



OPEN

## Soft threshold partial least squares predicts the survival fraction of malignant glioma cells against different concentrations of methotrexate's derivatives

Tahir Mehmood<sup>✉</sup> & Mudassir Iqbal

Chemotherapy appeared to be a significant advancement in cancer research, with fewer side effects. Methotrexate (MTX) is a widely used anticancer drug with strong activity but serious side effects. Several MTX derivatives have been reported, with modifications at various sites to reduce side effects and increase efficacy. The current study uses FTIR spectroscopy to predict the survival fraction of human malignant glioma U87 (MG-U87) cell lines against MTX derivatives. Together with Parent MTX several aldehydes viz. Benzaldehyde, Chlorobenzaldehyde, 2-Chlorobenzaldehyde, 3-Nitrobenzaldehyde, 5-Chloro-2-hydroxybenz-aldehyde, 2-Hydroxy-5-Nitrobenzaldehyde, 2-Thiocarboxyaldehyde, Trans-2-pentenal, and Glutaraldehyde are treated with MTX to obtain MTX derivatives. The prediction of survival fraction of malignant glioma cells is carried out by Lasso, Elastic net and Soft PLS at different concentration levels of synthesized derivatives, including 400  $\mu\text{M}$ , 200  $\mu\text{M}$ , 100  $\mu\text{M}$ , 50  $\mu\text{M}$ , 25  $\mu\text{M}$  and 12.5  $\mu\text{M}$ . The cross-validated prediction error is minimised to optimise spectral wavelength selection and model parameters. It appears that the RMSE computed from test data is significantly varying with the change of models ( $p = 0.012$ ), with the change of concentrations levels ( $p \leq 0.001$ ) and with the change of combination of models and concentration level ( $p \leq 0.001$ ). StPLS outperforms in predicting survival fraction of glioma cells at the concentration level 50  $\mu\text{M}$ , 100  $\mu\text{M}$  and 400  $\mu\text{M}$  respectively with relative RMSE = 0.1, 0.14 and 0.55. Lasso outperforms at the concentration level 12.5  $\mu\text{M}$ , and 200  $\mu\text{M}$  respectively with relative RMSE = 0.4 and 0.14. Elastic net outperforms at the concentration level 25  $\mu\text{M}$  with relative RMSE = 0.8. Consistently appeared influential wavelength identifies the influential functional compounds which best predicts the survival fraction. Hence FTIR appears potential candidate for estimating survival fraction of MTX derivatives.

Methotrexate (MTX) is certainly one of the broadly used antimetabolites in cancer chemotherapy that terminates DNA replication mainly by snooping with thymidylate synthesis pathway<sup>1,2</sup>. It encompasses several cancer types for example leukemia, gestational trophoblastic disease, lung cancer, breast cancer, osteosarcoma and lymphoma<sup>3</sup>. Different levels of MTX's dosages are recommended for different types of diseases. For instance, low dose MTX treats Rheumatoid arthritis and other types of inflammatory diseases<sup>4</sup>, and high doses MTX is used for intraocular lymphoma<sup>5</sup>. Hence relevant concentrations of MTX are of importance. Together with the wide applications of MTX, it has some drawbacks as it induces life-threatening side effects on sensitive organs. Specific side effects of MTX include diarrhea, feeling weak, cough, increased chances of infection, reduced counts of white blood cell, Within mouth skin deterioration, liver failure, lung disease, lymphoma, and extreme skin rash<sup>6</sup>.

Modification in MTX structure enhances its anticancer activity<sup>7</sup> and is expected to minimize its side effects. MTX structure has several sites, where synthesized modifications in these sites results in MTX derivatives<sup>8</sup>. One possible way to measure the MTX and its derivatives' biological effect is to measure the cancer cell survival fraction. It reflects the cancer cells that have retained their reproductive capacity after applying the cancer therapy<sup>9</sup>.

School of Natural Sciences (SNS), National University of Sciences and Technology (NUST), Islamabad, Pakistan.  
✉email: tahime@gmail.com

This is customary presented through the survival curve, which describes the link between the absorbed dose and the fraction of cells that survive.

Cancer medications are believed to interact with cell metabolism by many pathways. For characterization of MTX derivatives Fourier Transformation Infrared (FTIR) spectroscopy is potential candidate which is low in cost and fast in processing, which is sensitive to all kinds of molecules present in cells, hence can provide unique and accurate fingerprints of samples. For example, FTIR has been used to predict the migration of glioma cell lines in vitro<sup>10</sup>. FTIR characterize metabolic difference of cardiotoxic steroid family<sup>11</sup>, The subtoxic doses of gemcitabine, a cancer drug, can be monitored using FTIR spectroscopy<sup>12</sup>, it can model the metabolism disruption track cancer cells<sup>13,14</sup>, it can diagnose arthritis(rheumatoid) through serum<sup>15</sup>, it can recognize most of the cell types found in melanoma cancer<sup>16</sup>, it can discriminate the anticancer medication reference to their mode of operation for prostate cancer cells<sup>17</sup>, and it can provide label-free surveillance of therapeutic medicines that is Busulfan and Methotrexate in human serum<sup>18</sup>.

Chemometric analyses are mostly based on widely used partial least squares (PLS) regression. PLS is being extensively used for drug concentration prediction, survival and fitness response prediction<sup>14-19</sup>. Soft thresholding PLS (StPLS) is a potential variant adds the influential wavelength selection together with response prediction<sup>20</sup>. Together with PLS, elastic net (Elnet) and least absolute shrinkage and selection operator (Lasso) are also in practice for modeling cancer cell functions<sup>21,22</sup>. These algorithms implements cyclical coordinate descent for computing the regularization path. These approaches have the inner ability to integrate knowledge from all kinds of molecules found in cells and create a spectral signature specific to a cell group. As a result, using FTIR spectroscopy, StPLS, Lasso, and Elnet can be used to predict the survival fraction of MTX derivatives against human malignant glioma U87 (MG-U87) cell lines. These methods work in two steps: model construction and validation.

In the current study, FTIR spectroscopy on human malignant glioma U87 (MG-U87) cell lines reveals that MTX derivatives develop a reproducible FTIR spectral signature at various concentrations for methotrexate's unique inhibition of DNA synthesis.

## Methods

**Material.** In this study together with Parent MTX schiff base MTX derivatives with target compounds several aldehydes viz. Benzaldehyde, Chlorobenzaldehyde, 2-Chlorobenzaldehyde, 3-Nitrobenzaldehyde, 5-Chloro-2-hydroxybenz-aldehyde, 2-Hydroxy-5-Nitrobenzaldehyde, 2-Thiocarboxyaldehyde, Trans-2-pental, and Glutaraldehyde are considered. MTX is a commonly used drug for various types of cancer. All the new derivatives were synthesized for the first time in study<sup>23</sup>. These cell lines were used to check the initial response as anticancer potential. Initial response can be checked on any cell lines. All solvents were of analytical grade and dried solvents were used in reaction scheme. All reactions were carried out in inert atmosphere of nitrogen. By using the rotary evaporator R-210 solvents were dried. Thin-layer chromatography analytic was conducted on Baker 250F and Whatman MK6F silica gel plates. By using TLC on silica gel plates reaction products were purified. By using UV lamp chemical reactions were visualized. The melting point of reactant and product was determined. These MTX derivatives were dissolved in dry ethanol in round bottom flask, NaOH and given aldehyde was added and reaction mixture was refluxed until TLC indicated disappearance of aldehyde. After completion of reaction pH was adjusted to 5, precipitate was washed with hexane and dried.

**Evaluation for anticancer activities.** U87 cell lines were used to check the initial response as anticancer potential. These cell lines were imported from Merck, US supplier. Initial response can be checked on any cell lines. The established method for identification of the survival fraction of U87 is reported<sup>23</sup>. To evaluate the anticancer effect of all micro molar MTX derivatives, the growth inhibition assays on MG-U87 cell lines was observed over different MTX-derivative's incubation times using incremental doses from 12.5 to 400  $\mu\text{M}$ . The human malignant glioma U87 (MG-U87) cell lines were cultured in Dulbecco's Modified Eagle's Medium/F12 in 1:1 concentration, with addition of L-Glutamine and sodium bicarbonate (2.438 g/L). The media was added with 10% fetal bovine serum, 1% Penicillin/Streptomycin, 0.5% Fungizone (Invitrogen, USA) and grown at 37°C under 5% CO<sub>2</sub> supply. Following 80% confluency, the MG-U87 cell lines were seeded in 96 well plates at concentration of  $2.5 \times 10^4$  (200  $\mu\text{L}$ /well) and were allowed to grow for 24 h at 37 °C in 5% CO<sub>2</sub> supply. Meanwhile, drug dosages of methotrexate (MTX) and its derivatives were prepared in DMSO. After 24 h of growth, the cells were processed in triplicates along with several concentrations of MXT and its derivatives at dose of 0  $\mu\text{M}$  (untreated control), 12.5  $\mu\text{M}$ , 25  $\mu\text{M}$ , 50  $\mu\text{M}$ , 100  $\mu\text{M}$ , 200  $\mu\text{M}$  and 400  $\mu\text{M}$ . The untreated cells in triplicates were considered as control. The drug treatment was conducted for 24 h to determine dose concentration that can inhibit 70% or 50% growth of the cells (IC70 or IC50 value). Following 1 day of drug treatment, the cells were fixed with 4% formaldehyde (Scharlab S.L, Spain) for 10 min that was followed by staining with 0.1% crystal violet (BioM labs, USA). After staining the 96 wells plates, each well was washed with distilled water and 100  $\mu\text{L}$  of acetic acid was included to each well. The absorbance was measured at 630 nm wavelength through an ELISA plate reader (Z32HK: Germany). From, the growth inhibition assays, survival fraction of each MTX derivatives at different concentration levels is constituted in response matrix  $Y_{10 \times 6}$ .

**Spectroscopic experiment of MTX derivatives.** Attenuated Total Reflection-Fourier Transformation Infrared (ATR-FTIR) spectrometer ALPHA 200488 which covers mid-IR (4000–550  $\text{cm}^{-1}$  wavenumbers) was used. ATR-FTIR was maintained with UATR Diamond ATR (Single Reflection) and high linear room temperature detector. For each MTX derivative 10 scans were gathered with 1  $\text{cm/s}$  scan speed and 4  $\text{cm}^{-1}$  resolution. Before each measurement background spectra against each MTX derivative was obtained. The spectrum obtained from this spectroscopic experiment of MTX derivatives were used to contract the data matrix  $X_{10 \times 1676}$ .

**Baseline correction.** FTIR spectra includes linear or non-linear residuals from spectroscopic experiment results in non flat base line of spectra. For statistical modeling and analysis of spectral based data set flat line at zero is required<sup>24</sup>. For this use of baseline correction methods is in practice. For this asymmetric least squares (ALS)<sup>25</sup> is considered. ALS is based on least squares algorithm which weights explanatory variables with positive differences. For smoothing  $2^{\text{nd}}$  derivative restriction is incorporated through  $S = \sum w_i(x_i - b_i)^2 + \lambda \sum (\Delta^2 b_i)$  where  $b_i$  presents the estimated baseline  $x_i$  presents the original spectrum,  $w_i$  presents the asymmetric residual weights and  $\Delta^2$  presents second derivative of computed baseline. ALS is based on two parameters smoothing and weight denoted by  $\lambda$  and  $w$ . For optimal estimation of these parameters cross validation procedure introduced in<sup>24</sup> is used. For this several levels of these parameters are used. For each level of the MTX derivative's spectra baseline is corrected and is used for further analysis (Supplementary Information).

**Predicting the survival fraction.** Predicting the survival fraction of MTX's derivatives through FTIR spectral data results in multivariate data with much larger number of wavenumbers that is explanatory variables compared to the number of samples. Multivariate methods are considered as potential candidate<sup>26,27</sup> for modeling multivariate data. For prediction purpose two streams exists, one uses iterative loading based procedure called partial least squares (PLS1) and others are based on penalized linear models includes elastic net (Elnet) and least absolute shrinkage and selection operator (Lasso)<sup>28</sup>. Soft thresholding (PLS1) (StPLS) is a potential variant adds the influential wavelength selection together with response prediction<sup>20</sup>. In order to improve prediction and interpretability, these methods use variable selection and regularisation. As a result, StPLS, Lasso, and Elnet are used in this study to predict the survival fraction of MTX derivatives against human malignant glioma U87 (MG-U87) cell lines using FTIR spectroscopy.

*Soft-thresholding partial least squares (StPLS).* In basic PLS algorithm<sup>26</sup> the covariance between response  $y$  that is survival fraction of MTX derivatives and the explanatory variables' linear combination  $X$  that is MTX derivatives spectra are optimized through iterative components has

$$\max \text{Cov}(p_h^t X, q_h^t y)$$

where  $p_h$  and  $q_h$  are the  $h^{\text{th}}$   $X$ - and  $y$ - loadings respectively. In multivariate data sets larger number of explanatory variables compared to available sample size is very obvious. This scenario can increase the variation of estimated PLS coefficients<sup>29,30</sup>, hence motivates for variable selection. For variable selection in PLS soft-thresholding step in the PLS (StPLS) algorithm is introduced<sup>20</sup>. In each iteration components of StPLS loading-weights are computed as:

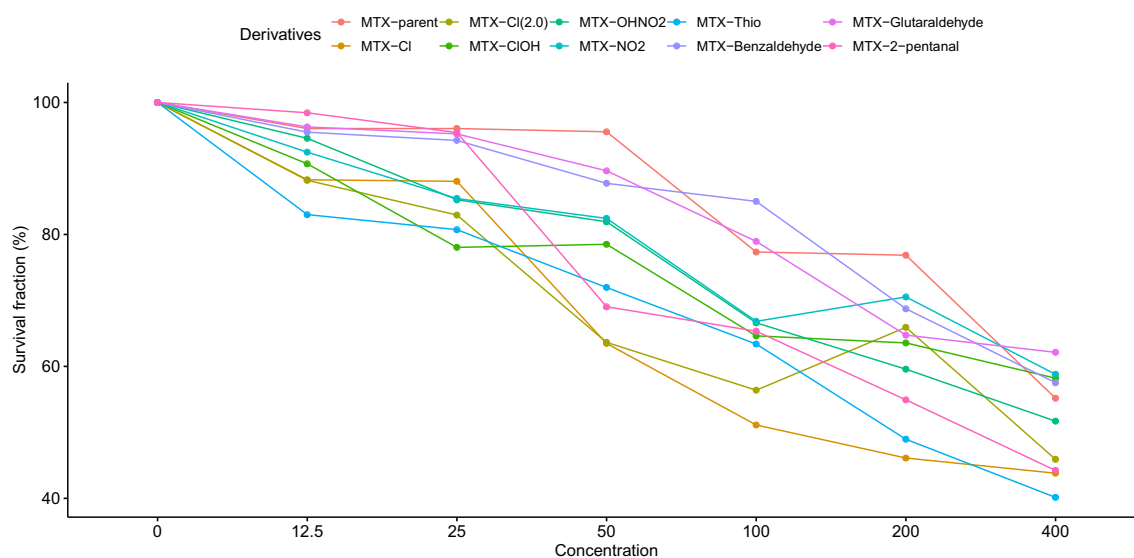
1. Scaling:  $w_h \leftarrow w_h / \max_j |w_{h,j}|$ , where  $j = 1, \dots, p$
2. Soft-thresholding:  $w_{h,j} \leftarrow \text{sign}(w_{h,j})(|w_{h,j}| - \delta)_+$ , where some  $\delta \in [0, 1)$  and  $(\dots)_+$  indicates  $\max(0, \dots)$
3. Normalizing:  $w_h \leftarrow w_h / \|w_h\|$

The shrinkage parameter  $\delta \in [0, 1)$  explains the level of shareholding in StPLS this means larger  $\delta$  will results in fewer selected set of variables. The optimal level of  $\delta$  is determined by cross validation<sup>20</sup>.

**Penalized model.** An alternative to iterative modeling procedure (PLS based) are penalized models. In this regard elastic net (Elnet) and least absolute shrinkage and selection operator (Lasso)<sup>28</sup> are potential candidates. These model implements the coordinate descent algorithm with certain penalty parameter called  $\theta$ . For computational purpose several levels are considered and optimal choice of these penalty parameter is obtained through cross validation. Here the multivariate regression coefficient  $\beta$  is penalized through the partial log-likelihood function defined by

$$L(\beta) - \sum_{j=1}^p p_{\theta^*}(|\beta_j|) \quad (1)$$

where  $L(\beta)$  denotes the partial log-likelihood for  $n$  samples.  $p_{\theta}(|\cdot|)$  presents the penalty function. In lasso<sup>31</sup>  $L_1$ -penalized model is used, which is defined as



**Figure 1.** At various concentrations, the survival fraction of these MTX derivatives against human malignant glioma U87 (MG-U87) cell lines is presented.

$$p_{\theta}(|\beta|) = \theta|\beta| \quad (2)$$

An alternative to  $L_1$  penalty is  $L_2$  penalty which shrinks some of the regression coefficients  $\beta$ , where  $L_2$  penalty is

$$p_{\theta}(|\beta|) = \theta\beta^2 \quad (3)$$

In elastic net (Elnet)<sup>32</sup> the mixture of both  $L_1$  and  $L_2$  penalty is used as

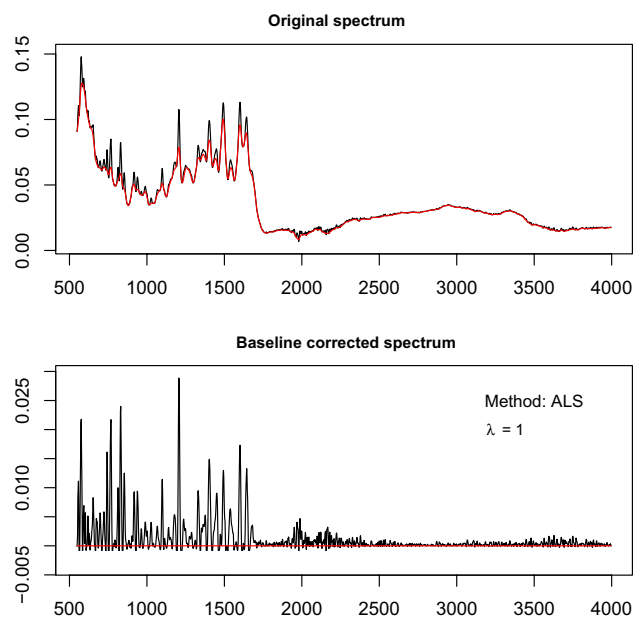
$$p_{\theta_1, \theta_2}(|\beta|) = \theta_1|\beta| + \theta_2\beta^2. \quad (4)$$

Both Lasso and Elnet do the variable selection by equating non significant variable's coefficients to zero. Here  $\theta_1$  is weight for  $L_1$  penalty and  $\theta_2$  is weight for  $L_2$  penalty, moreover  $\theta_1 + \theta_2 = 1$ . Small  $\theta_2$  level will result large number of variables and vice versa.

## Results

In current study 1 MTX parent and 9 MTX derivatives are considered, which include Schiff base derivatives MTX with Benzaldehyde, Chlorobenzaldehyde, 2-Chlorobenzaldehyde, 3-Nitrobenzaldehyde, 5-Chloro-2-hydroxybenzaldehyde, 2-Hydroxy-5-Nitrobenzaldehyde, 2-Thiocarboxyaldehyde, Trans-2-pentanal, and Glutaraldehyde. At various concentrations, the survival fraction of malignant glioma U87 (MG-U87) cell lines against these MTX derivatives is monitored. These concentrations include 400  $\mu\text{M}$ , 200  $\mu\text{M}$ , 100  $\mu\text{M}$ , 50  $\mu\text{M}$ , 25  $\mu\text{M}$  and 12.5  $\mu\text{M}$ . The survival fraction of malignant glioma cells at different concentration level of MTX derivatives is presented in Fig. 1. This indicates with 12.5  $\mu\text{M}$  concentration all MTX derivatives except MTX-Glutaraldehyde and MTX-2-pentanal compared to MTX-parent compound has better capability to kill MG-U87 cell lines. With 25  $\mu\text{M}$ , 50  $\mu\text{M}$  and 200  $\mu\text{M}$  concentration all MTX derivatives compared to MTX-parent has better capability to kill MG-U87 cell lines. With 100  $\mu\text{M}$  concentration all MTX derivatives except MTX-Glutaraldehyde, MTX-Benzaldehyde and MTX-parent drug has better capability to kill MG-U87 cell lines. With 400  $\mu\text{M}$  concentration all MTX derivatives except MTX-Benzaldehyde, MTX-Glutaraldehyde, MTX-NO<sub>2</sub>, and MTX-OHNO<sub>2</sub> compared to MTX-parent has better capability to kill MG-U87 cell lines.

The current study aims to predict the survival fraction of malignant glioma cell lines against several concentrations of MTX derivatives through FTIR spectroscopy. The 10 samples of MTX derivatives were subjected to FTIR spectroscopic experiments for this purpose. For the projected mathematical models, the threshold for FTIR spectra should be at 0. For ALS baseline adjustment, we have used algorithms. It is dependent on ALS to select their respective parameters. We also used the objective method for optimum parameter tuning, where combinations of various levels of the parameters are evaluated for the survival fraction by PLS. Cross validation establishes the optimal parameter estimates of ALS. For estimating the survival fraction at different stages of concentration, smoothing parameters  $\theta = 1$  and weights  $w = 0.001$  tend to optimally correct the spectral basis of the MTX derivatives. The relation between the continuum of MTX derivatives initial and baseline corrected is seen in Fig. 2. After the baseline correction, the survival fraction of MTX derivatives  $Y_{10 \times 6}$  is modeled with spectroscopic data matrix  $X_{10 \times 1676}$ . As preprocessing, the data matrix was scaled with mean zero and variance 1.

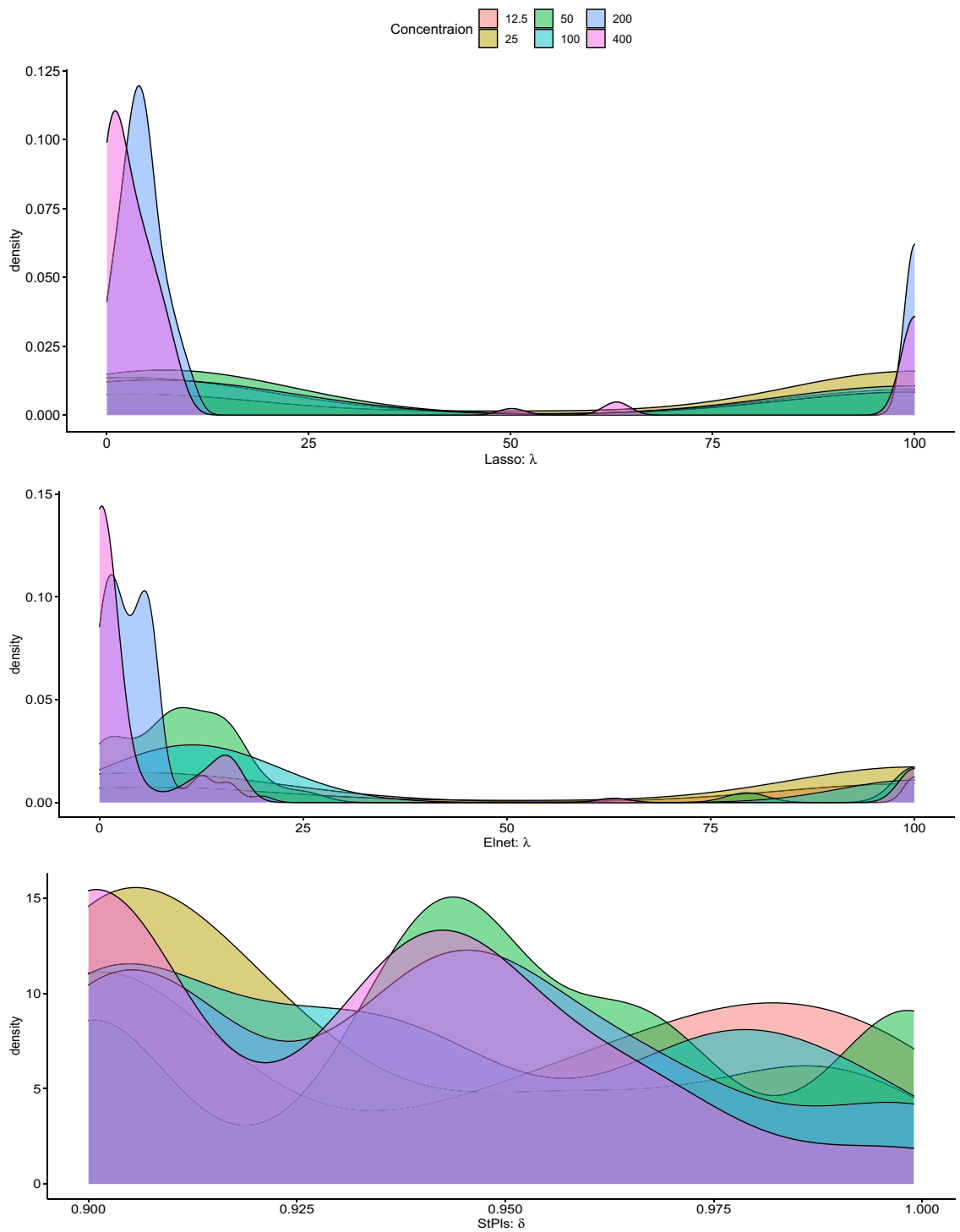


**Figure 2.** The comparison of original and baseline corrected spectral data is shown here. The base line is corrected using ALS with smoothing parameters ( $\lambda = 1$ ) and weights ( $w = 0.001$ ).

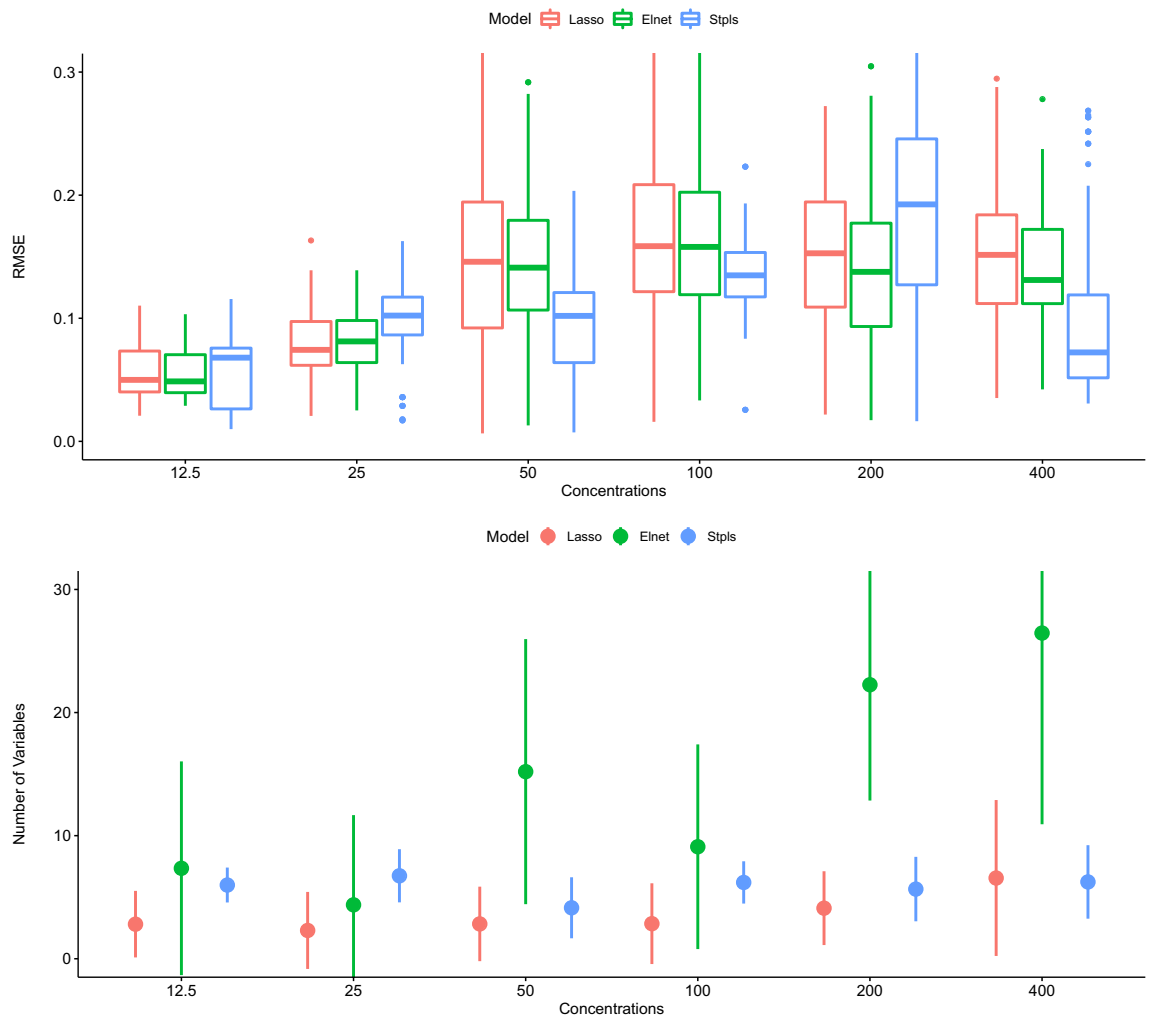
We use StPLS, Lasso and Elnet for the prediction of the survival fraction of MG-U87 against MTX derivatives at different concentrations. RMSE over the test data is used to measure the prediction performance of considered methods. We have randomly divided samples into test and training data for model building and comparison of the predicted model. Although test samples are used for model comparison or evaluation, training data fits the model. There is random bias. In order to remove this bias, the above data validation and calibration process will be replicated within 100 runs. The distribution of optimal parameter estimates for Lasso, Elnet and StPLs obtained against concentration level is shown in Fig. 3. The calculated ideal median Lasso  $\theta$  is 3.98, 100, 10, 10, 5 and 3.16 respectively for the MTX 12.5, 25, 50, 100, 200 and 400 concentrations. The calculated maximum Elnet  $\theta$  median for MTX concentrations of 12.5, 25, 50, 100, 200, and 400 is 7.94, 100.00, 11.29, 12.59, 3.98 and 0.23 respectively. Similarly, the approximate optimum StPLs  $\delta$  median is 0.96, 0.92, 0.94, 0.93, 0.94 and 0.93 respectively for the MTX 12.5, 25, 50, 100, 200, and 400 concentrations.

Lasso, Elnet and StPLs are designed to predict the survival fraction at each stage of the MTX concentration using the above approximate parameters. The validated prediction capabilities that are RMSE standardised over the test data are calculated in each sprint. The analysis of variance results indicating the significance of concentrations, models and their interactions for explaining the variation in RMSE over test data is presented in Table 1. It appears on the RMSE computed from test data is significantly varying with the change of models ( $p = 0.012$ ), with the change of concentrations levels ( $p \leq 0.001$ ) and with the change of combination of models and concentration level ( $p \leq 0.001$ ). In the upper panel of Fig. 4, the distribution of validated RMSE of all models over all concentrations is presented. The survival fraction of MTX derivatives with 12.5% concentrations seems to be reasonably expected by Lasso and Elnet. The survival fraction of MTX derivatives with a 25% concentration is better estimated by Lasso. The survival fraction of MTX derivatives with concentrations of 50%, 100% and 400% is better estimated by StPLs. With 200% concentrations, Elnet better estimates the survival fraction of MTX derivatives. The distribution in the lower panel of Fig. 4 of the number of selected variables (wavenumbers) of all models for all concentrations is provided. It seems that Lasso uses the least number of variables at all concentrations in estimating the survival fraction of MTX derivatives. Elnet uses the maximum number of variables to estimate the survival fraction of MTX derivatives at all concentrations except 25%, where StPLs uses the maximum number of variables.

Two model streams exist for prediction purposes, one uses iterative loading based method (StPLS) and others are based on penalised linear models (Lasso and Elnet). The sample covariance matrix's own value structure of the explanatory matrix is then used to describe the model prediction action's characteristics and the covariance between the main components and the response response<sup>20,33</sup>. If there are unrelated components with large eigenvalues, the forecast may be worse. As a result, poor prediction of variable selection methods based on PLS is predicted. Figure 5 depicts the data characteristics of the survival fraction of MTX derivatives and spectral data. The bars reflect the ionic fluids' spectral eigenvalue (scaled by the largest). The dots represent the covariance at different concentrations (scaled by the largest) between the main components and the survival fraction of MTX derivatives. This implies that between-variable dependencies are poor in the spectral data. The data heterogeneity



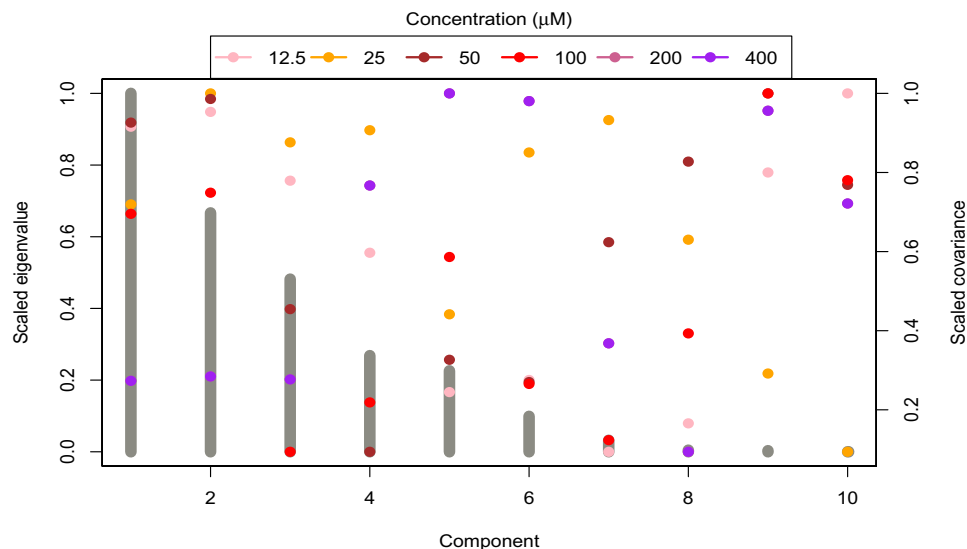
**Figure 3.** A comparison of the parameters of the fitted model is presented.



**Figure 4.** In the upper and lower panels, the distribution of RMSE computed from test data and the distribution of the number of selected variables through Lasso, Elnet and StPLS over all concentration levels is presented.

Factors	Df	Sum Sq	Mean Sq	F-value	p-value
Concentration	5	2.65	0.53	66.98	≤ 0.001
Model	2	0.07	0.03	4.41	0.012
Concentration: model	10	0.30	0.03	3.80	≤ 0.001
Residuals	1782	14.13	0.01		

**Table 1.** The analysis of variance results indicating the significance of concentrations, models and their interactions for explaining the variation in RMSE over test data is presented.



**Figure 5.** Data properties for the MTX derivative survival fraction as well as spectral data are provided. The bars represent the spectral eigenvalues of ionic fluids (scaled by the largest). The dots show the relationship between the main components and the survival fraction of MTX derivatives at different concentrations (scaled by the largest).

has to be clarified by a large number of latent elements, as we have a limited number of MTX derivatives, making it impossible for the PLS-based model to provide a better estimate of the survival fraction. We observe different behaviour of covariances between the main variable and survival fraction separately from the behaviour of eigenvalues. The own values and covariance are not aligned for relevant components with 12.5%, 25% and 200% concentrations, so a large number of PLS components are recommended for better prediction at these concentrations, which is therefore impossible to achieve here in predicting survival fraction, Lasso or Elnet outperforms.

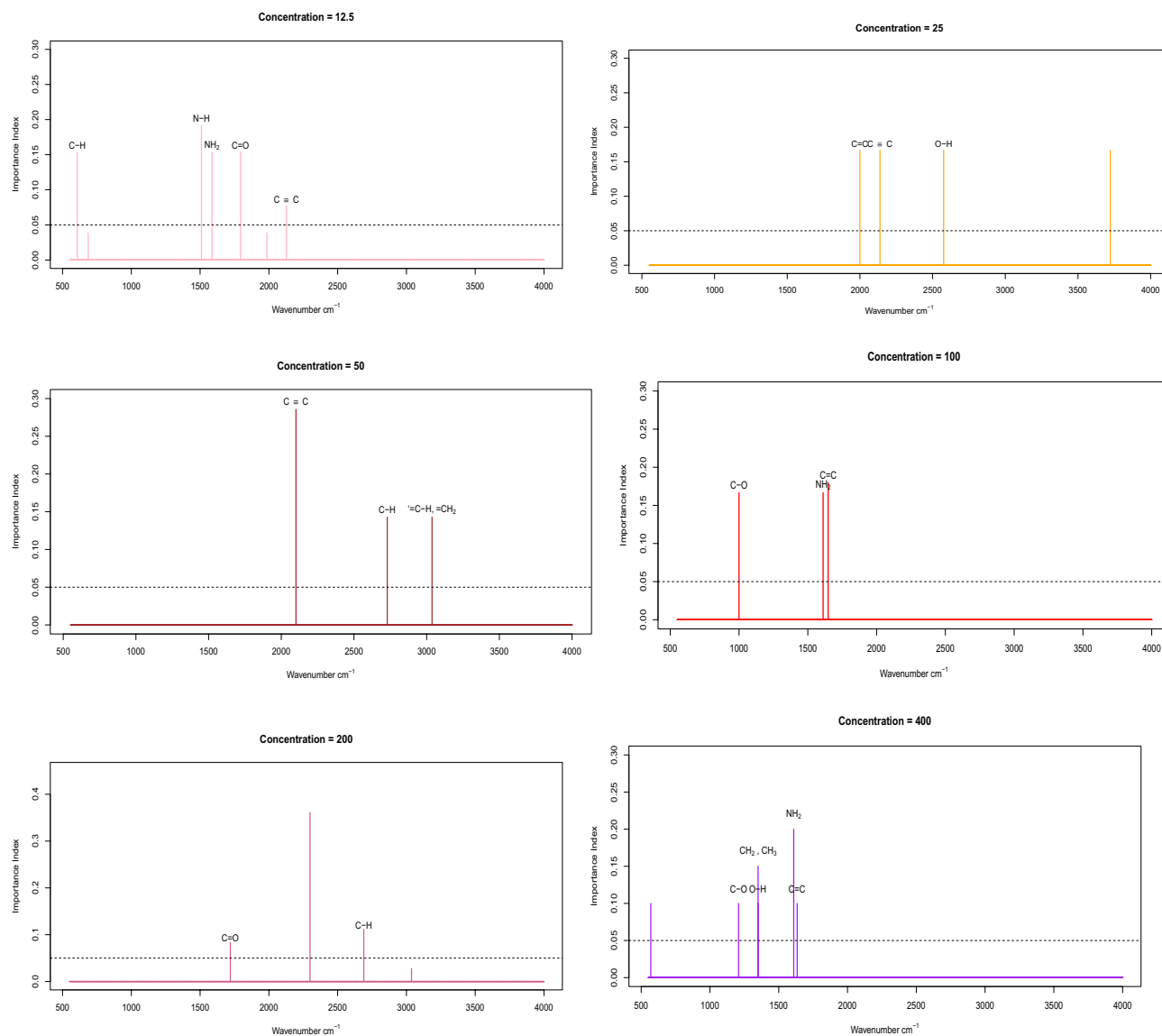
Influential wavenumbers are chosen by the optimum model defined in Fig. 4, which is seen in Fig. 6 at various concentration levels. Along with an influential functional compound, the strong regression coefficients of the best fitting model are shown. It seems that the prominent wavenumber corresponds to C–H, N–H, N–H<sub>2</sub>, C=O, C≡C for prediction of the MTX derivative survival fraction toward 12.5% concentrations. The prominent wavenumber corresponds to C–O, C≡C, O–H for calculation of the MTX derivative survival fraction against 25% concentrations. The prominent wavenumber corresponds to C–H, =CH<sub>2</sub>, C=O and C≡C for the prediction of the MTX derivatives survival fraction toward 50% concentrations. The prominent wavenumber corresponds to C–O, C≡C and N–H<sub>2</sub> for prediction of the survival fraction of the MTX derivatives against 100% concentrations. The prominent wavenumber refers to C–H and C–H for the estimation of the MTX derivative survival fraction against 200% concentrations. The influential wavenumber corresponds to C–O, C≡C, N–H<sub>2</sub>, O–H and CH<sub>2</sub>, CH<sub>3</sub> for prediction of the MTX derivative survival fraction against 400% concentrations.

**Computations.** For computations, modeling and figures R software is used<sup>34</sup>. For baseline correction R package 'baseline'<sup>35</sup> and for model fitting R packages 'glmnet' and 'plsVarSel'<sup>36,37</sup> are used.

## Conclusions

Via FTIR spectroscopy by StPLS, Lasso and Elnet, the study predicts the survival fraction of malignant glioma cells against many concentrations of methotrexate derivatives. ASL works well for base-line correction. The survival fraction prediction capabilities are linked with covariance between the survival fraction and spectrum data. Moreover Influential wavenumbers are picked by Lasso at concentrations of 12.5% and 25%, by StPLs at concentrations of 50%, 100% and 400%, and by Elnet at concentrations of 200% along with influential functional compounds.





**Figure 6.** Influential wavenumbers are picked at various concentration levels by the optimal model calculated as shown. Along with a powerful functional group, the relevant regression coefficients of the best suited model are shown.

Received: 11 June 2021; Accepted: 23 August 2021

Published online: 21 September 2021

## References

- Gubner, R. *et al.* Therapeutic suppression of tissue reactivity. 2. Effect of aminopterin in rheumatoid arthritis and psoriasis. *Am. J. Med. Sci.* **221**, 176–182 (1951).
- Kinsella, A., Smith, D. & Pickard, M. Resistance to chemotherapeutic antimetabolites: A function of salvage pathway involvement and cellular response to DNA damage. *Br. J. Cancer* **75**, 935–945 (1997).
- Abdel Bary, E., Harmal, A. N., Saeed, A. & Gouda, M. A. Design, synthesis, characterization, swelling and in vitro drug release behavior of composite hydrogel beads based on methotrexate and chitosan incorporating antipyrine moiety. *Polym. Plast. Technol. Eng.* **57**, 1906–1914 (2018).
- Kinder, A. *et al.* The treatment of inflammatory arthritis with methotrexate in clinical practice: Treatment duration and incidence of adverse drug reactions. *Rheumatology* **44**, 61–66 (2005).
- Batchelor, T. T., Kolak, G., Ciordia, R., Foster, C. S. & Henson, J. W. High-dose methotrexate for intraocular lymphoma. *Clin. Cancer Res.* **9**, 711–715 (2003).
- Wang, W., Zhou, H. & Liu, L. Side effects of methotrexate therapy for rheumatoid arthritis: A systematic review. *Eur. J. Med. Chem.* **158**, 502–516 (2018).
- Mehmood, T., Iqbal, M. & Hassan, R. Prediction of antibacterial activity in ionic liquids through FTIR spectroscopy with selection of wavenumber by pls. *Chemom. Intell. Lab. Syst.* **206**, 104124 (2020).
- Rubino, F. M. Separation methods for methotrexate, its structural analogues and metabolites. *J. Chromatogr. B Biomed. Sci. Appl.* **764**, 217–254 (2001).
- Hall, E. J. *et al.* *Radiobiology for the Radiologist* Vol. 6 (Lippincott Williams & Wilkins, 2006).

10. Gaigneaux, A. *et al.* The infrared spectrum of human glioma cells is related to their in vitro and in vivo behavior. *Exp. Cell Res.* **297**, 294–301 (2004).
11. Gasper, R., Dewelle, J., Kiss, R., Mijatovic, T. & Goormaghtigh, E. Ir spectroscopy as a new tool for evidencing antitumor drug signatures. *Biochim. Biophys. Acta (BBA) Biomemb.* **1788**, 1263–1270 (2009).
12. Draux, F. *et al.* Ir spectroscopy reveals effect of non-cytotoxic doses of anti-tumour drug on cancer cells. *Anal. Bioanal. Chem.* **395**, 2293–2301 (2009).
13. Kasemsumran, S., Du, Y.-P., Murayama, K., Huehne, M. & Ozaki, Y. Simultaneous determination of human serum albumin,  $\gamma$ -globulin, and glucose in a phosphate buffer solution by near-infrared spectroscopy with moving window partial least-squares regression. *Analyst* **128**, 1471–1477 (2003).
14. Perez-Guaita, D. *et al.* Evaluation of infrared spectroscopy as a screening tool for serum analysis: Impact of the nature of samples included in the calibration set. *Microchem. J.* **106**, 202–211 (2013).
15. Carvalho, C. D. S. *et al.* Study of advanced rheumatoid arthritis. *Revista Brasileira de Engenharia Biomédica* **30**, 54–63 (2014).
16. Wald, N. & Goormaghtigh, E. Infrared imaging of primary melanomas reveals hints of regional and distant metastases. *Analyst* **140**, 2144–2155 (2015).
17. Derenne, A., Gasper, R. & Goormaghtigh, E. The FTIR spectrum of prostate cancer cells allows the classification of anticancer drugs according to their mode of action. *Analyst* **136**, 1134–1141 (2011).
18. Parachalil, D. R. *et al.* Raman spectroscopy as a potential tool for label free therapeutic drug monitoring in human serum: The case of busulfan and methotrexate. *Analyst* **144**, 5207–5214 (2019).
19. Derenne, A., Gasper, R. & Goormaghtigh, E. Monitoring of metabolism perturbation in prostate pc-3 cancer cells by sub-lethal concentrations of methotrexate. *Spectroscopy* **24**, 55–60 (2010).
20. Sæbø, S., Almøy, T., Aaroe, J. & Aastveit, A. H. ST-PLS: A multi-dimensional nearest shrunken centroid type classifier via PLS. *J. Chemom.* **20**, 54–62 (2007).
21. Basu, A., Mitra, R., Liu, H., Schreiber, S. L. & Clemons, P. A. RWEN: Response-weighted elastic net for prediction of chemosensitivity of cancer cell lines. *Bioinformatics* **34**, 3332–3339 (2018).
22. Lu, G. *et al.* Predicting therapeutic antibody delivery into human head and neck cancers. *Clin. Cancer Res.* **26**, 2582–2594 (2020).
23. Nemat, A. *et al.* Synthesis, anticancer evaluation and molecular docking studies of methotrexate's novel Schiff base derivatives against malignant glioma cell lines. *J. Biomol. Struct. Dyn.*, 1–13 (2020).
24. Liland, K., Almøy, T. & Mevik, B. Optimal choice of baseline correction for multivariate calibration of spectra. *Appl. Spectrosc.* **64**, 1007–10016 (2010).
25. Eilers, P. H. & Boelens, H. Baseline correction with asymmetric least squares smoothing. *Leiden University Medical Centre Technical Report 1* (2005).
26. Wold, S., Martens, H. & Wold, H. The multivariate calibration problem in chemistry solved by the PLS method. *Conference Proceeding Matrix pencils*, 286–293 (1983).
27. Martens, H. & Næs, T. *Multivariate Calibration* (Wiley, 1989).
28. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1 (2010).
29. Keleş, S. & Chun, H. Comments on: Augmenting the bootstrap to analyze high dimensional genomic data. *TEST* **17**, 36–39 (2008).
30. Höskuldsson, A. Variable and subset selection in PLS regression. *Chemom. Intell. Lab. Syst.* **55**, 23–38 (2001).
31. Tibshirani, R. The Lasso method for variable selection in the Cox model. *Stat. Med.* **16**, 385–395 (1997).
32. Zou, H. & Hastie, T. Regularization and variable selection via the elastic-net. *J. R. Stat. Soc. B* **67**(2), 301–320 (2005).
33. Liland, K. H., Høy, M., Martens, H. & Sæbø, S. Distribution based truncation for variable selection in subspace methods for multivariate regression. *Chemom. Intell. Lab. Syst.* **122**, 103–111 (2013).
34. R Core Team *et al.* R: A language and environment for statistical computing (2013).
35. Liland, K. H., Mevik, B.-H. & Canteri, R. Package ‘baseline’ (2015).
36. Hastie, T. & Qian, J. Glmnet vignette. Retrieved June 9, 1–30 (2014).
37. Liland, K., Mehmood, T. & Sæbø, S. plsvarsel: Variable selection in partial least squares (2016).

## Acknowledgements

The work was supported by the School of Natural Sciences (SNS), National University of Sciences and Technology (NUST), Islamabad, Pakistan.

## Author contributions

Both authors wrote the main manuscript. T.M. conducted the analyses. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-97891-3>.

**Correspondence** and requests for materials should be addressed to T.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](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) 2021