Received: 2 April 2022 | Revised: 22 May 2022 | Accepted: 24 May 2022

DOI: 10.1002/jbio.202200097

RESEARCH ARTICLE

JOURNAL OF BIOPHOTONICS

Encoder–decoder neural networks for predicting future FTIR spectra – application to enzymatic protein hydrolysis

Miroslav Kuchta[1] | **Sileshi Gizachew Wubshet**[2] | **Nils Kristian Afseth**[2] | **Kent-André Mardal**[3,1] | **Kristian Hovde Liland**[4*]

[1]Department of Scientific Computing and Numerical Analysis, Simula Research Laboratory, Oslo, Norway

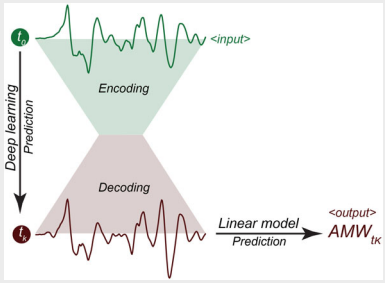
[2]Nofima - Norwegian Institute of Food, Fisheries and Aquaculture Research, Ås, Norway

[3]Department of Mathematics, University of Oslo, Oslo, Norway

[4]Department of Science and Technology, Norwegian University of Life Sciences, Ås, Norway

***Correspondence**
Kristian Hovde Liland, Department of Science and Technology, Norwegian University of Life Sciences, Ås, Norway.
Email: kristian.liland@nmbu.no

Abstract



In the process of converting food-processing by-products to value-added ingredients, fine grained control of the raw materials, enzymes and process conditions ensures the best possible yield and economic return. However, when raw material batches lack good characterization and contain high batch variation, online or at-line monitoring of the enzymatic reactions would be beneficial. We investigate the potential of deep neural networks in predicting the future state of enzymatic hydrolysis as described by Fourier-transform infrared spectra of the hydrolysates. Combined with predictions of average molecular weight, this provides a flexible and transparent tool for process monitoring and control, enabling proactive adaption of process parameters.

KEYWORDS

deep learning, encoder–decoder, enzymatic protein hydrolysis, FTIR, process control

1 | INTRODUCTION

Biotechnological solutions like fermentation and enzymatic protein hydrolysis (EPH) are recognized as essential tools in sustainable utilization of biomass, like the production of revenue streams from food processing by-products. In EPH, food grade proteases are used to digest protein-based biomasses, and this technology has attracted huge interest as a feasible tool for recovery of value-added ingredients from food-processing by-products [1]. Due to the high degree of variability of raw materials subjected to EPH, the need for a process monitoring and control tool is eminent in order to produce products with specific properties over time. Analytical measurements at critical points in the process are essential elements in controlling and optimizing a given biotechnological process. In this respect, the combined use of rapid spectroscopic measurements and data analytical methods is therefore particularly attractive.

The use of Fourier-transform infrared spectroscopy (FTIR) for bioprocess monitoring is currently gaining considerable attention. One of the intriguing research fields that builds on the extensive research related to FTIR and protein structure is the application of FTIR for probing protein modifications induced by enzymes, for example, in EPH. In one of the first FTIR studies to be reported on the monitoring of EPH Ruckebusch [2] and coworkers followed the degradation of bovine haemoglobin

by collecting transmission infrared spectra throughout the hydrolysis reaction. More recent studies have shown that FTIR spectroscopy can be used to follow the EPH process and thus the enzymatic degradation of proteins both qualitatively and quantitatively by predicting average molecular weights (AMW) as well as the degree of hydrolysis (DH), which are two well established process parameters ([3–7]). The rationale behind using FTIR spectroscopy to monitor the EPH process is as follows: When the peptide bond is hydrolysed during EPH, the result is an increasing number of C-terminal carboxylate (COO-) and N-terminal amino (NH3+) groups. This information is readily available in the FTIR spectra. In addition to the intrinsic changes in the primary structure, the general shortening of the protein chain also affects the secondary structure, which in turn also significantly affects the FTIR spectral fingerprints [4].

High-resolution spectroscopic tools like FTIR spectroscopy are most often used for high-throughput screening of products off-line (i.e. not directly in a process). Thus, in process control terms, this means that measurements are made after the process has taken place, and this information is then used to adjust processing parameters (i.e. the feed-backwards-scheme). The disadvantage of this approach is that control action can only be taken late, that is, a lot of product that is off-specification can be produced before any potential action is taken in the process. With rapid online FTIR analysers becoming available, so-called feed-forward control is starting to become a viable alternative. In this approach, a predefined model decides settings of the controllable process parameters, using measured raw material characteristics or early process data as input. This approach has been explored in a range of different studies, also related to EPH, for example, Wubshet et al. [8]. However, characterization of the future product, which steers the decisions of the feed-forward controller, is then often reduced to a single feature, and not to the entire information that can be provided by for instance an FTIR spectrum. Arguably, better control strategies could be reached with richer chemical characteristics. Motivated by feed-forward control of EPH, we are therefore in this study interested in models predicting FTIR spectra at arbitrary future time points based on a limited number of spectra from the initial stages of the process. To this end we wish to employ deep neural networks.

Deep neural networks (DNNs) have recently made significant contributions to a number of research problems and fields such as image classification [9], optimal control [10], weather prediction [11] and protein folding [12]. A particular class of DNNs relevant in the present study are encoder–decoder networks, which, broadly speaking, compute their output by first embedding the input into a (lower-dimensional) latent space. We note that in the special case when the networks are trained to reproduce their inputs, that is, approximate the identity mapping, the encoder–decoder networks are referred to as autoencoders (AE) [13, Ch 14]. Autoencoders, encoder–decoder networks and DNNs have also been widely applied in the context of bio-reactor monitoring, modelling and process control. For process monitoring, Jo et al. [14] and Jinadasa et al. [15] apply AE for feature/signal extraction in near-infrared and Raman spectroscopy. In several recent papers [16, 17] AEs and/or encoder–decoder networks [18, 19] are used for denoising and scattering removal in FTIR spectra or to obtain classifiers for spectral histopathology, for example, Raulf et al. [19]. In particular, the networks presented in Raulf et al.'s papers [18, 19] are obtained by fine-tuning the encoder–decoder models, where the encoders (architecture and initial parameters) originate from an autoencoder obtained in a pre-training step. That is, the pre-training provides an initial guess for embedding of the raw spectra suitable for a given problem (specified by the decoder), for example, regression [18] or classification [19]. Using pure data-driven approaches Mete et al. [20] construct surrogate models of bioreactors. Further, models based on simulations [21] (utilizing potentially expensive-to-solve models described by partial differential equations) or on hybrid approaches [22] combining simulations with measurements data have been developed. Finally, with the surrogate models available, optimal controllers are designed as DNNs [23, 24] trained by reinforcement learning, for example, Mnih et al. [25] and references therein.

Here we are interested in employing the encoder–decoder architecture for continuous-in-time inference of the FTIR spectra based on (a small number of) discrete spectra from the early stages of EPH.

The main aim of the present study was the combined use of FTIR spectra and DNNs to predict the potential outcome of an EPH process. One choice to be made when modelling in the feed-forward framework would be to either predict FTIR spectra at a future time point, given one or more spectra in an early phase of hydrolysis, or to directly predict future characteristics, such as AMW. In the current study, the former approach, that is, predicting future spectra, is used, as it has the benefit of providing a (feature)-rich and flexible view into the future. The generic approach of the study is illustrated in Figure 1, where a convolutional encoder–decoder network is used in order to predict the future spectra (i.e. the FTIR-to-FTIR mappings/models are represented as DNNs). With the process control application in mind, the spectral models we wish to construct should be capable of predicting the spectra at arbitrary future time points so that fine-grained (in time) control of the process can be

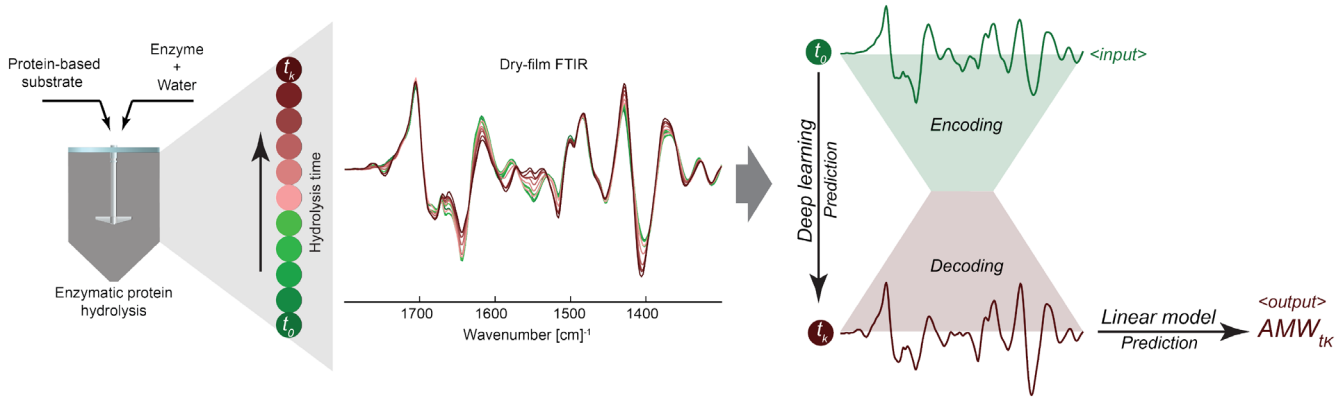


FIGURE 1 Process and quality monitoring of EPH. State of the reaction at time t_i, $i = 0, 1, \ldots$ is described by FTIR spectra. In turn, the product quality, here AMW, at time t_i can be obtained from the spectra by a dedicated model. Predictions of future product quality are accomplished by combining the AMW model with an encoder–decoder network predicting spectra at $t > t_i$ based on the input spectra at time t_i

potentially achieved. Convolutional AEs allowing for such continuous-in-time inference have been proposed, for example, by Vukotić et al. [26] who address the task of future video frame prediction. Specifically, their networks take as input the image at time t and an arbitrary offset* $\delta t > 0$ while the frame at $t + \delta t$ is the expected output. As commented by Vukotić et al., a discrete inference is more common, where δt is implicit and effectively fixed by the *equispaced* training data. We note that in the context of EPH, acquiring the equispaced data set is costly since early into the process small time steps (∼1 min) are required to capture accurately the dynamics while at later stages large steps (∼10 min) are sufficient. Thus, we argue that the continuous inference is more natural.

2 | METHODOLOGY

2.1 | FTIR encoder–decoder networks

In order to define our FTIR models, let $\eta_{t_i} \in R^n$ be a vector representing the FTIR spectrum at time t_i. We recall that a canonical autoencoder is a mapping $N : R^n \to R^n$ such that $\left\| \eta_{t_i} - N(\eta_{t_i}) \right\|$ is small and having a particular structure $N = D \circ E$ where $E : R^n \to R^l$ and $D : R^l \to R^n$. Here, E is the encoder(compression) network which takes the input into a latent space (typically $l \ll n$) while the decoder network D reconstructs the original input from the compressed representation. In the following the (reconstruction) norm is taken as the average l^2 norm, that is, $\left\| e \right\| = \left(\frac{1}{n} \sum_i |e_i|^2 \right)^{1/2}$.

To allow for the continuous-in-time inference of the FTIR spectra we follow Vukotić et al. [26] and introduce extra input variables (to be made more precise later) to the encoder part of the encoder–decoder network. To this end, let η_{t_j} be the spectrum at time $t_j > t_i$ and $\epsilon_{i,j} \in R^m$ be the additional inputs representing, for example, the offset/temporal distance $t_j - t_i$. We then wish to find $N = D \circ E$ which minimizes $\left\| \eta_{t_j} - D\left(E\left(\eta_{t_i}, \epsilon_{i,j}\right)\right) \right\|$ over the given input–output spectra.

In the following we will consider two different choices of the variables $\epsilon_{i,j} \in R^m$ leading to two different FTIR-to-FTIR models represented by encoder–decoder networks with different architectures. For the first architecture we set $m = 1$ and let the extra input be the offset time $t_j - t_0 = \delta t > 0$ between the *fixed* time t_0 at which the input spectrum η_{t_0} is taken and the inference time t_j (to be varied) at which the spectrum η_{t_j} is predicted. That is, the network attempts to predict the rest of the reaction from only one single state η_{t_0}, where t_0 may be both early and late in the process. For this reason we will further speak of *single input* time networks and denote the mapping as S_{t_0} to highlight the dependence of the network on the input time t_0. We remark that different t_0 yield different networks. Moreover, since only positive offsets are assumed the inference time for models with larger value of t_0 is shortened since we assume that all the EPH reactions (more precisely, our measurements of the process) have a fixed terminal time. In turn the dataset that can be used for training is shrunk for t_0 further away from the initial time of the reaction. At the same time, inference at $t > t_0$ then concerns only the later stages of the hydrolysis with possibly less dynamics (as reflected by the coarser measurement sampling in time).

A disadvantage of the S_{t_0} networks is that they do not utilize the spectra η_t for $t < t_0$. Indeed, the idea of predictions based on histories (as given, e.g., by batched sequences of reaction states) has already been proven useful in bioreactor modelling [27]. In the context of

encoder–decoder networks, predictions based on batched spectra can be realized in several ways. For example, the inputs to the convolutional encoder could be a matrix with entries formed by a fixed number of past spectra η_{t_j}, $t_j \leq t_c$. While simple to implement using standard layers, a disadvantage of this approach lies in the fixed input size. In particular, for t_c close to the onset the spectra might be too few to form the input thus restricting which cut-off time could be used. On the other hand, only part of the available history of the spectra is used when t_c is large. To enable variable-sized inputs, fully convolutional networks [28] or pyramid pooling layers [29] could be used. However, it is not evident how these approaches should be combined with temporal information about the spectra, which is needed especially if the spectral measurements are not equispaced. Here we therefore propose a simple network type, further referred to as *many input* time models, which maps the input time t, spectrum η_t and the offset δt to $\eta_{t+\delta t}$. This construction is rooted in the idea that by learning to predict the same spectra from different inputs, and especially times, the network can learn the dynamics between the inputs leading to more robust inference. We remark that for our *many inputs* model, the additional input space has dimension $m = 2$.

In the following we will restrict the input time such that $t \leq t_c$ for some cut-off-time t_c. Similarly to t_0 in S_{t_0} networks, the choice here is motivated by application in optimal control of EPH where we wish to predict the trajectory of the reaction based on the inputs close to its start. The many input networks which differ by the cut-off-time will be denoted as M_{t_c}. We note that a network M_{t_c} can predict the spectrum at some time t in a number of ways, reflecting the consistency property $\eta_t = \eta_{t_k+\delta t_k}$ for any input time and offset such that $t_k + \delta t_k = t$, $t_k < t_c$.

We construct both S_{t_0}, M_{t_c} as convolutional encoder–decoder networks where, following Vukotić et al. [26], in the encoding part the temporal variables are evolved (through dense layers) in branches independent of the input spectra. The encoding of the spectra is provided by convolutional layers while deconvolutional (transpose-convolutional) layers are used in the decoder. It is only in the decoder network that the time and the spectrum, which have both been embedded in their respective latent spaces, are combined and decoded together to yield the final inferred spectrum. We detail the networks' architectures in Figure 2. We note that for M_{t_c} the branches of time variables t and δ_t are combined before concatenating with the spectral branch (represented by an orange arrow in the figure). Furthermore, to prevent overfitting, the temporal branches in M_{t_c} use dropout layers [30]. Finally, we remark that the architectures of S_{t_0} and M_{t_c}, in particular the layer sizes, are chosen as independent of t_0 and t_c respectively. Thus all S_{t_0} and M_{t_c} networks are optimized through 12198 and 12484 trainable parameters, respectively.

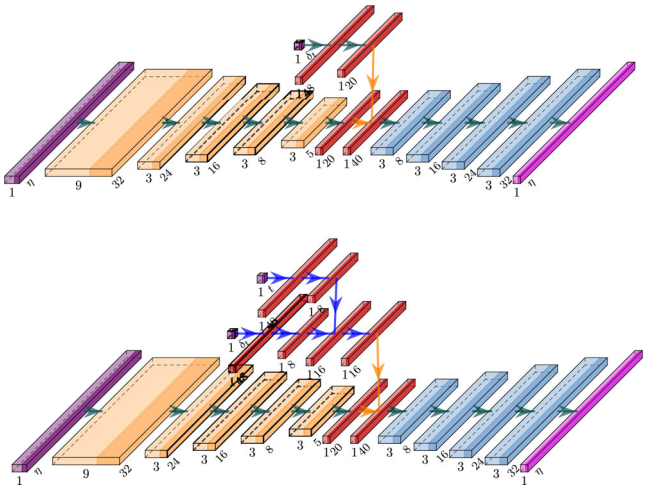


FIGURE 2 Architecture of convolutional encoder–decoder models for predicting future FTIR spectra. (Top) Single input time network. (Bottom) Many input time network. The architectures are chosen to be independent from t_0 and t_c. Convolutional and deconvolutional layers are represented in orange and blue colours where the first subscript denotes the number of filters and the second (rotated) subscript is the filter size. The red colour represents a dense layer. By blue arrows we denote a connection through a dropout layer (dropout probability 0.3-0.5 was used) while the orange arrows highlight the concatenation of the temporal branch with the (compressed) spectrum. All layers except the last are activated by rectified linear units ($ReLU(x) = max(0, x)$). Both network types have approximately 12 thousand weights

2.2 | FTIR dataset

The FTIR data are based on a recently published study [4] where 11 protein-rich food processing materials: chicken mechanical deboning residue (CR), heat treated chicken mechanical deboning residue (HC), chicken skin (CS), chicken bone (CB), turkey carcasses (TC), turkey mechanical deboning residue (TR), chicken muscle (CM), salmon heads (SH), salmon bone (SB), salmon skin (SS) and mackerel (Ma) were hydrolysed by a selection of five commercially available enzymes: Alcalase (A), Papain (Pa), Protamex (Pr), Flavourzyme (F) and Corolase (C), in addition to natural enzymes present in the raw materials (autolysis). A total of 28 different substrate–enzyme combinations were performed in the study, and FTIR spectra were obtained at different timepoints for all hydrolysis reactions, including {0.5, 2.5, 5, 7.5, 10, 15, 20, 30, 40, 50, 60, 80}min since onset of the reaction. Not all reactions were sampled for all substrate–

enzyme combinations. This resulted in a total of 885 FTIR spectra, as shown in Table 1.

The proposed networks will be trained to map between spectra from the FTIR dataset which are represented as real-valued vectors of size $n = 571$. To compensate for physical effects such as light scattering affecting the spectra, pre-processing is performed using Savitzky-Golay [31] filtering (with third-order polynomials, second-order derivatives and a local window size of 11) followed by Extended Multiplicative Signal Correction [32] (with quadratic baseline). The spectra are cut to the region 1800–700 cm^{-1} before modelling. For each of the 28 hydrolyses the set contains at most 12 sets of spectra which correspond to measurements at the different time points.

We note that neither of the networks S_{t_0}, M_{t_c} take as the input the encoding of the substrate or the enzyme. This network design choice is in contrast to the FTIR-to-AMW model presented in [4] where different regression models were used to predict AMW after an initial classification of the input substrate type. The reason why the two-step/hierarchical approach was need, was that the linear models used could not disentangle the variation due to substrate–enzyme and degree of hydrolysis. Such a hierarchical approach is challenging in our setting due to scarcity of data for some of the substrates. Moreover, explicit representation of substrate–enzyme can considerably increase the size of the input space and thus potentially the number of training parameters. This is in particular the case for the one-hot vector encoding used to code categorical inputs into binary format. However, the results of Måge et al. [33] suggest that FTIR spectra can be used to predict material composition (substrate) as well as processing factors (such as enzyme types). The observation is well aligned with our preliminary experiments where addition of one-hot encoded enzyme types did not translate to greater accuracy of the models. Finally, the inherent non-linearity means that the networks can learn these complex information mixtures with quite high precision without further aids. They also have the potential to be more robust and generalizable, since they are not dependent on knowing the substrate–enzyme combination at the time of prediction.

2.3 | Training set and training process

The training set used to optimize the FTIR networks is based on the FTIR dataset described in the previous section. However, we omit several measured reactions for testing purposes and augment the training set by 'synthesized' hydrolyses in order to increase robustness of the models (explained below).

To test generalization capabilities of the trained networks we leave out one hydrolysis of the CM-A and

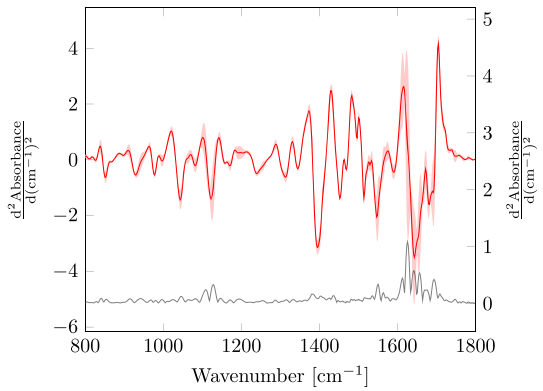


FIGURE 3 Variations in spectral measurements. The mean (solid red line) spectra $\bar{\eta}_{7.5}$ at time $t = 7.5$ min from eight hydrolyses of CM-A pair. Shaded regions denote values *three* standard deviations away from the mean. Standard deviation itself is plotted in black against the right axis. Large variations near 1100 cm^{-1} and 1700 cm^{-1} can be seen

TABLE 1 Composition of the FTIR dataset. For a given row (substrate type) and column (enzyme type) the numbers in the brackets count respectively the hydrolysis reactions and the FTIR spectra. Coloured pairs show the reactions left out for validation: either one reaction (orange) or all reactions (red) were removed. Absence of experimental data for the given pair is indicated by –

Subenz	A	Pa	Pr	F	C	Autolysis
CB	(1, 12)	(1, 12)	(1, 12)	(–, –)	(–, –)	(–, –)
CM	(8, 87)	(2, 23)	(1, 12)	(–, –)	(–, –)	(–, –)
CR	(8, 89)	(2, 24)	(3, 36)	(–, –)	(–, –)	(–, –)
CS	(2, 22)	(2, 24)	(1, 12)	(–, –)	(–, –)	(–, –)
TR	(2, 24)	(–, –)	(–, –)	(2, 24)	(2, 24)	(–, –)
Ma	(1, 12)	(1, 11)	(–, –)	(1, 12)	(–, –)	(1, 12)
SB	(1, 11)	(–, –)	(–, –)	(–, –)	(–, –)	(–, –)
SH	(12, 130)	(–, –)	(–, –)	(–, –)	(–, –)	(–, –)
SS	(12, 132)	(–, –)	(–, –)	(–, –)	(–, –)	(–, –)
TC	(2, 22)	(–, –)	(–, –)	(2, 23)	(2, 24)	(–, –)
HC	(2, 23)	(2, 24)	(1, 12)	(–, –)	(–, –)	(–, –)

SS-A data, which can be seen to be among the most frequent reactions in their respective groups (poultry and fish), see Table 1. With this choice we aim to see if the networks can generalize to new spectra drawn from a seen distribution. In addition, all hydrolyses with the Flavourzyme (F) enzyme and mackerel hydrolysis without a specific enzyme (Ma) were removed in order to test generalization to unseen enzymes. In summary, the dataset of measured reactions available for training has been reduced to 70 reactions counting 792 spectra.

The extended FTIR dataset used for training the encoder–decoder networks augments the measured hydrolyses with input–output pairs obtained by combining data from the different experiments available for most of the substrate–enzyme pairs in the original dataset. We remark that the aim of the data augmentation is to increase the robustness of the networks to raw material variation and measurement noise. The variability of spectra from different realizations of EPH with fixed substrate–enzyme pair is illustrated in Figure 3. Here, with CM-A, large standard deviations between the reactions can be seen for wavenumbers close to 1100 cm^{-1} and 1700 cm^{-1}. We remark that this localization is common to most pairs/reactants.

We will next describe the extended training set for S_{t_0} networks; for M_{t_c} networks the process is simply repeated for each $t_0 \le t_c$ and the resulting datasets are merged. To this end, let us fix the input time t_0. Then, for a substrate–enzyme pair the FTIR dataset contains $1 \le k \le k_{(substrate,enzyme)}$ measured input spectra, and for each k the set of spectra at time $t > t_0$ represents the target outputs. As an example, Table 1 gives seven possible input spectra† for CM-A hydrolysis. If the reactions were performed under identical external conditions, one can assume that the differences in spectra, see Figure 3, are primarily due to raw material variation and measurement error. In turn, new input–output pairs can be constructed by combining inputs from one reaction with the outputs from the remaining ones. We use approximately 80% of the available combinations in order to augment the measured data and thus finally form the training set for the S_{t_0} network. Note that the size of the training set depends on the input time t_0 with larger times leading to less data.

In order to train the network, the spectra in the extended datasets are normalized to have unit standard deviation while we normalize the temporal variables to lie roughly in the unit interval.‡ The networks are then trained to minimize the mean squared error using 5000 epochs of the ADAM optimizer with randomized mini-batches of size 40 for S_{t_0} and 64 for M_{t_c}. For both network types the weights were initialized by Xavier Glorot initialization [34]. During optimization, 20% of the training dataset is used for calculating validation error, and the learning rate is kept constant at 10^{-3}. Let us remark that the loss functions of the S_{t_0}, M_{t_c} networks are identical. In particular, we do not enforce the consistency property $M_{t_c}(t_i,\eta_i,\delta t_i) \approx M_{t_c}\left(t_j,\eta_j,\delta t_j\right)$ for t_i, $t_j \le t_c$ and $t_i + \delta_i = T = t_j + \delta_j$, see Section 2.1.

3 | SINGLE INPUT TIME NETWORKS

In the following we will discuss performance of networks S_{t_0} for input times $t_0 = 0.5$, 2.5, 5.0 min. The prediction error at time t will be defined as Mean Squared Error: $\text{MSE}_t = \frac{1}{n}\frac{1}{p}\sum\sum(\eta_t - \widehat{\eta}_t)^2$ and Root Mean Squared Error: $\text{RMSE}_t = \sqrt{\text{MSE}_t}$. The overall error across all times t for which inference can be performed is then $\text{MSE} = \frac{1}{T}\frac{1}{n}\frac{1}{p}\sum\sum(\eta_t - \widehat{\eta}_t)^2$ and Root Mean Squared Error: $\text{RMSE} = \sqrt{\text{MSE}}$ where T is the number of valid prediction times (which depends on t_0 and t_c). Finally, for the individual spectra we will also discuss the absolute error: $\max|\eta_t - \widehat{\eta}_t|$. Here, all the error measures have the unit $\frac{\text{d}^2\text{Absorbance}}{\text{d(cm)}^2}$, which we omit in the text in order to simplify the notation.

For each t_0 we have trained five networks which differ by the random seed value in the networks' weight initialization. Due to the low number of samples and high heterogeneity in the raw materials, some variation in prediction results of the models that are refitted is to be expected. The average standard deviation in predictions caused by varying the random seed ranged from 0.01 to 0.015. Based on the limited data available for this study, this is a low enough instability to be acceptable for practical usage. In the remainder of the article we will therefore for simplicity report results for a fixed choice of the seed.

Convergence of the training process for S_{t_0} networks is summarized in Figure A1 (see Appendix 7). At the final epoch the networks achieve similar accuracy on the training set while on the validation set the error of $S_{0.5}$ is slightly higher; $\text{MSE}_{0.5} = 0.016$ to be contrasted with $\text{MSE}_{2.5} = 0.013$ and $\text{MSE}_{5.0} = 0.011$. However, we recall that higher t_0 means that inference (at times $t > t_0$) avoids the initial stages of the reaction where most dynamics are expected. In this sense the learning problem is potentially simpler which could explain the observation that both in training and validation the errors $\text{MSE}_{5.0} < \text{MSE}_{2.5} < \text{MSE}_{0.5}$. Another potential reason is the fact that for large t_0 the temporal distance between input and output spectra is shorter.

To further analyse performance of the networks, Section 3.1 considers the accuracy on the *original* FTIR

dataset. That is, the networks will be evaluated using only the measured hydrolyses. Section 3.2 then addresses generalization of the networks as the analysis therein concerns the set withheld from training, see Table 1. We remark that the training data was obtained as random subset of combined reactions from the FTIR set. Thus it is possible that the *exact* measured sequence of spectra (a concrete measured hydrolysis) was not seen during training.

3.1 | Performance on FTIR dataset

To dissect the inference errors of the single input networks we choose to consider at first S_{t_0} with $t_0 = 0.5$ min since this network is the least accurate in the training metrics, see Figure A1.

Analysis of $S_{t_0=0.5}$

Fixing inference time at $t = 10$ min, Figure 4 shows the network error for CM-A reactions. We note that for this substrate–enzyme pair there are seven reactions, that is, seven input (and target) spectra are available for S_{t_0}. Picking one of the reactions, it can be observed that the network's absolute error is approximately 0.5 (which translates to relative error of approximately 12%) and the error is largely localized in the spectral range (1600, 1650) cm^{-1}, see also Figure A2 in Appendix A. However, the measured target spectra show large variance for these wavenumbers as can be seen in the bottom pane of Figure 4. Indeed, if the targets are averaged and compared with the mean prediction obtained by averaging network outputs for all the 7 measured input spectra the error drops to 0.1 (or relative error of 3%) and is rather delocalized, see also Figure A2. This suggests that the network does not overfit to the individual reactions but instead learns certain 'averaged' reactions. The claim can be further supported by the fact that the averaged predictions have a smaller standard deviation, 0.35, compared to the averaged standard deviation of 3.23 for the data at $t = 10$ min or the averaged standard deviation of the input values 0.96.

We next keep the inference time fixed at 10 min and vary the substrate–enzyme pairs. The results are summarized in Table A.1. It can be seen that the low/high prediction errors seem to correlate well with the number of hydrolysis experiments performed for the given combinations. For example, for CB-Pr and CB-Pa errors as large as 0.27 and 0.40 are observed and for these reactions we have only one experiment available, see Table 1. We remark that Table A.1 (as well as Tables 2, 3) *do not* include reactions for Ma or with the Flavourzyme enzyme since these pairs were removed from the training set for validation purposes, see Section 2.3.

To continue our dissection of the network prediction error, let us now fix the substrate–enzyme pair and vary the inference time. In Figure 5 we summarize the results for the CM-A combination. Similar to the stationary case we see that the error of a single reaction is larger than when the averages are compared. As in the stationary case, the error based on the single reaction is localized (close to the amide1 band) especially in the early stages of the reaction. For $t_i = 2.5$ min errors close to 0.9 can be seen. On the other hand, the error in averaged predictions seems to be rather constant in time and largely delocalized in the wavenumber space. In particular, the error in the amide1 band is approximately 0.1. Interestingly, the averaging had very little effect on the

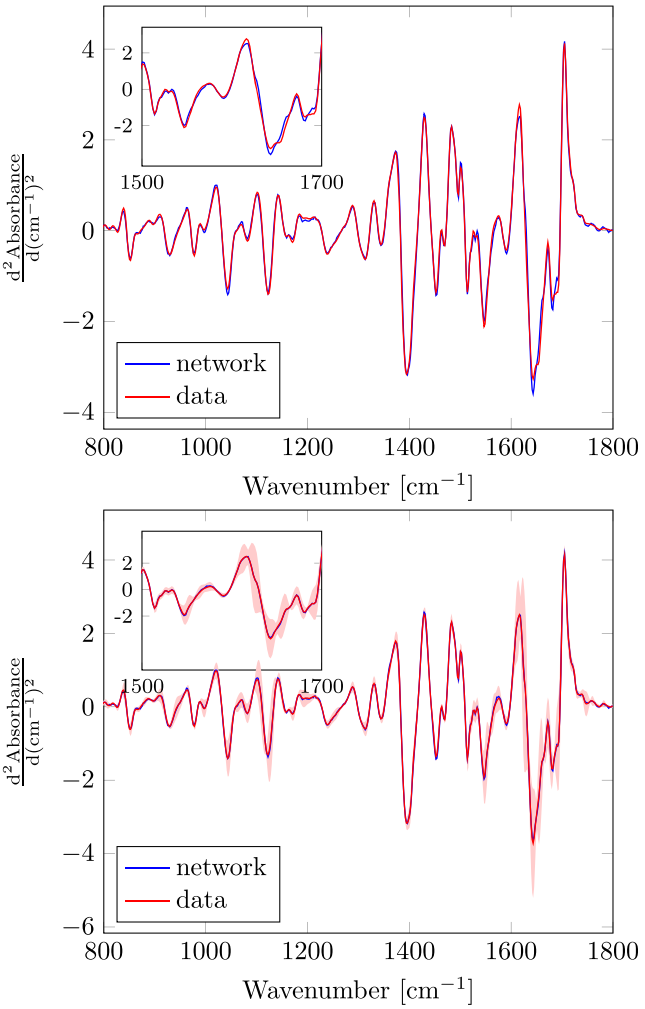


FIGURE 4 Prediction errors of single input network with time $t_0 = 0.5$ min. The inference time and substrate–enzyme pair are fixed at $t = 10$ min and CM-A. (Top) Prediction using one of seven available inputs. (Bottom) Mean prediction obtained as average of all seven measured input spectra. Shaded region shows values 3 standard deviations away from the mean

TABLE 2 Total prediction errors of single input network $S_{0.5}$. RMSE is reported based on all hydrolysis spectra of given substrate–enzyme pairs

Subenz	A	Pa	Pr	C
CB	0.16	0.32	0.22	—
CM	0.13	0.10	0.17	—
CR	0.11	0.11	0.15	—
CS	0.14	0.14	0.14	—
TR	0.15	—	—	0.12
Ma	0.15	0.18	—	—
SB	0.11	—	—	—
SH	0.06	—	—	—
SS	0.06	—	—	—
TC	0.21	—	—	0.19
HC	0.29	0.28	0.21	—

TABLE 3 Total prediction errors of single input networks $S_{2.5}$ (blue) and S_5 (green). The RMSE errors are computed based on all inferable hydrolysis spectra (i.e. η_t for $t > t_0$) of given substrate–enzyme pairs

Subenz	A	Pa	Pr	C
CB	0.12, 0.13	0.21, 0.20	0.15, 0.16	—
CM	0.12, 0.11	0.09, 0.09	0.11, 0.15	—
CR	0.10, 0.11	0.09, 0.10	0.14, 0.15	—
CS	0.12, 0.13	0.12, 0.12	0.14, 0.13	—
TR	0.14, 0.15	—	—	0.12, 0.12
Ma	0.14, 0.12	0.10, 0.13	—	—
SB	0.09, 0.09	—	—	—
SH	0.06, 0.07	—	—	—
SS	0.06, 0.05	—	—	—
TC	0.16, 0.17	—	—	0.17, 0.20
HC	0.25, 0.20	0.27, 0.18	0.33, 0.24	—

prediction error for wavenumbers close to 1425 cm^{-1}. In Table 2 we finally show the global prediction error of the $S_{0.5}$ network. The results are largely consistent with the errors at early inference time $t = 10$ min, see Table A.1. This observation again confirms the expectation that the error is localized in time close to the onset of the reaction.

Comparison across input times

In order to discuss the role of the input time t_0 we will compare performance of the different single input networks S_{t_0}, where $t_0 = 0.5, 2.5, 5$ min. We recall that the networks differ by their inference times, $t_i > t_0$ and therefore direct comparison requires predicting spectra at times that are common to all S_{t_0}, here $t > 5$ min.

Before the direct comparison we list in Table 3 the global prediction errors (similar to Table 2) of the new networks. Here the errors are computed on the maximum inferable part of the FTIR dataset, that is, $S_{2.5}$ will, in addition to all of the S_5 target spectra, predict also those corresponding to $t = 5$ min. In terms of the number of spectra, S_5 is evaluated on 585 samples while $S_{2.5}$ uses 683. Interestingly, we observe that the error of $S_{2.5}$ (on a larger dataset) is typically smaller than for S_5. In fact, of the 24 substrate–enzyme pairs S_5 is more accurate for HC-{A, Pa, Pr} in addition to Ma-A, SS-A, CM-A, CB-Pa and CS-Pr where it is only in the HC reactions where the difference is significant, for example, for HC-A the errors with $S_{2.5}$ and S_5 are respectively 0.25 and 0.25. At the same time, comparing with Table 2, the S_5 on average yields smaller errors than $S_{0.5}$. These observations suggest that there might be an optimal input time t_0 in terms of the accuracy of single-input networks.

We investigate the suggested optimality of $t_0 = 2.5$ min further in Figure 6. Therein the accuracy of predictions is measured using the *common* part of the training FTIR dataset corresponding to reaction times greater than 5 min. Then the errors achieved by the networks are 0.159, 0.141 and 0.137 for $S_{0.5}$, $S_{2.5}$ and $S_{5.0}$ respectively. However, as in the previous comparison, $S_{2.5}$ is most accurate in most reactions. In fact, we observe that $S_{2.5}$ is outperformed by S_5 only for Ma-A, SS-A, CB-Pa (where the differences between the networks' accuracy are small) and systematically in the hydrolyses of HC substrate. With the latter reactions removed, the performance of all networks improves to RMSE = 0.153, 0.132 and 0.133.

Performance on the validation set

We next evaluate the ability of trained single-input networks to generalize to data unseen during training. Similar to the training set performance, we will first consider a fixed input time t_0 and address accuracy for a given inference time or material. Here $S_{2.5}$ is considered since it showed highest accuracy on most reactions in the training set. Comparison of the networks is presented at the end of the section.

Analysis of $S_{t_0=2.5}$

We recall that the validation set consists of reaction spectra from one or more hydrolysis using the substrate–enzyme pairs CM-A, SS-A, TC-F, TR-F, MA-F and Ma. Here the hydrolysis with Flavourzyme and mackerel

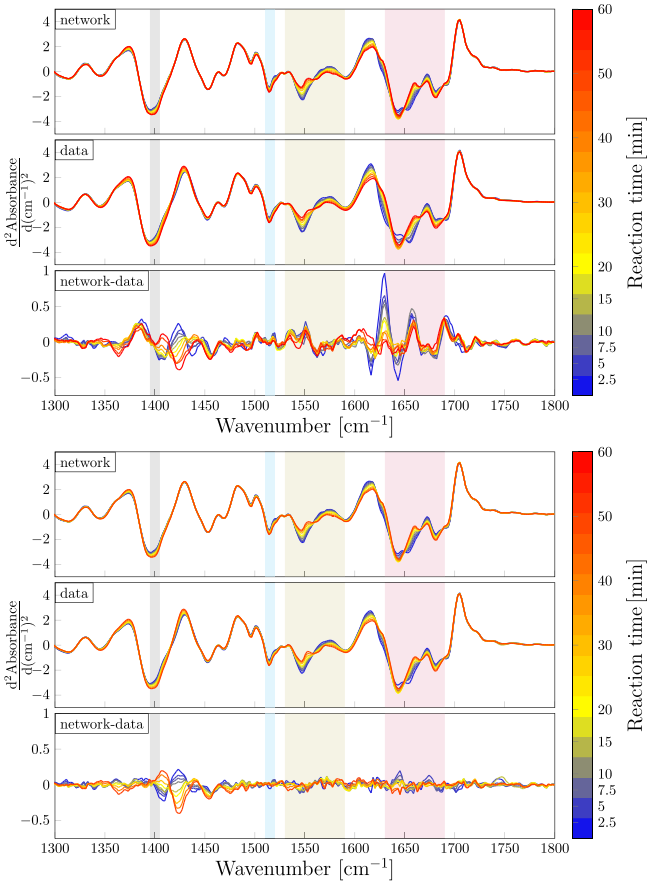


FIGURE 5 Prediction errors for network $S_{0.5}$ and the entire hydrolysis process with CM-A pair. (Top) A single reaction is compared with corresponding network predictions based on the reaction spectrum at $t_0 = 5$ min. (Bottom) Mean of the reactions is compared with the averaged predictions from the possible inputs. The mean concerns only sampling times which are common to *all* reactions. Here, some reactions were missing data for $t = 60$ min. The coloured stripes show the approximate spectral bands of amide1 (1630, 1690) cm^{-1}, amide2 (1530, 1590) cm^{-1}, nh3 (1510, 1520) cm^{-1} and coo (1395, 1405) cm^{-1}

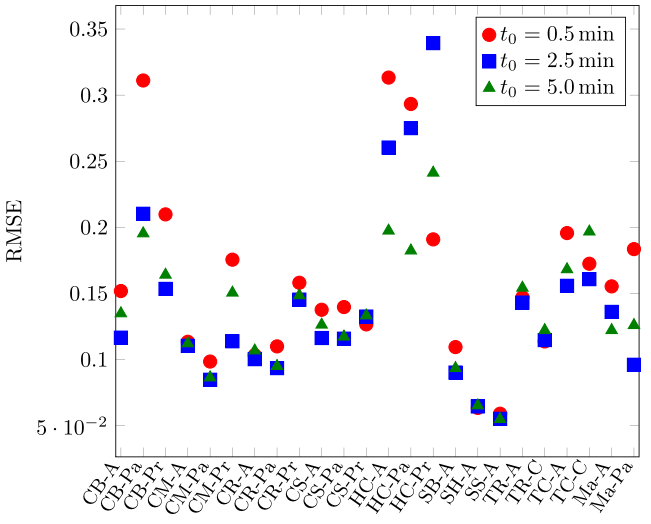


FIGURE 6 Comparison of total accuracy of single-time input networks S_{t_0}, $t_0 = 0.5, 2.5, 5$ min on a common part of the FTIR dataset. Spectra for $t > 5$ min are inferable for all the networks. Network $S_{2.5}$ is the most accurate

were not seen in the training, both of which showing substantially differences from the other FTIR spectra.

In Figure 7 we consider generalization of the network to inputs from the CM-A and SS-A reactions, i.e., data drawn from a distribution used in the training, see Table 1. Comparing with Figure 5 we observe that the error on unseen CM-A input is similarly localized to wavenumbers close to 1600 cm^{-1} with large overshoots in the spectra for inference times $t < 30$ min. However, the error is almost twice as large compared to the spectrum drawn from the training set. On the other hand, for SS-A the RMSE error does not exceed 0.06 for any of the input times. We hypothesize these observations are due to the choice of the training set for S_{t_0} which effectively forces the network to predict averages of several hydrolyses. For SS-A the unseen reaction is then captured well-enough by the learned average.

Typical features of the generalization error of the single-input networks with unseen reactants are illustrated in Figure 8 where predictions for TC-F and Ma are shown. For fixed inference time $t = 10$ min we observe that the error is largely due to shifts between the spectra (e.g. coo- band for TC-F, Ma or amide2 band for Ma) and smoothing (e.g. amide2 TC-F). However, smoothing in general does not appear to be associated with high frequency oscillations in the spectra but rather with wavelengths and the spectral intensity. In particular, for Ma small amplitude and high frequency features in the amide1 are well resolved. On the other hand, for TC-F the features that are smoothed/ignored by the network have longer wavelengths but higher amplitudes. Concerning predictions for different time points, we notice that the errors are in general not localized in space as was the case with the reactants seen during the training. However, for SS-A and CM-A we see that the error does not decrease for later stages of the reaction.

The generalization error is further analysed in Table 4 where the RMSE for the individual reactions are shown as a function of time. Comparing with Table A.2 we conclude that the error for seen reactions is comparable to those drawn from the training set. For SS-A the performance is practically identical (RMSE ~ 0.05) for all time points. In case of unseen CM-A spectra, the network generalizes well in later stages $t > 20$ min while the error is large close to the onset (RMSE ~ 0.20 vs. RMSE ~ 0.10). For the reactants that were not part of the training set we notice that the error is large in particular for the turkey

substrate (RMSE $\sim$ 0.6 while the largest error on the training set was RMSE $\sim$ 0.3).

We conclude the section by comparing the networks $S_{0.5}$, $S_{2.5}$ and $S_{5.0}$. Using target spectra common to all networks ($t > 5$ min), Table 5 reveals that the networks perform rather similarly on the seen substrate–enzyme pairs. In case of the unseen reactants, there is a notable difference between mackerel and turkey substrate. With the latter the accuracy of the networks improves for larger t_0, while with Ma reactions $S_{0.5}$ can yield smaller error that the other networks. We remark that on the training set $t_0 = 0.5$ min yielded consistently the least accurate predictions, see Figure 6.

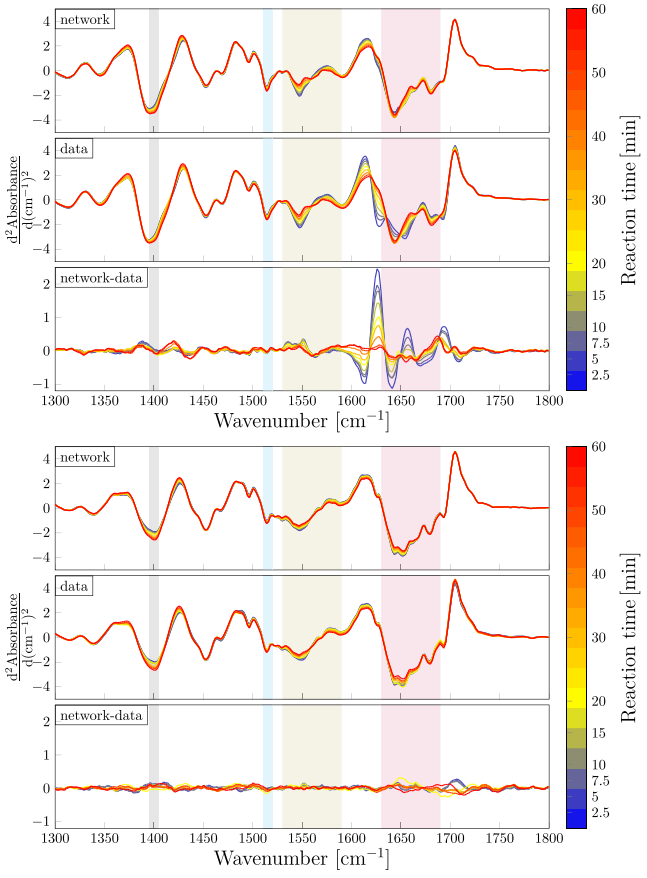


FIGURE 7 Generalization error of $S_{t_0=2.5}$ network to reactants seen in the training. (top) CM-A and (bottom) SS-A

4 | MANY-INPUT TIME NETWORKS

In this section we discuss performance of the many-input networks M_{t_c} which aim to improve the accuracy of S_{t_0}. Here cut-off times $t_c = 7.5$, 15, 30 min will be considered. We recall that the inputs to the network form a triplet $(\eta_{t_i}, t_i, \delta_{t_i})$ where $t_i < t_c$ is the time point corresponding to input spectrum η_{t_i} and $\delta_{t_i} > 0$ is the temporal offset such that $t = t_i + \delta_{t_i}$ is the inference time; $\eta_t = M_{t_c}(\eta_{t_i}, t_i, \delta_{t_i})$. Therefore, if $t_{c_0} < t_{c_1}$ the network $M_{t_{c_1}}$ can predict η_t from more inputs than $M_{t_{c_0}}$ and correspondingly, the size of the training set of $M_{t_{c_1}}$ is larger. Convergence of the optimizer for training M_{t_c} networks is summarized in Figure A1. All three networks§ achieve similar RMSE on their respective training/validations sets, where $M_{t_c=30}$ is slightly more accurate $\text{RMSE}_{\text{valid}} = 7.12 \cdot 10^{-3}$, $\text{RMSE}_{\text{train}} = 9.14 \cdot 10^{-3}$ (to be contrasted with $\text{RMSE}_{\text{valid}} = 7.43 \cdot 10^{-3}$, $\text{RMSE}_{\text{train}} = 9.78 \cdot 10^{-3}$ for M_{15}). We remark that, unlike with single-input networks S_{t_0}, the inference time for all M_{t_c} networks here starts at $t = 2.5$ min.

In order to assess accuracy of M_{t_c} we will next address the issue of non-uniqueness of predictions by many-input networks. Given network M_{t_c} with $t_c > 0.5$ min, observe that using the measured input spectra the prediction at a given time can be obtained in multiple ways. More precisely, let $t_i \leq t_c$ be possible input times, and for inference time t, let δ_{t_i} such that $t = t_i + \delta_{t_i}$ be the temporal offsets. Note that by construction of the training process only positive offsets $\delta_{t_i} > 0$ constitute a valid input to M_{t_c}. For any triplet $(t_i, \eta_{t_i}, \delta_{t_i})$ the network M_{t_c} then computes a

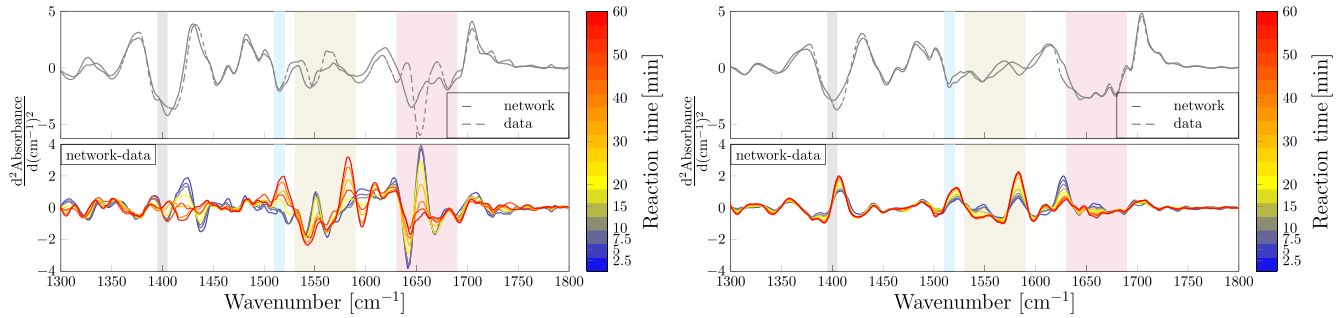


FIGURE 8 Generalization error of $S_{t_0=2.5}$ network to two of the reactants (left) TC-F and (right) Ma unseen during training. In each plot comparison of instantaneous spectra at time $t = 10$ min is shown in the top panel (grey curves). The bottom panel compares predictions with measured spectra at different times during the hydrolysis

(possibly different) approximation η_t. This property of many-input networks is illustrated in Figure A3 using M_{15} as an example. Recalling that in the analysed FTIR dataset the typical measurement time points are 0.5, 2.5, 5.0, 7.5, 10, 15, ... min the network can thus predict $\eta_{t>15}$ from 6 different inputs. However, we observe that the predictions are consistent in the sense that the they have a well-defined mean.

Based on the observation in Figure A3 we will in the following obtain a unique predicted spectrum based on all measured admissible inputs using an averaged network whose value at time $t = t_i + \delta_{t_i}$ is given as

$$\overline{M_{t_c}}(t) = \text{mean}_{\delta_{t_i}>0, t=t_i+\delta_{t_i}} M_{t_c}\left(t_i, \eta_{t_i}, \delta_{t_i}\right). \quad (1)$$

Thus predictions at $t > t_c$ are computed based on *all* input times $t_i \leq t_c$, while predictions at earlier times use fewer input spectra. We remark that different constructions are possible in order to obtain the single spectrum, for example, $\overline{M_{t_c}}(t) := M_{t_c}\left(\min_{t_i}, \eta_{t_i}, \delta_{t_i}\right)$.

4.1 | Performance

We proceed to evaluate prediction errors of the many-time input networks on the raw FTIR dataset similar to S_{t_0} in Section 3.1. For the sake of brevity the cut-off time will be fixed as $t_c = 15$ min and we consider the error of the averaged output $\overline{M_{15}}$.

Broken down by substrate–enzyme pairs, Table 6 shows the RMSE error computed from spectra at $t \geq 5$ min and all the reactions of a given pair. This set allows for a direct comparison with performance of $S_{0.5}$ (see Table 2); with $S_{0.5}$ the overall RMSE error is 0.137 while $\overline{M_{15}}$ yields RMSE = 0.129. Comparing the individual reactions we observe that the many-input network is more accurate for hydrolyses with Ma, HC and TC substrates for which at most two set of measured spectra were available. We recall that, having approximately 2% more degrees of freedom, the M_{t_c} networks are slightly larger than S_{t_0}. However, the size difference is small and thus the improved accuracy is not likely to be due to overfitting.

We compare the accuracy of the different many-input networks in Figure 9. Similar to S_{t_0}, the larger cut-off times lead to greater accuracy; the observed RMSE errors for $\overline{M_{t_c}}$, t_c = 7.5, 15, 30 min are 0.126, 0.129 and 0.113 respectively. Each of the networks is then more accurate than the most performant single-input network S_5 (RMSE = 0.137). In agreement with Table 6, the improvements of $\overline{M_{t_c}}$ can be traced to the reactions with turkey substrates. We recall that the three networks are identical in their architectures but greater t_c yields larger sets used for training.

Taking M_{30} as the most accurate of the many-time input networks we finally consider the generalization error. For the six reactions withheld during training, the RMSE evolution over time is listed in Table 7. Comparing with $S_{t_0=2.5}$ shown in Table 4 the two networks perform similarly in general. In particular, for TC-F, Ma and SS-A hydrolyses the relative difference between the networks' prediction is below 10%. The largest difference appears for Flavourzyme hydrolyses of TR where the many-input RMSE = 0.75 and the single-input RMSE = 0.63, and the error of M_{30} is consistently large(r) for all the time points.

5 | PREDICTING FUTURE AMW

Having demonstrated the abilities of encoder–decoder models to predict the FTIR fingerprint of EPH, we will now consider applying the networks in modelling the future product quality characteristics. To this end we will

TABLE 4 Generalization (RMSE) error of $S_{t_0=2.5}$ network to reactants unseen during training

Time (min)	CM-A	TR-F	TC-F	Ma-F	Ma	SS-A
5	0.27	0.73	0.73	0.22	0.33	0.05
7.5	0.22	0.67	0.70	0.23	0.32	0.05
10	0.20	0.62	0.70	0.23	0.31	0.04
15	0.16	0.53	—	0.24	0.33	0.04
20	0.12	0.50	0.58	0.24	0.34	0.06
30	0.11	0.53	0.59	0.27	0.37	0.03
40	0.09	0.62	0.59	0.28	0.40	0.04
50	0.09	0.67	0.64	0.32	0.41	0.04
60	0.09	0.69	0.68	0.34	0.41	0.05
80	—	0.71	0.75	0.39	0.42	—

TABLE 5 Generalization (RMSE) errors for single input networks S_{t_0} with $t_0 = 0.5, 2.5, 5.0$ min. Maximal set concerns inference times $t > t_0$ while $t > 5.0$ min for the common set

t_0[min] Sub-enz	0.5	2.5	5	0.5	2.5	5
	Maximal set			Common set		
Ma	0.28	0.37	0.30	0.29	0.37	0.30
TR-F	0.80	0.63	0.59	0.79	0.62	0.59
TC-F	0.80	0.66	0.62	0.78	0.66	0.62
Ma-F	0.26	0.28	0.28	0.28	0.29	0.28
CM-A	0.18	0.16	0.14	0.13	0.14	0.14
SS-A	0.05	0.05	0.04	0.04	0.04	0.04

TABLE 6 Prediction errors of Equation (1)-averaged many-input network M_{15} measured on the raw FTIR dataset, see Table 2 for $S_{0.5}$. The inference covers times $t \geq 5$ min

Subenz	A	Pa	Pr	C
CB	0.17	0.23	0.22	—
CM	0.13	0.08	0.11	—
CR	0.12	0.10	0.16	—
CS	0.12	0.12	0.12	—
TR	0.17	—	—	0.13
Ma	0.11	0.09	—	—
SB	0.09	—	—	—
SH	0.07	—	—	—
SS	0.06	—	—	—
TC	0.14	—	—	0.13
HC	0.10	0.12	0.11	—

use the average molecular weight (AMW) and utilize the hierarchical model of Kristoffersen et al. [4], which uses spectra together with categorical information about the substrate–enzyme pair, to predict the current AMW, at the hydrolysis time corresponding to input spectrum. Denoting AMW_t the average molecular weight at time t, and $\mathcal{S}$ the set of substrate–enzyme pairs, the model thus predicts $\mathrm{AMW}_t = H(\eta_t, s)$ for η_t the spectrum of hydrolysis with a reactant $s \in \mathcal{S}$. Using the future spectra η_t, $t = \delta_t + t_0$ predicted by the S_{t_0} network as the input for the hierarchical model we then aim to predict AMW_t as $\mathrm{AMW}_t = H\left(S_{t_0}\left(\eta_{t_0}, \delta t\right), s\right)$. For the many-time input networks the construction is analogous once a unique η_t is determined. Here we will use the average from Equation (1). We remark that in the following the two models are trained separately using their independent loss functions. This simple choice is due to practical considerations where it is potentially useful to obtain a modular quality controller by composition. That is, a predictor for a different quality indicator could be constructed by providing its dedicated model (to replace the H function above). Alternatively, variational encoders/decoders could be modified to include elements of $\mathcal{S}$ as additional input and output both η_t and AMW_t using predictions of the H model as part of the loss function. The combined approach is clearly specific to the given quality indicator and while not modular the resulting model is expected to be more accurate in predicting AMW than the combined model.

To evaluate the combined model we next replace the spectra at time t from the FTIR dataset by outputs of $S_{t_0=0.5}$ and averaged $\overline{M}_{t_c=30}$ networks. More precisely, for fixed s, we use all initial spectra η_{t_0} for $S_{0.5}$ and η_{t_i}, $t_i \leq t_c$ for M_{15} to predict the spectra η_t, $t > t_0$. The accuracy of the combined models are evaluated in Figure 10. Here the pairs CM-A and SH-A are used since they are among the most frequent reactions in the dataset. In addition, the reactants illustrate sufficiently the typical features of

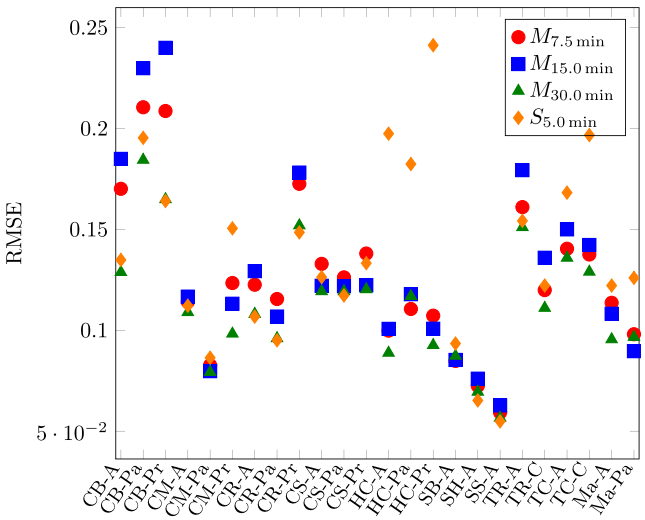


FIGURE 9 Comparison of total accuracy of many-time input networks M_{t_c}, $t_c = 7.5, 15, 30$ min on FTIR dataset inferring spectra η_t, $t \geq 5$ min

the combined models. Focusing first on the single input network, we notice that for CM-A (second and fourth row of Figure 10) the model over-predicts AMW_t and the difference is approximately constant in time. On the other hand, for SH-A the model predictions are inaccurate for the early stages of the reaction ($t < 20$ min) and improve afterwards. The predictions from the model based on M_{30} are then more accurate for both CM-A and SS-A. More precisely, in terms of R^2 measured on all reaction times, the latter model achieves 80.5% and 70.4% on the respective pairs, while accuracy using $S_{0.5}$ is 71.1% and 50.9%. Since in the context of control of the hydrolysis, one is mostly interested in AMW at later times, we

TABLE 7 Generalization error of $M_{t_c=30}$ network to reactions unseen during training

Time (min)	CM-A	TR-F	TC-F	Ma-F	Ma	SS-A
2.5	0.39	0.90	0.87	0.18	0.37	0.07
5.0	0.29	0.78	0.75	0.18	0.36	0.05
7.5	0.23	0.71	0.71	0.17	0.32	0.04
10	0.20	0.68	0.71	0.17	0.31	0.04
15	0.16	0.61	—	0.17	0.30	0.04
20	0.11	0.60	0.61	0.17	0.30	0.06
30	0.10	0.65	0.62	0.22	0.32	0.04
40	0.08	0.73	0.58	0.24	0.34	0.04
50	0.09	0.79	0.62	0.30	0.35	0.04
60	0.10	0.83	0.67	0.33	0.35	0.05
80	—	0.91	0.77	0.38	0.36	—

also evaluate the accuracy of the combined models using spectra from reaction times $t > 10$ min. In Figure 10 it can be seen that in this case the predictions appear more tightly clustered around the diagonal.

We remark that the AMW predictions of the combined model are rather sensitive to the input spectra. In particular, we recall that the variations in η_t obtained from different S_{t_0} are small, see Figure 6, however, they might translate to large difference in predicted AMW. To obtain more robust models than the composite ones discussed here possible alternative strategies include (1) tuning the encoder–decoder networks with added AMW, substrate and enzyme information (as mentioned above), (2) retraining the AMW prediction on the output from the networks and (3) combining the FTIR networks with AMW prediction modelling for a joint optimization. We remark that for the combined approach (3), exploring the pre-training and fine-tuning strategy [18, 19] based on FTIR autoencoders appears fruitful.

6 | SUMMARY AND DISCUSSION

In a biotechnological process, such as enzymatic protein hydrolysis, feed-forward predication of product characteristics is a powerful and essential element for process control [8]. One choice to be made when modelling in the feed-forward framework would be to either predict the entire FTIR spectrum at a future time point in the process using one or more spectra in an early phase of hydrolysis, or to directly predict future characteristics, such as AMW. The former route, predicting the FTIR spectrum, has the benefit of providing a (feature)-rich and flexible view into the future.

In this article we have constructed deep neural networks for predicting the future FTIR spectra of EPH based on single spectra from the early stages of the reaction and the temporal offsets. Two types of networks were proposed: the single-input networks are trained using input spectra at fixed input time and the temporal offset, the many-input networks use different spectrum-offset pairs in their training and were found to yield more accurate predictions. Due to a relatively small amount of data available, both architectures were trained on a dataset where the measured reactions were augmented by combining spectra from different EPH (of the same substrate–enzyme pair). The data augmentation resulted in models which do not overfit to individual spectra in the measured reactions but instead capture certain averaged features. For greater accuracy larger dataset of EPH reactions with more uniform representation of the reactions would be required. In addition, the learning problem could be modified/simplified by restricting the predictions (and inputs) in wavenumbers $> 1000\ \mathrm{cm}^{-1}$ or by considering different weighting for key parts of the spectra (e.g. amide1, amide, coo and nh3 bands) in the loss function.

In order to predict future product characteristic we have combined the neural network models predicting the future spectra with a linear FTIR-to-AMW model. While AMW is one product characteristic that can be derived from FTIR spectra, it has been shown that the IR signature of a given hydrolysate can serve as a chemical signature of a given product [33]. Therefore, given that the IR signature of the desired product is known, the approach presented here can serve as a powerful tool that can predict in the beginning of the hydrolysis process if product specification will be met. In addition, similar combinations of deep learning and a linear prediction model (demonstrated in the current study for AMW) could be developed for other important product characteristics such as degree of hydrolysis and bioactivity.

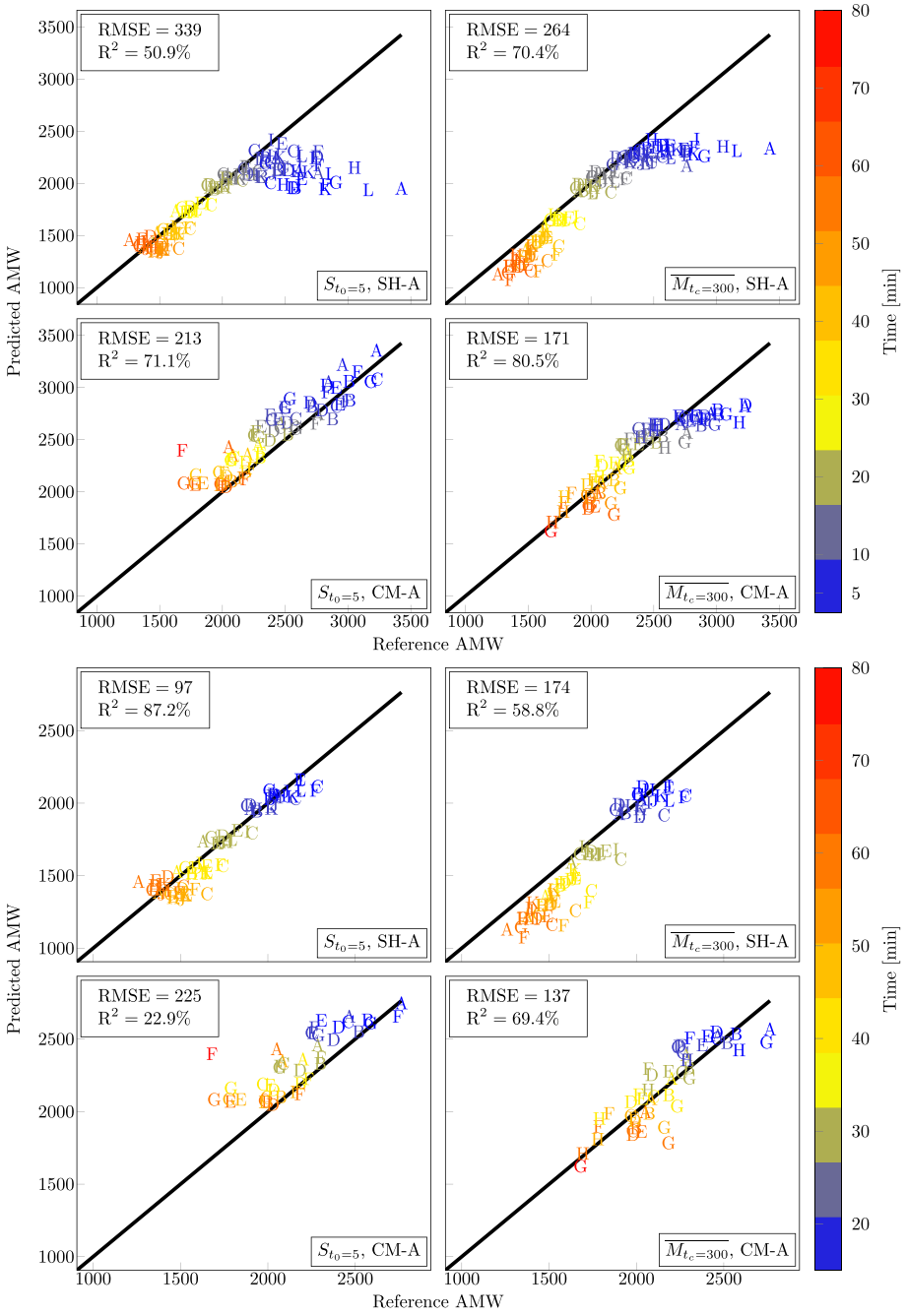


FIGURE 10 Accuracy of the composite models combining a hierarchical model [4] with FTIR encoder–decoder networks. (Left panes) Single time input network $S_{t_0=0.5}$. (Right panes) Many-time input network $\overline{M_{t_c=30}}$. Each model predicts AMW for all reactions (represented by markers) and time points (colour-encoded) of SH-A (first/third row) and CM-A (second/fourth row). The bottom four panes show outputs restricted to $t > 10$ min

The use of the resulting combined model predicting the spectra and different product characteristics in a process control setting can be envisioned in various scenarios and levels of granularity and automation. The most coarse analysis would be to use the trajectories of predictions to estimate the time point at which the reaction has reached sufficient conversion of proteins. Process parameters like temperature, mixing speeds, etc. may be adjusted in an attempt to reach optimal time usage. If the trajectories indicate that sufficient hydrolysation will not be achieved at any time point, addition of more or different enzymes may be considered. Furthermore, if predicted spectra give indications that the hydrolysis product is likely to end up outside quality specifications, preventive measures including the above may be taken early in the hydrolysis process or the process may be guided towards a lower grade end-product. Regardless of the scenario, the possibility of early intervention and decision support could be beneficial. We finally mention that if process control parameters were passed as inputs to the FTIR networks, the trained models could be used for optimal control of EPH. That is, strategies for obtaining the product of desired quality would be obtained by optimization in terms of the controls.

Overall, EPH of complex biomasses is a complex process prone to high degree of raw material and, as a result, product quality variation. Therefore, the feed-forward process control strategy presented here holds a promising potential that can ensure a stable product quality over time.

FUNDING INFORMATION

Research Council of Norway (NFR) under grant 280709.

DATA AVAILABILITY STATEMENT

The data used in this publication is available upon request.

ORCID

Miroslav Kuchta https://orcid.org/0000-0002-3832-0988
Sileshi Gizachew Wubshet https://orcid.org/0000-0001-7423-4043
Kent-André Mardal https://orcid.org/0000-0002-4946-1110
Kristian Hovde Liland https://orcid.org/0000-0001-6468-9423

ENDNOTES

* The offset time may be arbitrary, however, Vukotić et al. [26] observe that the accuracy of predictions deteriorates for large values.

† We recall that one reaction (of the 8 listed in the table) was omitted for validation.

‡ We normalize t, t_0, δt in minutes by 60 min which is the most common terminal time in the FTIR dataset.

§ The networks are identical in terms of architecture and number of parameters.

REFERENCES

[1] T. Aspevik, Å. Oterhals, S. B. Rønning, T. Altintzoglou, S. G. Wubshet, A. Gildberg, N. K. Afseth, R. D. Whitaker, D. Lindberg, Chemistry and chemical technologies in waste valorization, **2017**, 123–150.
[2] C. Ruckebusch, L. Duponchel, J.-P. Huvenne, P. Legrand, N. Nedjar-Arroume, B. Lignot, P. Dhulster, D. Guillochon, *Anal. Chim. Acta* **1999**, *396*(2-3), 241.
[3] U. Böcker, S. G. Wubshet, D. Lindberg, N. K. Afseth, *Analyst* **2017**, *142*(15), 2812.
[4] K. A. Kristoffersen, K. H. Liland, U. Böcker, S. G. Wubshet, D. Lindberg, S. J. Horn, N. K. Afseth, *Talanta* **2019**, *205*, 120084.
[5] K. A. Kristoffersen, A. van Amerongen, U. Böcker, D. Lindberg, S. G. Wubshet, H. d. V.-v. den Bosch, S. J. Horn, S. J. Afseth, *Sci. Rep.* **2020**, *10*(1), 1.
[6] S. G. Wubshet, I. Måge, U. Böcker, D. Lindberg, S. H. Knutsen, A. Rieder, D. A. Rodriguez, N. K. Afseth, *Anal. Methods* **2017**, *9*(29), 4247.
[7] N. A. Poulsen, C. E. Eskildsen, M. Akkerman, L. B. Johansen, M. S. Hansen, P. W. Hansen, T. Skov, L. B. Larsen, *Int. Dairy J.* **2016**, *61*, 44.
[8] S. G. Wubshet, J. P. Wold, N. K. Afseth, U. Böcker, D. Lindberg, F. N. Ihunegbo, I. Måge, *Food Bioproc. Tech.* **2018**, *11*(11), 2032.
[9] A. Krizhevsky, I. Sutskever, G. E. Hinton, Advances in neural information processing systems **2012**, *25*, 1097–1105.
[10] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, *Nature* **2016**, *529*(7587), 484.
[11] S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, R. Prudden, A. Mandhane, A. Clark, A. Brock, K. Simonyan, R. Hadsell, N. Robinson, E. Clancy, A. Arribas, S. Mohamed, *Nature* **2021**, *597*(7878), 672.
[12] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*(7873), 583.
[13] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, Cambridge, MA **2016**. http://www.deeplearningbook.org.
[14] S. Jo, W. Sohng, H. Lee, H. Chung, *Food Chem.* **2020**, *331*, 127332.
[15] M. W. N. Jinadasa, A. C. Kahawalage, M. Halstensen, N. Skeie, K. Jens, in *Recent Developments in Atomic Force Microscopy and Raman Spectroscopy for Materials Characterization*, (Eds: C. S. Pathak, & S. Kumar). IntechOpen, **2021**. https://doi.org/10.5772/intechopen.99770
[16] X.-G. Fan, Y. Zeng, Y.-L. Zhi, T. Nie, Y.-J. Xu, X. Wang, *J. Raman Spectrosc.* **2021**, *52*(4), 890.
[17] E. A. Magnussen, J. H. Solheim, U. Blazhko, V. Tafintseva, K. Tøndel, K. H. Liland, S. Dzurendova, V. Shapaval, C. Sandt, F. Borondics, et al., *J. Biophotonics* **2020**, *13*(12), e202000204.
[18] A. P. Raulf, J. Butke, L. Menzen, C. Küpper, F. Großerueschkamp, K. Gerwert, A. Mosig, *J. Biophotonics* **2021**, *14*(3), e202000385.
[19] A. P. Raulf, J. Butke, C. Küpper, F. Großerueschkamp, K. Gerwert, A. Mosig, *Bioinformatics* **2020**, *36*(1), 287.
[20] T. Mete, G. Ozkan, H. Hapoglu, M. Alpbaz, *Comput. Appl. Eng. Educ.* **2012**, *20*(4), 619.
[21] E. A. del Rio-Chanona, J. L. Wagner, H. Ali, F. Fiorelli, D. Zhang, K. Hellgardt, *AIChE J.* **2019**, *65*(3), 915.
[22] P. Patnaik, *Bioprocess Biosyst. Eng.* **2001**, *24*(3), 151.
[23] Y. Ma, W. Zhu, M. G. Benton, J. Romagnoli, *J. Process Control* **2019**, *75*, 40.
[24] Y. Ma, D. A. Noreña-Caro, A. J. Adams, T. B. Brentzel, J. A. Romagnoli, M. G. Benton, *Comput. Chem. Eng.* **2020**, *142*, 107016.
[25] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, *Nature* **2015**, *518*(7540), 529.

[26] V. Vukotić, S.-L. Pintea, C. Raymond, G. Gravier, J. C. Van Gemert, *International conference on image analysis and processing*, Springer, Cham **2017**, p. 140. https://doi.org/10.1007/978-3-319-68560-1_13
[27] P. C. Lopez, I. A. Udugama, S. T. Thomsen, C. Roslander, H. Junicke, M. Mauricio-Iglesias, K. V. Gernaey, *Biofuels Bioprod. Biorefin.* **2020**, *14*(5), 1046.
[28] J. Long, E. Shelhamer, T. Darrell, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2015**, 3431. https://doi.org/10.1109/CVPR.2015.7298965
[29] K. He, X. Zhang, S. Ren, J. Sun, *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*(9), 1904.
[30] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, *J. Mach. Learn Res.* **2014**, *15*(56), 1929.
[31] A. Savitzky, M. J. Golay, *Anal. Chem.* **1964**, *36*(8), 1627.
[32] H. Martens, E. Stark, *J. Pharm. Biomed. Anal.* **1991**, *9*(8), 625.
[33] I. Måge, U. Böcker, S. G. Wubshet, D. Lindberg, N. K. Afseth, *LWT* **2021**, *152*, 112339.
[34] X. Glorot, Y. Bengio, JMLR Workshop and Conference Proceedings. in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, PMLR, Chia, Sardinia **2010**, p. 249.

How to cite this article: M. Kuchta, S. G. Wubshet, N. K. Afseth, K.-A. Mardal, K. H. Liland, *J. Biophotonics* **2022**, *15*(9), e202200097. https://doi.org/10.1002/jbio.202200097

APPENDIX A

In this section we collect additional results supporting the findings reported in the main article. Figure A1 shows convergence of the optimization process for training the different single- and many-input neural networks. Tables A.1 and A.2 analyse the prediction errors of the single-input network $S_{t_0=0.5}$ for fixed inference time and evolution of the error in time. Figure A2 demonstrates the property of the obtained models in predicting the mean of the measured EPH reactions (for the given substrate–enzyme pair) rather than fitting to individual spectra. Finally, Figure A3 illustrates the consistency property of the many-input networks; as desired in the application, the different input pairs are mapped to practically identical outputs.

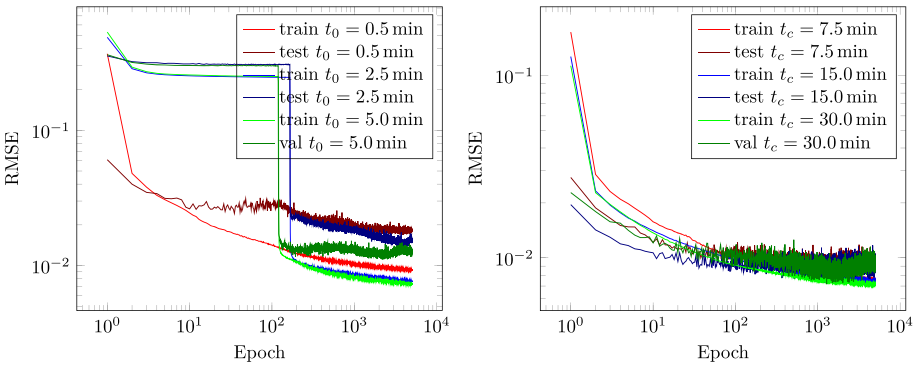


FIGURE A1 Convergence of ADAM iterations in training (left) single input networks S_{t_0} with $t_0 = 0.5$, 2.5, 5.0 min and (right) many input networks M_{t_c} with $t_c = 7.5$, 15, 30 min. Training/validations sets differ between the networks

TABLE A.1 Prediction errors for $S_{0.5}$ network at fixed inference time $t_i = 10$ min. RMSE is reported based on all measured reactions for given substrate–enzyme pairs

Subenz	A	Pa	Pr	C
CB	0.17	0.40	0.27	—
CM	0.14	0.09	0.16	—
CR	0.11	0.11	0.11	—
CS	0.12	0.14	0.16	—
TR	0.21	—	—	0.11
Ma	0.13	0.16	—	—
SB	0.08	—	—	—
SH	0.08	—	—	—
SS	0.06	—	—	—
TC	0.17	—	—	0.18
HC	—	0.21	0.16	—

TABLE A.2 Time evolution of the prediction error of the $S_{0.5}$ network. RMSE is reported based on all measured reactions for given substrate–enzyme pairs

Time (min)	CM-A	SS-A	Time (min)	CM-A	SS-A
2.5	0.19	0.08	20	0.10	0.05
5	—	0.06	30	0.09	0.05
7.5	0.15	0.07	40	0.09	0.06
10	0.14	0.06	50	0.10	0.06
15	0.12	0.06	60	0.10	0.06

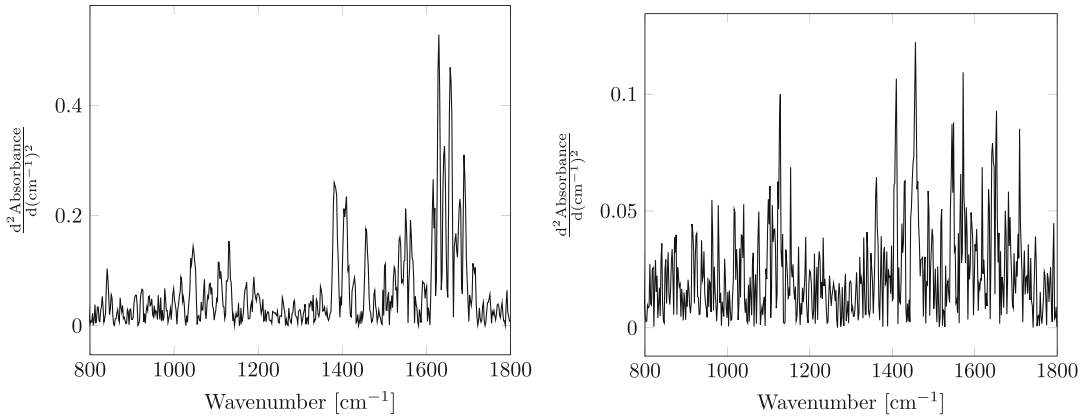


FIGURE A2 Prediction errors of single input network with time $t_0 = 0.5$ min. The inference time and substrate–enzyme pair are fixed at $t = 10$ min and CM-A. (Left) Prediction using one of seven available inputs. The error is noticeably larger in the band between 1600 cm^{-1} and 1800 cm^{-1} (Right) Mean prediction obtained as average of seven inputs is compared with the mean of the targets. The error is rather delocalized

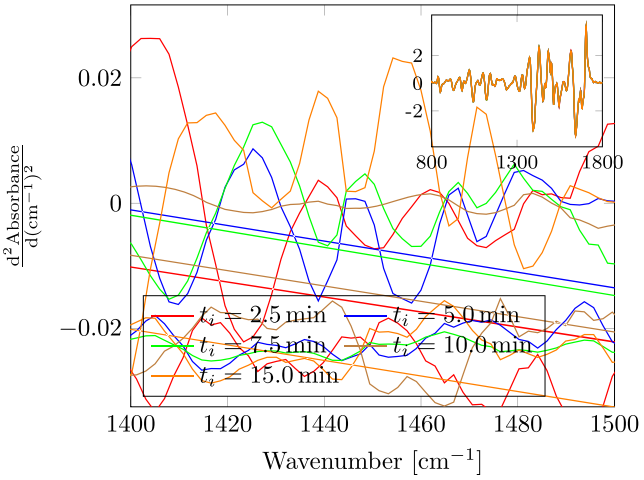


FIGURE A3 Non-uniqueness of many-input network predictions. (Left) Predicted spectra of a single CM-A reaction at time $t = 40$ min by network M_{t_c}, $t_c = 15$ min using input spectra at times $t_i = 2.5, 5.0, 7.5, 10, 15$ min. Predictions are consistent with respect to t_i as can be seen through small variations in difference between the mean prediction (taken over all t_i) and the individual spectra. In the inset figure showing the entire wavenumber range the predicted spectra are practically identical. (Right) RMSE for all seven CM-A reactions in the FTIR dataset and the different input times t_i