

<https://doi.org/10.1038/s41698-024-00772-x>

# Prediction of post-treatment recurrence in early-stage breast cancer using deep-learning with mid-infrared chemical histopathological imaging

Check for updates

Abigail Keogan<sup>1,2</sup>, Thi Nguyet Que Nguyen<sup>3</sup>, Pascaline Bouzy<sup>4</sup>, Nicholas Stone<sup>4</sup>, Karin Jirstrom<sup>5</sup>, Arman Rahman<sup>6,7</sup>, William M. Gallagher<sup>6</sup> & Aidan D. Meade<sup>1,2</sup>✉

Predicting long-term recurrence of disease in breast cancer (BC) patients remains a significant challenge for patients with early stage disease who are at low to intermediate risk of relapse as determined using current clinical tools. Prognostic assays which utilize bulk transcriptomics ignore the spatial context of the cellular material and are, therefore, of limited value in the development of mechanistic models. In this study, Fourier-transform infrared (FTIR) chemical images of BC tissue were used to train deep learning models to predict future disease recurrence. A number of deep learning models were employed, with champion models employing two-dimensional and two-dimensional-separable convolutional networks found to have predictive performance of a ROC AUC of approximately 0.64, which compares well to other clinically used prognostic assays in this space. All-digital chemical imaging may therefore provide a label-free platform for histopathological prognosis in breast cancer, opening new horizons for future deployment of these technologies.

Approximately 2.2 million women worldwide were diagnosed with breast cancer in 2020, rendering it the most frequently occurring cancer in women<sup>1</sup>. Of these, approximately 80% are hormone receptor positive, and therefore candidates for cytotoxic adjuvant chemotherapy; the clinical decision to undergo this type of therapy is influenced by the risk profile of each patient<sup>2</sup>. Human epidermal growth factor receptor 2 (HER2)-negative, hormone receptor positive patients with early disease may have their risk profile determined by a range of clinicopathological risk factors including patient age, tumor size and grade, and lymph node (LN) involvement. The Nottingham Prognostic Index (NPI)<sup>3,4</sup>, which combines tumor size, grade and LN involvement, classifies early and locally advanced breast cancer cases into three or more prognostic groups.

For early-stage ER+ LN− patients, several tests are available on the market that attempt to stratify based on the likelihood of future disease recurrence. These include but are not limited to, the Oncotype DX test<sup>5</sup> (a qRT-PCR assay measures the gene expression levels of 16 cancer-related

genes plus 5 control genes) and the MammaPrint test<sup>6</sup> (which uses a 70 gene profile). Another test, OncoMasTR, has recently emerged that calculates a Molecular Score (OMm), which contains information relating to solely 3 master transcription regulators (MTRs) and 3 reference genes, and its composite OncoMasTR Risk Score (OM), which combines OMm with LN involvement and tumor size and categorizes patients as having either low or high risk of recurrence. This test has been analytically validated<sup>7</sup> and clinically validated to Level 1 evidence via two category B studies as significantly prognostic for breast cancer recurrence in two independent patient cohorts<sup>8,9</sup>.

Despite the clinical use and development of next-generation transcriptomic assays to predict breast cancer treatment outcome, discordance in the risk classification of individual patients with different tests confounds the choice of an optimal molecular-based signature for a given patient<sup>10</sup>. Moreover, despite their differential approaches to prognose treatment outcome, each test has a very similar predictive power, with OncoTypeDX having a

<sup>1</sup>Radiation and Environmental Science Centre, Physical to Life Sciences Research Hub, Technological University Dublin, Dublin, Ireland. <sup>2</sup>School of Physics, Clinical and Optometric Sciences, Technological University Dublin, City Campus, Dublin, Ireland. <sup>3</sup>Digital Futures Research Hub, Technological University Dublin, Dublin, Ireland. <sup>4</sup>Department of Physics and Astronomy, University of Exeter, Exeter, UK. <sup>5</sup>Division of Oncology and Therapeutic Pathology, Department of Clinical Sciences, Lund University, Lund, Sweden. <sup>6</sup>UCD School of Biomolecular and Biomedical Science, UCD Conway Institute, University College Dublin, Dublin, Ireland. <sup>7</sup>UCD School of Medicine, UCD Conway Institute, University College Dublin, Dublin, Ireland. ✉e-mail: [aidan.meade@tudublin.ie](mailto:aidan.meade@tudublin.ie)

ROC AUC of 0.69 and MammaPrint 0.59<sup>11,12</sup>. This indicates that each test, while powerful, also fails to identify a significant proportion of patients in whom cancer recurs. Additionally, tests such as OncoType-DX are expensive (~\$3000 per patient in the UK) and have a 7–10 day turnaround time.

Aside from molecular tests, tumor organization is a key factor for disease progression, affecting the ability of the tumor to interact with the microenvironment<sup>13,14</sup>. Recent evidence suggests that robust predictive assays for recurrence which employ machine learning and artificial intelligence on haematoxylin and eosin stained pathological images have a performance approaching that achieved with genomic and proteomic tests<sup>15</sup>. Chemical imaging (via either Raman or FTIR microspectroscopy) offers an alternative, label-free, non-destructive approach to histopathological image generation which profiles histopathological tissue in both a spatial and biochemical context, with demonstrated applications both in the pathology laboratory<sup>16</sup> and in surgery<sup>17,18</sup>, owing to its potential to image both preserved and fresh tissues. As a result it has seen rapid development towards devices for diagnosis of cancer and pre-cancerous lesions<sup>19</sup>, with some examples of the development of clinical support tools for diagnostic histopathology<sup>20</sup>, and more recently moves towards the development of prognostic tests<sup>13</sup>. FTIR imaging spatially captures a combined spectrum of the vibrations of molecular moieties including  $-\text{NH}$ ,  $-\text{C}=\text{O}$ ,  $-\text{CN}$ ,  $-\text{CH}_3$ ,  $-\text{CH}_2$ , and  $-\text{PO}_2$ <sup>21</sup>. It has been demonstrated that spectral data are altered as a result of pre-translational (DNA methylation) and post-translational effects (e.g., phosphorylation, acetylation, glycosylation, etc.) in disease states<sup>22,23</sup>, such that objective histopathological classification models may be constructed. Recent evidence has also highlighted the connectivity between chemical imaging data and deep spatial tissue biology<sup>24</sup>. Crucially, analysis pipelines for chemical imaging data typically require the application of segmentation and dimensionality reduction steps before application of ML techniques, which may result in the loss of important spectral and spatial information, and therefore a pivot towards whole image analytical pipelines for this data is required.

The advent of deep-learning (DL) now offers a pipeline directly from image acquisition to classification either via digital pathology or whole image classification<sup>25</sup>. DL is an application of neural networks with many hidden layers that internally identify patterns in data without the need to recourse to feature engineering<sup>26,27</sup>. DL-convolution neural networks (CNNs) have demonstrated astounding abilities to resolve class across many computer vision tasks, including recognition of skin cancer subtype with simple RGB image data with performance exceeding that of Board accredited clinicians<sup>28</sup>. Essentially DL-CNNs may be presented with raw data and automatically discover features or their combinations which can learn, classify or detect objects within them<sup>26,27</sup>. Despite these advances, the development of network learning approaches specifically designed to work with chemical or hyperspectral imaging data, comprising of hundreds to thousands of channels measured spatially, is in its infancy<sup>29</sup>.

In the present study, an FTIR chemical-imaging approach to segregation of early stage BC patients according to downstream recurrence (or not) has been developed in a mixed cohort of patients with varying ER, PR and HER2 status. The study represents the first of its kind to use whole-image deep-learning pipelines for development of recurrence-segregation models with chemical imaging data. Several ML/DL pipelines employing both traditional algorithms and DL approaches were developed to model the connection between chemical images and all breast cancer recurrence (BCR) events (i.e., both local and distant). We demonstrate that chemical-imaging deep learning pipeline was found to have a prognostic performance ROC AUC of approximately 0.64 for prediction of BCR, with additional results demonstrating network capability in predicting other treatment endpoints such as BCR survival (up to 7 years post therapy). This BCR prediction performance agrees excellently with current transcriptomic tests for recurrence of this disease, pointing to the potential wider opportunities for deployment of this technology in the future.

## Methods

### Cohort characteristics and survival analysis

This study used formalin-fixed paraffin preserved tissue micro-array specimens from a consecutive Swedish cohort, which is described elsewhere<sup>30,31</sup>.

A total of 1–2 tissue cores were taken from full-face tissue sections from 144 BC patients with invasive breast cancer who were diagnosed at Malmö University Hospital, Malmö, Sweden. Patients within this cohort did not receive neoadjuvant therapy and were treated with radical local mastectomy or wide local excision<sup>32</sup>.

Full details on the cohort characteristics are shown in Table 1 (nodal status was known for 130 patients). Due to extensive core loss, image data from a number of patients were excluded from analysis (5 of the ER+ patients, 3 HER2+).

In total, after imaging, a dataset from 139 patients was used for machine learning analysis, with 29 experiencing future recurrence of the disease (all but one of which were ER+).

Kaplan–Meier survival analysis was conducted on the expression levels for a set of hormone receptors and molecular markers which were available for the cohort. This analysis was conducted in Python v. 3.9 and employed the lifelines package (v.0.29.0). The log-rank test was used for the calculation of chi-squared and *p* values. A visualization of the survival functions with respect to a selection of targets is provided in Fig. 1a.

### FTIR chemical imaging

A 5  $\mu\text{m}$  thick slice of each TMA block was taken for FTIR imaging, with two contiguous sections taken for haematoxylin-eosin and immunohistochemical staining.

FTIR imaging was conducted using an Agilent Cary 620 IR microscope coupled with an Agilent Cary 670 FTIR spectrometer at the University of Exeter. This instrument uses a  $128 \times 128$  pixel focal plane array, with  $2 \times 2$  mosaic tiles capturing the full TMA core (typically 750 mm diameter). Spectra were collected at a spectral resolution of  $16 \text{ cm}^{-1}$  at 4 scans per pixel over the fingerprint region from  $1000\text{--}1800 \text{ cm}^{-1}$ , where the number of scans was chosen to optimize the signal-to-noise ratio. A total of 245 images were selected to form the dataset in the study.

All samples were imaged without chemical dewaxation. Samples were excluded based on visual examination during collection; more intrusive collections of paraffin wax on the tissue resulted in exclusion and those with especially sparse tissue cores were not imaged. A sample FTIR spectrum of the tissue measured in this work is presented in Fig. 1b.

### FTIR image and spectral preprocessing

Prior to using FTIR spectral data for analysis, various pre-processing steps were implemented as shown in Fig. 2B. Specifically, the spectral baseline contribution from resonant Mie scattering was removed using a single iteration of the resonant Mie scattering-EMSC algorithm employing Matrigel as the reference spectrum<sup>33–35</sup>. This step was conducted in Matlab (R2021b, the Mathworks Inc.), with all subsequent processing and modeling steps conducted in Python 3.9, with sci-kit learn v.1.0 and tensorflow v.2.8.

Detection of pixels from the slide substrate around the image of the TMA core was performed using a *k*-means cluster algorithm with *k* = 2, with the image array padded with zeros for those pixels in which no tissue was present. A collage of spectral images from the TMA are provided in Fig. 1c, together with images of haematoxylin-eosin stained tissue sections which were cut contiguously to the section for spectral imaging, and which are paired with each corresponding spectral image. False-colored spectral images were generated from the application of crisp-cluster validity indices as described previously<sup>36,37</sup>.

### ML and DL pipeline development

The following represents a summary of the ML and DL strategies used in the study. Figure 2C, D provides a pictorial visualization of the set of ML and DL strategies employed in this study. For each individual model and modeling strategy training was conducted using images randomly selected from 60% of the patient cohort, validation with 20% and the remaining 20% held out for testing. A number of complete training-validation-testing rounds, within which image data were randomly sorted between training, validation and testing sets, was conducted for each model to assess model robustness. The sections following detail the approach in respect of each model type.

To mine information from the spatial and spectral dimensions, one significant obstacle is the sheer size of each chemical image in data terms. As each image has an average size of 25 Mb, RAM drains quickly during training of ML and DL algorithms, even in scenario's where high memory

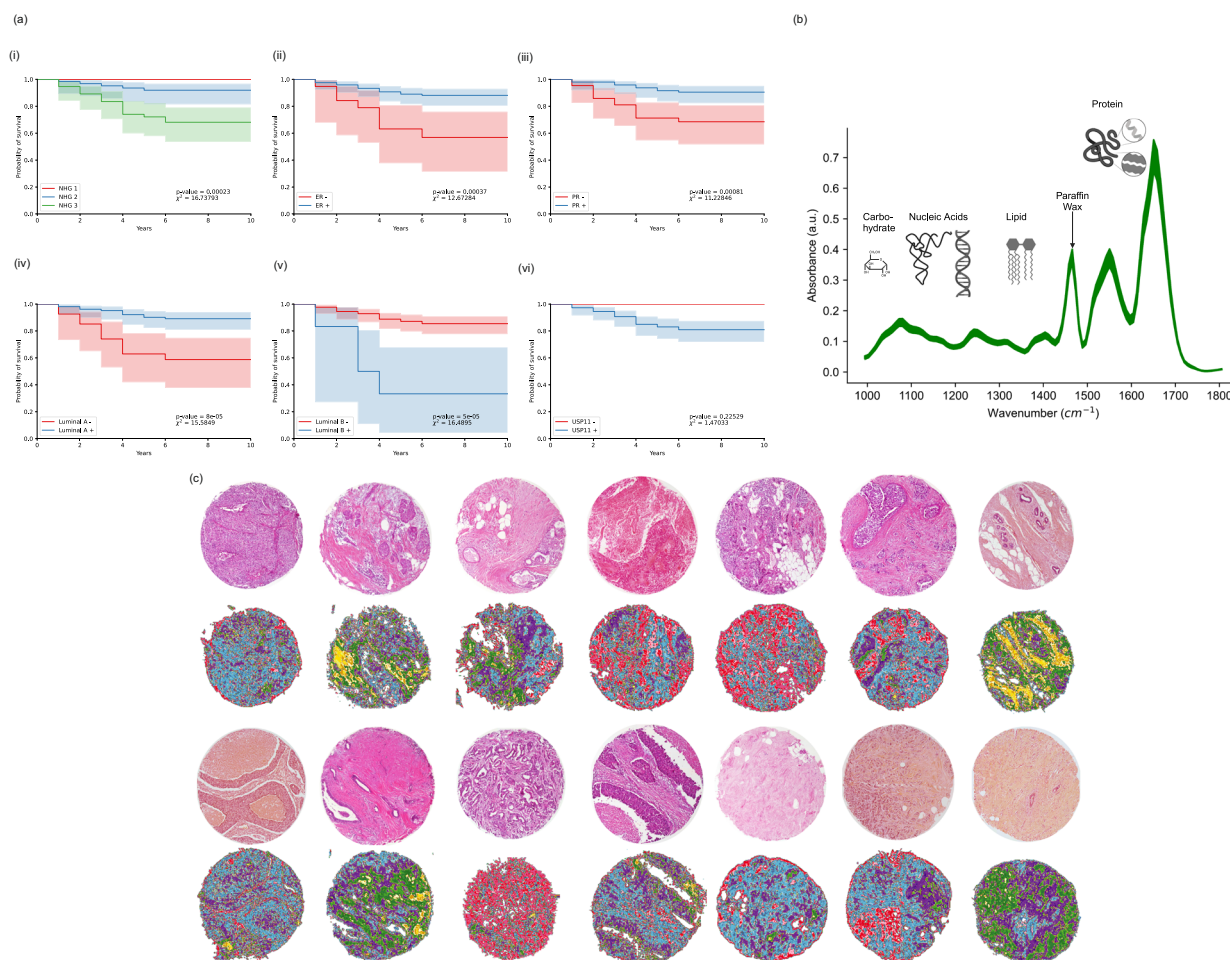
**Table 1 | Cohort characteristics including endocrine receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor (HER2) positivity, together with lymph node (LN) status, recurrence free survival (RFS) and overall survival**

| Characteristic                                             | Value       |
|------------------------------------------------------------|-------------|
| Age at diagnosis (years)                                   | 65 (13.5)   |
| ER+/-                                                      | 125/19      |
| PR+/-                                                      | 100/44      |
| HER2+/-                                                    | 51/93       |
| LN+/-                                                      | 56/74       |
| RFS (years)                                                | 5.45 (2.2)  |
| OS (years)                                                 | 5.78 (1.88) |
| NHG (1/2/3)                                                | 22/64/57    |
| Therapy type (radiotherapy/chemotherapy/endocrine therapy) | 60/113/96   |

Tumor grade is encapsulated as Nottingham Histological Grade (NHG) score.

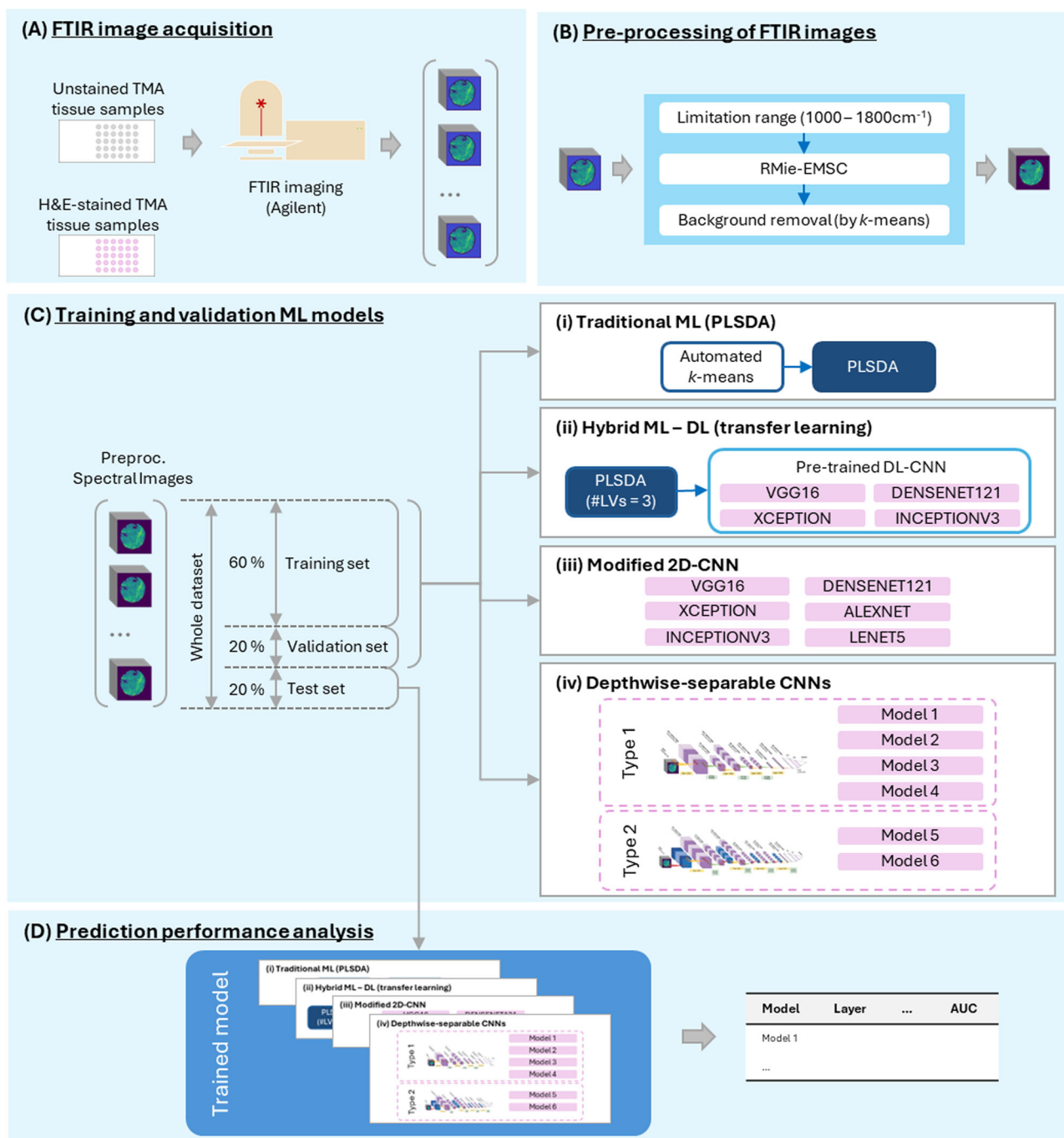
and parallelization of execution are available. Of course, the high dimensionality of the samples also complicates the learning process for a ML algorithm. As a consequence in the present work we have adopted several approaches to the implementation and training of ML models, limited by the aforementioned factors, for prediction of future recurrence with chemical imaging data:

- The first was to apply a traditional ML approach to prediction of recurrence with chemical imaging data<sup>36</sup>. Here images were segmented using an automated hierarchical double  $k$ -Means clustering algorithm which employed crisp cluster validity indices and which was applied across the entire dataset employ with  $k_{\min} = 2$  and  $k_{\max} = 5^{38,39}$ . Subsequently spectra from each cluster were used to train individual partial-least squares discriminant analysis (PLSDA) models for the predictive task (Fig. 2C (i)) with the PLSDA models built using 1–50 latent variables, and the optimal model complexity found for each iteration using an F1-score on the validation set. This process was repeated in its entirety on 10 occasions.
- The second was to employ a hybrid ML-DL transfer learning approach, where freely available 2-dimensional DL- CNN network architectures (VGG16<sup>40</sup>, Xception<sup>41</sup>, InceptionV3<sup>42</sup>, and DenseNet121<sup>43,44</sup>) were employed. For compression of the spectral dimension of these images to 3 channel variables a PLSDA model with 3 latent variables was built across a training set



**Fig. 1 | Cohort survival analysis, with introduction to Fourier-Transform infrared chemical imaging for histopathology. a** Kaplan-Meier survival analysis for a selection of histological, physiological and molecular markers within the cohort, including (i) Nottingham histological grade, (ii) estrogen receptor status (iii) progesterone receptor status (iv) luminal type A, (v) luminal type B and USP11 positivity. **b** Sample FTIR spectra of BC tissue, with important bands of interest annotated biochemically. **c** Collage of

haematoxylin-eosin stained images of TMA cores with matched FTIR chemical images of corresponding serial, unstained section (First and third row—H&E stained cores; second and fourth rows—serial chemical image). FTIR chemical images are false colored on the basis of a double application hierarchical auto-k-means clustering (with  $k$  varying from 2 to 5). This method employed the Pakhira-Bandyopadhyay-Maulik index for optimization of the number of clusters across all FTIR images acquired in the study.

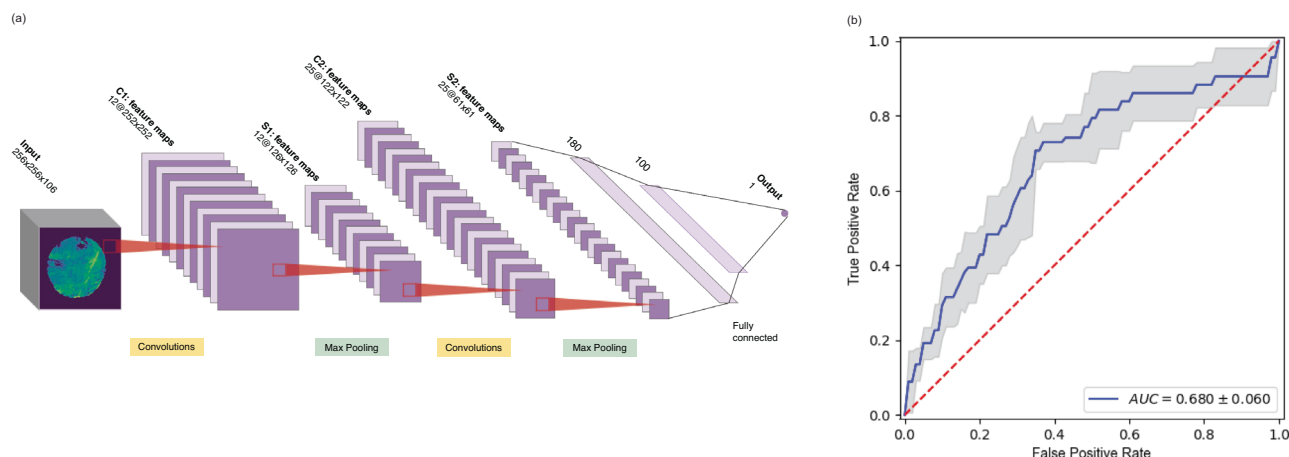


**Fig. 2 | Visualization of data acquisition to modeling pipelines employed within the study.** In (A), Fourier Transform Infrared (FTIR) imaging data were acquired from unstained TMA sections using a serial H&E section as ground truth. In (B), image data were preprocessed to truncate the spectral region to the fingerprint region, then to remove a spectral background and finally, to remove pixels

surrounding the tissue which did not contain tissue spectra. In (C), the set of ML + DL modeling approaches for the prediction of recurrence are detailed, with prediction performance evaluated across all modelling approaches using the receiver-operator-characteristic area-under-the curve (D).

of spectra in the image dataset and the optimal model at training was used to construct a secondary image dataset. The PLS embeddings for each spectrum were resolved spatially such that each pixel of the image data was made up of the three PLS scores for each spectrum at that point. The test set was transformed using this model. The 3-channel image dataset was then passed through the early layers of pre-trained deep learning networks listed previously, with 2 fully connected layers added at the end for classification of recurrence (Fig. 2C (ii)).

(c) In the third approach the networks in (b) were modified to allow input of raw, hyperspectral, chemical imaging data for binary classification, and were trained from scratch. This approach and approach (b) investigated the hypothesis that coupling spectral information to spatial information would provide improved predictive performance (Fig. 2C (iii)). Full FTIR spectral images were fed directly into networks with architectures similar to well-known networks including those listed in (b) above, and both AlexNet<sup>45</sup> and LeNet5<sup>46</sup>. The initial layers of each of these networks were modified and adapted to the dimensions



**Fig. 3 | Modified LeNet5 deep learning model architecture with performance in predicting BC survival post-therapy. a** Modified LeNet5 deep learning architecture for binary classification with chemical imaging data. **b** Typical performance of the

model in (a) in predicting BC survival post therapy. Variability in performance is depicted as the shaded area around the curve.

of our chemical images ( $256 \times 256$  pixels  $\times$  106 channels). Figure 3a depicts the modifications employed for one of the architectures, the LeNet5 network. A fully connected layer replaces the third convolutional layer, and the first two convolutional layers have a variable number of filters. In all the networks the convolution filters used were of dimension  $5 \times 5$ , and pooling windows were size  $2 \times 2$ , with a stride of 1 in each spatial dimension. Convolution layer kernel weights and biases were initialized using the *keras* built-in uniform He-initializer, and activated using the ReLu function, with training initially for 500 epochs. The data was again split by patient in a random, stratified manner into training, validation and test sets in a 60:20:20 ratio. Model performance was expressed in terms of the receiver-operator characteristic area-under-curve of each test set, and this splitting, training and testing was performed 5 times. Following the emergence of a leading model (LeNet5), the effects of batch size on the performance of the modified version of LeNet5 were examined. Batch sizes of 8, 16, 20, and 32 were examined for prediction of recurrence, along with batch normalization (BN) between convolution and pooling layers. A *k*-fold train test split with 5 folds, with a constant random seed, was used to form training and testing sets from here on. These models, and all subsequent models (unless otherwise stated) were trained for 150–200 epochs. For regularization, if the training loss did not decrease over 60 epochs, training was halted using the *keras* callback early stopping option, and the model parameters at the lowest loss value were saved. The average test ROC AUC over the 5 folds is reported.

- (d) In the final approach, two avenues for deeper mining of spectral information of approaches were analyzed, involving (i) the implementation of 3D CNN layers<sup>47</sup> and (ii) the implementation of depth-wise separable CNN (DSCNN) layers<sup>41,48</sup> (Fig. 2C (iv)).

A 3D filter is a cube which convolves over the entire image resulting a 3D feature map of the same or of a slightly reduced size. This introduces two major sources of RAM drainage. Firstly, layers in a 3D-DL-CNN comprise of several filters, each of which outputs several 3D feature maps which must then be input into the layer following, all of which absorbs substantial amounts of memory. Secondly, the number of learnable parameters in a 3D convolution layer is increased significantly. Here the implementation of 3D-CNN layers proved infeasible owing to RAM drainage and this aspect was not explored further.

As an alternative depth-wise separable convolution (DSC) layers were employed, replacing the 2D convolution layers in the champion network emerging from approach (c) to allow deeper mining of the spectral information. Additionally, other network architectures built from DSC layers were trialed for this task, with modifications made to result in lighter models.

Approach (a) trained a model solely on spectral data, and examined one subset of data at a time. Approach (b) clearly does not leverage the rich spectral information within the chemical image, but does allow mining of the spatial distribution of features within the image. Similarly in approach (c) each convolutional filter is passed over all channels individually, and the resulting 3D feature map is flattened to a single channel array by vector addition such that a single channel input is provided to the remaining layers.

Approach (d) on the other hand does allow mining of the spectral information and spatial information in which the 3D feature map is not flattened by simple vector addition, but instead is used as a sublayer input for the convolution in the spectral dimension which follows, by passing a 1D-convolution kernel with learnable weights over the spectral information, or image channels (Fig. 4). We have employed two implementations of this approach, which we here refer to as *Type 1* (Fig. 4a) and *Type 2* (Fig. 4b).

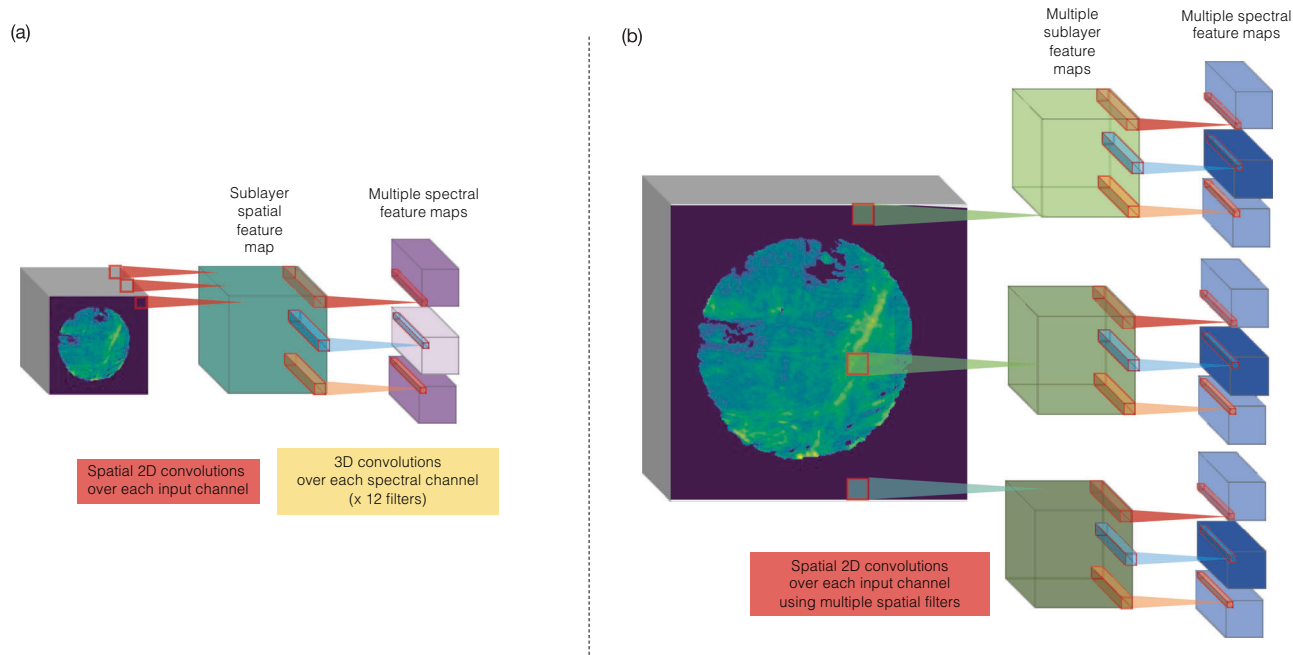
In *Type 1*, a single filter ( $x$ - $y$ ) is passed over the spatial image dimensions as a standard 2D filter. At this step, several 1D filters are passed over the spectral dimension of this feature map, to produce several 3D feature maps, similar to the operation of a standard 3D convolution layer. This layer will then mine mostly spectral information with some spatial components. Type 1 DSCs are implemented as shown within the network in Fig. 5a.

In the *Type 2* approach, several ( $x$ - $y$ ) filters are passed over the input chemical image, producing sublayer feature maps. Again, 1D filters are then passed over these filter maps in the spectral ( $z$ )-dimension. This results in rapidly increasing amounts of high dimensional feature maps and provides an architecture which provides deeper mining of both the spatial and spectral information within the image. Type 2 DSCs are implemented as shown within the network in Fig. 5b.

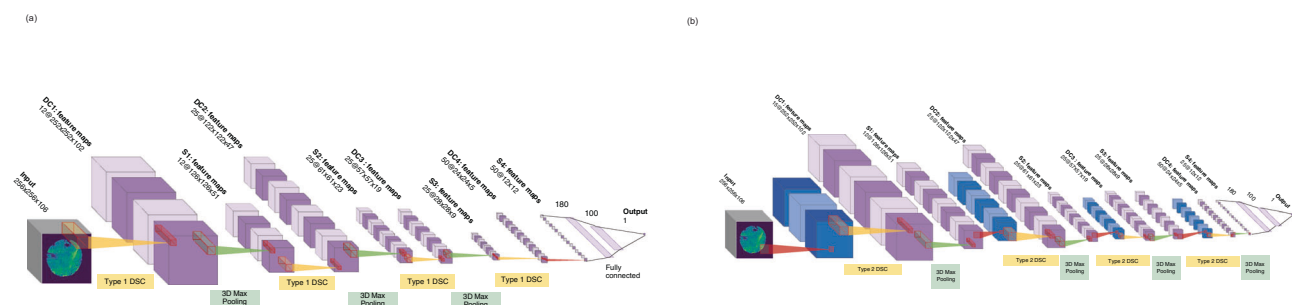
Ultimately both of these approaches allowed us to increase the degree to which the spectral information was mined by our DL algorithms using various approaches:

- (1) Initially, a single adjustment was made to the modified LeNet5 architecture, whereby the input layer 2D convolutions were replaced with a separable convolution layer (Model 1). Following that, Models 2, 3, and 4 were constructed using Type 1 depthwise-separable convolution layers (Fig. 5a). These models varied in the rate at which the spectral dimension was reduced, by varying pooling types and window size, as well as the number of feature maps produced per layer. The first layers of each model were the same, consisting of a depthwise-separable convolution (DSC) with 12 individual filters of dimension  $5 \times 5 \times 5$  pixels, followed by 3D Max pooling with dimension  $2 \times 2 \times 2$  pixels, another DSC layer with 25 individual  $5 \times 5 \times 5$  pixel filters, and a 3D Max pooling layer.

Models 2 and 3 had a third DSC with 25 individual  $5 \times 5 \times 5$  pixel filters. The latter half of Model 2 was made up of a 3D max pooling



**Fig. 4 | Operation of depth wise separable convolutions.** Visualization of the operation of (a) Type 1 and (b) Type 2 depthwise separable convolution operation employed in this work.



**Fig. 5 | Depth-wise separable CNN architectures for prediction of treatment outcomes with chemical imaging data.** a Depthwise separable convolutional neural network with Type 1 separable convolutions. Depthwise separable convolutional (DSC) operations are shown in yellow with max-pooling operations shown in green.

b Depthwise separable convolutional neural network with Type 2 separable convolutions. Depthwise separable convolutional (DSC) operations are shown in yellow with max-pooling operations shown in green.

layer, a DSC layer with 50 individual  $5 \times 5 \times 5$  pixel filters and a second 3D max pooling layer. Model 3 had a 1D average pooling layer following the third DSC, collapsing the spectral dimension of the feature maps entirely, and a 2D max pooling layer ( $2 \times 2$ ) to condense the spatial dimension. Model 4 had one DSC with 50 individual  $5 \times 5 \times 5$  pixel filters following the 2nd pooling layer, after which a 3D global average pooling layer was employed. For each of these models, the remaining features were flattened to one dimension, which were fed into a 180 node fully connected layer, followed by a fully connected layer of dimension 100, followed by the output layer that had dimension 1, and which utilized the sigmoid activation function. Each model was trained using a binary cross-entropy loss function and an adam optimizer. All models were trained with accuracy used as the metric of performance, with weighting for class imbalance.

- (2) Model 5 was constructed using Type 2 depthwise-separable convolution layers, as shown in Fig. 5b. The first layer of Model 5 is a type 2 DSCNN made up of  $5 \times 3$  filters, with 5 of dimension  $5 \times 5 \times 1$  pixels, followed by 3 of dimension  $1 \times 1 \times 5$  pixels, resulting in  $15 \times 3$ D feature maps used to input to the subsequent 3D max pooling layer, with dimension  $2 \times 2 \times 2$  pixels. The next type 2 DSC layer was made up of  $5 \times 5$  pixel filters, of the same dimensions, and second 3D max pooling

layer. This was repeated for the 3rd DSC and pooling layers. The final type 2 DSC layer consisted of  $5 \times 10$  pixel filters, resulting in 50 feature maps to be pooled using a 3D max pooling layer with window size  $2 \times 2 \times 5$  pixels. The resulting feature maps were flattened and fed into two fully connected layers of 180 and 100 nodes. The output was from a single node activated by a sigmoid function.

To further combat RAM issues, a model designed to reduce the size of feature maps was then designed. The input was reduced using a 3D average pooling window of size  $4 \times 4 \times 4$  pixels, before DSC type 1 convolutional filters of greater dimension ( $8 \times 8 \times 8$  pixels and  $3 \times 3 \times 5$  pixels) were applied. After two DSC layers and pooling layers, the spectral dimension was reduced such that a 2D convolution layer and max pooling layer was placed before the feature maps were flattened and fed into the fully connected layers and class was output.

## Results

### Hybrid ML (PLSDA)—TL modeling

Models developed using an image segmentation—traditional ML pipeline with PLSDA were overall poor at predicting BCR, with no image segment or spectral subset observed as a prominent marker of recurrence.

**Table 2 | Performance of transfer learning approach with image dimensionality reduction via PLSDA**

| DL network  | ROC AUC       |
|-------------|---------------|
| VGG16       | 0.500 ± 0.000 |
| Xception    | 0.524 ± 0.060 |
| InceptionV3 | 0.499 ± 0.001 |
| DenseNet121 | 0.500 ± 0.000 |

**Table 3 | Performance of modified 2D CNN networks in predicting BC recurrence**

| DL network  | ROC AUC              |
|-------------|----------------------|
| VGG16       | 0.515 ± 0.030        |
| Xception    | 0.524 ± 0.028        |
| InceptionV3 | 0.577 ± 0.054        |
| DenseNet121 | 0.536 ± 0.075        |
| AlexNet     | 0.483 ± 0.023        |
| LeNet5      | <b>0.639 ± 0.081</b> |

The value in bold denotes the highest classification performance achieved across all network architectures.

Similarly, using the primary PLS-fed transfer learning approach, no model was successful in classifying BC recurrence (BCR). The performances of CNN models trained for classification of BCR using this dataset are listed in Table 2.

### Modified 2-D CNNs

The performances of the networks in classifying both BCR and survival after therapy BC are presented in Table 3, and model ROCs across all modeling instances are visualized in Figs. 3b and 6. InceptionV3 and LeNet5 were capable of classifying recurrence with ROC-AUCs of  $0.577 \pm 0.054$  and  $0.639 \pm 0.081$ , respectively (an average receiver operator characteristic (ROC) for the LeNet5 architecture is depicted in Fig. 4b). The LeNet5 architecture has  $\sim 16 \times 10^6$  learnable parameters.

As the modified LeNet5 architecture was found to be the champion model here (Fig. 3b), an interrogation of the effect of batch size during training was considered, for classifying both BCR and early (up to 7-year) survival after therapy for BC. These results are shown in Table 4.

### Depthwise-separable CNNs

After substituting the first layer of the modified LeNet5 model with a 3D separable convolution layer, the average ROC AUC was  $0.641 \pm 0.095$ , similar to the purely 2D convolutional neural network. Models 2–6, which were built entirely using DSC layers produced a drop in performance. Tables 5 and 6 depict the overall results of these investigations, demonstrating the average receiver operator characteristic curves over 5 separate executions of each model. Figure 6 depicts the typical ROC profiles and AUCs observed for each of these models.

In 2D convolutional layers, the image channels (in this case the IR spectral dimension) collapses to a single value per pixel by vector addition. In Model 1, this was done in a learnable manner, by way of a separable convolutional layer. This did not have a significant deleterious effect on performance. Following this, maintaining the spectral dimension as a minable aspect of the data deeper in the networks was a focus of network design. Even the use of 3D depthwise-separable convolutions over 3D convolutions did not completely eradicate the RAM constraints associated with this type of network. While the number of learnable parameters within each of the DSC models were orders of magnitude lower than both the 2D CNNs, maintaining several 3D feature maps of a batch of samples in computer memory during training presented a significant obstacle. Batch

size could not be increased beyond 8 in models 2–5, which would have had an affect on the variability of the model performances.

Each of the 3D-DSCNN models were smaller in terms of parameters, with model 3 being larger than model 2 ( $\sim 3.5 \times 10^6$  versus  $1.3 \times 10^6$  parameters), and Model 4 being the smallest at  $\sim 32 \times 10^4$  parameters. Model 5 with type 2 DSC had  $\sim 1 \times 10^6$  parameters, and the model designed for greater dimensionality reduction, Model 6, had  $\sim 1 \times 10^5$  parameters. Each of these models were trained using batches of 8 samples, except Model 6 which was trained using a batch size of 16.

Of the type 1 DSCNN models, Model 3, despite being the largest DSCNN model type, exhibited the lowest performance. Model 4, the smallest model overall, had 3 DSC layers. Notably this model had the same depth as model 3, but exhibited superior performance. Here, instead of flattening the final feature maps for input to fully connected layers, which uses a lot of RAM, global average pooling layer was used to reduce these to single values. Model 2 was the deepest of the type 1 DSC models, but not the most complex, and was the best performing model by a slight margin.

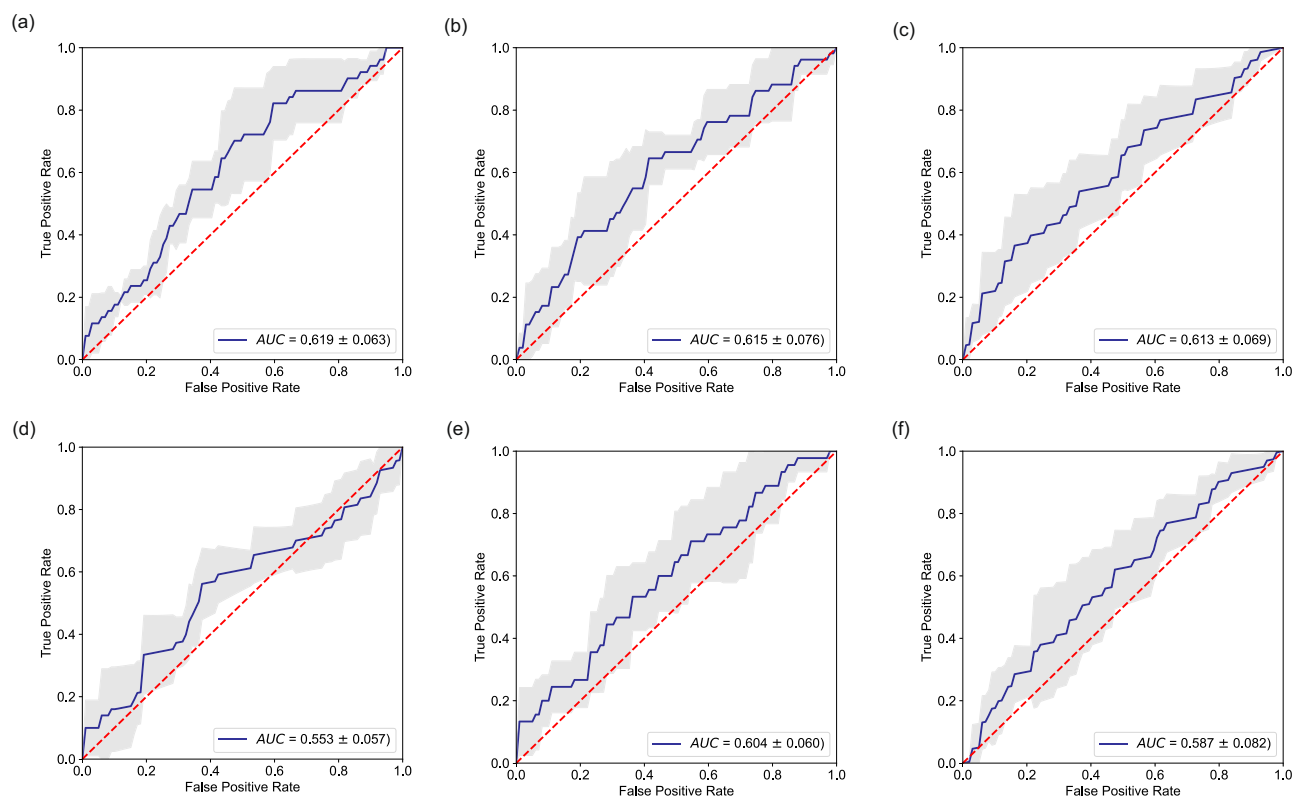
Using the type 2 DSCNN, Model 5, had a poorer performance, and took much longer to train, though this model was the only type to mine both the spatial and spectral dimensions to an equal extent. Reducing both spatial and spectral dimensions to produce more manageably sized feature maps was the goal of implementation of Model 6. This model had slightly poorer performance, and also had fewer learnable parameters than the best performing type 1 DSCNN model, Model 2.

## Discussion

Chemical imaging in histopathology is relatively well established as an approach for the development of label-free diagnostic pipelines for cancer<sup>22,25</sup>, with a recent focus being applied on the development of methods for prediction of post-treatment response<sup>13</sup>. Typically, pipeline development in the field has employed image segmentation followed by the application of machine learning approaches on spatially binned spectral data, for classification of disease type. This is a time consuming process requiring pooling or separation of groups of spectra, or intense dimensionality reduction. The central hypothesis of this study is that the use of deep learning approaches on chemical image data, where spectral information is not separated from its spatial context, will lead to improved performance in predicting disease recurrence, where spatial context within the tumor microenvironment is known to be a contributory factor<sup>49–51</sup>.

Here the performances of ML and hybrid ML-DL-TL models validate this hypothesis, suggesting that a more biochemically holistic image of the tissue sample is required. As clusters of FTIR spectra correspond loosely to subtypes of tissue, of which none individually could be used as a predictor of recurrence here, this result overall suggests that separating the spectral and spatial components of chemical imaging datasets is insufficient for the development of models of treatment outcome. However while it is acknowledged that a from-scratch approach to training of models with PLS embedded images might have provided improved performance, this aspect was not explored.

While the development of deep learning models with image data from scratch requires the use of very large datasets of the order of  $>10^6$  images, it has been demonstrated in many instances that robust models which do not overfit can be developed using small datasets<sup>52,53</sup>. In the context of the present study, it is firstly notable that, of the widely cited image based 2D-CNN architectures trained from scratch on chemical images, the shallowest, LeNet5, was the champion model, with Inception-V3 as the next best model. As the dataset used in this study is small compared to the datasets typically used for training of DL models, which typically comprise millions of images, the performance of the network was quite dependent on the training batch size, resulting in wide measures of uncertainty. Additionally, the sample set is imbalanced with respect to recurrence, with samples mainly confined to the negative class. This was mitigated to an extent during training by weighting samples of the recurrence class, but this imbalance still affected the training loss gradient. Overall, after varying the batch size from 8 to 16 to 32, a slight increase in average ROC AUC value was associated with smaller



**Fig. 6 | Sample receiver operator characteristics for DSCNN models.** **a** Model 2 (with batch normalization and a batch size of 8); **(b)** Model 2 (with batch normalization and a batch size of 16); **(c)** Model 3; **(d)** Model 4; **(e)** Model 5; **(f)** Model 6. The

reader is referred to the main text to obtain details on the modeling parameters for each model variant.

**Table 4 | Modified LeNet 5 model performance with classification task and batch normalization approach**

| Batch size | Class label  | ROC AUC              |
|------------|--------------|----------------------|
| 8          | BCR survival | 0.576 ± 0.080        |
| 8          | BCR survival | <b>0.680 ± 0.060</b> |
| 20         | BCR survival | 0.639 ± 0.081        |
| 8          | BCR          | 0.600 ± 0.149        |
| 16         | BCR          | <b>0.607 ± 0.094</b> |
| 32         | BCR          | 0.572 ± 0.063        |

The values in bold highlight the maximal classification performance observed.

batch sizes, but a slight reduction in uncertainty was associated with larger batch sizes. The task re-definition also made a small difference to the predictive performance of the models, where models predicting BC survival (with a batch size 8 and batch normalization) were most successful, having an average ROC AUC of  $0.68 \pm 0.06$ . This hints at the possible benefit of re-examining the definition of the task, perhaps excluding those patients with mortality by causes other than breast cancer prior to any potential recurrence event.

Again, we observed that while the batch training performance of the 2D-CNN model did not vary substantially, the result of this part of the analysis confirms that the coupling of spectral and spatial information within DL models leads overall to superior predictive performance for BCR, approaching that of validated proteomic and genomic tests in this space.

This result is again echoed in the continuation work which looked to develop depthwise-separable CNNs for more effective mining of the spectral information within FTIR chemical images. Overall that analysis suggests that an optimal depth and complexity exists for network models for this application, below which model performance is reduced. Type 1 DSCNN

models appear to be superior to Type 2 models for this application, with Model 2 producing a ROC AUC performance for prediction of BCR of  $0.628 \pm 0.095$ . This model was comparable in performance to the modified 2D-DCNN LeNet5 architecture (with ROC AUC of  $0.639 \pm 0.081$ ) and while it was the deepest of the Type 1 models, it was not the most complex. Therefore, the coupling between spectral and spatial information offered by chemical images is best leveraged through relatively shallow DL networks for prediction of future treatment outcomes, with no loss of performance seen when layer outputs are pooled harshly.

For contextualization of the performance of the networks designed in this work, equivalent studies in digital pathology provide a useful reference. Within the field of digital pathology substantial advances have been made towards the use of digital images acquired from standard H&E specimens within automated pathological support systems<sup>40,54</sup>. Within the breast cancer field the approach has successfully demonstrated the capability to classify disease grade<sup>55,56</sup>, subtype<sup>56,57</sup> and for automated image segmentation, including the identification of the location and distribution of tumor infiltrating lymphocytes, which is itself an important prognostic indicator<sup>49,58–60</sup>. More recently the field has demonstrated potential to classify survival probability<sup>61,62</sup> and risk of recurrence using OncoType-DX genetic risk scores as ground truth<sup>63</sup>. In particular Whitney et al.<sup>64</sup> have demonstrated the capability of digital pathological metrics to classify OncoType-DX scores at low versus intermediate/high levels with ROC AUC of 0.58 and low+intermediate versus high with a ROC AUC of 0.6. Similarly Howard et al. have classified high OncoType DX scores using a purely image-based model with ROC AUC of 0.797 (95% CI 0.68–0.901) and a model incorporating a clinical nomogram with a ROC AUC of 0.814 (0.709 to 0.901)<sup>65</sup>. As we have developed models for prediction of recurrence, including patients across all risk profiles, these studies are not directly comparable with our own, though the work of Whitney et al. does demonstrate a similar level of performance to our own for a similar predictive objective.

**Table 5 | Performance of modified LeNet5 model architecture with a separable convolutional layer as the first layer, and a Type 1 DSCN model with 4 convolutional layers for prediction of BCR and BC survival**

| Model   | Layer type | Depth | No. params         | Batch size | Class label  | ROC AUC                             |
|---------|------------|-------|--------------------|------------|--------------|-------------------------------------|
| Model 1 | S-2D CONV  | 4     | $16.7 \times 10^6$ | 8          | BCR survival | <b><math>0.641 \pm 0.095</math></b> |
| Model 1 | S-2D CONV  | 4     | $16.7 \times 10^6$ | 8          | BCR          | $0.600 \pm 0.122$                   |
| Model 2 | DSC1       | 6     | $1.3 \times 10^6$  | 8          | BCR survival | $0.608 \pm 0.068$                   |
| Model 2 | DSC1       | 6     | $1.3 \times 10^6$  | 8          | BCR          | <b><math>0.628 \pm 0.098</math></b> |

Here, depth is number of layers with learnable parameters, excluding the output single node layer.

Values in bold denote the highest classification performance achieved for each network architecture and classification task.

**Table 6 | Performance of Type 1 and Type 2 depthwise separable convolution models for prediction of BCR**

| Model   | Layer type | Depth | No. params        | Batch size | Class label | ROC AUC           |
|---------|------------|-------|-------------------|------------|-------------|-------------------|
| Model 3 | DSC1       | 5     | $3.5 \times 10^6$ | 8          | BCR         | $0.521 \pm 0.063$ |
| Model 4 | DSC1       | 5     | $3.2 \times 10^4$ | 8          | BCR         | $0.557 \pm 0.068$ |
| Model 5 | DSC2       | 6     | $1.3 \times 10^6$ | 8          | BCR         | $0.588 \pm 0.131$ |
| Model 6 | POOL-DSC1  | 5     | $1.0 \times 10^5$ | 16         | BCR         | $0.546 \pm 0.065$ |

Here, depth is number of layers with learnable parameters, excluding the output single node layer.

As genetic alternations of tissue precede morphological changes that are diagnostic or prognostic<sup>66,67</sup>, downstream chemical changes to tissue owing to alterations in transcription and protein translation will be visible spectroscopically prior to their morphological manifestation. Critically, H&E based digital pathology only assesses spatial and morphological metrics for diagnostic and prognostic purposes, while chemical imaging measures the intrinsic biochemistry of tissues spatially. Our approach develops models whose performance exceeds those of purely AI-based histomorphometric approaches, suggesting the existence of spatial chemical signatures which may offer a performance advantage for prediction of recurrence. Practically this approach may also have application in a number of contexts, as (1) a first stage triage for recurrence prediction, or (2) in surgery in a similar fashion to that seen recently with stimulated Raman imaging<sup>17</sup>, allowing for near real-time decision making in a surgical context. Ultimately this label-free all digital approach may hold promise in improving estimates of recurrence within breast cancer, particularly for those patients with low risk of recurrence.

Ultimately, while bulk molecular assays of pathological samples do provide an avenue for the development of validated prognostic tests for cancer treatment outcome, their explanatory utility for cell function is limited given the absence of spatial context. Chemical imaging, and similar deep spatial biochemical profiling technologies do provide a novel opportunity for the development of prognostic technologies that also provide insights into spatial context in a label-free manner<sup>24</sup>. Recent investigations have demonstrated that mainstream two-dimensional convolutional networks do not effectively mine the rich, exquisitely sensitive, molecular information within these images, rather mining the spatial information within the images<sup>68</sup>.

As exemplified by this study, analytical pipelines can be developed that have prognostic value in early stage breast cancer, which have demonstrated performance in prediction of all breast cancer recurrence events (ROC AUC of approximately 0.64) which aligns well with bulk molecular tests and AI-detected histomorphometric tests in this space. Our work demonstrates that appropriately designed convolutional neural networks are capable of learning the spectral and spatial characteristics within breast tissue pathology to provide robust prognostics in a strongly heterogeneous context. Importantly this study demonstrates for the first time that vibrational spectra of histopathological samples, and their spatial context, are signatures with prognostic value which may have potential for more widespread use in objective, label-free all-digital chemical imaging driven histopathological prognostics. Future work will examine the potential of this pipeline for prediction of post-treatment survival probability, and other treatment

endpoints in BC and other cancers, for further enhancement of the power of the technology.

### Data availability

Data associated with this work may be shared with interested parties upon reasonable request.

### Code availability

Code associated with this work may be shared with interested parties upon reasonable request.

Received: 30 March 2024; Accepted: 25 November 2024;

Published online: 17 January 2025

### References

- Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Harnan, S. et al. Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer: a systematic review and economic analysis. *Health Technol. Assess.* **23**, 1–327 (2019).
- Galea, M. H., Blamey, R. W., Elston, C. E. & Ellis, I. O. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res. Treat.* **22**, 207–219 (1992).
- Gray, E., Donten, A., Payne, K. & Hall, P. S. Survival estimates stratified by the Nottingham Prognostic Index for early breast cancer: a systematic review and meta-analysis of observational studies. *Syst. Rev.* **7**, <https://doi.org/10.1186/S13643-018-0803-9> (2018).
- Paik, S. et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J. Clin. Oncol.* **24**, 3726–3734 (2006).
- Glas, A. M. et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* **7**, 278 (2006).
- Loughman, T. et al. Analytical validation of a novel 6-gene signature for prediction of distant recurrence in estrogen receptor-positive, HER2-negative, early-stage breast cancer. *Clin. Chem.* **68**, 837–847 (2022).
- Buus, R. et al. Validation of the OncoMasTR risk score in estrogen receptor-positive/HER2-negative patients: a TransATAC study. *Clin. Cancer Res.* **26**, 623–631 (2020).
- Filipits, M. et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin. Cancer Res.* **17**, 6012–6020 (2011).

10. Bartlett, J. M. S. et al. Comparing breast cancer multiparameter tests in the OPTIMA prelim trial: no test is more equal than the others. *J. Natl. Cancer Inst.* **108**, <https://doi.org/10.1093/JNCI/DJW050> (2016).
11. Goldstein, L. J. et al. Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features. *J. Clin. Oncol.* **26**, 4063–4071 (2008).
12. Marchionni, L., Afsari, B., Gernan, D. & Leek, J. T. A simple and reproducible breast cancer prognostic test. *BMC Genomics* **14**, 336 (2013).
13. Tiwari, S. et al. INFORM: INFRared-based ORganizational Measurements of tumor and its microenvironment to predict patient survival. *Sci. Adv.* **7**, [https://doi.org/10.1126/SCIADV.ABB8292/SUPPL\\_FILE/SCIADV.ABB8292\\_SM.PDF](https://doi.org/10.1126/SCIADV.ABB8292/SUPPL_FILE/SCIADV.ABB8292_SM.PDF) (2021).
14. Pavlova, N. N. & Thompson, C. B. The emerging hallmarks of cancer metabolism. *Cell Metab.* **23**, 27–47 (2016).
15. Shi, Y. et al. Predicting early breast cancer recurrence from histopathological images in the Carolina Breast Cancer Study. *NPJ Breast Cancer* **9**, 92 (2023).
16. Bhargava, R. Digital histopathology by infrared spectroscopic imaging. *Annu. Rev. Anal. Chem.* **16**, 205–230 (2023).
17. Hollon, T. C. et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat. Med.* **26**, 52–58 (2020).
18. Hollon, T. C. & Orringer, D. A. An automated tissue-to-diagnosis pipeline using intraoperative stimulated Raman histology and deep learning. *Mol. Cell. Oncol.* **7**, 1736742 (2020).
19. Groberueschkamp, F., Jütte, H., Gerwert, K. & Tannapfel, A. Advances in digital pathology: from artificial intelligence to label-free imaging. *Visc. Med.* **37**, 482–490 (2021).
20. Diem, M. et al. Molecular pathology via IR and Raman spectral imaging. *J. Biophotonics* **6**, 855–886 (2013).
21. Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta* **1767**, 1073–1101 (2007).
22. Mittal, S. et al. Simultaneous cancer and tumor microenvironment subtyping using confocal infrared microscopy for all-digital molecular histopathology. *Proc. Natl Acad. Sci. USA* **115**, E5651–E5660 (2018).
23. Wald, N. & Goormaghtigh, E. Infrared imaging of primary melanomas reveals hints of regional and distant metastases. *Analyst* **140**, 2144–2155 (2015).
24. Sigle, M. et al. Translating genomic tools to Raman spectroscopy analysis enables high-dimensional tissue characterization on molecular resolution. *Nat. Commun.* **14**, 1–16 (2023).
25. Berisha, S. et al. Deep learning for FTIR histology: leveraging spatial and spectral features with convolutional neural networks. *Analyst* **144**, 1642–1653 (2019).
26. Kelleher, J. D. *Deep Learning* 1st edn (MIT Press, 2019).
27. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
28. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
29. Signoroni, A., Savardi, M., Baronio, A. & Benini, S. Deep learning meets hyperspectral image analysis: a multidisciplinary review. *J. Imaging* **5**, 52 (2019).
30. Magnusson, K. et al. ANLN is a prognostic biomarker independent of Ki-67 and essential for cell cycle progression in primary breast cancer. *BMC Cancer* **16**, 904 (2016).
31. Medrek, C., Pontén, F., Jirstrom, K. & Leandersson, K. The presence of tumor associated macrophages in tumor stroma as a prognostic marker for breast cancer patients. *BMC Cancer* **12**, 306 (2012).
32. Svensson, K. J. et al. Chondroitin sulfate expression predicts poor outcome in breast cancer. *Int. J. Oncol.* **39**, 1421–1428 (2011).
33. Lima, F. A. D. et al. Digital de-waxing on FTIR images. *Analyst* **142**, 1358–1370 (2017).
34. Bassan, P. et al. Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples. *Analyst* **135**, 268–277 (2010).
35. Bassan, P. et al. RMieS-EMSC correction for infrared spectra of biological cells: extension using full Mie theory and GPU computing. *J. Biophotonics* **3**, 609–620 (2010).
36. Keogan, A. et al. Chemical imaging and machine learning for sub-classification of oesophageal tissue histology. *Transl. Biophotonics* **3**, e202100004 (2021).
37. Nguyen, T. N. Q. et al. Prediction of pathological response to neo-adjuvant chemoradiotherapy for oesophageal cancer using vibrational spectroscopy. *Transl. Biophotonics* **3**, 1–13 (2021).
38. Nguyen, T. N. Q. et al. Fully unsupervised inter-individual IR spectral histology of paraffinized tissue sections of normal colon. *J. Biophotonics* **9**, 521–532 (2016).
39. Nguyen, T. N. Q., Jeannesson, P., Groh, A., Guenot, D. & Gobinet, C. Development of a hierarchical double application of crisp cluster validity indices: a proof-of-concept study for automated FTIR spectral histology. *Analyst* **140**, 2439–2448 (2015).
40. Mandair, D., Reis-Filho, J. S. & Ashworth, A. Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology. *Npj Breast Cancer* **9**, 1–11 (2023).
41. Chollet, F. Xception: deep learning with depthwise separable convolutions. in *Proc. — 30th IEEE Conference on Computer Vision and Pattern Recognition CVPR*, 1800–1807. <https://doi.org/10.1109/CVPR.2017.195> (2017).
42. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. in *33rd International Conference on Learning Representations, ICLR 2015 — Conference Track Proceedings*. <https://arxiv.org/abs/1409.1556v6> (2015).
43. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2818–2826. <https://doi.org/10.1109/CVPR.2016.308> (2016).
44. Huang, G., Liu, Z., Maaten, L. V. D. & Weinberger, K. Q. Densely connected convolutional networks. in *Proc. — 30th IEEE Conference on Computer Vision and Pattern Recognition CVPR*, 2261–2269 (2017).
45. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017).
46. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2323 (1998).
47. Tran, D., Bourdev, L., Fergus, R., Torresani, L. & Paluri, M. Learning spatiotemporal features with 3D convolutional networks. in *2015 IEEE International Conference on Computer Vision (ICCV)* 4489–4497 (IEEE Computer Society, 2015).
48. Dong, H., Zhang, L. & Zou, B. PolSAR image classification with lightweight 3D convolutional networks. *Remote Sens.* **12**, 396 (2020).
49. Page, D. B. et al. Spatial analyses of immune cell infiltration in cancer: current methods and future directions: a report of the International Immuno-Oncology Biomarker Working Group on Breast Cancer. *J. Pathol.* **260**, 514–532 (2023).
50. Jahangir, C. A. et al. Image-based multiplex immune profiling of cancer tissues: translational implications. A report of the International Immuno-oncology Biomarker Working Group on Breast Cancer. *J. Pathol.* **262**, <https://doi.org/10.1002/PATH.6238> (2024).
51. Fassler, D. J. et al. Spatial characterization of tumor-infiltrating lymphocytes and breast cancer progression. *Cancers* **14**, 9 (2022).
52. Brigato, L. & Iocchi, L. A close look at deep learning with small data. *arXiv* <http://arxiv.org/abs/2003.12843> (2020).
53. Vabalas, A., Gowen, E., Poliakov, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **14**, e0224365 (2019).
54. Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V. & Madabhushi, A. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **16**, 703–715 (2019).

55. Basavanthally, A. et al. Multi-field-of-view framework for distinguishing tumor grade in ER+ breast cancer from entire histopathology slides. *IEEE Trans. Biomed. Eng.* **60**, 2089 (2013).
  56. Araujo, T. et al. Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE* **12**, <https://doi.org/10.1371/JOURNAL.PONE.0177544> (2017).
  57. Couture, H. D. et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *Npj Breast Cancer* **4**, 1–8 (2018).
  58. Heindl, A. et al. Relevance of spatial heterogeneity of immune infiltration for predicting risk of recurrence after endocrine therapy of ER+ breast cancer. *J. Natl Cancer Inst.* **110**, 166–175 (2018).
  59. Bejnordi, B. E. et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Mod. Pathol.* **31**, 1502–1512 (2018).
  60. Yuan, Y. Modelling the spatial heterogeneity and molecular correlates of lymphocytic infiltration in triple-negative breast cancer. *J. R. Soc. Interface* **12**, <https://doi.org/10.1098/RSIF.2014.1153> (2015).
  61. Kather, J. N. et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med.* **16**, <https://doi.org/10.1371/JOURNAL.PMED.1002730> (2019).
  62. Beck, A. H. et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. Transl. Med.* **3**, <https://doi.org/10.1126/SCITRANSLMED.3002564> (2011).
  63. Romo-Bucheli, D., Janowczyk, A., Gilmore, H., Romero, E. & Madabhushi, A. A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers. *Cytom. Part J. Int. Soc. Anal. Cytol.* **91**, 566–573 (2017).
  64. Whitney, J. et al. Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer. *BMC Cancer* **18**, <https://doi.org/10.1186/S12885-018-4448-9> (2018).
  65. Howard, F. M. et al. Integration of clinical features and deep learning on pathology for the prediction of breast cancer recurrence assays and risk of recurrence. *Npj Breast Cancer* **9**, 1–6 (2023).
  66. Moynihan, F. et al. Concurrent and independent genetic alterations in the stromal and epithelial cells of mammary carcinoma: implications for tumorigenesis. *Cancer Res.* **60**, 2562–2566 (2000).
  67. Castro, N. P. et al. Evidence that molecular changes in cells occur before morphological alterations during the progression of breast ductal carcinoma. *Breast Cancer Res.* **10**, 1–14 (2008).
  68. Müller, D. et al. Dimensionality reduction for deep learning in infrared microscopy: a comparative computational survey. *Analyst* **148**, 5022–5032 (2023).
- and is co-funded under the European Regional Development Fund (ERDF) through Grant # 13/RC/2106\_P2. The authors also acknowledge funding via Royal Society Exchanges (grant IE170330). This work was also part-supported by the North-South Research Programme administered by the Higher Education Authority on behalf of the Department of Further and Higher Education, Research, Innovation and Science and the Shared Island Fund (AICRIstart: A Foundation Stone for the All-Island Cancer Research Institute (AICRI): Building Critical Mass in Precision Cancer Medicine), Irish Cancer Society (Collaborative Cancer Research Centre BREAST-PREDICT; CCRC13GAL; <https://www.breastpredict.com>), the Science Foundation Ireland Investigator Programme (OPTi-PREDICT; 15/IA/3104), the Science Foundation Ireland Strategic Partnership Programme (Precision Oncology Ireland; 18/SPP/3522; <https://www.precisiononcology.ie>).

## Author contributions

A.D.M. conceived the study with W.M.G. and A.R., and acquired funding. K.J. supplied the tissue for analysis. T.N.Q.N. and P.B. collected chemical images of the tissue. N.S. provided access to FTIR imaging facilities and advice on spectral sampling methods. A.K. and T.N.Q.N. performed various aspects of data analysis under the direction of A.D.M. A.K. and A.D.M. wrote and edited the manuscript. All authors contributed to manuscript review.

## Competing interests

W.M.G. declares that he is a part-time employee of OncoAssure Limited, as well as shareholder and chair of the OncoAssure Board.

## Additional information

**Correspondence** and requests for materials should be addressed to Aidan D. Meade.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

## Acknowledgements

This research was conducted with the financial support of Technological University Dublin (Fiosraigh Award), Science Foundation Ireland (Grant Agreement No. 19/FFP/6772), and that of the Health Research Board (Grant agreement no. HRA-POR-2015-1314) at the ADAPT SFI Research Centre at TU Dublin. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centers Programme