Image sequence based prediction of the temporal evolution of fresh concrete properties under realistic conditions

Max Meyer[1], Amadeus Langer[1], Max Mehltretter[1], Dries Beyer[2], Max Coenen[3], Bastian Strybny[3], Tobias Schack[4], Michael Haist[4], Christian Heipke[1]

[1]Institute of Photogrammetry and GeoInformation, Leibniz University Hannover, Germany
(meyer, langer, mehltretter, heipke)@ipi.uni-hannover.de
[2]Feist Construct GmbH, Bad Pyrmont, Germany
beyer@feist-construct-gmbh.de
[3]Institute of Building Materials Science, Leibniz University Hannover, Germany
(m.coenen, b.strybny)@baustoff.uni-hannover.de
[4]Institute of Construction Materials, University of Stuttgart, Germany
(tobias.schack, michael.haist)@iwb.uni-stuttgart.de

Keywords: Image sequences, Deep learning, Fresh concrete properties, Building materials, Time dependency.

Abstract

Advancing the level of digitalization and automation in concrete manufacturing can substantially contribute to lowering CO_2 emissions associated with the concrete production. This work introduces a new methodology for predicting the time-dependent properties of fresh concrete during mixing. For the prediction, a deep learning network is created which uses stereoscopic image sequences of the flowing material together with tabular data as input. Besides mix design parameters and process state data, like energy consumption, moisture and fresh concrete temperature, temporal information is included in the tabular data. The temporal information represents the time interval between image acquisition and the time for which the properties should be predicted. During training, this interval corresponds to the difference between the image acquisition and the time at which reference measurements are taken, allowing the network to implicitly learn the temporal evolution of the material properties, namely the slump flow diameter, yield stress, and plastic viscosity. Incorporating time-dependent prediction enables the forecasting of property changes throughout the mixing process, offering a valuable tool for real-time process control. This capability allows timely adjustments whenever deviations from the desired material behavior are detected. The experimental investigations presented in this paper demonstrate the feasibility of this method under realistic conditions.

1. Introduction

The construction industry is one of the largest producers of CO_2 emissions, with the production of cement alone being responsible for approximately 7–8 % of the global anthropogenic CO_2 emissions (Haist et al., 2022a). To reduce the CO_2 emissions of concrete, research seeks to lower the cement content by incorporating supplementary materials. As a result, modern concrete often contains multiple constituents in addition to the original constituents water, cement and aggregate. This development leads to more complex mix designs, which can compromise the robustness of the concrete (González-Taboada et al., 2018) and makes the control of fresh concrete properties more challenging.

The existing quality assurance methods are insufficient to cope with these new challenges: Quality assurance is usually performed manually and on a limited number of samples, which both increase the measurement uncertainty. Since the properties of fresh concrete at the time of placement influence the quality of the resulting construction component, quality assurance methods are carried out right before placement, but adjustments are no longer feasible at this point. The property deviations from the target may result in the rejection of entire batches, causing material waste and project delays, which hinders the adoption of CO_2-reduced concretes. Advancing sustainability in the concrete production requires the digitization and automation of production processes (Haist et al., 2022b), and, in particular, the sensor-based monitoring of material behavior.

As the construction area remains one of the least digitized industries (Green, 2016), the goal of the ReCyCONtrol[1] project is it to advance digitization and automation in the concrete sector. This paper focuses on the part of the project, which aims to predict the properties of fresh concrete. Since the mixing stage provides the last opportunity in the production chain for significantly adjusting the fresh concrete properties, the prediction of the properties is done during mixing. Furthermore, as hydration causes concrete to become considerably stiffer between mixing and placement (Liu et al., 2024), it is essential to capture the time-dependent development of the concrete properties. Our vision is to enable prediction of the properties at the time of placement based on observations during mixing. If deviations from the desired properties are detected, corrective actions, such as adding superplasticizers or accelerators, can still be taken.

In this work, the fresh concrete properties to be predicted are the slump flow diameter, describing consistency, and two rheological values, yield stress (minimum stress required for flow) and plastic viscosity (resistance at higher shear rates), as concrete is generally modeled as a Bingham fluid (Yahia et al., 2016). Notably, slump flow diameter and yield stress are partially correlated (Wallevik, 2006). So far, the slump flow

[1] https://www.recycontrol.uni-hannover.de/en/

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-427-2026 | © Author(s) 2026. CC BY 4.0 License.

diameter is typically measured through the slump test (EN 12350-5, 2019), while rheological parameters are derived from a flow curve which is determined by rheometer measurements. The goal of this work is to replace these manual measures with an automated process. For this purpose, an optical sensing system combined with a photogrammetric processing pipeline has been developed. A stereo camera setup monitors the flow behavior of the fresh concrete during mixing, and the recorded image sequences serve as input to a deep learning model that predicts the fresh concrete properties. The image sequences are complemented by tabular data which consist of temporal information, mix design information and process state data. Temporal information enables the network to model property evolution over time, allowing predictions for arbitrary future points, including the time of placement. Furthermore, the fusion of image data with mix design parameters and process state data such as energy consumption, moisture, and fresh concrete temperature are investigated. The novelties of the presented work are the following ones:

- For training the network, a new loss function is developed which is calculated depending on the different value ranges of the three outputs, thus achieving a balanced influence of the three outputs on the loss.
- The time-dependent prediction based on images and tabular data, including temporal information, is demonstrated using a data set recorded with an industrial mixer under realistic conditions.

2. Related work

In recent years, interest in automating quality assurance in concrete production has steadily increased. Numerous machine and deep learning approaches have been proposed, to be applied across different stages of the production chain. The present work concentrates on methods for automating the quality control of fresh concrete.

Traditional quality assurance methods such as the slump test (EN 12350-5, 2019) and rheometer measurements are labor-intensive and prone to uncertainties. To address these issues, Tuan et al. (2021) propose an automated slump test using a stereo camera to record the spreading process, replacing manual measurements with image analysis. Similarly, Yoon et al. (2023) investigate cement paste as a model material, capturing 3D point clouds with a depth camera and deriving geometrical features used in a deep learning model to estimate rheological properties and behaviors such as yield stress, viscosity, superplasticizer adsorption, and bleeding. Extending this line of research, Schack et al. (2023a,b,c) analyze surface images of spread concrete, enabling not only slump flow measurement but also insights into mixture composition, for example, aggregate content. Coenen et al. (2024) introduce an approach based on channel flow at truck outlets: Derived from optical flow images, spatio-temporal flow fields represent the flow behaviour of the concrete over time and serve as input for a Convolutional Neural Network (CNN) for predicting consistency and rheological parameters. However, as conventional testing, these methods are applied after production. Consequently, any detected deviations still require discarding and replacing the whole batch, resulting in inefficiencies and waste.

To overcome this limitation, the prediction of fresh concrete properties must take place before or during mixing. Two main strategies have emerged: Prediction based on mix design data and prediction based on live data recorded during the mixing process.

The first approach relies on information about the raw materials of the concrete. Classical models exist, particularly for plastic viscosity (Chidiac and Mahmoodzadeh, 2009), but show high variability between different models. More recent work employs a wide range of machine learning and deep learning methods, including approaches like least squares support vector machines (LSSVM) with particle swarm optimization (Nguyen et al., 2020), random forests and XGBoost (Inqiad et al., 2025), multilayer perceptrons (MLP) (Navarrete et al., 2023), gradient boosting (Kumar et al., 2024), and deep neural network (DNN) regressors (Babaei et al., 2025). Nguyen et al. (2020), Navarrete et al. (2023) and Inqiad et al. (2025) also incorporate temporal information to achieve a time-dependent prediction. Nonetheless, the accuracy of theses approaches strongly depends on the reliability of the mix design data, which is often compromised by raw material variability, although improvements have recently been achieved in this field (Coenen et al., 2023; Langley et al., 2025).

The second approach can be divided into two subgroups: One is based on sensors that measure parameters of the mixing process, and the other one is based on images. The first subgroup of the approach based on live data exploits measurements on the mixer, like energy consumption and hydraulic pressure, which accounts for uncertainties from raw material variability. Several studies demonstrate correlations between these measurements and fresh concrete properties (Khayat et al., 2014; Wallevik and Wallevik, 2020; Malekipour and Moodi, 2021; Rong et al., 2022). Catti et al. (2025) use a bidirectional Long Short-Term Memory (Bi-LSTM) network that predicts the properties of fresh concrete based on live data from the mixing drum of a truck mixer. While such methods are already used in practice, to the authors’ knowledge no method yet predicts the time-dependent development of properties from such data.

The second subgroup uses image data obtained during mixing. Like live data from the mixer, these images provide live insights independent of mix design uncertainties. Early studies apply classical image processing techniques to predict flow-related parameters (Li and An, 2014), while later work combine CNNs with LSTMs to predict the slump or viscosity values from image sequences (Ding and An, 2018; Yang et al., 2021; Guo et al., 2022). Other contributions use semantic segmentation and a residual network for slump classification (Gao and Yan, 2023), a CNN with images from the hopper discharge (Idrees et al., 2024), or a CNN coupled with stereo image pairs of a surrogate material for flow curve prediction (Ponick et al., 2022). 3D texture features from stereo vision have also been used with XGBoost for slump class prediction (Ojala and Punkki, 2024). In (Meyer et al., 2024), a method is presented which uses a CNN based on stereoscopic images, temporal information and mix design information to generate a time-dependent prediction of fresh concrete properties. However, the functionality of the method is only demonstrated on the basis of a substitute device which simulates the mixing process of an industrial mixer.

3. Deep learning for fresh concrete property prediction

3.1 Input data

3.1.1 Image data A calibrated and synchronized grayscale stereo camera system captures image pairs of the concrete surface at regular intervals. From each pair, an orthophoto O and a digital elevation model (DEM) D are generated using standard photogrammetric procedures. The orthophoto O provides surface texture, while the DEM D represents the geometric surface structure. For the generation of O and D, the mixing paddle is masked, when it appears in the image pair, as it does not contain any information about the concrete. Furthermore, it may happen that no valid values can be calculated in some areas of O and D, as only observations from one camera are available in these areas due to occlusions. The pixel values of these invalid areas (paddle mask and occlusions) are set to zero.

To capture the flow behavior of the concrete, the optical flow $OF_{i,i+1}$ is calculated between consecutive orthophotos O_i and O_{i+1}, where i represents the time step, by using the algorithm of Farnebäck (2003), implemented in OpenCV (Bradski, 2000). To calculate $OF_{i,i+1}$, the invalid areas in O_i and O_{i+1} are set to the mean value of the valid pixels of O_i and O_{i+1}, respectively. For each pixel in O_i the displacement in x and y direction is calculated. The calculated displacements for non valid pixels in O_i are then set to zero in $OF_{i,i+1}$. Each sample thus forms a four-channel input $[O_i, D_i, OF_{i,i+1}]$, with two channels of OF representing pixel displacements in the x and y directions. D and OF are standardized to zero mean and unit standard deviation with the mean and standard deviation of the training data set for the respective image channel in order to preserve the characteristic differences between the images. Each O is standardized individually to account for variable lighting. The non valid pixels remain zero in all channels.

3.1.2 Tabular data The image-based input is complemented by tabular input which consists of temporal information, mix design information and process state data. The temporal information, which is decisive for the time-dependent prediction, is represented by Δ_t. Δ_t denotes the time difference between the image acquisition and the point in time for which the fresh concrete properties should be predicted. Δ_t consists of two time intervals: The interval $\Delta_{t,mix}$ between image acquisition and mixing end, and the interval $\Delta_{t,ref}$ between mixing end and the point in time at which the fresh concrete properties are to be predicted. This separation is important because the structural build-up, as the result of the hydration process, is broken up during mixing, causing the concrete to stiffen more slowly than it would after the mixing process has ended. During training, the point in time at which the fresh concrete properties are to be predicted, corresponds to the time of a reference measurement.

Furthermore, the mix design m is used as input. The mix design contains information about cement type, water content, cement content, stone powder content, aggregate content by size (0–2 mm, 2–8 mm, and 8–16 mm) and superplasticizer content (the only additive used in this study). The cement type (three categories: CEM II/A-LL 42,5; CEM I 42.5 R; CEM I 42.5 (PUR4)) is included by using One Hot Encoding.

In addition, process state data is included in the input. The energy consumption of the mixer and the moisture of the concrete are recorded continuously during mixing. For each concrete, the mean and standard deviation of both quantities are computed during the image recording resulting in a total of four parameters. The temperature of the fresh concrete is measured after each mixing process. These five process state parameters are represented by p.

Analogous to the procedure for D and OF, the tabular

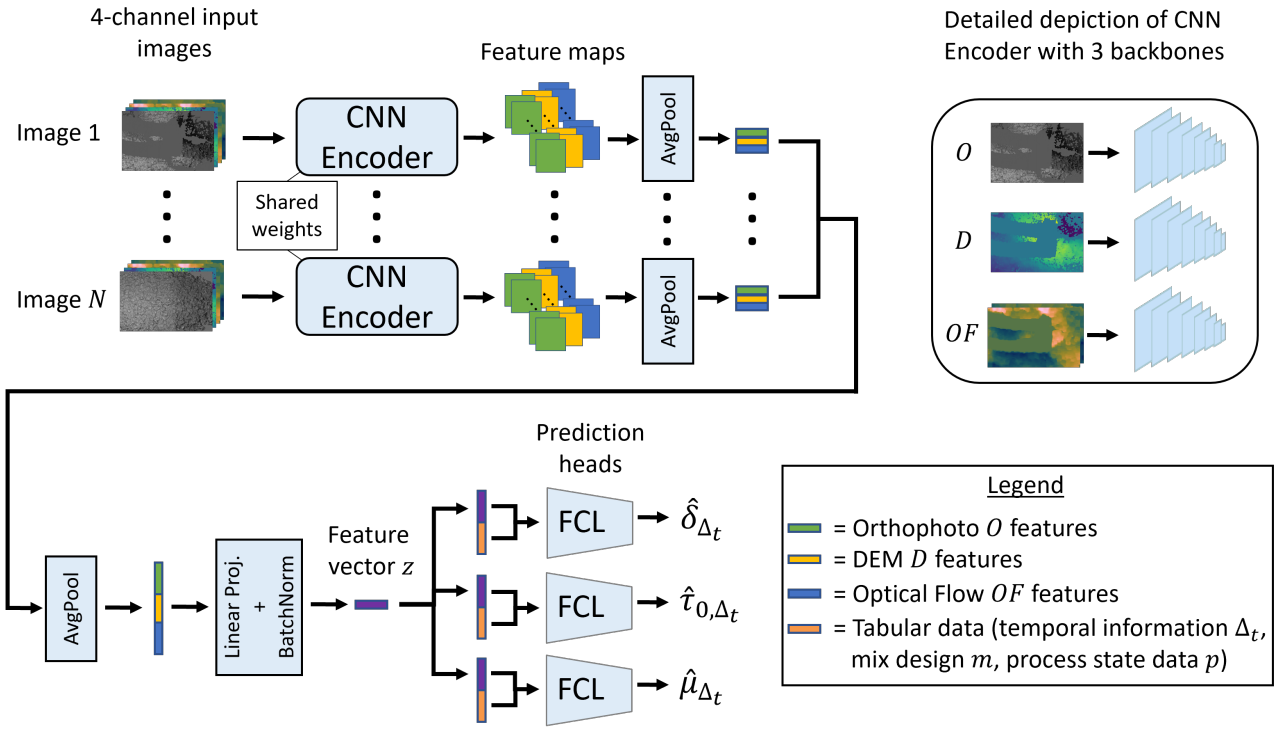


Figure 1. Architecture of the developed network.

inputs (except of the cement type) are standardised to a mean value of zero and a standard deviation of one using the mean value and standard deviation from the training data set of each input parameter.

3.2 Network architecture

The architecture of our network is depicted in Fig. 1 and is chosen based on preliminary tests. The network consists of a CNN Encoder, with three CNN backbones (one for each modality of the image-based input (O, D, OF)), and three prediction heads. Each head predicts one of the three time-dependent output parameters slump flow diameter δ_{Δ_t}, yield stress τ_{0,Δ_t} and plastic viscosity μ_{Δ_t}. For a prediction, an image sequence is used, which consists of N images representing one full mixer rotation. Every image of the image sequence is fed to the CNN Encoder. Each CNN backbone extracts eight feature maps, which are then each reduced to one value by using average pooling and then concatenated to a vector of dimension 8. This leads to three vectors for each image (one for each modality), which results in $3N$ vectors (N vectors per modality) for an input image sequence consisting of N images. For each modality average pooling is used to reduce the N vectors to one. The resulting three vectors, one for each modality, are concatenated and passed through a linear projection and a batch normalization to generate a feature vector z. The linear projection ensures a more compact representation, while the batch normalization ensures that z has a similar value range as the standardised tabular features. The feature vector z is then concatenated with the tabular inputs (Δ_t, m, p) and is used as input for the prediction heads. As the tabular data stays constant for one concrete, unlike the image data, a dropout (p=0.5) is used for the inputs m and p to prevent overfitting. There is no dropout for Δ_t as it is decisive for the time-dependent prediction. The three prediction heads consist each out of three fully connected layers (FCL). The first FC layer expands the input dimension from 27 to 40 neurons and is followed by a batch normalization and a ReLU activation. The second FC layer reduces the dimensionality from 40 to 20 neurons, followed again by a ReLU activation. Finally, the third FC layer, namely the output layer, maps the 20 neurons to a single output neuron.

The CNN backbones in the CNN Encoder consist of eight gated convolutional layers (Yu et al., 2019). Gated Convolution extends the conventional operation by introducing a learnable gating mechanism that modulates the feature activation at each spatial position and is therefore particularly useful for incomplete data, such as the partially masked images used here. Each layer has for the feature and the gated convolution a kernel size of 5×5 and a stride of 2, except for the first layer, where a stride of 1 is used. Each layer is followed by a group normalization and a ReLU activation function.

The proposed network architecture yielded the best results in preliminary studies and is relatively lightweight compared to those used in computer vision tasks. In general, preliminary studies have shown that the use of complex operations with a high number of weights led to overfitting and thus poorer results. This was to be expected, as the data set used here is significantly smaller than in other computer vision applications and the problem is less complex than typical classification and segmentation problems. The relatively small number of weights therefore serves as regularisation.

The CNN Encoder is pretrained using the masked autoencoder (MAE) approach presented in (He et al., 2022), which leads to a more robust generalisation with incomplete data. In this approach, the encoder processes images with masked areas and extracts features from it, while a lightweight decoder reconstructs the original image from these extracted features. The decoder consists of three blocks (one for each modality). The blocks receive the extracted feature maps of the respective CNN backbone as input. Each block consists of three layers and an output layer. One layer consists of an upsampling block followed by a convolution with a 3×3 kernel and a stride of 1, group normalisation and a ReLU activation function. The output layer consists of a convolution with a 3×3 kernel and a stride of 1.

3.3 Training

The training process starts with the MAE approach, which pretrains the CNN Encoder. During training and validation with MAE, pixels are masked with a probability of 50 %, which has led to better results than patch masking in preliminary tests. For the MAE the mean absolute error (L1) is used as loss, which determines the average distance between the reconstructed pixel values and the original pixel values. Only pixels that are randomly masked and have valid values are taken into account in the loss. After pretraining with MAE, training for the time-dependent prediction of fresh concrete properties is carried out. The model weights ω are optimized using a modified Mean Absolute Percentage Error (MAPE) loss (see eq. 1). Instead of dividing by the reference value y_m^k, the difference between prediction and reference value is normalized by the value range of training data $y_{max}^k - y_{min}^k$ for each target variable. This normalization ensures balanced weighting among outputs with different value ranges, which is important in our case, because slump flow diameter, yield stress and plastic viscosity differ significantly in their value ranges. For a batch of M samples and K target variables the loss is defined as:

$$L(\omega) = \frac{1}{M \cdot K} \sum_{k=1}^{K} \sum_{m=1}^{M} \left| \frac{y_m^k - \hat{y}_m^k}{y_{max}^k - y_{min}^k} \right|, \tag{1}$$

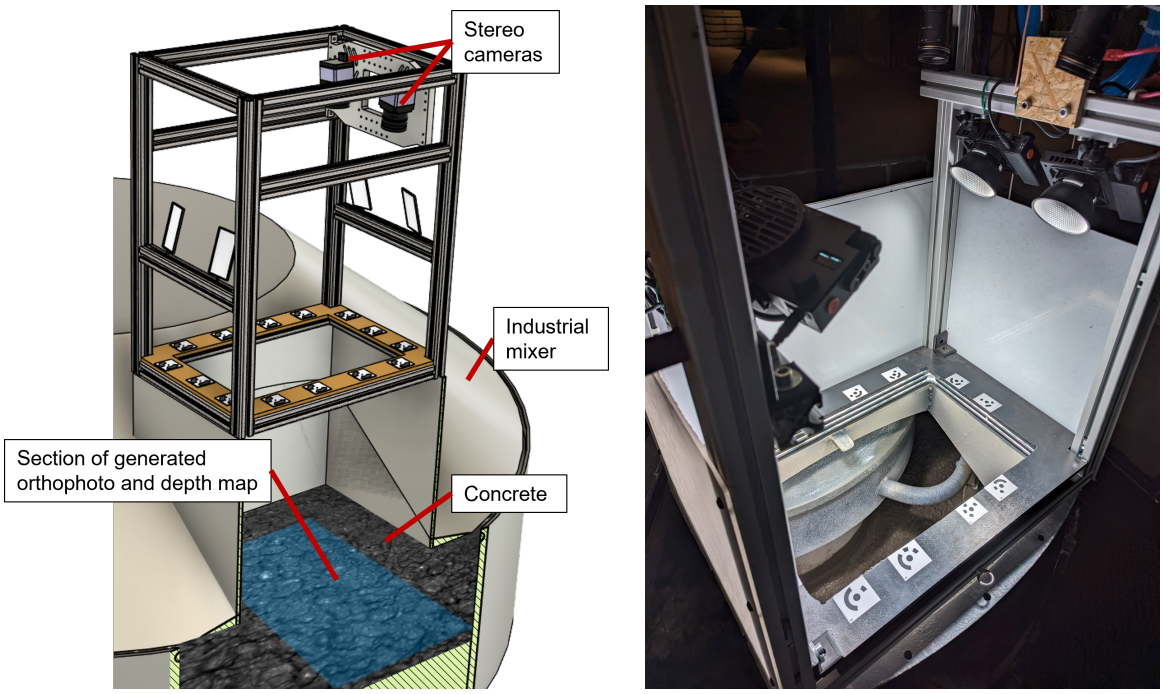
where y represents the reference and $\hat{y}$ the prediction. To limit overfitting, weight decay regularization is applied to the model weights in the training (including pretraining):

$$L_{\text{total}}(\omega) = L(\omega) + \lambda \sum \omega^2, \tag{2}$$

where λ denotes the weight decay coefficient. The output layer of the prediction heads are excluded from regularization to not restrict or bias the prediction towards smaller values.

3.4 Prediction based on tabular data only

For comparison, three separate multilayer perceptrons (MLPs), one for each output parameter, are trained using only Δ_t, m an p as input. The MLP comprises three layer: The first layer expands the number of neurons from 17 to 40 and is followed by a leaky ReLU activation (slope 0.2) and a dropout (p=0.5). The second layer decreases the number of neurons from 40 to 20 and is followed by a ReLU. The third layer, namely the output layer, maps the 20 neurons to the output neuron. This structure showed the best performance in preliminary tests. The training procedure follows Section 3.3 without the MAE pretraining.



(a) Schematic overview. (b) Real setup.

Figure 2. Experimental setup for generating the data set.

4. Data generation

4.1 Data acquisition

For training and evaluation, a dataset is created using an industrial mixer. A *PM 500* planetary mixer from *Pemat*[2] with a maximum capacity of 500 liters is employed, controlled via software from *Bikotronic*[3]. The mixer lid features a rectangular opening through which a stereo camera system records the mixing process. The fresh concrete surface is captured by two geometrically calibrated and synchronized *pco.edge 3.1 USB* cameras from *Excelitas*[4], recording at 50 frames per second (fps) with a resolution of 2048×1536 px. Due to memory limitations, the images are captured in 4 consequtive runs (for some concretes, the number of runs is lower due to technical difficulties). Each run has a duration of 50 seconds which results in 10,000 image pairs for each concrete. The stereo system is mounted on a rack at the opening, with markers in the field of view to determine the relative position of the camera position and viewing direction to the mixer opening. This allows to consider possible movements of the camera with respect to the mixer opening. Furthermore, spotlights on the rack ensure adequate illumination, which is varied during the tests to approximate realistic conditions. Figure 2 shows the schematic setup and the real setup during the experiments.

In parallel to the image recordings, the control unit provides the mixer’s energy consumption at a frequency of 10 Hz, while a *Type 6* microwave sensor from *Bikotronic*[3], mounted at the mixer bottom, measures moisture at the same frequency. In order to use the energy consumption and moisture as input, the mean and standard deviation are calculated for each image recording run. The duration of 50 seconds per run is a realistic duration of the wet mixing phase in practice.

In total, 34 unique concrete mixtures (120 liter each) are tested, differing in water–cement ratio, paste content, grading curve, cement type, lime–sand content, and superplasticizer dosage. The rotation speed of the mixer is kept constant during the experiments. One whole rotation takes two seconds. The concrete is mixed by six paddles, which differ in form and position. In addition the paddles are arranged in a way that there is a larger gap between two paddles. Consequently, the flow behavior of the concrete in the images depends strongly on the paddle type or whether there is a paddle at all. Therefore the recorded images are sorted by the different paddle positions which are visible in the images with a simple automatic procedure based on cross correlation. This leads to seven categories (six paddles and one paddle gap) in which the images are assigned to per rotation. For each category are about 3-4 images for each rotation recorded. Only images which can be clearly assigned to one category are used. The images which lie between two categories are not used. This categorisation ensures that the input for the network is more structured and that differences in flow behaviour occur only due to the properties of the fresh concrete and not due to the paddle position.

4.2 Generation of orthophotos and DEMs

Orthophotos O and Digital elevation models (DEMs) D are generated with the image-matching algorithm of *Agisoft Metashape*[5]. For masking the paddles the SAM segmentation network ViT-H (Kirillov et al., 2023) is used. To reduce computational load, O and D (1866×1066 px) are cropped to 760×492 px. The cropped region is the region which lies directly under the cameras, as this region contains the least occlusions.

[2] https://www.pemat.de/en/
[3] https://www.bikotronic.de/en/
[4] https://www.excelitas.com/
[5] https://www.agisoft.com/

https://doi.org/10.5194/isprs-annals-XI-1-2026-427-2026

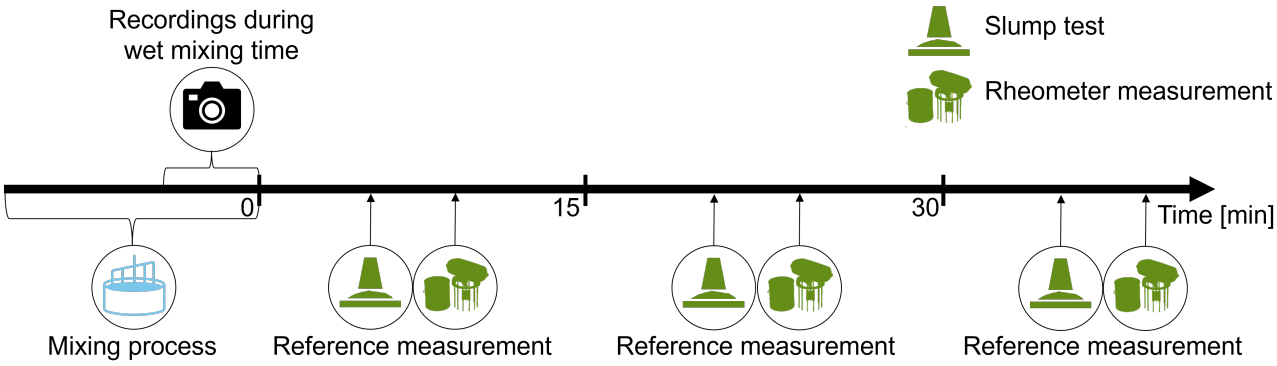


Figure 3. Timeline of the experimental workflow.

4.3 Reference Values

After mixing, reference measurements are taken to determine the slump flow diameter δ, yield stress τ_0, and plastic viscosity μ. The slump flow diameter is measured manually with the standard slump test EN 12350-5 (2019), while τ_0 and μ are obtained with an *eBT-V* rheometer from *Schleibinger*[6] following Feys et al. (2013). To replicate continuous truck mixing, which reduces but does not prevent structural build-up caused by the chemical hydration process, the concrete samples are remixed manually before each reference measurement.

4.4 Timeline

The timeline of the workflow is illustrated in Figure 3. The mixing starts with cement and aggregate. Later in the mixing process, water and then admixtures are added. The stereo recordings are started 20 s after the addition of the admixtures in order to ensure homogeneity of the fresh concrete. The following mixing time extends up to approx. 5 minutes (in two cases up to 10 minutes) to capture more image data. The time zero in the timeline is set at mixing end.

After mixing, the concrete is sampled for reference measurements. The first slump test is carried out about 5 minutes later, followed immediately by rheometer measurements and a manual temperature check. The slump test and rheometer measurement are repeated after approx. 20 minutes and 35 minutes, giving three slump tests and three rheometer measurements per concrete. The temporal information $\Delta_{t,mix}$ for each input sequence is defined as the time interval between the recording of the input sequence and the mixing end. $\Delta_{t,ref}$ is defined as the time interval between mixing end and the reference measurement, separately for slump test and rheometer measurement values. Therefore $\Delta_{t,ref}$ contains two values.

5. Experiments and Results

The experiments aim to examine the effect of using image based data and tabular data on the prediction accuracy. Image based variants of our method are trained with and without mix design m and process state data p and are compared to the MLP based on tabular data described in Sec. 3.4. The results and reasons for poor predictions in specific cases are discussed, and the ability of our method to predict temporal changes in fresh concrete properties is demonstrated.

[6] http://www.schleibinger.com/cmsimple/en/

5.1 Training Configuration

The input sequence consists of seven images which originate from the same rotation. The generation of the input sequences is done by the following procedure: The first image is selected from each paddle position, and these selected images then form the input sequence. The procedure continues analogue for the second, third and fourth images in the paddle positions. If a paddle position contains fewer images than the others, the image that is closest to the missing image is reused to complete the input sequence. This approach ensures that the distances between the images in an input sequence stay approximately constant. The image dataset contains 74,468 images from 2760 mixer rotations which results in 11,040 input sequences. Note, that, as mentioned before, only images which are assigned to one paddle position category are included in the dataset.

For training, a five-fold cross-validation is used, consisting of five cross-validations groups, each containing 6-7 concretes. The data splitting is carried out as follows: First, the concretes are sorted by initial slump flow diameter δ_1 and are assigned to high, medium, and low δ_1-categories. Each cross-validation group randomly receives 2–3 concretes per δ_1-category, ensuring a balanced distribution. Five trainings are conducted, in which each of the cross-validation groups is used once as a test set. Furthermore, each cross-validation group is randomly selected once as validation set. To do so, 2-3 concretes are randomly removed from the selected cross-validation group so that the validation set always consists of 4 concretes, of which at least one concrete from each δ_1-category (high, medium, low) comes. The removed concretes and the concretes from the remaining cross-validation groups form the training set. To ensure that the network sees a wide range during training, several conditions are imposed for data splitting: The two concretes with the highest and second highest value for each property - initial slump flow diameter, initial yield stress, initial plastic viscosity, and superplasticiser content - must not be assigned to the same cross-validation group. Furthermore, none of these concretes may be assigned to the validation set. These conditions are formulated analogously for the concretes with the lowest and second lowest values for the mentioned properties. If these conditions are not met, the data splitting process is repeated until they are met. By fulfilling these conditions, the network sees a much wider range of values for these important properties during training.

As the reference values originate from separate measurements (slump test, rheometer measurement) each input sample combines one slump flow diameter δ_{Δ_t} with one pair of

rheological values τ_{0,Δ_t}, and μ_{Δ_t}. For each concrete, three slump tests and three rheometer measurements yield nine reference combinations. For the training phase, the input sequences are randomly assigned to one of these reference combinations in each training epoch. For the validation, all input sequences are assigned evenly in a deterministic manner to the reference combinations. For testing, a different approach is used: To enhance the robustness of the predicted target parameters, the final prediction is obtained by averaging the predictions from 100 distinct input sequences, per concrete. To do so, prediction groups of 100 subsequent input sequences are assigned to a reference combination in a deterministic manner until every reference combination is used once. If there are more reference combinations than prediction groups, the already assigned prediction groups are reused. If there are more prediction groups than reference combinations, the already assigned reference combinations are reused. Recording 100 input sequences takes 50 s, which aligns with the time interval in which the mean and standard deviation from the energy consumption and moisture are computed. $\Delta_{t,mix}$ is calculated on the basis of the first image of the input sequence and $\Delta_{t,ref}$ is the time intervals between the end of the mixing process and the corresponding reference measurements.

The mix design m and the temperature contained in the process state data p are constant for each concrete, while energy consumption and moisture vary. When image-based inputs and p are used together, the mean and standard deviation of the run, in which the images were recorded, are used for p.

For technical reasons, the data set has some peculiarities that are taken into account during the experiments: Two concretes produced implausible rheometer data due to segregation, while one produced implausible rheometer data because its consistency was too stiff. Consequently, only 31 concretes are available for training and testing the prediction of τ_{0,Δ_t} and μ_{Δ_t}. The corresponding reference values and predictions of these three concretes are excluded from the loss calculation, increasing the weight of the slump flow diameter for these concretes (for the concretes affected by segregation only one reference measurement of the slump flow diameter is available). Furthermore, four concretes were recorded at 49 fps, their optical flow images are therefore scaled by a factor 0.98 to harmonise them with the images from the remaining concretes.

For the pretraining with the MAE approach, we use the AdamW optimizer with a learning rate of $5 \cdot 10^{-3}$, a batch size of 16 in combination with gradient accumulation to effectively obtain a batch size of 32, and He initialization. Regularization is applied by weight decay ($\lambda = 10^{-4}$). Data augmentation is used during training to prevent the network from overfitting. For the augmentation of the brightness, for each image and each channel an offset is determined with an uniform distribution in an interval of $[-0.1, 0.1]$. Analogously, a factor for the contrast is determined with an uniform distribution in an interval $[0.95, 1.05]$. To reduce computational load, the images are randomly cropped to 380 $\times$ 246 px during training, whereas for validation and testing a centered crop of the same size is applied. Training runs for 5 epochs with validation after every epoch. The following main training uses the same configurations with some extension: The training runs for 40 epochs, the learning rate halves every 5 epochs, early stopping is carried out after seven validations without improvement and the images are not cropped. The pretrained layers of the CNN Encoder use only 10 % of the learning rate. Furthermore, tabular data is augmented by a random offset, which is determined with an uniform distribution in an interval of $[-0.1, 0.1]$.

The training of the MLPs is analogous to the training of the CNN with a few adjustments: As the three target values are predicted separately, the references are not combined. In addition, every input sample is combined with every reference, since the MLPs only process tabular inputs (Δ_t, m, p). Furthermore, there are multiple means and standard deviations of the energy consumption and moisture per concrete. As a consequence there are multiple input samples per concrete where only the energy consumption, moisture, $\Delta_{t,mix}$ and $\Delta_{t,ref}$ vary and the remaining parameters stay constant. $\Delta_{t,mix}$ is calculated for each input sample as the time interval between the middle of the image recording run, for which energy consumption and moisture are computed, and the mixing end. In total, the dataset contains 371 samples for δ_{Δ_t} and 353 samples for τ_{0,Δ_t} and μ_{Δ_t} (note that τ_{0,Δ_t} and μ_{Δ_t} have three concretes less than δ_{Δ_t}). The input is augmented by a random offset, which is determined with an uniform distribution in an interval of $[-0.07, 0.07]$. Training extends to 7000 epochs with a halving of the learning rate every 1400 epochs and a validation every 50 epochs. Early stopping is applied after 25 validations without improvement. The batch size equals the size of the training data set, which means that one training iteration equals one epoch.

5.2 Evaluation Metrics

The performance of the network is assessed with the mean absolute error

$$\epsilon_{abs} = \frac{1}{A}\sum_{a=1}^{A}\frac{1}{J}\sum_{j=1}^{J}|y_j^a - \hat{y}_j^a|\,, \tag{3}$$

the mean relative error

$$\epsilon_{rel} = \frac{1}{A}\sum_{a=1}^{A}\frac{1}{J}\sum_{j=1}^{J}\frac{|y_j^a - \hat{y}_j^a|}{y_j^a}\cdot 100\,, \tag{4}$$

and the mean absolute error normalized by the training value range

$$\epsilon_{rel,range} = \frac{\epsilon_{abs}}{y_{max,train} - y_{min,train}}\cdot 100\,. \tag{5}$$

Here, A is the number of concretes in a set (6–7 for testing, 4 for validation), and J the number of samples per concrete. Metrics are first computed per concrete and then averaged to ensure equal weighting for each concrete. ϵ_{abs} and ϵ_{rel} are useful to asses the prediction quality for practical applications, while $\epsilon_{rel,range}$ enables comparison across outputs with differing value ranges. Weight selection is based on the mean $\epsilon_{rel,range}$ from validation. The weight selection during pretraining is based on the average L1 loss from the validation.

5.3 Results and discussion

5.3.1 Influence of different input combinations The developed network is tested with the input combinations images + Δ_t (images + temporal information) and images + Δ_t + m + p (images + tabular data) to examine their effect on model performance. Furthermore the results are compared to

Table 1. Mean absolute error ϵ_{abs}, mean relative error ϵ_{rel}, and mean absolute error normalized by the reference range during training $\epsilon_{rel,range}$ (standard deviations in brackets) for various input combinations using the described [1]: proposed image-based network and [2]: MLP, respectively (Δ_t = temporal information; m = mix design information; p = process state data).

		Images + Δ_t [1]	Images + Δ_t + m + p [1]	Δ_t + m + p [2]
δ_{Δ_t}	ϵ_{abs} ↓ [cm]	2.78 (0.13)	3.02 (0.25)	3.04 (0.10)
	ϵ_{rel} ↓ [%]	5.56 (0.27)	6.23 (0.54)	6.26 (0.21)
	$\epsilon_{rel,range}$ ↓ [%]	11.36 (0.52)	12.44 (1.03)	12.45 (0.41)
τ_{0,Δ_t}	ϵ_{abs} ↓ [Pa]	38.62 (1.89)	41.60 (0.89)	45.48 (3.41)
	ϵ_{rel} ↓ [%]	24.73 (1.28)	28.56 (0.59)	28.87 (2.76)
	$\epsilon_{rel,range}$ ↓ [%]	10.34 (0.52)	11.27 (0.29)	11.64 (0.77)
μ_{Δ_t}	ϵ_{abs} ↓ [Pa·s]	6.51 (0.15)	7.33 (0.51)	7.52 (0.84)
	ϵ_{rel} ↓ [%]	22.81 (0.74)	26.73 (1.78)	24.72 (2.30)
	$\epsilon_{rel,range}$ ↓ [%]	11.58 (0.23)	12.89 (0.85)	13.36 (1.27)

the MLPs which only use Δ_t + m + p (tabular data) as input.

The results are shown in Tab. 1. For δ_{Δ_t}, the predictions based on images + Δ_t perform slightly better than the predictions based on images + Δ_t + m + p, and the predictions based on Δ_t + m + p, if ϵ_{abs} is considered. Similar patterns are observed for τ_{0,Δ_t} and μ_{Δ_t}, except that the predictions based on Δ_t + m + p for τ_{0,Δ_t} perform slightly worse than those based on images + Δ_t + m + p. When comparing ϵ_{rel} across targets, δ_{Δ_t} is predicted with higher precision than τ_{0,Δ_t} and μ_{Δ_t}. This difference diminishes when results are evaluated by $\epsilon_{rel,range}$, which allows a better comparison between the targets as it takes the value ranges into account.

Since each concrete has a distinct mix design, no identical concretes are present in both training and test sets. Therefore, the network demonstrates the ability to generalize to unseen concretes with comparable composition. For the interpretation of the results the precision of the reference measurements must be considered, which is 2.46 cm for the slump test (EN 12350-5, 2019). The achieved accuracy for δ_{Δ_t} can already be regarded as sufficient for practical applications. The assessment of τ_{0,Δ_t} and μ_{Δ_t} is more challenging due to limited usage in practice. According to (Feys et al., 2023), the *eBT-V* rheometer typically shows standard deviations of 12.2 Pa for yield stress and 3.5 Pa·s for plastic viscosity. However, the cementitious materials studied there are more fluid and not directly comparable to the stiffer concretes in this work. Consequently, the transferability of these uncertainty estimates is limited. In general, a limitation of the presented results is that the dataset only contains mixtures with the same fill level.

The usage of images + Δ_t + m + p does not improve the prediction compared to images + Δ_t. The reason for this could be the very different variability of images and tabular data. During training, hundreds of images per concrete are used, whereas only four p and one m per concrete are used. Therefore, p and m are repeating very often, which probably causes the network to recognise the image features as noise and therefore ignores them. Δ_t is not connected to this issue as Δ_t is similar for every concrete and therefore can not be used exclusively by the network to generate a reasonable prediction. Another reason could be that images and tabular data contain redundant information. Furthermore, preliminary investigations have shown that moisture information is particularly important for the prediction if only tabular data is used.

The fresh concrete properties of certain concretes are more accurately predicted than others. The reasons for this is, in general, that the network has learned to interpolate within certain value ranges but struggles with extrapolation. Therefore, the concretes with significantly higher or lower values for the slump flow diameter, yield stress and plastic viscosity are in general the ones with the highest prediction errors. Furthermore, a few concretes with a high superplasticizer content are predicted very poorly. The reason for this is probably that these concretes react in a more sensitive manner to environmental parameters and are therefore more difficult to predict, especially with a small training set.

5.3.2 Time-dependent prediction model for fresh concrete properties As the network receives the time interval between mixing end and target prediction time, it implicitly learns how the material properties change over time after the mixing. This enables continuous prediction. Fig. 4 illustrates such continuous predictions of the slump flow diameter based on the input combination images + Δ_t. For each concrete depicted in the figure, 100 input sequences and a varying $\Delta_{t,ref}$ are used to generate a prediction for each minute between the first and the last reference measurement. The x-axis represents the concrete age, defined by $\Delta_{t,ref}$, and the y-axis shows the predicted slump flow diameter. Furthermore, the reference values and their uncertainty (EN 12350-5, 2019) are shown in the figure.

For the shown concretes the continuous predictions reveal that the network correctly learned the decrease of the slump flow diameter over time, which aligns with the expected behavior. However, this continuous prediction over time has some limitations. The prediction does not yet work as well as shown in the figure for all concretes. In addition, the data set is relatively small and only contains concretes with decreasing consistency over time.

6. Conclusion and outlook

The presented experiments show that it is possible to generate time-dependent predictions of fresh concrete properties during the mixing process under realistic conditions. The best results are achieved when images and temporal information are used for the prediction. Nevertheless, good results are also achieved if only tabular data is used. In addition, it is demonstrated that it is possible to predict fresh concrete properties as a function of time, thus representing a step towards our vision of predicting,

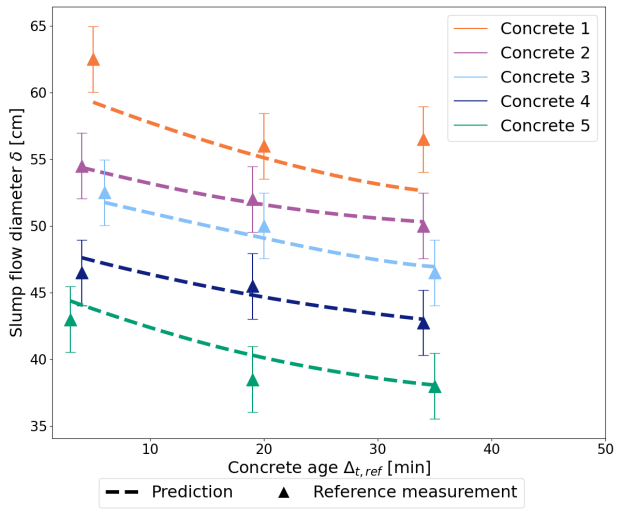


Figure 4. Examples for the prediction of the slump flow diameter over time.

during mixing, the fresh concrete properties for a specific point in time in the future (e.g. at the time of placement).

Future work will focus on further investigating the influence of the various inputs on the prediction. Furthermore, we will examine to which extent the uncertainty of each prediction can be determined in order to identify out-of-distribution concretes already during mixing. In addition, the proposed method will be evaluated using other datasets that are recorded during operation in a concrete plant and are significantly larger and more diverse than the dataset used here and also include varying fill levels.

Acknowledgments

This work is supported by the Federal Ministry of Education and Research of Germany (BMBF) as part of the research project ReCyControl [Project number 0336260A], `https://www.recycontrol.uni-hannover.de/en/` and by the LUIS computing cluster funded by the German Research Foundation (DFG) - INST 187/742-1 FUGG.

References

Babaei, H., Zamani, M., Mohammadi, S., 2025. The impact of data splitting methods on machine learning models: A case study for predicting concrete workability. *Machine Learning for Computational Science and Engineering*, 1(1), 1–22.

Bradski, G., 2000. The OpenCV Library. *Dr. Dobb’s Journal of Software Tools*.

Catti, P., Nikolakis, N., Ntoulmperis, M., Lakkas-Pyknis, V., Alexopoulos, K., 2025. Real-Time Concrete Workability Estimation in Transit via an IoT-Enabled Cyber-Physical System. *Electronics*, 14(16), 3289.

Chidiac, S., Mahmoodzadeh, F., 2009. Plastic viscosity of fresh concrete–A critical review of predictions methods. *Cement and Concrete Composites*, 31(8), 535–544.

Coenen, M., Beyer, D., Haist, M., 2023. Granulometry Transformer: Image-based Granulometry of Concrete Aggregate for an Automated Concrete Production Control. *Proceedings of the 2023 European Conference on Computing in Construction (EC3)*, 4.

Coenen, M., Vogel, C., Schack, T., Haist, M., 2024. Deep Concrete Flow: Deep learning based characterisation of fresh concrete properties from open-channel flow using spatio-temporal flow fields. *Construction and Building Materials*, 411, 134809.

Ding, Z., An, X., 2018. Deep learning approach for estimating workability of self-compacting concrete from mixing image sequences. *Advances in Materials Science and Engineering*, 2018, 1–16.

EN 12350-5, 2019. Testing Fresh Concrete - Part 5: Flow Table Test. European Committee for Standardization.

Farnebäck, G., 2003. Two-frame motion estimation based on polynomial expansion. *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13*, Springer, 363–370.

Feys, D., Keller, H., El Cheikh, K., Secrieru, E., Vanhove, Y., 2023. RILEM TC 266-MRP: round-robin rheological tests on high performance mortar and concrete with adapted rheology—a comprehensive flow curve analysis. *Materials and Structures*, 56(5), 105.

Feys, D., Wallevik, J. E., Yahia, A., Khayat, K. H., Wallevik, O. H., 2013. Extension of the Reiner–Riwlin equation to determine modified Bingham parameters measured in coaxial cylinders rheometers. *Materials and structures*, 46, 289–311.

Gao, X., Yan, H., 2023. Numerical detection of concrete slump by fusion of target segmentation and image classification network. *Journal of Physics: Conference Series*, 2562number 1, IOP Publishing, 012023.

González-Taboada, I., González-Fonteboa, B., Martínez-Abella, F., Roussel, N., 2018. Robustness of self-compacting recycled Concrete: Analysis of Sensitivity Parameters. *Materials and Structures*, 51(8).

Green, B., 2016. *Productivity in Construction: Creating a Framework for the Industry to Thrive*. Chartered Institute of Building (CIOB).

Guo, P., Du, J., Bao, Y., Meng, W., 2022. Real-time video recognition for assessing plastic viscosity of ultra-high-performance concrete (UHPC). *Measurement*, 191, 110809.

Haist, M., Bergmeister, K., Curbach, M., Forman, P., Gaganelis, G., Gerlach, J., Mark, P., Moffatt, J., Müller, C., Müller, H. S. et al., 2022a. Nachhaltig konstruieren und bauen mit Beton. *BetonKalender 2022: Nachhaltigkeit, Digitalisierung, Instandhaltung*, 421–531.

Haist, M., Heipke, C., Beyer, D., Coenen, M., Vogel, C., Schack, T., Ponick, A., Langer, A., 2022b. Digitization of the Concrete Production Chain using Computer Vision and Artificial Intelligence. *Proceedings of the 6th fib Congress*, 434–443.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R., 2022. Masked autoencoders are scalable vision learners. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 16000–16009.

Idrees, S., Nugraha, J. A., Tahir, S., Choi, K., Choi, J., Ryu, D.-H., Kim, J.-H., 2024. Automatic concrete slump prediction of concrete batching plant by deep learning. *Developments in the Built Environment*, 18, 100474.

Inqiad, W. B., Javed, M. F., Alsekait, D. M., Khan, N. M., Khan, M., Aslam, F., Elminaam, D. S. A., 2025. Soft-computing models for predicting plastic viscosity and interface yield stress of fresh concrete. *Scientific Reports*, 15(1), 10740.

Khayat, K. H., Libre, N. A. et al., 2014. Automated measurement and control of concrete properties in a ready mix truck with verifi. Technical report, Missouri University of Science and Technology. Center for Transportation ....

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment Anything. *arXiv:2304.02643*.

Kumar, R., Rathore, A., Singh, R., Mir, A. A., Tipu, R. K., Patel, M., 2024. Prognosis of flow of fly ash and blast furnace slag-based concrete: leveraging advanced machine learning algorithms. *Asian Journal of Civil Engineering*, 25(3), 2483–2497.

Langley, A., Lonergan, M., Huang, T., Azghadi, M. R., 2025. Analyzing mixed construction and demolition waste in material recovery facilities: Evolution, challenges, and applications of computer vision and deep learning. *Resources, Conservation and Recycling*, 217, 108218.

Li, S., An, X., 2014. Method for estimating workability of self-compacting concrete using mixing process images. *Computers and Concrete*, 13(6), 781–798.

Liu, L., Cao, G., Shi, Y., Jiang, S., Deng, D., 2024. Effect of time-dependence on the concrete transportation process. *Powder Technology*, 437, 119535.

Malekipour, M., Moodi, F., 2021. A novel approach to improve quality of delivered concrete using slump estimations of the ready-mixed concrete (RMC) truck mixer. *Journal of Building Engineering*, 44, 103361.

Meyer, M., Langer, A., Mehltretter, M., Beyer, D., Coenen, M., Schack, T., Haist, M., Heipke, C., 2024. Fresh Concrete Properties from Stereoscopic Image Sequences. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 92(5), 517–529.

Navarrete, I., La Fé-Perdomo, I., Ramos-Grez, J. A., Lopez, M., 2023. Predicting the evolution of static yield stress with time of blended cement paste through a machine learning approach. *Construction and Building Materials*, 371.

Nguyen, T.-D., Tran, T.-H., Hoang, N.-D., 2020. Prediction of interface yield stress and plastic viscosity of fresh concrete using a hybrid machine learning approach. *Advanced Engineering Informatics*, 44.

Ojala, T., Punkki, J., 2024. Estimating the Workability of Concrete with a Stereovision Camera during Mixing. *Sensors (Basel, Switzerland)*, 24(14), 4472.

Ponick, A., Langer, A., Beyer, D., Coenen, M., Haist, M., Heipke, C., 2022. Image-Based Deep Learning for Rheology Determination of Bingham Fluids. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 711–720.

Rong, X., Liu, H., Li, C., 2022. A proposed method and monitoring system for evaluating workability of Portland cement concrete during mixing. *Heliyon*, 8(11).

Schack, T., Coenen, M., Haist, M., 2023a. Bildbasierte Frischbetonprüfung – Teil 1: Konsistenz und Leimgehalt des Frischbetons. *Beton- und Stahlbetonbau*, 118(4), 220-228.

Schack, T., Coenen, M., Haist, M., 2023b. Bildbasierte Frischbetonprüfung – Teil 2: Granulometrische Eigenschaften der Gesteinskörnung. *Beton- und Stahlbetonbau*, 118(8), 556-564.

Schack, T., Coenen, M., Haist, M., 2023c. Bildbasierte Frischbetonprüfung: Teil 3: Homogenität des Frischbetons. *Beton- und Stahlbetonbau*, 118(10), 716-724.

Tuan, N. M., Van Hau, Q., Chin, S., Park, S., 2021. In-situ concrete slump test incorporating deep learning and stereo vision. *Automation in Construction*, 121, 103432.

Wallevik, J. E., 2006. Relationship between the Bingham parameters and slump. *Cement and concrete research*, 36(7), 1214–1221.

Wallevik, J. E., Wallevik, O. H., 2020. Concrete mixing truck as a rheometer. *Cement and Concrete Research*, 127, 105930.

Yahia, A., Mantellato, S., Flatt, R. J., 2016. Concrete rheology: A basis for understanding chemical admixtures. *Science and Technology of Concrete Admixtures*, Elsevier, 97–127.

Yang, L., An, X., Du, S., 2021. Estimating workability of concrete with different strength grades based on deep learning. *Measurement*, 186.

Yoon, J., Kim, H., Ju, S., Li, Z., Pyo, S., 2023. Framework for rapid characterization of fresh properties of cementitious materials using point cloud and machine learning. *Construction and Building Materials*, 400, 132647.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T. S., 2019. Free-form image inpainting with gated convolution. *Proceedings of the IEEE/CVF international conference on computer vision*, 4471–4480.