Domain-adversarial neural networks for cross-instrument harmonization in blood-based infrared spectroscopy

Flóra B. Németh[a,b,c] and Kosmas V. Kepesidis[a,b,c]

[a]Center for Molecular Fingerprinting (CMF), Czuczor Street 2-10, 1093 Budapest, Hungary
[b]John von Neumann Institute for Data Science (JNI), Semmelweis University, Üllői Street 26, 1085 Budapest, Hungary
[c]Faculty of Physics, Ludwig-Maximilians-Universität München (LMU), Am Coulombwall 1, 85748 Garching, Germany

ABSTRACT

Fourier transform infrared spectroscopy of human blood enables rapid, label-free biochemical profiling; however, predictive models trained on spectra from one instrument often fail to generalize to another due to device-specific differences in measurement characteristics. Addressing this domain shift is essential for reliable cross-device analysis in biomedical applications. In this study, we investigated domain-adversarial representation learning using FTIR spectra acquired with two instruments in different laboratories. The training and validation sets comprised 3,000 and 335 samples per device, respectively, obtained from longitudinal measurements, while an independent cross-sectional test set of 260 samples per device was used for evaluation. A Domain-Adversarial Neural Network was trained with auxiliary prediction tasks based on routine blood parameters reliably inferred from FTIR spectra, thereby encouraging the model to preserve biologically meaningful information while suppressing device-specific variation. Regression models trained on the learned feature representations demonstrated substantially improved cross-device consistency compared to models trained directly on the original spectra, while maintaining comparable within-device performance. Furthermore, the learned representations improved generalization in an independent classification task not used during training, indicating that the model captured domain-invariant biochemical structure. These results demonstrate the potential of domain-adversarial representation learning to enable robust, transferable feature representations, supporting scalable integration of data acquired across different measurement devices and facilitating broader deployment of machine learning in real-world biomedical applications.

Keywords: Infrared spectroscopy, Domain adaptation, Domain-Adversarial Neural Networks

1. INTRODUCTION

Fourier-transform infrared (FTIR) spectroscopy enables rapid, label-free characterization of the molecular composition of biological samples and has emerged as a promising tool for clinical and translational research.[1–4] FTIR spectra capture information about proteins, lipids, carbohydrates, and other biomolecular components, allowing predictive models to infer clinically relevant biochemical and physiological parameters directly from spectral measurements. As a result, FTIR spectroscopy has been increasingly explored for applications including metabolic profiling, disease detection, and biomarker discovery.[5–16]

Despite this promise, the widespread clinical adoption of FTIR-based predictive models is hindered by limited generalizability across measurement devices. Spectra acquired using different instruments or at different laboratories often exhibit systematic differences arising from variations in optical components, calibration procedures, environmental conditions, and preprocessing pipelines. These device-specific effects introduce domain shifts that can substantially degrade predictive performance when models trained on one instrument are applied

Further author information: (Send correspondence to Kosmas V. Kepesidis.)
Kosmas V. Kepesidis: E-mail: kosmas.kepesidis@cmf.hu
Flóra B. Németh: E-mail: flora.nemeth@cmf.hu

Data Science for Photonics and Biophotonics II, edited by Thomas Bocklitz
Proc. of SPIE Vol. 14099, 1409909 · 2026 Published by SPIE
0277-786X · doi: 10.1117/12.3099135

to data from another. In a previous work, we demonstrated that such cross-device domain shifts can lead to pronounced performance degradation, underscoring the need for domain adaptation to enable reliable cross-device generalization in biologically meaningful predictive applications.[17]

To address this challenge, we previously introduced a data-augmentation–based domain-adaptation strategy that modeled device-specific variability and generated synthetic visit data to improve model robustness. This approach proved effective and, importantly, remained usable even in small-data settings where neural network–based approaches would be difficult to utilize reliably. However, the method required longitudinal measurements to estimate within-individual variability and relied on a dedicated calibration set consisting of samples measured on both devices. These requirements limit its applicability in many real-world scenarios, in which longitudinal follow-up or paired calibration measurements may be unavailable.

Domain-adversarial neural networks (DANNs) offer an alternative approach by directly learning feature representations that are invariant to domain-specific differences while preserving task-relevant information.[18,19] Through adversarial training, the model simultaneously optimizes predictive performance on auxiliary tasks and minimizes its ability to distinguish between domains, encouraging the learned representations to retain biologically meaningful structure while suppressing device-specific variation. Importantly, this framework does not require paired measurements or longitudinal data and can operate on independent datasets collected across different instruments.

In this work, we propose a domain-adversarial neural network framework to learn device-invariant representations from large-scale FTIR spectral data acquired across multiple instruments. The model is trained using auxiliary prediction tasks based on routine clinical blood parameters that can be reliably inferred from FTIR spectra, thereby ensuring that the learned features preserve biologically relevant information. We demonstrate that the resulting representations substantially improve cross-device predictive performance while maintaining comparable within-device performance. Furthermore, we show that the learned representations generalize to downstream tasks not explicitly included during training, indicating that the model captures robust, domain-invariant biochemical information.

Together, these results demonstrate the feasibility of adversarial representation learning for harmonizing FTIR spectral data across measurement devices and highlight its potential to enable robust, scalable integration of multi-instrument spectral datasets for biomedical and clinical applications.

2. METHODS

2.1 Measurements

For this work, we used data from the Health for Hungary (H4H) study, a large-scale longitudinal cohort of participants classified as healthy at enrollment (study code: H4H_HU_2020_Sample Collection; ethics approval reference: 2754-11/2020/EÜIG).[20,21] As part of the study protocol, certified external laboratories measured 27 routine blood parameters, and FTIR spectra of human blood plasma were acquired at two locations: Szeged, Hungary (device A), and Garching, Germany (device B). Both FTIR devices were MIRA Analyzers (CLADE GmbH), differing only in their auto-sampler units; their optical components and software versions were otherwise identical.

2.2 Data preprocessing

FTIR spectra were subjected to a structured preprocessing pipeline to retain biologically informative regions and ensure suitability for machine learning analysis. First, spectral ranges below 1000 cm^{-1} and above 3000 cm^{-1} were excluded, as these peripheral regions consistently showed low absorbance and lacked meaningful molecular information. Second, outlier detection was performed in two stages. Spectra containing an abnormal absorption feature near 1045 cm^{-1}—likely caused by recurring but unidentified contamination—were removed. In addition, statistical outliers were identified using the Local Outlier Factor (LOF) algorithm applied to the truncated spectra, and only spectra classified as inliers were retained. Third, L2 normalization was applied to each spectrum to reduce inter-sample intensity variability. Finally, the spectral interval between 1800 and 2800 cm^{-1}, which lacks biochemical information in this context, was excluded. The resulting spectra consisted of the ranges 1000–1800 cm^{-1} and 2800–3000 cm^{-1}, corresponding to a total of 519 spectral features.

Clinical blood parameters were preprocessed by standardizing measurement units and removing outliers. Biological plausibility filters were applied using predefined physiological reference ranges for each parameter. In addition, statistical outliers were identified using DBSCAN clustering applied separately to each marker, and detected outliers were marked as missing values. Samples containing any missing blood parameter after preprocessing were excluded from subsequent analyses.

2.3 Data splitting

The dataset consisted of paired FTIR measurements acquired using Devices A and B, with longitudinal follow-up of individuals across up to five visits. To train and validate the domain-adversarial neural network, we selected subjects with complete five-visit records, resulting in 667 individuals corresponding to 3,335 samples per device.

To prevent information leakage arising from strong individual-specific spectral signatures, data were split at the subject level using a 90:10 ratio. This resulted in 3,000 samples per device from 600 individuals in the training set and 335 samples per device from 67 individuals in the validation set. Subject-level splitting ensured that no measurements from the same individual were present in both sets. Importantly, although measurements were available in paired form across devices, this pairing information was not used during training, and samples from each device were treated as independent observations.

An independent test set was constructed using only the first visit from separate individuals, yielding 260 samples per device. Using cross-sectional test data avoided bias from longitudinal correlations and provided an unbiased, realistic evaluation of generalization performance.

2.4 DANN architecture and training procedure

To learn domain-invariant yet biologically informative representations from FTIR spectra, we implemented a domain-adversarial multi-task neural network consisting of a shared feature encoder, multiple regression task heads, and an adversarial domain classification head (Figure 1).

The feature encoder transforms the input FTIR spectrum into a latent feature representation. As illustrated in Figure 1, the encoder applies consecutive one-dimensional convolutional layers to extract local spectral patterns, followed by adaptive average pooling and a fully connected projection to obtain a compact latent feature vector. Residual fully connected blocks further refine this representation while preserving stability through skip connections.

The resulting latent representation is shared between two parallel branches. The task heads consist of independent regression networks that predict selected blood parameters, encouraging the encoder to retain biologically meaningful information. In parallel, a domain classification head predicts the measurement device. A gradient reversal layer is placed between the encoder and domain classifier, which acts as an identity function during forward propagation but reverses gradients during backpropagation, encouraging the encoder to learn domain-invariant features.

To stabilize adversarial training, a short warmup phase was applied at the beginning of training, during which gradient reversal was disabled, enabling the encoder to first learn predictive features. After this phase, full adversarial training was performed using alternating optimization: the domain classifier was updated while keeping the encoder fixed, followed by joint optimization of the encoder and task heads using both regression and adversarial domain losses.

This training strategy enables the encoder to learn latent representations that preserve clinically relevant biochemical information while minimizing domain-specific variation, improving generalization across measurement devices.

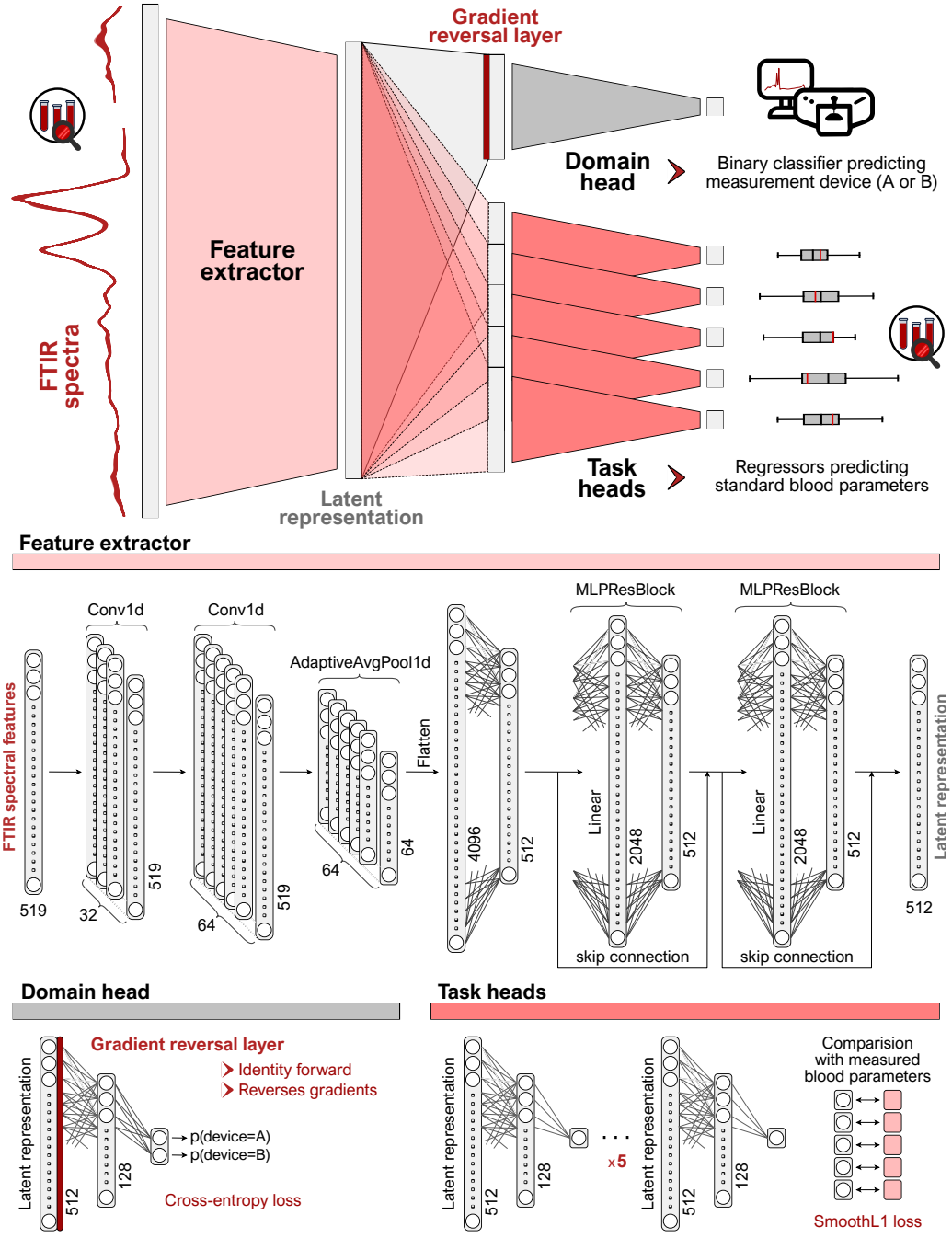

Figure 1. Domain-adversarial multi-task neural network architecture for learning domain-invariant representations from FTIR spectra. The spectra are processed by a shared feature extractor consisting of convolutional layers, adaptive pooling, and residual multilayer perceptron blocks, producing a latent feature representation. This representation is used by multiple regression heads to predict standard blood parameters. In parallel, a domain classification head predicts the measurement device. A gradient reversal layer connects the encoder and the domain head, reversing gradients during training and encouraging the encoder to learn domain-invariant features while preserving predictive information relevant to the regression tasks.

2.5 Selection of task heads

To define the prediction tasks used in the multi-task architecture, we first evaluated the predictability of all 27 available routine blood parameters from FTIR spectra using simple linear regression models. Each parameter was predicted independently from spectra measured on device A, and model performance was quantified using Pearson's correlation coefficient between predicted and measured values. As shown in Figure 2, several blood parameters exhibited strong correlations, indicating that they can be reliably inferred from spectral measurements. Based on this analysis, the five parameters with the highest correlation coefficients were selected as targets for the task heads. Notably, the ranking of parameters was highly consistent across devices, and the same top five parameters were identified when performing the analysis using spectra from device B. This selection ensured that the model focused on biologically meaningful targets with a strong spectral signature, while avoiding an excessive number of task heads that could increase model complexity and reduce training stability. Since each blood parameter was measured only once per sample and independently of the FTIR acquisition, this procedure also avoided any bias related to repeated measurements across devices.

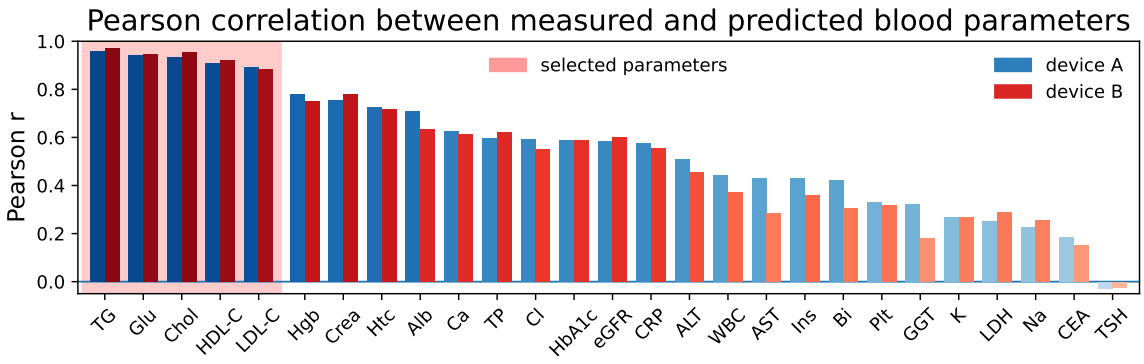

Figure 2. Pearson correlation coefficients between measured and predicted routine blood parameters using linear regression models trained on FTIR spectra acquired with device A (blue) and device B (red). Each parameter was predicted independently, and correlation coefficients quantify the strength of the relationship between spectral features and biochemical measurements. The shaded region highlights the five parameters with the highest correlations, which were selected as prediction targets for the task heads in the domain-adversarial neural network. The highly consistent ranking across devices indicates that these parameters have robust spectral signatures and can be reliably inferred from FTIR measurements.

3. RESULTS

3.1 Improved cross-device prediction using domain-adversarial features

To evaluate the effectiveness of the learned domain-invariant representations, we compared prediction performance using the original FTIR spectra and the DANN-encoded feature space. Linear regression models were trained to predict blood parameters using the training set and evaluated on the independent test set, both for within-device and cross-device scenarios. Prediction accuracy was quantified using the root mean squared error (RMSE), which provides a scale-sensitive measure of deviation between predicted and measured values.

When using the original spectral data, cross-device predictions showed substantially higher RMSE compared to within-device predictions (Fig. 3a). Notably, prediction performance was more severely degraded in the device A to device B direction than in the reverse case. This asymmetry is consistent with the differences in spectral variability between devices, where device B exhibits broader distributional variability than device A (see Appendix Sec. A). This confirms the presence of strong domain-specific differences that limit cross-device generalization.

After training the DANN and encoding both training and test samples into the learned feature space, the same regression procedure was repeated (Fig. 3b). In this encoded feature space, cross-device prediction errors were markedly reduced and became comparable to within-device errors across all examined blood parameters.

Importantly, this improvement in cross-device generalization was achieved without meaningful degradation of within-device prediction performance, which remained at a similar level as in the original spectral space.

Notably, the benefit of the learned representation extended beyond the five blood parameters used as auxiliary tasks during DANN training. Improved cross-device consistency was also observed for additional parameters that were not directly used as supervision signals, indicating that the encoder captured general domain-invariant biochemical information rather than task-specific mappings.

These results demonstrate that the domain-adversarial training successfully learned feature representations that preserve biologically relevant predictive information while substantially reducing device-specific variability.

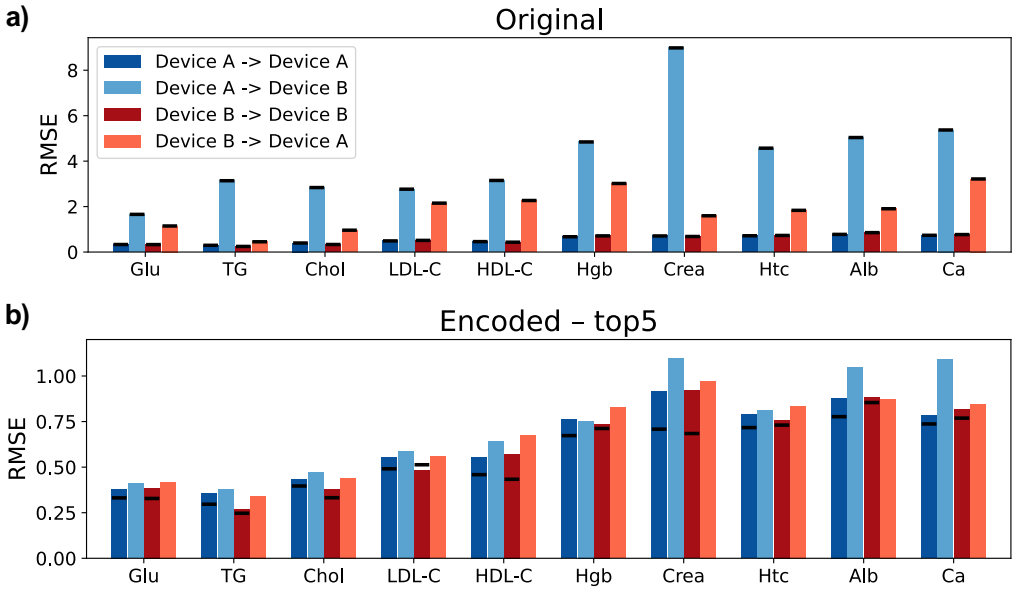

Figure 3. Comparison of prediction performance using original FTIR spectra and DANN-encoded features. (a) Root mean squared error (RMSE) of linear regression models trained and evaluated within and across devices using the original spectral data. Cross-device predictions exhibit substantially higher error, indicating limited cross-device generalization. (b) Prediction performance using the DANN-encoded feature representation. Cross-device prediction errors are markedly reduced and become comparable to within-device performance. Black horizontal lines indicate the corresponding within-device RMSE values obtained using the original spectral data, serving as baseline references. Results are shown for the top ten most predictable blood parameters ranked by Pearson correlation.

3.2 DANN features improve cross-device classification consistency

To further evaluate the generalization properties of the learned feature representations, we examined a binary classification task not directly used during DANN training. Logistic regression classifiers were trained to predict biological sex from FTIR spectra using both the original spectral data and the DANN-encoded feature space.

Using the original spectral data, classification performance showed a pronounced cross-device degradation (Fig. 4b, left panel). While within-device classification achieved high accuracy, models trained on one device performed substantially worse when evaluated on the other device, indicating strong domain dependence of the learned decision boundaries.

When using the encoded feature representations obtained from the DANN trained with the top five most predictable blood parameters, cross-device classification performance improved considerably, and within-device and cross-device receiver operating characteristic (ROC) curves became closely aligned (Fig. 4b, middle panel). This demonstrates that the domain-adversarial training successfully reduced device-specific bias, even for a downstream task that was not explicitly included during training.

However, this improved cross-device consistency was accompanied by a moderate reduction in within-device classification performance. To investigate whether this could be explained by the selection of auxiliary prediction tasks, we examined the distribution of all 27 blood parameters stratified by sex (Fig. 4a). Among the originally selected top five parameters, only HDL-cholesterol showed a noticeable sex-dependent difference, suggesting that the auxiliary supervision may not sufficiently capture sex-related biological variation.

To address this, we defined an alternative set of auxiliary tasks by including three strongly sex-dependent parameters—hemoglobin, creatinine, and hematocrit—alongside triglycerides and glucose. After retraining the DANN model using this modified task set, classification performance remained largely unchanged (Fig. 4b, right panel), indicating that the observed trade-off between domain invariance and task-specific separability was not primarily driven by the choice of auxiliary targets. Similarly, increasing the number of auxiliary tasks from five to ten did not substantially alter the overall domain adaptation performance (see Appendix Sec. B), indicating that including additional tasks does not necessarily yield further benefit. This suggests that selecting five auxiliary targets provide sufficient supervisory signal while maintaining a favorable balance between informative guidance and model complexity.

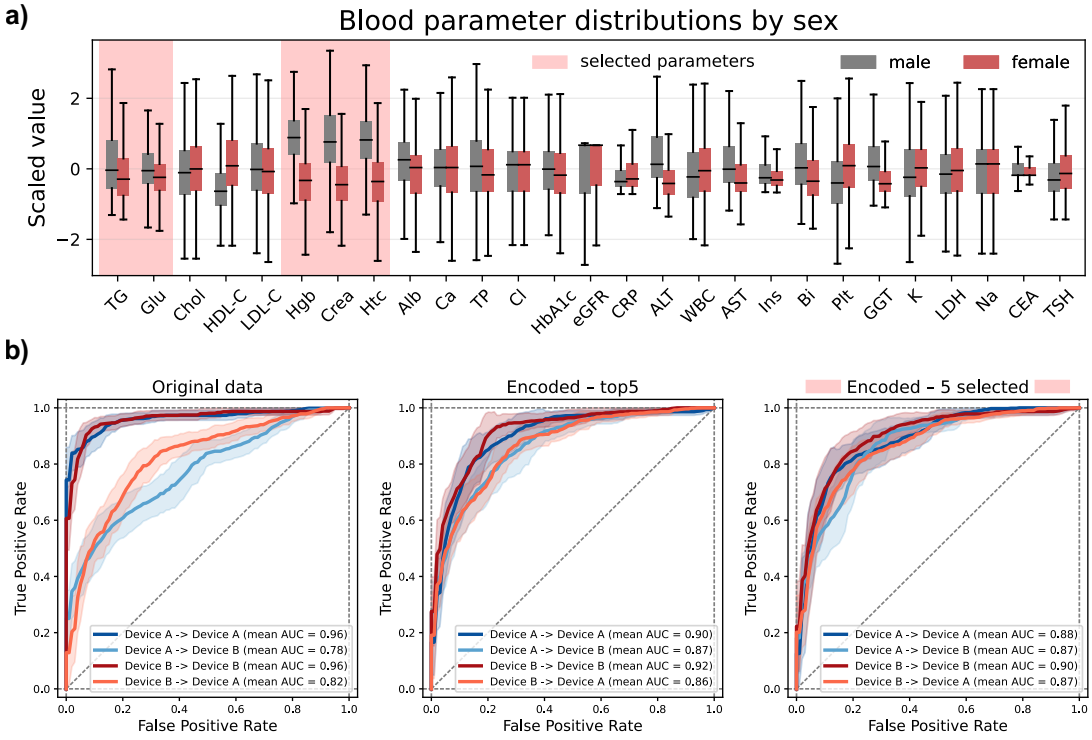

Figure 4. Evaluation of sex classification performance using original and domain-adversarial feature representations. (a) Distribution of routine blood parameters stratified by sex, highlighting parameters with strong sex-dependent differences. The shaded region indicates selected auxiliary prediction targets used during DANN training. (b) Receiver operating characteristic (ROC) curves for logistic regression classifiers trained to predict sex using original FTIR spectra (left), DANN-encoded features trained with the top five most predictable parameters (middle), and DANN-encoded features trained with an alternative set of sex-dependent parameters (right). Using encoded features substantially improves cross-device consistency, reducing the performance gap between within-device and cross-device classification, although within-device performance is moderately reduced compared to the original spectral representation.

These results confirm that the proposed domain-adversarial framework effectively learns device-invariant representations that generalize to independent downstream tasks. At the same time, the observed reduction in within-device classification performance suggests that certain components of biologically relevant variation may not be fully preserved during domain alignment. This likely reflects incomplete disentanglement between

biological and domain-related sources of variability, particularly for biological signals that are not explicitly reinforced by the auxiliary prediction tasks. As a result, biological variation that is weakly represented or unrelated to the selected auxiliary targets may be partially suppressed together with domain-specific variation. These findings highlight the importance of carefully selecting auxiliary tasks that comprehensively capture biologically meaningful structure to enable more effective separation of domain and biological effects.

4. DISCUSSION

In this study, we demonstrated that domain-adversarial neural networks can effectively reduce device-specific variability in FTIR spectral data while preserving biologically relevant predictive information. Models trained on the learned feature representations showed substantially improved cross-device generalization compared to models trained directly on raw spectra, both for regression of routine blood parameters and for an independent downstream classification task. These findings highlight the feasibility of learning domain-invariant representations that enable robust integration of spectral data acquired using different measurement devices.

At the same time, the results indicate that domain-adversarial training involves an inherent trade-off between removing device-specific variation and preserving task-relevant biological signal. While cross-device prediction performance improved markedly, within-device performance for certain tasks showed a modest reduction. This suggests that some biologically meaningful information may be partially entangled with domain-specific variation, and aggressive domain alignment may attenuate signal components that are informative for specific prediction tasks. Careful selection of auxiliary supervision is therefore critical to ensure that biologically relevant structure is preserved during training.

The choice of auxiliary prediction targets plays a particularly important role in shaping the learned representation. In this study, routine clinical blood parameters were used as auxiliary tasks because they provide biologically interpretable supervision and were readily available in the H4H study. However, these measurements capture only a subset of the molecular information reflected in FTIR spectra. Incorporating additional modalities, such as proteomic or metabolomic measurements, could provide richer and more specific supervisory signals. Such tasks may enable the model to better preserve individual-specific and disease-related biochemical information while simultaneously removing device-specific variation. This multi-modal supervision represents a promising direction for future work.

Further improvements may also be achieved through architectural refinements and training strategies. Although the current encoder successfully learned domain-invariant features, alternative architectural and training strategies may further improve the separation between biological and domain-related sources of variation. Representation regularization techniques, such as reconstruction losses or feature variance constraints, could help ensure that biologically meaningful information is preserved in the latent space. In addition, explicit disentanglement approaches that separately model domain-invariant and domain-specific components may enable more precise removal of device-related variation while retaining clinically relevant biological signal. In addition, the present framework can be naturally extended to scenarios involving more than two measurement devices by using a multi-class domain classifier instead of a binary one. This would allow the model to learn representations that are invariant across multiple instruments simultaneously, further increasing its practical applicability in heterogeneous clinical environments.

Importantly, the domain-adversarial framework used in this study does not rely on paired measurements of the same samples across devices, nor does it require longitudinal data. This makes the approach applicable to real-world clinical scenarios where sufficient numbers of samples are available from multiple instruments, but paired calibration measurements are not available. Such situations are common in large-scale clinical studies and multi-center data collections, where independent cohorts are measured using different devices.

More broadly, the proposed approach is not limited to FTIR spectroscopy. With appropriate modifications to the feature encoder, the same domain-adversarial learning framework can be applied to other types of biomedical data, including other spectroscopic modalities, imaging data, or multi-omics measurements. This flexibility highlights the general potential of adversarial representation learning as a tool for harmonizing heterogeneous biomedical datasets.

Overall, while the proposed approach does not fully eliminate domain-related differences, it substantially improves cross-device generalization and demonstrates a viable pathway toward harmonizing FTIR spectral data acquired using different measurement systems. These findings support the potential of domain-adversarial representation learning as a practical tool for enabling robust, scalable, and device-independent spectral analysis in biomedical and clinical applications.

APPENDIX A. DOMAIN DIFFERENCES

Substantial distributional differences were observed between spectra acquired using device A and device B. In particular, device B exhibited consistently higher spectral variability across most wavenumbers, as evidenced by the broader variance envelope compared to device A (Fig. 5a). While the mean spectra of the two devices were largely similar, the increased dispersion of device B measurements indicates a wider domain distribution.

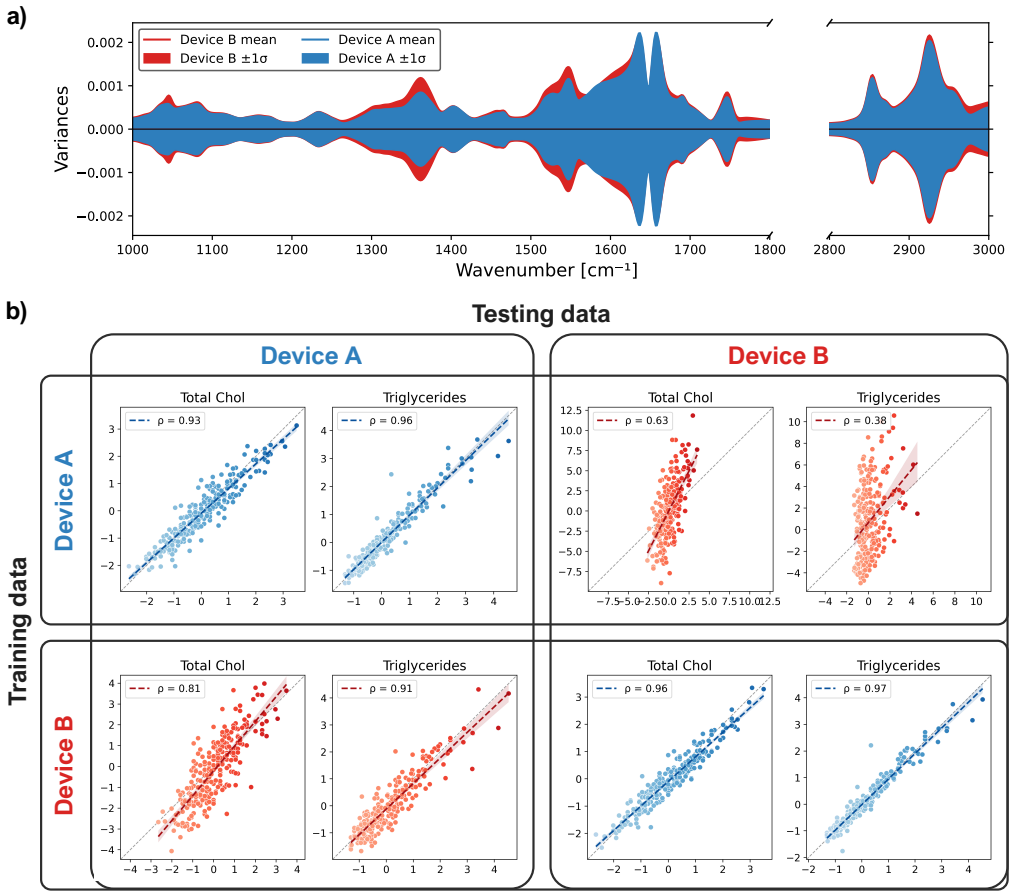

Figure 5. Domain differences between FTIR measurement devices and their impact on predictive performance. (a) Mean-centered spectral variance profiles for device A and device B across the measured wavenumber range. Device B exhibits consistently higher variability, indicating a broader domain distribution despite similar mean spectral structure. (b) Predictive performance for representative blood parameters under different training and testing device combinations. Models trained and evaluated on the same device achieve high accuracy, whereas cross-device predictions, especially models trained on device A and tested on device B, show substantially reduced performance. This asymmetric degradation is consistent with an extrapolation effect, as device B exhibits greater variability than device A.

This difference in variability provides a plausible explanation for the observed asymmetry in cross-device predictive performance (Fig. 5b). Models trained on device A and evaluated on device B showed substantially

reduced predictive accuracy compared to the opposite direction. Since device A represents a narrower distribution, models trained on device A are exposed to a more limited range of spectral variation. When applied to device B, which exhibits greater variability, these models are required to extrapolate beyond the range of patterns observed during training, leading to degraded performance. In contrast, models trained on device B are exposed to a broader distribution of spectral variability and are therefore better able to generalize to device A, which represents a subset of this broader domain.

APPENDIX B. INCREASING THE NUMBER OF TASK HEADS

To assess the effect of increasing the number of auxiliary prediction targets, we repeated the DANN training procedure using the top ten most predictable blood parameters instead of the top five. This experiment aimed to determine whether expanding the task set improves domain adaptation, particularly for parameters ranked between sixth and tenth that were not directly included as auxiliary tasks in the original model.

As shown in Fig. 6, overall performance using ten auxiliary tasks was very similar to that of the original five-task model. Within-device prediction accuracy remained comparable, and most cross-device results showed only minor differences relative to the baseline.

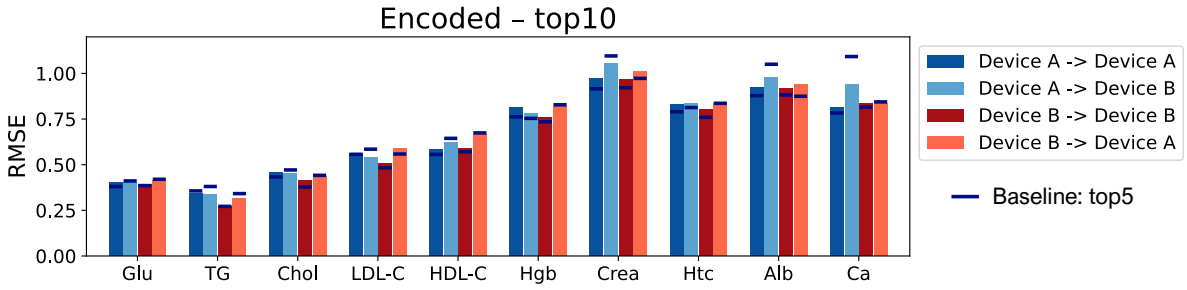

Figure 6. Comparison of prediction performance using DANN models trained with five versus ten auxiliary prediction tasks. Bars show root mean squared error (RMSE) for within-device and cross-device prediction of routine blood parameters using encoded features from the model trained with ten tasks. Dark horizontal markers indicate baseline performance obtained using the five-task model.

A modest improvement was observed specifically for predictions trained on device A and evaluated on device B for parameters ranked seventh, ninth, and tenth (creatinine, albumin, and calcium), which were newly included as auxiliary targets. For these parameters, cross-device prediction error decreased relative to the five-task baseline, while no consistent improvement was observed for other parameters or in the opposite cross-device direction.

These findings indicate that increasing the number of auxiliary tasks provides only limited additional benefit, and that a small set of carefully selected auxiliary targets is sufficient to capture the domain-invariant structure relevant for downstream prediction.

ACKNOWLEDGMENTS

The work on the H4H study data was part of project no. 2020-2.1.1-ED-2022-00213. Project no. 2020-2.1.1-ED-2022-00213 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the 2020-2.1.1-ED funding scheme. Flóra B. Németh was supported by the 2025-2.1.2-EKÖP-KDP University Research Scholarship Programme of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. We gratefully acknowledge the many individuals who volunteered to participate in the H4H clinical study used in this work.

REFERENCES

[1] Baker, M. J., Trevisan, J., Bassan, P., Bhargava, R., Butler, H. J., Dorling, K. M., Fielden, P. R., Fogarty, S. W., Fullwood, N. J., Heys, K. A., et al., "Using fourier transform ir spectroscopy to analyze biological materials," *Nature protocols* **9**(8), 1771–1791 (2014).

[2] Butler, H. J., Brennan, P. M., Cameron, J. M., Finlayson, D., Hegarty, M. G., Jenkinson, M. D., Palmer, D. S., Smith, B. R., and Baker, M. J., "Development of high-throughput atr-ftir technology for rapid triage of brain cancer," *Nature communications* **10**(1), 4501 (2019).

[3] Voronina, L., Leonardo, C., Mueller-Reif, J. B., Geyer, P. E., Huber, M., Trubetskov, M., Kepesidis, K. V., Behr, J., Mann, M., Krausz, F., et al., "Molecular origin of blood-based infrared spectroscopic fingerprints," *Angewandte Chemie* **133**(31), 17197–17206 (2021).

[4] Paraskevaidi, M., Matthew, B. J., Holly, B. J., Hugh, B. J., Thulya, C. P., Loren, C., StJohn, C., Peter, G., Callum, G., Sergei, K. G., et al., "Clinical applications of infrared and raman spectroscopy in the fields of cancer and infectious diseases," *Applied Spectroscopy Reviews* **56**(8-10), 804–868 (2021).

[5] Huber, M., Kepesidis, K. V., Voronina, L., Božić, M., Trubetskov, M., Harbeck, N., Krausz, F., and Žigman, M., "Stability of person-specific blood-based infrared molecular fingerprints opens up prospects for health monitoring," *Nature communications* **12**(1), 1511 (2021).

[6] Huber, M., Kepesidis, K. V., Voronina, L., Fleischmann, F., Fill, E., Hermann, J., Koch, I., Milger-Kneidinger, K., Kolben, T., Schulz, G. B., et al., "Infrared molecular fingerprinting of blood-based liquid biopsies for the detection of cancer," *Elife* **10**, e68758 (2021).

[7] Martin, F. L., Kelly, J. G., Llabjani, V., Martin-Hirsch, P. L., Patel, I. I., Trevisan, J., Fullwood, N. J., and Walsh, M. J., "Distinguishing cell types or populations based on the computational analysis of their infrared spectra," *Nature protocols* **5**(11), 1748–1760 (2010).

[8] Ghimire, H., Garlapati, C., Janssen, E. A., Krishnamurti, U., Qin, G., Aneja, R., and Perera, A. U., "Protein conformational changes in breast cancer sera using infrared spectroscopic analysis," *Cancers* **12**(7), 1708 (2020).

[9] Ollesch, J., Theegarten, D., Altmayer, M., Darwiche, K., Hager, T., Stamatis, G., and Gerwert, K., "An infrared spectroscopic blood test for non-small cell lung carcinoma and subtyping into pulmonary squamous cell carcinoma or adenocarcinoma," *Biomedical Spectroscopy and Imaging* **5**(2), 129–144 (2016).

[10] Eissa, T., Leonardo, C., Kepesidis, K. V., Fleischmann, F., Linkohr, B., Meyer, D., Zoka, V., Huber, M., Voronina, L., Richter, L., et al., "Plasma infrared fingerprinting with machine learning enables single-measurement multi-phenotype health screening," *Cell Reports Medicine* **5**(7) (2024).

[11] Kepesidis, K. V., Stoleriu, M.-G., Feiler, N., Gigou, L., Fleischmann, F., Aschauer, J., Eiselen, S., Koch, I., Reinmuth, N., Tufman, A., et al., "Assessing lung cancer progression and survival with infrared spectroscopy of blood serum," *BMC medicine* **23**, 101 (2025).

[12] Kepesidis, K. V., Bozic-Iven, M., Huber, M., Abdel-Aziz, N., Kullab, S., Abdelwarith, A., Al Diab, A., Al Ghamdi, M., Hilal, M. A., Bahadoor, M. R., et al., "Breast-cancer detection using blood-based infrared molecular fingerprints," *BMC cancer* **21**, 1–9 (2021).

[13] Backhaus, J., Mueller, R., Formanski, N., Szlama, N., Meerpohl, H.-G., Eidt, M., and Bugert, P., "Diagnosis of breast cancer with infrared spectroscopy from serum samples," *Vibrational Spectroscopy* **52**(2), 173–177 (2010).

[14] Elmi, F., Movaghar, A. F., Elmi, M. M., Alinezhad, H., and Nikbakhsh, N., "Application of ft-ir spectroscopy on breast cancer serum analysis," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **187**, 87–91 (2017).

[15] Ollesch, J., Heinze, M., Heise, H. M., Behrens, T., Brüning, T., and Gerwert, K., "It's in your blood: spectral biomarker candidates for urinary bladder cancer from automated ftir spectroscopy," *Journal of biophotonics* **7**(3-4), 210–221 (2014).

[16] Anderson, D., Anderson, R., Moug, S., and Baker, M., "Liquid biopsy for cancer diagnosis using vibrational spectroscopy: systematic review," *BJS open* **4**(4), 554–562 (2020).

[17] Nemeth, F. B., Leopold-Kerschbaumer, N., Debreceni, D., Fleischmann, F., Borbely, K., Mazurencu-Marinescu-Pele, D., Bocklitz, T., Žigman, M., and Kepesidis, K. V., "Bridging spectral gaps: Cross-device model generalization in blood-based infrared spectroscopy," *Analytical Chemistry* (2025).

[18] Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., and Marchand, M., "Domain-adversarial neural networks," *arXiv preprint arXiv:1412.4446* (2014).

[19] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., and Lempitsky, V., "Domain-adversarial training of neural networks," *Journal of machine learning research* **17**(59), 1–35 (2016).

[20] "H4h study homepage." https://h4h.hu/en/ [Accessed: 2026/03/20].

[21] Kepesidis, K. V., Zarandy, Z. I., Nemeth, F. B., Gigou, L., Žigman, M., and Krausz, F., "Integration of infrared molecular fingerprinting data in a longitudinal health profiling cohort," in [*International Conference on Artificial Intelligence in Medicine*], 180–190, Springer (2025).