EXPLOSIVE DETECTION WITH FTIR SPECTROSCOPY: THE ROLE OF INDUCTIVE BIAS IN SPECTRAL GENERALIZATION

Anis Dhifi[1,2], *Oscar Olarte*[1], *Bart De Clerck*[1], *Sam Leroux*[2], *Rob Haelterman*[1]

[1]Department of Mathematics, Royal Military Academy, Brussels, Belgium
[2]Ghent University, Ghent, Belgium

ABSTRACT

Explosive detection is a critical defence and public-safety task, where spectroscopic screening increasingly relies on machine-learning classifiers trained on a finite spectral library. The most challenging scenario is chemistry outside that library: threats evolve faster than curated collections, and the costliest failure is an unknown explosive classified as harmless. Using FTIR spectra of five bulk explosives, their mixtures, and a heterogeneous benign set, we evaluate generalization capability across four explosive classes under increasingly severe holdouts, up to excluding an entire chemical family from training. Closed-set accuracy is not informative: all models reach near-ceiling performance. However, when entire families are withheld, the models diverge sharply: some detect every unseen spectrum, others none. This divergence arises from inductive bias: models built on vibrational features shared across chemical families transfer to unseen chemistry, whereas models depending on family-specific features fail. The study also outlines a blind band selection that recovers the shared nitro bands, and a Flag-on-Unknown gate that routes unclassifiable spectra to inspection instead of clearing them as benign.

***Index Terms*—** FTIR spectroscopy, explosive detection, inductive bias, band selection, anomaly detection, open-set recognition, chemometrics

1. INTRODUCTION

Vibrational spectroscopy with compact, field-deployable instruments is a standard modality for the identification of hazardous materials [1–3], and machine-learning classifiers now automate the explosive/benign decision. The dominant evaluation practice trains and tests on the same compounds, differing only in replicate or batch. It therefore answers one question: can the model recognize the compounds it was trained on? The operationally decisive question is different: what does the model do with a chemistry it has never seen? New threat chemistries appear faster than any spectral library can grow, and a detector that silently classifies an unseen explosive class as benign fails in the costliest possible way. Methodologically, these challenges are hyperspectral signal-processing questions: spectral-library design, band selection, and open-set anomaly detection. Generalizing across chemical families in spectroscopy is directly analogous to cross-scene generalization in hyperspectral imaging, where classifiers trained on one scene often fail on another [4, 5]. The findings are isolated from background and instrument effects: one point spectrometer, one acquisition protocol, and one fixed processing chain, so any difference between models is attributable to the models alone. They are also instrument-agnostic: the selected bands are vibrational features of the chemistry, not of the sensor; the novelty gate needs only benign reference spectra; and the class holdouts are defined on compounds, not on measurements.

This paper makes four contributions on a fully controlled compact-FTIR dataset with the processing chain held fixed. (i) We introduce an evaluation ladder of increasing distribution shift: closed-set replicate splits, leave-one-compound-out (LOCO), and leave-one-chemical-class-out (LCCO), defined in Section 3.3. Closed-set accuracy is uninformative for model selection, while the class holdouts dissociate sharply: PLS-DA and k-NN detect every spectrum of the two unseen classes; four other model families detect none. (ii) The study also contributes methodological controls. A carbon-attachment control verifies that generalization is not driven by trivial functional-group cues. A predefined robustness sweep demonstrates that the failures are intrinsic, not caused by tuning: increasing capacity improves the performance of the subspace model, but no configuration succeeds in rescuing the more flexible model families. (iii) Key-band discrimination: a blind, leakage-checked variable-importance analysis recovers the shared NO_2/ring bands. (iv) A Flag-on-Unknown (FoU) safety gate: spectra with no confident classifier assignment and no benign nearest-neighbour match are routed to inspection, so no unseen explosive is cleared as benign. This work is a controlled methodological study focused on training protocols and model inductive biases. It is not intended as an operational benchmark or a claim of deployment-ready performance.

2. DATA

Attenuated-total-reflectance (ATR) spectra (%R) were acquired with an Agilent Cary 630 compact FTIR spectrometer over 399.8–4000.4 cm^{-1} (3865 channels, $\approx$0.93 cm^{-1} spacing). Each physical sample was measured six times (replicates R1–R6) to capture measurement variability across repositioning and contact. This gives 198 spectra from 33 samples in 21 compound groups, where a group is a pure compound or an intimate mixture family. The explosive side comprises the pure organics DNT and TNT (nitroaromatic class) and PETN (nitrate-ester class); their three binary mixture families at mass ratios 1:3, 1:1, and 3:1; seven ternary compositions (1:1:1 and all permutations of 3:1:1 and 2:2:1); and nitroguanidine (nitramine class) and ammonium nitrate (AN, inorganic nitrate class), each pure only. Twelve benign confusants complete the set: Al, Mg, $KClO_3$, $KClO_4$, the KI/NaCl salt blend KiNaCl, the sodium carboxylate salt RCOONa, sugars, and foodstuffs. In total there are 126 explosive and 72 benign spectra. The four explosive classes differ in the bonding environment of the nitro oxygen: C–NO_2 on a ring (nitroaromatic), C–O–NO_2 (nitrate ester), N–NO_2 (nitramine), and ionic NO_3^- (inorganic nitrate). The benign set is deliberately heterogeneous and includes the inorganic oxidizers $KClO_3$/$KClO_4$

Table 1. Model Panel: Hyperparameters Fixed A Priori, Never Tuned

Model	Fixed values	Why
PLS-DA [10, 11]	12 components; threshold 0.5	standard rank; any rank ≥ 6 identical (Sec. 6)
PCA-LDA	20 PCs	conventional variance-retention choice
RBF-SVM	γ=scale, C=1; balanced weights	library defaults; weights for the 126/72 imbalance
k-NN	$k = 5$, Euclidean	textbook default
Random forest	300 trees; balanced weights	enough trees for stable votes
1-D CNN	conv 8/16 ch. (k=9), GAP, drop 0.3; Adam 10^{-3}, wd 10^{-3}, 80 ep.	1,265 params, capacity-matched to $n = 198$

(labelled benign per the dataset convention). This closes a shortcut in the unseen-AN fold: a detector that simply flags any unfamiliar inorganic salt would flag the benign chlorates as well, so detecting AN cleanly requires the genuine spectral contrast between the NO_3^- and ClO_3^-/ClO_4^- fundamentals.

3. METHODS

3.1. Preprocessing

All preprocessing is per-signature, so no statistic couples training and test data: (i) Hampel despiking (window 9; points >100 %R or $> 6\times$ the local median absolute deviation are interpolated); (ii) asymmetric least squares (AsLS) baseline removal [6, 7]; (iii) Savitzky–Golay first derivative [8] (window 15, order 2); (iv) standard normal variate (SNV) scaling [9]. The AsLS baseline $\mathbf{b}$ of a spectrum $\mathbf{z} \in \mathbb{R}^P$ minimizes the roughness-penalized criterion (1),

$$S(\mathbf{b}) = \sum_{i=1}^{P} w_i \left(z_i - b_i\right)^2 + \lambda \sum_{i=3}^{P} \left(\Delta^2 b_i\right)^2, \qquad (1)$$

with $w_i = p$ if $z_i > b_i$ and $1 - p$ otherwise ($\lambda = 10^5$, $p = 0.01$, ten iterations). For the band analysis (Section 7) we use baseline-corrected pseudo-absorbance $A = -\log_{10} R$, where $R = \mathrm{clip}(\%R, 1, 100)/100$ maps reflectance to the interval $(0, 1]$. Channel-wise standardization, the only data-estimated step, is fit inside every training fold.

3.2. Model panel

We evaluate one representative classifier from each major model family, with standard hyperparameters fixed in advance and never tuned on any evaluation split (Table 1). This design isolates inductive bias rather than tuning; the pre-specified sweep in Section 6 confirms no conclusion depends on these choices. All reported detection fractions carry exact Clopper–Pearson 95% confidence intervals. Each classifier operates at its native decision rule: a 0.5 threshold on the PLS-DA output, the natural decision boundary for LDA and SVM, majority vote for k-NN and the random forest, and a logit-zero threshold for the CNN.

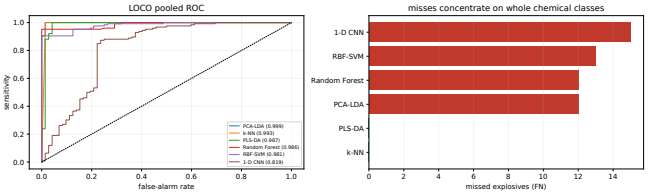


Fig. 1. LOCO evaluation (21 compound folds, pooled out-of-fold scores). Left: ROC curves with pooled AUC. Right: missed explosive spectra per model.

3.3. Evaluation ladder

Closed-set: each compound's six replicates are split into four for training and two for testing, exhaustively over all $\binom{6}{2} = 15$ test pairs, so every compound appears in training in every split. *LOCO* (leave-one-compound-out): 21 folds, one per compound group; all replicates and mixing ratios of the held-out group leave training together, and metrics are pooled over the 198 out-of-fold predictions. *LCCO* (leave-one-chemical-class-out): three class holdouts, constructed so that no holdout repeats a LOCO fold (Section 6). Both LOCO and LCCO remove samples, not features; they are unrelated to leave-one-covariate-out inference.

4. CLOSED-SET RECOGNITION

Under the randomized replicate splits, five of six models are at or near ceiling. Mean accuracy over the 15 splits is 1.000 (random forest), 0.999 (PCA-LDA), 0.995 (RBF-SVM), 0.985 (PLS-DA), and 0.979 (k-NN), with at most one missed explosive in the worst split. Only the CNN lags (0.865 ± 0.025; F1 0.900 ± 0.016). The ranking among the classical models is within noise, and nothing here predicts the failure structure that follows: closed-set evaluation is uninformative for model selection in this application.

5. LEAVE-ONE-COMPOUND-OUT: THE DISSOCIATION

Under LOCO (Table 2; Fig. 1), aggregate metrics remain high for all classical models. Where the models make errors is more informative than how often: the misses concentrate on two compounds. Nitroguanidine and AN are the only mixture-free representatives of their chemical classes, so holding either out removes the entire class from training. Their LOCO folds are therefore de facto class holdouts, and detecting them requires zero-shot transfer from the remaining chemistry. The models dissociate sharply at their native operating points, identically on both classes. PLS-DA and k-NN detect all 12 unseen-class spectra (6/6 nitroguanidine, 6/6 AN; exact 95% CI [0.74, 1.00]), while PCA-LDA, the random forest, the RBF-SVM, and the CNN detect none (0/12, [0.00, 0.26]). The remaining errors of the transferring models are benign false alarms (8 for PLS-DA, 6 for k-NN), concentrated on salts, sugars, and metal powders. Aggregate F1 spans only 0.946–0.977, yet within that band some models silently pass every nitramine and inorganic-nitrate sample while others catch all of them at the cost of a few false alarms. For safety-critical screening, the compound-level miss structure, not aggregate accuracy, is what matters.

6. LEAVE-ONE-CLASS-OUT: LIMITS AND ANATOMY

Three non-redundant holdouts. Nitroguanidine and AN already behave as class-level holdouts under LOCO, because each is the only

Table 2. LOCO Detection (21 Folds, Pooled; 126 Explosive / 72 Benign Spectra)

Model	Sensitivity [95% CI]	Specificity	F1	ROC-AUC	Missed explosives
PCA-LDA	0.905 [0.84, 0.95]	0.986	0.946	0.999	12 (6 nitroguanidine, 6 AN)
k-NN	1.000 [0.97, 1.00]	0.917	0.977	0.993	0
PLS-DA	1.000 [0.97, 1.00]	0.889	0.969	0.987	0
Random forest	0.905 [0.84, 0.95]	1.000	0.950	0.986	12 (6 nitroguanidine, 6 AN)
RBF-SVM	0.897 [0.83, 0.94]	1.000	0.946	0.981	13 (6 nitroguanidine, 6 AN, 1 mixture)
1-D CNN	0.881 [0.81, 0.93]	0.708	0.860	0.819	15 (6 nitroguanidine, 6 AN, 3 mixture)

Note: all decisions at the models' native operating points.

Table 3. LCCO: Three Non-Redundant Class Holdouts (Native Operating Points)

Held-out chemistry	Test n	Excl. n	Train. expl. n	PLS-DA	k-NN	PCA-LDA	RBF-SVM	RF	1-D CNN
Nitroaromatic (DNT, TNT, DNT/TNT)	30	78	18	0/30	0/30	0/30	0/30	0/30	0/30
Nitrate ester (PETN)	6	78	42	0/6	0/6	0/6	0/6	0/6	0/6
Both non-C-attached (nitramine + inorg. nitrate)	12	0	114	12/12	12/12	0/12	0/12	0/12	0/12

Note: RF, random forest; Excl., entangled mixed-family spectra excluded from training and test. Exact 95% CIs: 0/30 $[0, 0.12]$; 0/6 $[0, 0.46]$; 0/12 $[0, 0.26]$; 12/12 [0.74, 1].

mixture-free representative of its class. Separate LCCO folds would be redundant, so LCCO adds only three new splits (Table 3). Rows 1–2 hold out the multi-compound organic classes, which also appear in mixtures. Removing them excludes mixed-family spectra from both training and test (78 spectra), leaving a smaller residual explosive library of 18–42 spectra. None of the models detects the held-out class in either split: model structure cannot compensate for missing chemistry. The third split holds out both classes whose nitro oxygen is not bonded to carbon (N–NO_2 nitramine, ionic NO_3^-), leaving only carbon-attached chemistry in training. This split therefore acts as a carbon-attachment control, asking whether transfer survives when no non-carbon-attached chemistry remains. Despite this, PLS-DA and k-NN still detect all 12 held-out spectra. This suggests they may rely on N–O features shared across bonding environments. The result alone does not prove the mechanism, but Section 7 supports it: blind band selection recovers exactly those bands.

Not a tuning problem. Table 4 sweeps every model's main hyperparameters on this control. For the SVM and the random forest, no margin or partition setting ever produces reliable detection. The CNN remains unstable and never exceeds 6/12 across seeds, training lengths, learning rates, and widths up to 74.6k parameters. These models fail consistently, indicating that the limitation lies in the model families themselves rather than in their tuning. In contrast, PCA-LDA improves steadily with dimensionality: detection rises from 0/12 at 20 PCs to 11/12 at 60 and 12/12 at 80. PLS-DA succeeds once its rank spans the shared spectral structure (rank ≥ 6), and k-NN transfers for all reasonable k. This pattern is not a simple linear-versus-nonlinear distinction: PCA-LDA is linear and fails at low capacity, while k-NN is nonlinear and transfers reliably. The operative factor appears to be whether representation and decision rule are expressed in features shared across the chemistry. This is consistent with shortcut learning [12], underspecification [13], and open-set recognition [14].

Table 4. Robustness Sweep on the Carbon-Attachment Control: Detections /12 at Native Operating Points; All Settings Pre-Specified and Reported

Model (parameter)	Settings	Detected (/12)
RBF-SVM (γ)	$10^{-5} \ldots 10^{-3}$, scale	0, 0, 0, 0, 0, 0
Random forest (depth)	1, 2, 3, 5, 10, ∞	0, 0, 0, 0, 0, 0
1-D CNN (seed)	0–5	0, 0, 6, 0, 0, 0
1-D CNN (epochs)	20, 40, 80, 160, 320, 640	0, 0, 0, 0, 0, 2
1-D CNN (learning rate)	$3{\cdot}10^{-4} \ldots 10^{-1}$	0, 0, 0, 6, 0, 0
1-D CNN (width)	1, 2, 3, 4, 6, 8 ($\leq$74.6k par.)	0, 0, 0, 0, 0, 0
CNN minibatch ctrl. (seed)	0–5	2, 0, 2, 0, 1, 0
PCA-LDA (PCs)	10, 20, 40, 60, 80, 100	0, 0, 6, 11, 12, 12
PLS-DA (rank)	1, 4, 6, 12, 30	0, 6, 12, 12, 12
k-NN (k)	1, 3, 5, 11, 21	10, 12, 12, 12, 12

Note: Upper block: no setting transfers reliably. Lower block: transfer once the subspace or neighbourhood suffices. γ and learning-rate grids log-spaced; below $3{\cdot}10^{-4}$ the CNN does not converge in 80 epochs. Minibatch control: batch-32 Adam, balanced loss, 300 epochs.

7. KEY-BAND DISCRIMINATION

To identify the channels that drive the explosive/benign contrast, we apply supervised band selection [15]. We compute the PLS-DA variable importance in projection (VIP) [16] on the pseudo-absorbance representation, using $k = 12$ latent variables. Let P be the number of channels, and let $\mathbf{w}_a$, $\mathbf{t}_a$, and q_a denote the unit-norm weight vector, score vector, and y-loading of latent variable a. The VIP score

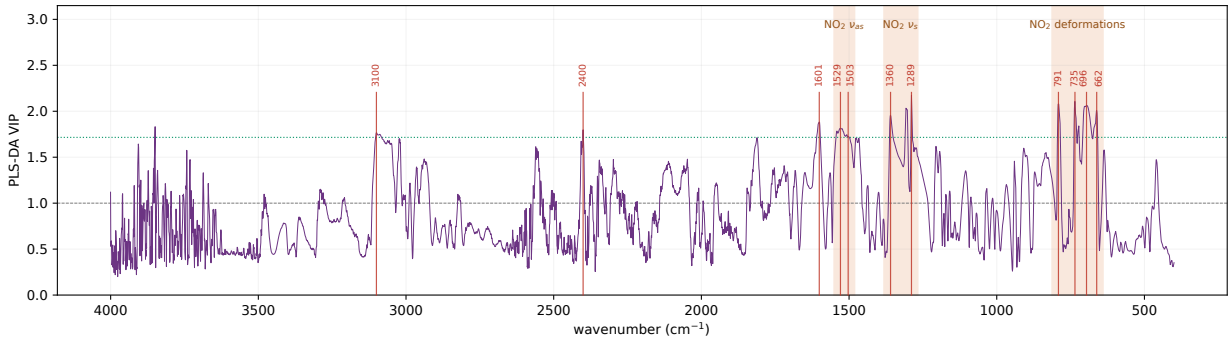


Fig. 2. PLS-DA VIP spectrum ($k = 12$ latent variables, pseudo-absorbance; post-hoc). Shading highlights the peaks attributed to NO_2 modes [17]. Red: blindly selected peaks. Dotted: 95th-percentile cutoff (1.715).

of channel j is then given by (2):

$$\mathrm{VIP}_j = \sqrt{P \, \frac{\sum_{a=1}^{k} q_a^2 \, \mathbf{t}_a^\top \mathbf{t}_a \, w_{ja}^2}{\sum_{a=1}^{k} q_a^2 \, \mathbf{t}_a^\top \mathbf{t}_a}}. \tag{2}$$

This analysis is a post-hoc interpretation and is not part of any detector.

Peaks are selected blind: local maxima of the VIP curve above the 95th percentile, at least 25 cm^{-1} apart, with a positive explosive-minus-benign class-mean difference. Eleven peaks qualify. Eight are attributed to NO_2 modes [17] (Fig. 2): 662, 696, 735, and 791 cm^{-1} to the deformations, 1289 and 1360 cm^{-1} to the symmetric stretches, and 1503 and 1529 cm^{-1} to the asymmetric stretch. The remaining three peaks are 1601 cm^{-1}, which also matches the aromatic C=C position, the aromatic C–H stretch at 3100 cm^{-1}, and 2400 cm^{-1}. The last peak coincides with the atmospheric CO_2 absorption band and is therefore most likely due to residual CO_2 in the instrument or imperfect background subtraction, rather than the sample itself.

To verify that the selection is robust and free from data leakage, the VIP scores are recalculated on the 186-spectrum training set of the carbon-attachment control (Table 3, row 3). All eight attributed peaks remain above the 93rd percentile, so the selection does not rely on the unseen classes. Importantly, these stable NO_2 bands are consistent with the zero-shot transfer to the unseen classes observed in Section 6. In contrast, no peak is selected near 1640 cm^{-1}, corresponding to the O–NO_2 asymmetric stretch (ν_{as}), despite this being the most intense spectral band of PETN. The VIP score at this wavenumber reaches only the 76th percentile, below the selection threshold, in both fits. Because this band is specific to nitrate esters, it distinguishes the PETN subset rather than contributing to the broader discrimination between explosive and benign classes.

8. FLAG-ON-UNKNOWN: AN OPEN-SET REJECT OPTION

Zero-shot transfer presupposes shared spectroscopy; a chemistry sharing no bands with the training set (e.g., peroxide explosives) could evade any supervised representation. To mitigate this vulnerability, we use the Flag-on-Unknown (FoU) framework as a target-agnostic anomaly detector and novelty gate [14, 18]. In this decision chain (Fig. 3), spectra flagged by the classifier raise an alarm, and only the remainder reach the gate, which routes unfamiliar spectra to "unknown → inspect" instead of silently clearing them.

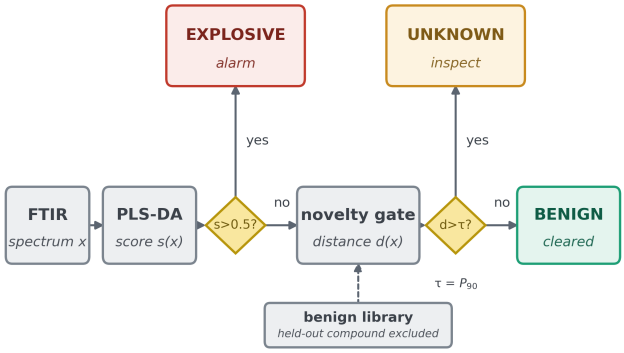


Fig. 3. Proposed Flag-On-Unknown (FoU) decision chain.

The anomaly gate evaluates incoming samples using two alternative distance-based scores. The first, $d(x)$, is the mean Euclidean distance to the five nearest benign training spectra, computed on standardized channels with the training benigns scored leave-one-out. The second is the reconstruction error of a PCA(10) model of the benign library. Both are thresholded at the 90th percentile (P_{90}) of the training-benign scores, and no explosive data enter the gate.

We test the gate on the three LCCO holdouts of Section 6, where it succeeds in routing every held-out spectrum to inspection, even the organic classes that all six classifiers miss (Table 5). Its cost is the benign inspection burden, the fraction of genuine benigns it also flags. The distance gate flags fewer benigns, so it is the better operating choice, while the reconstruction gate is the more threshold-robust alternative. This burden is the price of a zero-miss reject option on a small benign library.

9. LIMITATIONS AND FUTURE WORK

(i) Two unseen classes, one instrument, laboratory conditions: the dissociation is a controlled case study, not evidence of general cross-chemistry coverage. (ii) Transfer presupposes shared spectroscopy; FoU is the fallback where no bands are shared. (iii) Each unseen class is a single specimen, so its six replicates capture measurement variability, not chemical diversity. The per-class intervals are therefore optimistic, and wider sampling would consolidate the contrasts. (iv) Twelve benign compounds bound the specificity and burden estimates. (v) The CNN result is established for compact 1-D CNNs trained from scratch at this data scale; pretrained or physics-informed deep spectral models remain untested. Ongoing

Table 5. FoU on the LCCO Holdouts (τ = P90 of Training-Benign Novelty)

Gate	Nitroarom.	Nitr. ester	Non-C	Burden
kNN-5 distance	30/30	6/6	12/12	27/72
PCA-10 reconstruction	30/30	6/6	12/12	36/72
best supervised base	0/30	0/6	12/12	—

Note: Routing fractions give the proportion of held-out spectra intercepted and sent to inspection. The burden is the out-of-sample benign false-alarm rate under the nested leave-one-benign-compound-out rotation. Last row: best of the six classifiers at native operating points.

work transfers the evaluation ladder to the reflective domain (350–2500 nm): ASD FieldSpec 4 point spectra and airborne HySpex Mjolnir VS-620 VNIR–SWIR imagery. Two questions carry over: whether the dissociation persists when mid- and long-wave IR fundamentals give way to SWIR overtone/combination bands, and whether spectral unmixing enables per-pixel abundance mapping for standoff screening.

10. CONCLUSION

This controlled study examined how spectroscopic explosive detectors behave on chemistry outside their training library. Three results stand out. First, closed-set accuracy is uninformative, with every model family near ceiling on familiar compounds. Second, under class-level holdout the panel divides cleanly. The subspace and instance-based models detect every spectrum of two unseen explosive classes, while the discriminant, margin, partition, and convolutional families detect none. The pre-specified sweep shows this divide is intrinsic, not a tuning artifact, and the carbon-attachment control ties the transfer to shared nitro-group chemistry. Third, blind band selection independently recovers those same NO_2 bands. The Flag-on-Unknown gate covers the remaining risk, routing spectra that match nothing in the library to inspection rather than clearing them. Class-level holdout therefore belongs in every evaluation, and detector choice should weight inductive bias over in-distribution margins.

REPRODUCIBILITY AND ETHICS

All analyses were implemented in Python (Python 3.13.5; scikit-learn 1.9.0; torch 2.12.1). The dataset and source code are available from the corresponding author, and their release depends on institutional clearance. This work supports defensive screening and contains no synthesis, formulation, or operational information.

ACKNOWLEDGEMENT

This work was supported under project DFR-SpectrEx-DAP25/13. The authors would like to thank Bart Simoens and Joris Flebus of the Department of Chemistry, Royal Military Academy, for their invaluable assistance in providing the chemical materials and their technical guidance on FTIR instrumentation.

11. REFERENCES

[1] M. López-López and C. García-Ruiz, "Infrared and Raman spectroscopy techniques applied to identification of explosives," *TrAC Trends Anal. Chem.*, vol. 54, pp. 36–44, 2014.

[2] L. C. Pacheco-Londoño, W. Ortiz-Rivera, O. M. Primera-Pedrozo, and S. P. Hernández-Rivera, "Vibrational spectroscopy standoff detection of explosives," *Anal. Bioanal. Chem.*, vol. 395, no. 2, pp. 323–335, 2009.

[3] I. M. van Damme *et al.*, "Rapid and on-scene chemical identification of intact explosives with portable near-infrared spectroscopy and multivariate data analysis," *Sensors*, vol. 23, no. 8, art. 3804, 2023.

[4] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, 2016.

[5] J. Yue, L. Fang, and M. He, "Spectral–spatial latent reconstruction for open-set hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 31, pp. 5227–5241, 2022.

[6] P. H. C. Eilers, "A perfect smoother," *Anal. Chem.*, vol. 75, no. 14, pp. 3631–3636, 2003.

[7] P. H. C. Eilers and H. F. M. Boelens, "Baseline correction with asymmetric least squares smoothing," Leiden University Medical Centre, Tech. Rep., 2005.

[8] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.

[9] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Appl. Spectrosc.*, vol. 43, no. 5, pp. 772–777, 1989.

[10] M. Barker and W. Rayens, "Partial least squares for discrimination," *J. Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.

[11] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemom. Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, 2001.

[12] R. Geirhos *et al.*, "Shortcut learning in deep neural networks," *Nat. Mach. Intell.*, vol. 2, pp. 665–673, 2020.

[13] A. D'Amour *et al.*, "Underspecification presents challenges for credibility in modern machine learning," *J. Mach. Learn. Res.*, vol. 23, no. 226, pp. 1–61, 2022.

[14] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1757–1772, 2013.

[15] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, 2019.

[16] I.-G. Chong and C.-H. Jun, "Performance of some variable selection methods when multicollinearity is present," *Chemom. Intell. Lab. Syst.*, vol. 78, no. 1–2, pp. 103–112, 2005.

[17] R. W. Beal and T. B. Brill, "Vibrational behavior of the $-NO_2$ group in energetic compounds," *Appl. Spectrosc.*, vol. 59, no. 10, pp. 1194–1202, 2005.

[18] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, 2014.