Comment on "Proteoliposomes on 2D-MoS_2 Plasmonic Nanocavities for Enhanced Raman Spectroscopy with Machine Learning-based Identification and Classification"

XinYu Liu*

School of Mathematics and Physics, Hebei University of Engineering, Handan 056038, npj 2D Materials and Applications 2025, 9, 115. DOI: 10.1038/s41699-025-00636-3

The integration of engineered lipid vesicles with surface-enhanced Raman spectroscopy (SERS) and supervised machine learning represents a productive strategy for label-free molecular profiling of nanoparticle biomarkers. The recent paper by Shiekh et al.1 describes a library of synthetic proteoliposomes functionalized with cancer-associated surface markers and a 2D molybdenum disulfide (MoS_2) plasmonic nanocavity microchip (MoSERS) for SERS classification. The authors report Random Forest Classifier (RFC) and Support Vector Machine (SVM) test accuracies of 82% and 76%, respectively, with area under the ROC curve (AUC) values of approximately 0.97–0.98. While the experimental platform and its concept are noteworthy, this comment identifies several methodological concerns — encompassing biomarker identification, machine learning evaluation practice, enhancement factor definition, and the impact of partial protein incorporation—whose resolution would substantially strengthen the reproducibility and comparability of the reported findings.

A straightforward but consequential discrepancy is present in the paper. The Introduction states that "EGFR and integrins α6β1 and αvβ5" were reconstituted into POPC vesicles. However, in the Results section, the Methods, and all figures, the second integrin is consistently designated α6β4, not α6β1. These are structurally and functionally distinct proteins. The α6β4 integrin contains a uniquely large cytoplasmic β4 subunit of approximately 1000 amino acids and localizes to hemidesmosomes,2 whereas α6β1 is a conventional integrin involved in laminin-mediated cell adhesion and migration. The two integrins carry different charges, different extracellular domain architectures, and would be expected to yield distinct SERS signatures. Because the machine learning classification hinges on spectral discrimination between these surface markers, clarity about which integrin was used is essential for assessing and reproducing the work. This inconsistency should be resolved in a correction.

Several aspects of the machine learning analysis warrant discussion. The total spectral dataset comprises 75–85 spectra per class, of which 80% (approximately 60–68 spectra) are used for training and the remaining 20% (approximately 15 – 17 spectra) for testing. Performance estimates derived from such small held-out test sets exhibit large variance.3 While the authors appropriately report results across five independent train/test splits, the boxplots in Figures 4f and 4g confirm substantial run-to-run variability, particularly for the SVM. The use of stratified 5-fold cross-validation for hyperparameter selection is good practice, but the cross-validation folds and the final test set are not cleanly separated, raising the possibility of indirect data leakage in hyperparameter selection.4 Collecting additional spectra and adopting a nested cross-validation scheme4 would yield more reliable performance estimates.

More critically, both the RFC and SVM models achieve 100% accuracy on the training data (Fig. 4f, 4g). Perfect training accuracy, combined with test accuracy of approximately 76 – 82%, signals a substantial generalization gap consistent with overfitting. While RFC is an ensemble method that tends to generalize better than individual decision trees, the small training set size and high spectral dimensionality (wavenumber range 100 – 3000 cm^{-1}) predispose it to overfitting. The Elements of Statistical Learning5 and related references discuss the relationship between sample size, feature dimensionality, and overfitting in ensemble classifiers in detail.

Finally, the comparison is limited to RFC versus SVM. These two classifiers are not benchmarked against simpler baselines such as linear discriminant analysis (LDA) or κNN, nor against modern deep learning approaches that have demonstrated strong performance on SERS spectral data.6 Seifert7 has shown that RFC performance on SERS data depends substantially on preprocessing choices; the sensitivity of the reported accuracy to baseline correction and normalization parameters is not explored. Without comparison to established methods, it remains difficult to assess whether the observed accuracy ceiling of ~82% reflects a fundamental limitation of the MoSERS spectral data or an artifact of the classifier choice and dataset size.

The authors report an enhancement factor (EF) of 54.45, derived by comparing Raman intensities collected from nanocavity regions against adjacent flat silver film regions on the same substrate. While this metric is informative for characterizing the confinement advantage of the nanocavity geometry, it does not

correspond to the standard analytical SERS enhancement factor as defined in the literature. The widely adopted definition compares the SERS signal per molecule to the normal Raman signal per molecule from the same analyte in bulk solution, explicitly accounting for the difference in the number of molecules in the probe volume.8 A nanocavity-to-flat-film intensity ratio underestimates the true analytical EF because the flat silver film itself supports localized surface plasmon resonance and provides non-negligible signal enhancement relative to a non-metallic surface. Reporting the standard analytical EF would allow the MoSERS platform to be meaningfully compared with other SERS substrates described in the literature, and is important context for the claimed sensitivity.

Dual-channel fluorescence co-localization analysis revealed that approximately 34% of particles showed a protein signal, meaning that the nominally "proteoliposome" preparations contain roughly two-thirds bare, protein-free POPC liposomes. This heterogeneity has direct implications for the machine learning experiment. If SERS spectra were acquired by scanning nanocavities loaded with the ensemble population, then each proteoliposome spectral class contains a majority contribution from unmodified POPC vesicles that are spectrally indistinguishable from the "liposome only" control class. The classifier would thus be operating partly on irrelevant background spectra rather than on biomarker-specific signal, which could inflate apparent accuracy or, conversely, suppress it.

Determining protein reconstitution efficiency and its dependence on lipid-to-protein ratio is well established in the literature.9,10 Restricting SERS measurements to vesicles identified as protein-positive through co-localization, or increasing reconstitution efficiency before building the spectral library, would substantially clarify the source of spectral discrimination. The authors' pneumatic confinement device already provides a route to single-vesicle interrogation11; applying it in combination with fluorescence gating would directly address this concern.

The MoSERS platform and the proteoliposome engineering strategy described by Shiekh et al. represent creative contributions to the field of SERS-based vesicle profiling.1,11,12 Correcting the biomarker nomenclature error, adopting the standard analytical EF definition, expanding the spectral training set with nested cross-validation, and accounting for partial protein reconstitution in the spectral analysis would provide a stronger and more reproducible evidential basis for the classification claims, and would facilitate direct comparison with other SERS-based vesicle characterization platforms described in the literature.6,12

AUTHOR INFORMATION

Corresponding Author
*E-mail:2208936483@qq.com.
Notes
The authors declare no competing financial interest.

References

1. Shiekh, S.; Zhou, Y.; del Real Mata, C.; Jalali, M.; McCormack-Ilersich, J.; Hosseini, I. I.; Liu, Z.; Pessoa, M. A. S.; Hamidi, S. V.; Montermini, L.; Rak, J.; Reisner, W.; Mahshid, S. Proteoliposomes on 2D-MoS_2 Plasmonic Nanocavities for Enhanced Raman Spectroscopy with Machine Learning-based Identification and Classification. npj 2D Mater. Appl. 2025, 9, 115.
2. Sonnenberg, A.; Linders, C. J. T.; Daams, J. H.; Kennel, S. J. The α6β1 (VLA-6) and α6β4 Protein Complexes: Tissue Distribution and Biochemical Properties. J. Cell Sci. 1990, 96, 207–217.
3. Varoquaux, G. Cross-validation failure: Small sample sizes lead to large error bars. NeuroImage 2018, 166, 518–534.
4. Filzmoser, P.; Liebmann, B.; Varmuza, K. Repeated double cross validation. J. Chemom. 2009, 23, 160–171.
5. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed.; Springer: New York, 2009.
6. Ho, C.-S.; Jean, N.; Hogan, C. A.; Blackmon, L.; Jeffrey, S. S.; Holodniy, M.; Banaei, N.; Bhatt, D. L.; Boxer, S. G.; Bhatt, D. Rapid Identification of Pathogenic Bacteria Using Raman Spectroscopy and Deep Learning. Nat. Commun. 2019, 10, 4927.
7. Seifert, S. Application of Random Forest Based Approaches to Surface-Enhanced Raman Scattering Data. Sci. Rep. 2020, 10, 5436.
8. Le Ru, E. C.; Blackie, E.; Meyer, M.; Etchegoin, P. G. Surface Enhanced Raman Scattering Enhancement Factors: A Comprehensive Study. J. Phys. Chem. C 2007, 111, 13794–13803.
9. Rigaud, J.-L.; Lévy, D. Reconstitution of Membrane Proteins into Liposomes. Methods Enzymol. 2003, 372, 65–86.
10. Veit, S.; Paweletz, L. C.; Pomorski, T. G. Determination of Membrane Protein Orientation upon Liposomal Reconstitution down to the Single Vesicle Level. Biol. Chem. 2023, 404, 647–661.
11. Jalali, M.; Lu, Y.; del Real Mata, C.; Rak, J.; Mahshid, S. Nanoscopic Technologies toward Molecular Profiling of Single Extracellular Vesicles for Cancer Liquid Biopsy. Appl. Phys. Rev. 2025, 12, 011312.
12. Jalali, M.; Mahshid, S. et al. MoS_2-Plasmonic Nanocavities for Raman Spectra of Single Extracellular Vesicles Reveal Molecular Progression in Glioblastoma. ACS Nano 2023, 17, 12052–12071.