Journal of Biophotonics

JOURNAL OF BIOPHOTONICS

RESEARCH ARTICLE

External Validation of Raman Spectroscopy for Lyme Disease Diagnostics

Isaac D. Juárez[1,2] | Aidan P. Holman[1,2] | Elizabeth J. Horn[3] | Artem S. Rogovskyy[4] | Dmitry Kurouski[1,2]

[1]Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas, USA | [2]Interdisciplinary Faculty of Toxicology, Texas A&M University, College Station, Texas, USA | [3]Lyme Disease Biobank, Portland, Oregon, USA | [4]Department of Pathobiology and Diagnostic Investigation, College of Veterinary Medicine, Michigan State University, East Lansing, Michigan, USA

Correspondence: Artem S. Rogovskyy (rogovsky@msu.edu) | Dmitry Kurouski (dkurouski@tamu.edu)

Received: 22 November 2024 | **Revised:** 3 February 2025 | **Accepted:** 6 February 2025

Funding: This study was supported by the Bay Area Lyme Foundation and the Texas A&M AgriLife Research.

Keywords: *Borreliella burgdorferi* | chemometrics | Lyme disease | PLS-DA | Raman spectroscopy

ABSTRACT

Lyme disease (LD), caused by *Borreliella burgdorferi*, is the most common tick-borne illness in the United States, yet early-stage diagnosis remains challenging due to the limitations of current serological diagnostics. Raman spectroscopy (RS), paired with partial least squares discriminant analysis (PLS-DA), showed promise as an alternative diagnostic tool. Using RS, we analyzed 107 coded human blood samples (42 LD-positive and 65 LD-negative) obtained from the Lyme Disease Biobank. PLS-DA models showed nearly perfect internal validation performance with a sensitivity and specificity of 97.1% and 100.0%, respectively, indicating robust predictive capabilities. External validation of the developed chemometrics model with 80/20 training/validation split of all spectra gave true positive rates of 92.7% and 87.3% for serological positive and negative spectra, respectively. These findings highlight the potential of RS as a rapid and noninvasive diagnostic platform for LD, particularly when integrated with machine learning.

1 | Introduction

Lyme disease (LD) is a bacterial infection caused by spirochete *Borreliella burgdorferi* (*B. burgdorferi*) primarily in the temperate regions of North America. The disease was first recognized in New England in the 1970s, but it is now prevalent down to the mid-Atlantic and Great Lakes regions. It is the most prevalent tick-borne illness in the United States [1, 2]. The pathogen is vectored by *Ixodid* ticks. The disease progresses through three stages, with the telltale sign being erythema migrans (colloquially referred to as the bulls-eye rash); however, the slow progression of the disease combined with flu-like symptoms makes it often difficult to diagnose [3–5]. The disease is readily treated with antibiotics but can pose prolonged symptoms and chronic status if not treated appropriately within the first stages [6].

Currently, three LD diagnostic approaches have been cleared by the FDA: a two-tiered serology system using an enzyme immunoassay (EIA) followed by Western blot, a modified two-tiered serology system using two EIAs, or the new stand-alone iDart Lyme IgG ImmunoBlot Kit [7–9]. There are several limitations, however, notably low sensitivity during the early stage, even for LD patients with erythema migrans [10]. Similarly, cross-reactivity and background seropositivity complicate findings. Most importantly, however, serological tests also cannot distinguish between active and past infections or reinfections [11]. These problems underscore a major need for more accurate and robust diagnostic methods, especially in early-stage disease. Current research focuses on direct and indirect approaches to this, such as cytokine-based immunoassays, the detection of biomarkers from serum protein patterns, or spectroscopic techniques [12–14].

Recently, two studies on Raman spectroscopy (RS) for LD diagnostics were conducted by our laboratories [15, 16]. In our previous studies, the researchers collected and analyzed spectra of blood from mice infected with three strains from all the major clades of *B. burgdorferi* found within the United States as well as infected with two European genospecies *Borreliella afzelii* and *Borreliella garinii*. Our results had true positivity rates ranging from 83% to 100% accuracy of identification when using partial least squares discriminant analysis (PLS-DA). Additionally, we collected spectra of 90 human blood samples. These results indicate the major potential RS has as an LD diagnostic tool when paired with machine learning algorithms. The major advantage of using RS for diagnostics is the rapid detection of metabolic patterns present within a blood sample [17, 18]. RS is a form of vibrational spectroscopy that works by using a laser to probe vibrational bonds present with a sample, thereby providing information about the chemical composition of a sample. Changes within spectra between serologically negative and positive samples can then be used to diagnose LD. It should be noted that in addition to LD, RS was used to diagnose numerous diseases in plants, animals, and humans. For instance, Farber and Kurouski showed that RS could be used to detect fungal infections in corn [19], while Sanchez et al. demonstrated that changes in the plant biochemistry caused by the gram-negative *Candidatus* Liberibacter spp. bacteria could be used to detect and identify citrus greening disease [20]. Giuseppe et al. found that RS could be used to differentiate between healthy and Leishmania-infected dogs [21], whereas Vyas recently reported that changes in saliva probed by RS could be used to diagnose Sjögren's disease [22].

The approach presented in the current study is an expansion of our previous investigations [15, 16], which is focused on assessing our chemometrics model's capability when used with external validation of unidentified human blood samples. Specifically, we have collected spectra from a new set of 107 human blood samples obtained from anonymous donors. We built and evaluated three PLS-DA models that were first constructed using the spectra previously generated from the 90 human blood samples in our earlier study [16] and then refined by analyzing an additional set of human blood samples in the present investigation. These models were then either externally or internally validated against the second portion of unidentified human blood samples. We focused on PLS-DA due to the broad acceptance of this supervised classification method in the field of spectroscopy, as well as its low demand for computational resources. It should be noted that in comparison to other supervised classification algorithms, PLS-DA typically performs similarly or better than linear discriminant analysis (LDA) or soft independent modeling by class analogy (SIMCA) [23].

2 | Methods

2.1 | Blood Acquisition

A total of 107 human whole blood samples (EDTA tubes) were provided by the Lyme Disease Biobank (LDB) (Table S1) [24]. Of the 107 specimens, 42 samples originated from LD patients, whose diagnosis was serologically confirmed by the two-tier system or by 2 positive EIAs (LD-confirmed samples), and the other 65 samples represented LD-free individuals, who tested negative by serology (controls). First, all 107 samples were coded by the LDB to mask their LD status (positive and negative control samples), which allowed us to perform the blind study in two steps. First, 26 specimens, which consisted of 7 LD-confirmed and 19 control samples, were analyzed, and the results were submitted to the LDB for decoding. In our second blind study, the remaining still-unrevealed 81 specimens (35 LD-confirmed and 46 control samples) were subjected to RS, after which the results were submitted to the LDB for decoding.

2.2 | Spectroscopy

For use in RS, blood samples were thawed and vortexed before 50 μL were spread onto aluminum foil-wrapped microscope slides and scanned directly wet [16]. Between 30 and 60 spectra were acquired from all 107 human blood samples. Spectra for the study were collected using a Nikon inverted confocal microscope (model TE-2000 U) equipped with the 20× dry Nikon objective; spectral acquisition time was 30 s per spectrum. Spectra were acquired during 6 months. A 785 nm solid-state laser was directed toward the sample, passed through a 50/50 beam splitter, and then collected using backscattering geometry, with the Rayleigh scattering traveling through a long-pass filter (Semrock, LP03-785RS-25) before finally entering an IsoPlane-320 spectrometer (Princeton Instruments) equipped with a 600 groove/mm grating. The collected light lastly entered the CCD detector (PIX-400BR). The laser power was set to 8 mW for all acquisitions.

2.3 | Chemometrics

PLS_toolbox (Eigenvector Research Inc) was used in MATLAB to perform all machine learning algorithms. All spectra were baselined using Automatic Weighted Least Squares with an eighth-order polynomial and then background subtracted. This differs from the preprocessing done by Goff et al, who used a sixth-order polynomial [16]. Validation spectra were also smoothed using the Savitzky–Golay filter. The model spectra were averaged in groups of 2–12 and evaluated according to PLS-DA internal validation to improve the signal-to-noise ratio. The two groupings' averages with the best true positive rates (TPRs) for each model were then kept as preliminary Models A, B, and C. These six models were then compared against 81 unidentified human blood samples as external validation, and the three best models were kept as final Models A, B, and C. From binary comparison with the 81 blood specimens, every sample was assigned a value between 0 and 1 by each model. This value was used to determine if the model predicted a sample as positive or negative. After model completion, the status of each of the 81 samples was revealed, and all models were calculated for true positives, true negatives, false positives, and false negatives. All remaining model statistics were performed via formulas in Microsoft Excel.

3 | Results

3.1 | Raman Spectra

The Raman spectra collected from the blood contained peaks related primarily to heme (562, 676, 752, 962, 1226, 1376, and 1532 cm^{-1}), protein (1002, 1172, 1275, and 1340 cm^{-1}), and other

aromatic molecules (1604 and 1622 cm^{-1}). These peaks closely corresponded with the Raman spectra acquired from murine blood in our previous experiments [15, 16]. Small visual changes between the confirmed positive blood and the serologically negative blood could be observed throughout the spectra (Figure 1). Averaged Raman spectra of human blood samples with the mean and standard deviations for serologically negative or positive status are shown in Figure S1. The largest changes were increases in the protein-related peaks (1275 and 1340 cm^{-1}) correlating with positive infection, an increase in the heme-related peak (1376 cm^{-1}), and a decrease in the heme-related peak (562 cm^{-1}), both correlated again with positive infection. Previously reported analysis of metabolic profiles of blood obtained from early-stage LD patients using mass-spectroscopy revealed changes in the concentration of over 30 different biomolecules, including proteins, lipids, and their derivatives such as cholesterol, cholesteryl acetate, diacylglycerol, phospholipids, sphingolipids, and triglycerides [25].

3.2 | Model Construction

Using our preliminary database of human blood scans, along with the new scans of the 107 human blood samples obtained in the present study, we have built three PLS-DA models to analyze RS's potential for accurate LD diagnosis (Figure 2). Model A was built first, using the 8198 spectra previously collected by Goff et al. [16]. This model was largely similar to the model reported, albeit with an alternative preprocessing of the spectra. Model B was built by adding in 1630 spectra from 26 of the now-classified samples to the spectra from Model A. Finally, Model C contained all 9828 spectra from Model B, along with 1785 spectra from the remaining 81 human blood samples. Models A and B had their performance externally validated against the unknown 81 blood samples, while Model C was only internally validated against the 81 samples due to being composed of spectra from all 107 samples but was briefly externally validated using an 80/20 split of the spectra comprising the model. Additionally, Models A and B were trimmed at 1700 cm^{-1} to reduce noise and improve model performance. Model C performed well and was left untrimmed.

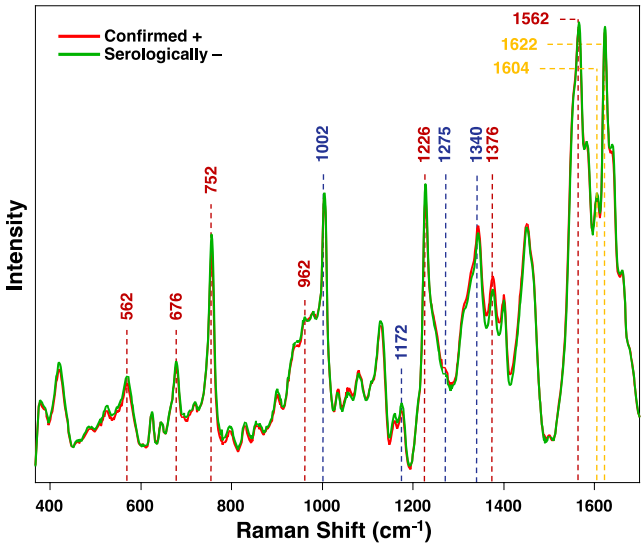

FIGURE 1 | Averaged Raman spectra of human blood samples with serologically negative or positive status for Lyme disease. Red bands indicate heme-related peaks, blue bands indicate protein-related peaks, and yellow bands indicate aromatic-related peaks.

To improve the signal-to-noise ratio of the Raman spectra, each group was randomly averaged in groups of 2–12 spectra according to their serological status (Figure 3, Table 1). For all models, as the grouping size increased, the TPR for spectral classification improved as determined by internal validation. TPR of spectra obtained from the negative controls trailed TPR of confirmed positive spectra in all instances. After TPR had been plotted, the two best-performing averaging sizes for all preliminary models were 9 and 11 spectra per averaging group. Both grouping sizes were externally validated for each preliminary model, and the stronger preliminary model was kept as our final Models A, B, or C. The three final models are based on the 9, 11, and 9 averaging groups, respectively, all of which had internally validated TPRs between 83% and 88%. Latent variables for each PLS-DA model were selected by plotting calculated and cross-validation classification errors for TPR against the number of latent variables. The latent variable with the lowest classification error and the lowest root mean square error of cross-validation (RMSECV) was then selected for the model. Final models of A, B, and C used 8, 15, and 13 latent variables, respectively.

3.3 | Model Evaluation

After selecting the final models, Models A and B were externally validated using PLS-DA, while Model C was only internally validated (Figure 4, Table 2). The results for Model A showed a large selection bias towards negative predictions,

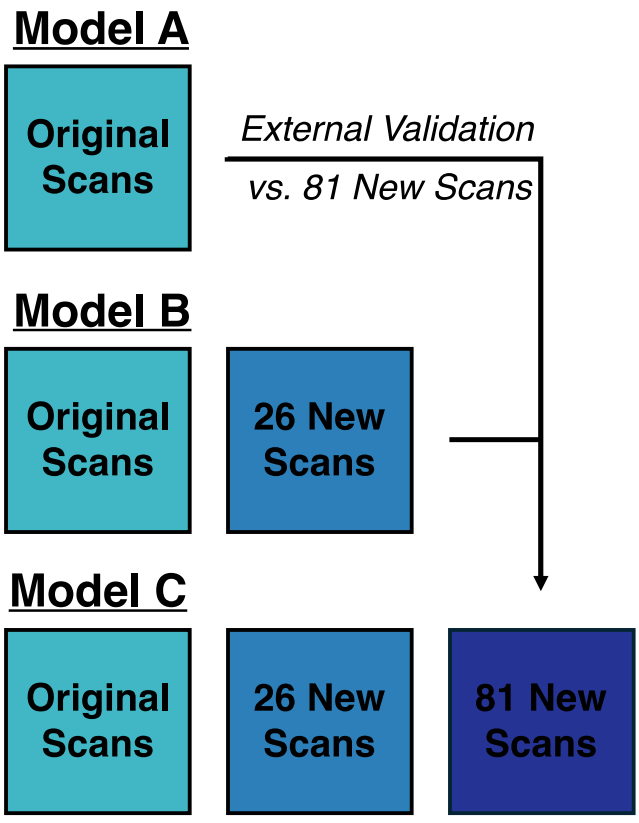

FIGURE 2 | Workflow of model design and evaluation.

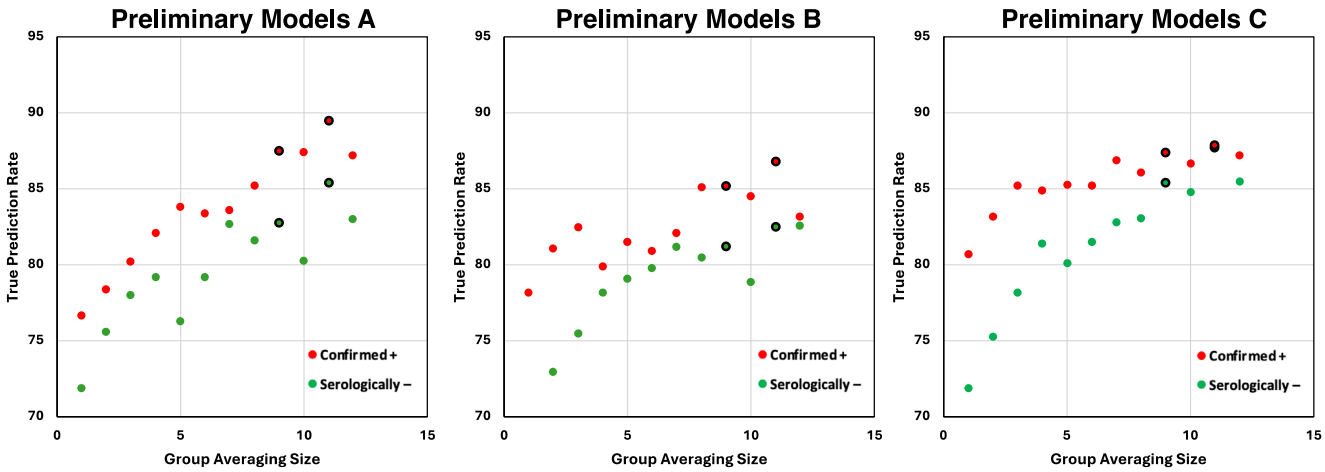

FIGURE 3 | Relationship between group averaging size and true prediction rates for preliminary models as determined by internal validation.

TABLE 1 | Relationship between group averaging size and true prediction rates for preliminary models as determined by internal validation.

	Preliminary Model A			Preliminary Model B			Preliminary Model C		
Group average size	Number of spectra	TPR (+)	TPR (–)	Number of spectra	TPR (+)	TPR (–)	Number of spectra	TPR (+)	TPR (–)
1	8198	77%	72%	9828	78%	69%	11 613	81%	72%
2	4098	78%	76%	4914	81%	73%	5805	83%	75%
3	2734	80%	78%	3275	83%	76%	3871	85%	78%
4	2050	82%	79%	2457	80%	78%	2904	85%	81%
5	1640	84%	76%	1967	82%	79%	2323	85%	80%
6	1367	83%	79%	1639	81%	80%	1937	85%	82%
7	1172	84%	83%	1405	82%	81%	1660	87%	83%
8	1025	85%	82%	1229	85%	81%	1452	86%	83%
9	912	88%	83%	1093	85%	81%	1291	87%	85%
10	820	87%	80%	984	85%	79%	1162	87%	85%
11	747	90%	85%	894	87%	83%	1057	88%	88%
12	684	87%	83%	819	83%	83%	969	87%	86%

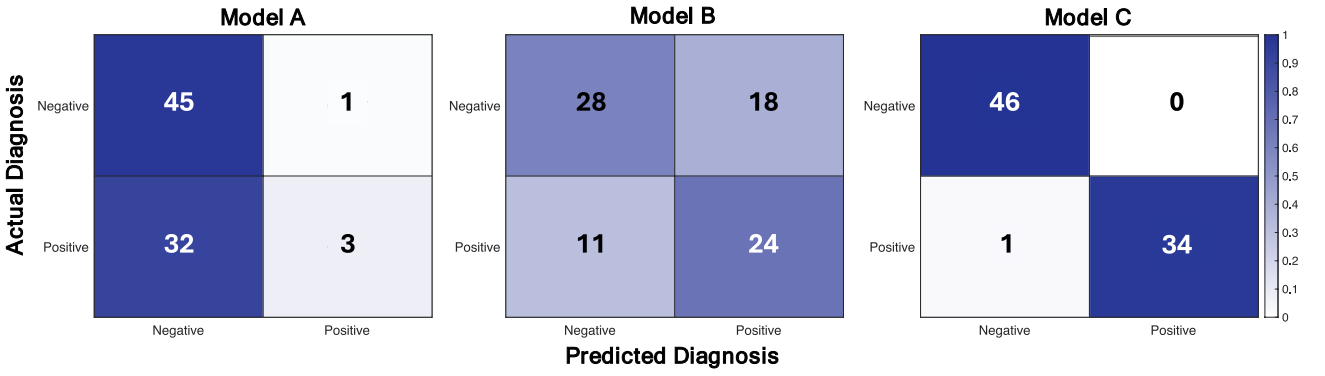

FIGURE 4 | Confusion matrices for Models A, B, and C. Color intensity represents classification accuracy given as the proportion of correct predictions relative to the actual diagnosis.

TABLE 2 | Evaluation parameters for Models A, B, and C.

	Model A	Model B	Model C
Sensitivity	8.6%	68.6%	97.1%
Specificity	97.8%	60.9%	100%
PPV	75%	57.1%	100%
NPV	58.4%	71.8%	97.9%
Accuracy	59.3%	64.2%	98.8%
F1 score	15.4%	62.3%	98.6%
Brier score	0.384	0.314	0.034
MCC	0.146	0.292	0.989
AUC	0.542	0.623	0.986

while Model B showed much improved performance, with overall more true positives and true negatives than false positives and false negatives. Model C prediction was almost perfect, except for a singular false negative. Using the values in the confusion matrix, we calculated the sensitivity (recall) and specificity for the models, which evaluated the models' ability to correctly identify positive samples and negative samples, respectively. The large bias toward negative predictions from Model A results in a high specificity of 97.8%; however, this was offset by the incredibly low sensitivity of 8.6%. Model B, being composed of more spectra, had better sensitivity, and Model C had even more, leading to the conclusion that the models' sensitivity would continue improving as more spectra were given to it for training.

The positive predictive value (PPV or precision) and negative predictive value (NPV) indicate the likelihood that a positive or negative prediction in the model is correct.

Model B shows a significant improvement in NPV compared to Model A; however, its PPV is lower as a lack of confirmed positive assignments in Model A artificially inflated its PPV. Both values were near 100% in Model C. The accuracy parameter provides us with an understanding of the total correctness of the model across both positive and negative predictions. Here, we see clear improvements in model performance across Model A to Model B to Model C, indicating that the model becomes more accurate as it trains on more spectra.

In clinical diagnostics, both sensitivity and PPV are important metrics, and the F1 score combines these two values into one metric. The low sensitivity of Model A is reflected in its F1 score of 15.4%, but in Models B and C, the F1 score is at 62.3% and 98.6%, reflecting the model's much greater diagnostic capabilities. The Matthews correlation coefficient (MCC) measures the model's capabilities when accounting for all true and false positives and negatives. Neither Model A nor Model B performed very well in this regard, likely due to the higher amounts of false positives and false negatives. Finally, the Brier score provides insight into the model's confidence in binary classification assignment, with a score of 0 indicating complete confidence. From this, we can understand that Models A and B had a similar amount of confidence (0.384 and 0.314) but that Model C had the strongest confidence (0.034) when making classification assignment.

When looking into how the model classifies spectra, we can measure the AUC (area under the curve) of the ROC (receiver operating characteristic) curve. By plotting the models' TPR and FPR across threshold levels from 0 to 1, we can evaluate the effectiveness of the model in class discrimination compared to random chance in what is called the ROC curve (Figure 5). The results of this curve show that the discrimination abilities of Model A are only slightly above random chance, while Model B has significantly better ability. Model C's ROC curve shows that the model had an almost near-perfect discrimination ability. To quantify these curves, we calculated the AUC for the models. These values reflect the previous assessment about the models' different discrimination abilities.

Finally, a major advantage of PLS-DA is the ability to plot loadings or latent variables from the model to understand how a decision is being made about class assignment (Figure 6). A latent variable analysis (LVA) plot corresponding to the PLS-DA loadings for Model C is shown in Figure S2. Based on the first three latent variables for Model C, we can see that intensity at the 752, 1002, 1226, 1340, and 1376 cm^{-1} peaks, as well as the 1562–1622 cm^{-1} region; all play key roles in the model's assignment of classes. Taken together, these three latent variables constitute 78.76% of the model's variance within class distinction.

4 | Discussion

The results of this study align well with our previous studies and help visualize them in the context of LD diagnostics. The PLS-DA loadings highlight that changes in heme and protein content heavily contribute to the model's classification capabilities, but that changes are not specific to any one metabolite.

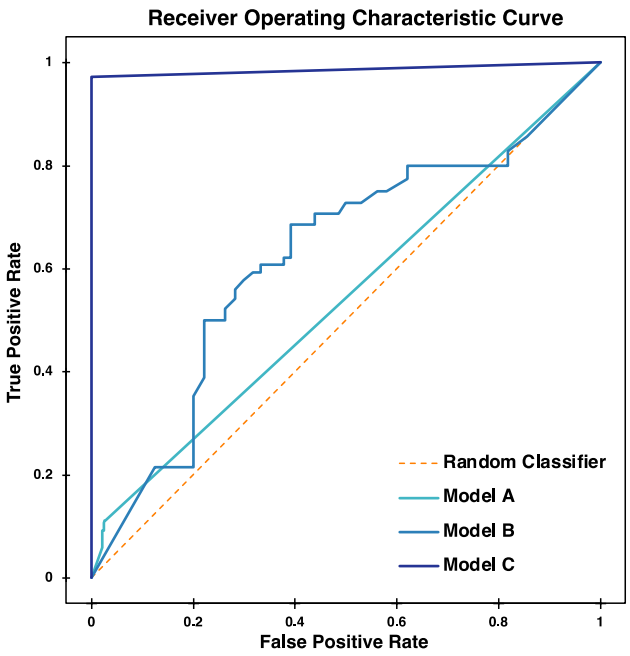

FIGURE 5 | Receiver operating characteristic curve for Models A, B, and C.

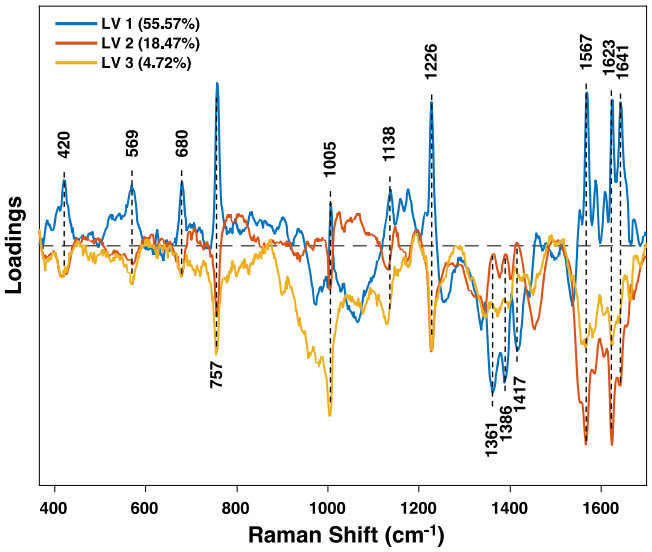

FIGURE 6 | PLS-DA loadings plot from Model C comprised the first three latent variables (LVs). Greater deviation from the dotted line indicates regions with a stronger influence on the PLS-DA model.

From a statistical viewpoint, it is clear Model A is not reliable in any capacity for LD diagnostics. This model is purely built from previously acquired human spectra, collected by different researchers. The limitation of Model A, when externally validated, reflects the complexity of human blood and performing human diagnostics. This is most evident from the low F1 score and sensitivity, despite high specificity and mediocre accuracy. This is even clearer when Model A's ROC curve is visualized, where the curve nears the line of random classification.

However, once Model A was expanded by the extra 26 samples to create Model B, we could see a clearly marked increase in model performance across most measures. Notably, Model B had large increases in F1 score and sensitivity, and the increase in accuracy to 64.2%. Similarly, the ROC curve and AUC of Model B show a clear improvement in discrimination ability, showing that the predictive capabilities of Model B are beyond random classification, unlike what can be said for Model A. From a clinical standpoint, the improvements in sensitivity and NPV but decreases in specificity and PPV indicate the model's strength lies in identifying true positives and accurately predicting negative samples when they appear. However, while the model is adept at finding positives, it also classifies several negative cases incorrectly as positive, leading to the lower specificity and PPV. This reflects a bias towards positive detection, which is useful for minimizing missed cases but reduces confidence in positive predictions.

Finally, Model C was evaluated primarily using internal validation, as it included spectra from all available samples. Since Model C was built using a group averaging size of 9 with spectra randomly grouped by classification, the resultant training spectra do differ from the original noisier spectra used for validation. Nevertheless, external validation was also performed for Model C using an 80/20 training/validation split of all spectra, giving TPRs of 92.7% and 87.3% for serological positive and negative spectra, respectively. Still, the strong performance of this model leads us to believe, that further external validation of this model using more clinical samples would lead to a better performance than that of Model B. Similarly, while Model B is not in a state ready for clinical implementation, the increases in performance as the model trains on more spectra indicate that stronger and more robust models can be built with access to more human blood samples.

The greatest limitation of results in this study was the number of samples available. With access to more samples, an optimal study would be able to separate training, validation, and test sets to evaluate model performance. Similarly, the largest set of data (the 8198 original spectra) is singularly sourced, while the additions come from a separate source. Relying on one population source for samples can lead to a prediction bias that causes findings to be applicable to samples from another source, as can be seen in Model A. This is why, as more spectra were added in from the new sample source, the model became better at predicting the serological status of each sample. Therefore, a more optimal model should be developed with samples from a wide range of sources. This should be the primary focus of any future work aimed at fully developing an RS-based test for LD diagnostics.

5 | Conclusion

In sum, this study demonstrates the potential of combining RS with PLS-DA for LD diagnostics. Our findings largely exhibit that model performance is tied both to the diversity and quantity of samples tested. Model A had low statistical reliability, likely due to its dependence on a single sample source, causing a bias toward negative predictions. Model B had marked improvements in sensitivity and predictive capabilities due to the inclusion of new samples. Model C had the greatest statistical performance but was only evaluated using internal validation. The models overall demonstrate the importance of sample variety and quantity to improve diagnostic accuracy, and importantly stress the need for thorough external validation of built diagnostic models. Future efforts should focus on refining these models to bring RS closer to practical LD diagnostics.

Acknowledgments

This study was supported by the Bay Area Lyme Foundation grant titled "Validating Raman spectroscopy-based approach for the diagnosis of Lyme disease" and the Texas A&M AgriLife Research insect vector disease seed grant titled "Development of a new diagnostic test for Lyme disease patients." We would like to thank the Lyme Disease Biobank for providing us with human blood samples.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. A. M. Gardner, N. C. Pawlikowski, S. A. Hamer, et al., "Landscape Features Predict the Current and Forecast the Future Geographic

Spread of Lyme Disease," *Proceedings of the Royal Society B* 287, no. 1941 (2020): 20202278.

2. P. Mead, A. Hinckley, and K. Kugeler, "Lyme Disease Surveillance and Epidemiology in the United States: A Historical Perspective," *Journal of Infectious Diseases* 230, no. Suppl_1 (2024): S11–S17.

3. C. Flynn and A. Ignaszak, "Lyme Disease Biosensors: A Potential Solution to a Diagnostic Dilemma," *Biosensors* 10, no. 10 (2020): 137.

4. A. R. Marques, F. Strle, and G. P. Wormser, "Comparison of Lyme Disease in the United States and Europe," *Emerging Infectious Diseases* 27, no. 8 (2021): 2017–2024.

5. J. D. Radolf, K. Strle, J. E. Lemieux, and F. Strle, "Lyme Disease in Humans," *Current Issues in Molecular Biology* 42, no. 1 (2021): 333–384.

6. K. H. Wong, E. D. Shapiro, and G. K. Soffer, "A Review of Post-Treatment Lyme Disease Syndrome and Chronic Lyme Disease for the Practicing Immunologist," *Clinical Reviews in Allergy & Immunology* 62, no. 1 (2022): 264–271.

7. IGeneX Inc, "Lyme ImmunoBlot Receives FDA Clearance," (2024).

8. J. A. Branda, K. Strle, L. E. Nigrovic, et al., "Evaluation of Modified 2-Tiered Serodiagnostic Testing Algorithms for Early Lyme Disease," *Clinical Infectious Diseases* 64, no. 8 (2017): 1074–1080.

9. A. Pegalajar-Jurado, M. E. Schriefer, R. J. Welch, et al., "Evaluation of Modified Two-Tiered Testing Algorithms for Lyme Disease Laboratory Diagnosis Using Well-Characterized Serum Samples," *Journal of Clinical Microbiology* 56, no. 8 (2018): e01943-17, https://doi.org/10.1128/jcm.01943-01917.

10. J. A. Branda and A. C. Steere, "Laboratory Diagnosis of Lyme Borreliosis," *Clinical Microbiology Reviews* 34, no. 2 (2021): e00018-19, https://doi.org/10.1128/cmr.00018-19.

11. E. Chou, A. Minor, and N. C. Cady, "Quantitative Multiplexed Strategies for Human Lyme Disease Serological Testing," *Experimental Biology and Medicine* 246, no. 12 (2021): 1388–1399.

12. A. Choopani, F. Matoofi, A. Karami, and R. Alizadeh, "Lyme Disease and New Molecular Biological Detection Methods," *Novel Clinical Medicine* 2, no. 1 (2023): 11–23.

13. T. A. Leth, R. B. Dessau, and J. K. Møller, "Discriminating Between Lyme Neuroborreliosis and Other Central Nervous System Infections by Use of Biomarkers CXCL13 and IL-6," *Ticks and Tick-borne Diseases* 13, no. 5 (2022): 101984.

14. B. K. Patterson, J. Guevara-Coto, J. Mora, et al., "Long COVID Diagnostic With Differentiation From Chronic Lyme Disease Using Machine Learning and Cytokine Hubs," *Scientific Reports* 14, no. 1 (2024): 19743.

15. C. Farber, R. Morey, M. Krimmer, D. Kurouski, and A. S. Rogovskyy, "Exploring a Possibility of Using Raman Spectroscopy for Detection of Lyme Disease," *Journal of Biophotonics* 14, no. 5 (2021): e202000477.

16. N. K. Goff, T. Dou, S. Higgins, et al., "Testing Raman Spectroscopy as a Diagnostic Approach for Lyme Disease Patients," *Frontiers in Cellular and Infection Microbiology* 12 (2022): 1006134.

17. K. Hanna, E. Krzoska, A. M. Shaaban, D. Muirhead, R. Abu-Eid, and V. Speirs, "Raman Spectroscopy: Current Applications in Breast Cancer Diagnosis, Challenges and Future Prospects," *British Journal of Cancer* 126, no. 8 (2022): 1125–1139.

18. N. M. Ralbovsky and I. K. Lednev, "Towards Development of a Novel Universal Medical Diagnostic Method: Raman Spectroscopy and Machine Learning," *Chemical Society Reviews* 49, no. 20 (2020): 7428–7453.

19. C. Farber and D. Kurouski, "Detection and Identification of Plant Pathogens on Maize Kernels with a Hand-Held Raman Spectrometer," *Analytical Chemistry* 90, no. 5 (2018): 3009–3012, https://doi.org/10.1021/acs.analchem.8b00222.

20. L. Sanchez, S. Pant, Z. Xing, K. Mandadi, and D. Kurouski, "Rapid and Noninvasive Diagnostics of Huanglongbing and Nutrient Deficits on Citrus Trees With a Handheld Raman Spectrometer," *Analytical and Bioanalytical Chemistry* 411, no. 14 (2019): 3125–3133, https://doi.org/10.1007/s00216-019-01776-4.

21. A. Giuseppe, F. Annastella, C. Giannetto, et al., "Preliminary Study for the Application of Raman Spectroscopy for the Identification of Leishmania Infected Dogs," *Scientific Reports* 12, no. 1 (2022): 7489, https://doi.org/10.1038/s41598-022-11525-w.

22. B. Vyas, A. Khatiashvili, L. Galati, et al., "Raman Hyperspectroscopy of Saliva and Machine Learning for Sjogren's Disease Diagnostics," *Scientific Reports* 14, no. 1 (2024): 11–135, https://doi.org/10.1038/s41598-024-59850-6.

23. C. Farber and D. Kurouski, "Raman Spectroscopy and Machine Learning for Agricultural Applications: Chemometric Assessment of Spectroscopic Signatures of Plants as the Essential Step Toward Digital Farming," *Frontiers in Plant Science* 13 (2022): 887511, https://doi.org/10.3389/fpls.2022.887511.

24. E. J. Horn, G. Dempsey, A. M. Schotthoefer, et al., "The Lyme Disease Biobank: Characterization of 550 Patient and Control Samples From the East Coast and Upper Midwest of the United States," *Journal of Clinical Microbiology* 58, no. 6 (2020): e00032-20, https://doi.org/10.1128/JCM.00032-20.

25. C. R. Molins, L. V. Ashton, G. P. Wormser, et al., "Development of a Metabolic Biosignature for Detection of Early Lyme Disease," *Clinical Infectious Diseases* 60, no. 12 (2015): 1767–1775, https://doi.org/10.1093/cid/civ185.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.