Starting from the results presented here, more work could be performed along several lines of research. First, the reasons for the differences observed in the measured values of the same specimen between the different experimental setups should be further explored. These differences may be accredited to changes in the instrument configurations, acquisition parameters, limits of detection and sample fragment size, shape and orientation. Furthermore, the impact of these parameters on the model derivation process and quality should be studied. Next, as noted above, model performances and generality should be improved. For example, the data in Figure 7 clearly suggests that glass fragments retrieved from several manufacturers have only small differences in composition thereby challenging classification and degrading model performances. Either a larger data set or better ML algorithms are required to cope with this problem. Related to this is the fact that classifiers of the sort developed in this work will always classify a new sample into one of the pre-defined classes. However, introducing the concept of applicability domain, will allow the model to produce a new class, namely, a “none of known classes” class. Finally, models also incorporating data on trace elements should be derived and the relative contributions of major vs. trace elements to model performances should be investigated. This information may prove useful for making the workflow accessible to law agencies that use SEM-EDX or for the design of other types of low-cost instrumentations for measuring element composition.

REFERENCES

[1] Q. Rossy, S. Ioset, D. Dessimoz, and O. Ribaux, "Integrating forensic information in a crime intelligence database," *Forensic Sci. Int.*, vol. 230, no. 1–3, pp. 137–146, 2013, doi: 10.1016/j.forsciint.2012.10.010.

[2] Y. Erlich, T. Shor, I. Pe'er, and S. Carmi, "Identity inference of genomic data using long-range familial searches," *Science (80-.).*, vol. 362, no. 6415, pp. 690–694, 2018, doi: 10.1126/science.aau4832.

[3] S. A. A. Salman Iqbal, "Advancing Automation in Digital Forensic Investigations Using Machine Learning Forensics," *IntechOpen*, no. Forensics, p. 13, 2016, [Online]. Available: https://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics.

[4] B. Caddy, *Forensic examination of glass and paint: analysis and interpretation*. CRC press, 2001.

[5] V. M. Maxwell, "Forensic Interpretation of Glass Evidence," *J. Forensic Identif.*, vol. 51, no. 6, p. 597, 2001.

[6] C. G. G. Aitken and D. Lucy, "Evaluation of trace evidence in the form of multivariate data," *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 53, no. 1, pp. 109–122, 2004.

[7] C. G. G. Aitken, G. Zadora, and D. Lucy, "A two-level model for evidence evaluation," *J. Forensic Sci.*, vol. 52, no. 2, pp. 412–419, 2007.

[8] C. G. G. Aitken, D. Lucy, G. Zadora, and J. M. Curran, "Evaluation of transfer evidence for three-level multivariate data with the use of graphical models," *Comput. Stat. Data Anal.*, vol. 50, no. 10, pp. 2571–2588, 2006.

[9] P. Embrechts, E. T. H. Zurich, N. L. Johnson, and S. Kotz, "Statistics and the Evaluation of Evidence for Forensic Scientists by C . G . G . Aitken Review by : MW Journal of the American Statistical Association , Vol . 91 , No . 434 (Jun ., 1996), p . 915 Published by : American Statistical Association American ," vol. 91, no. 434, 2014.

[10] G. Zadora, "Examination of the refractive index of selected samples of glass for forensic

purposes," *Z Zagadnien Nauk Sadowych*, vol. 45, pp. 36–51, 2001.

[11] M. Pawluk-Kołc, J. Zieba-Palus, and A. Parczewski, "The effect of re-annealing on the distribution of refractive index in a windscreen and a windowpane. Classification of glass samples," *Forensic Sci. Int.*, vol. 174, no. 2–3, pp. 222–228, 2008, doi: 10.1016/j.forsciint.2007.04.229.

[12] M. Pawluk-Kołc, J. Zieba-Palus, and A. Parczewski, "Application of false discovery rate procedure to pairwise comparisons of refractive index of glass fragments," *Forensic Sci. Int.*, vol. 160, no. 1, pp. 53–58, 2006, doi: 10.1016/j.forsciint.2005.08.016.

[13] S. Park and A. Carriquiry, "Learning algorithms to evaluate forensic glass evidence," *Ann. Appl. Stat.*, vol. 13, no. 2, pp. 1068–1102, 2019, doi: 10.1214/18-AOAS1211.

[14] S. Park and S. Tyner, "Evaluation and comparison of methods for forensic glass source conclusions," *Forensic Sci. Int.*, vol. 305, p. 110003, 2019, doi: 10.1016/j.forsciint.2019.110003.

[15] A. J. Tallón-Ballesteros and J. C. Riquelme, "Data mining methods applied to a digital forensics task for supervised machine learning," *Stud. Comput. Intell.*, vol. 555, no. January, pp. 413–428, 2014, doi: 10.1007/978-3-319-05885-6_17.

[16] M. A. Kraus and M. Drass, "Artificial intelligence for structural glass engineering applications—overview, case studies and future potentials," *Glas. Struct. Eng.*, pp. 1–39, 2020.

[17] H. Liu, Z. Fu, K. Yang, X. Xu, and M. Bauchy, "Machine learning for glass science and engineering: A review," *J. Non-Crystalline Solids X*, vol. 4, p. 100036, 2019, doi: https://doi.org/10.1016/j.nocx.2019.100036.

[18] O. Kaspi, O. Girshevitz, and H. Senderowitz, "PIXE Based Machine-Learning (PIXEL)," *Talanta*, p. 122608, 2021, doi: https://doi.org/10.1016/j.talanta.2021.122608.

[19] T. K. Ho, "Random decision forests," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 1, pp. 278–282, 1995, doi: 10.1109/ICDAR.1995.598994.

[20] J. V. Stone, "Principal Component Analysis and Factor Analysis," *Indep. Compon. Anal.*,

2018, doi: 10.7551/mitpress/3717.003.0017.

[21] L. Van Der Maaten and G. Hinton, "Visualizing Data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

[22] N. Civici, E. V.-R. R. Phys, and undefined 2013, "Analysis of automotive glass of various brands using EDXRF spectrometry," *rrp.infim.ro*, vol. 65, no. 4, pp. 1265–1274, 2013, Accessed: Sep. 14, 2021. [Online]. Available: http://www.rrp.infim.ro/2013_65_4/A14.pdf.

[23] S. Scholes, "Modern glass practice," 1975, Accessed: Sep. 14, 2021. [Online]. Available: https://www.bcin.ca/bcin/detail.app?id=51892&wbdisable=false.

[24] H. Pfaender, *Schott guide to glass*. 2012.