Øyvind S. Sortland
Mohammed M'Hamdi
Moez Jomâa

Assessment of feature engineering and long short-term memory for structure loss identification from process data in monocrystalline silicon growth by the Czochralski method

Report

NTNU
Norwegian University of Science and Technology
Faculty of Faculty of Natural Sciences
Department of Materials Science and Engineering

NTNU
Norwegian University of Science and Technology

NTNU
Norwegian University of Science and Technology

Report

Assessment of feature engineering and long short-term memory for structure loss identification from process data in monocrystalline silicon growth by the Czochralski method

VERSION
1st version

DATE
28.11.2019

AUTHOR
Co-authors (optionally)

PROJECT NUMBER
257639

CLIENT(S)

NUMBER OF PAGES AND ATTACHMENTS
12

CLIENTS REF.

ABSTRACT
Structure loss impairs the quality of monocrystalline ingots and represents a major productivity loss for crystal growers. If structure loss could be predicted in advance it could help reduce the lost production time or preventive measures could be initiated. For this reason, feature engineering and machine learning by long short-term memory (LSTM) is used to assess if structure loss could be predicted from sensor data collected during growth of ingots. The method is not able to predict structure loss in advance, and the predictions may likely be based on the length of the signal, which is shortened for structure loss ingots as their growth is interrupted, and not actual features in the sensor readings.

CORRESPONDING AUTHOR
Øyvind S. Sortland

SIGNATURE

CONTROLLED BY

SIGNATURE

APPROVED BY

SIGNATURE

REPORT NUMBER	ISBN	CLASSIFICATION	CLASSIFICATION THIS PAGE

NTNU
Norwegian University of Science and Technology

Background

VERSION	DATE
1	28.11.2019

VERSION DESCRIPTION
1st published version

Contents

Introduction **5**

Methods **6**

A. Data *6*

B. Calculation of features *8*

C. LSTM network training and testing *9*

Results and discussion **10**

Conclusion **15**

Acknowledgements **15**

References **15**

Introduction

The Czochralski method is the main process for producing monocrystalline silicon for high-efficiency solar cells. The successful ingots are considered to be dislocation free [1]. In around one third of the ingots however, dislocation generation starts and the ingot starts to grow multicrystalline in an event called structure loss (SL). This represents lost production time or yield for monocrystalline silicon producers as the crystal may be remelted or the affected part may be cut away.

In the Czochralski process, high-purity silicon is molten in a silica crucible (Figure 1a-b) and a monocrystalline seed crystal is dipped into the top of the melt (Figure 1c-e). The crucible and the seed are rotated in opposite directions. When the seed is dipped into the melt, it experiences thermal shock and dislocations are generated. First a thin neck is grown at high pull speed to make dislocations move out to the surface so that the rest of the ingot can grow dislocation free [1]. The diameter is increased in the crown to the desired diameter of the useful body of the crystal (Figure 1f-g). A tail is grown to make a thin tip at separation of the ingot from the melt, in order to avoid slip and dislocation generation into the body (Figure 1h-j).

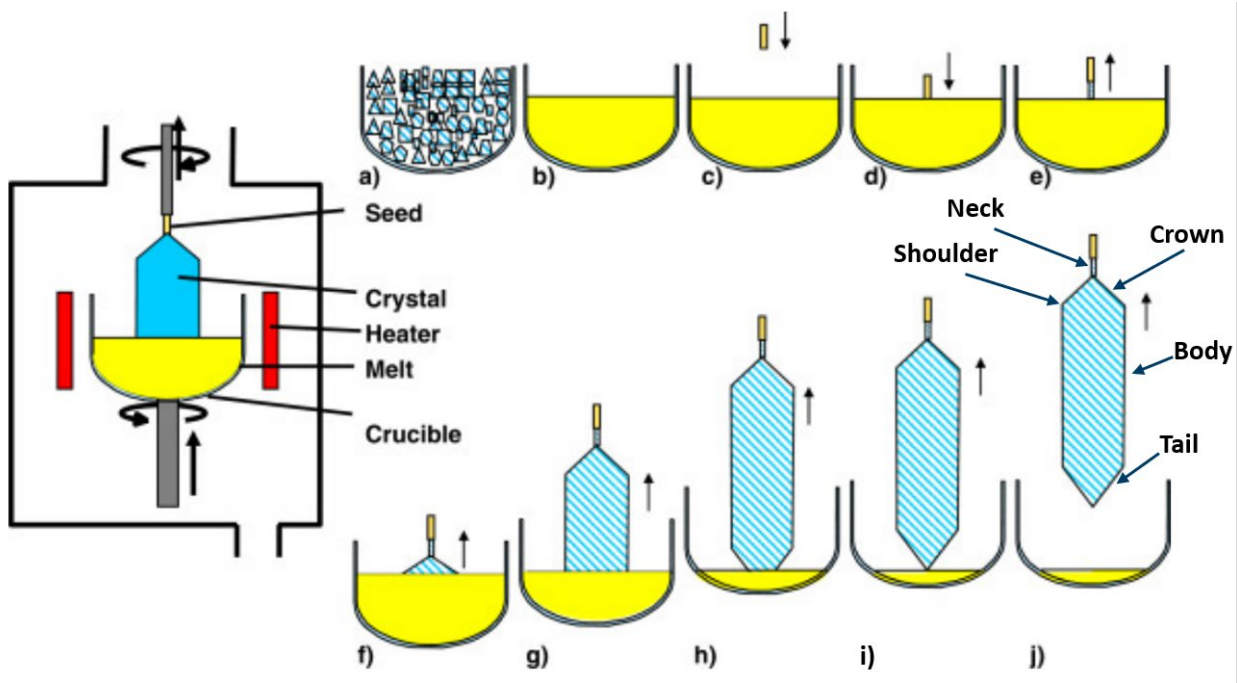

Figure 1: Schematic of the principle of the Czochralski method (left) and illustration of the different steps (a–j) of the Cz process for growing a monocrystalline ingot. (a) The polycrystalline feedstock is melted (b) in a crucible. (c, d) Seeding procedure: The seed crystal is dipped into the melt, followed by Dash necking (e), crown growth (f), cylindrical body growth (g), growth of tail cone (h), lift off (i), cooling down and removing of the crystal (j) (modified from [2]).

The process is continuously monitored and data are logged every second in the NorSun plant in Årdal, Norway. This allows for mining of big data, and the main objective of this study is to test statistical learning and investigate whether the data can be used to predict structure loss. There is a potential to use real-time or periodically sampled data during growth of an ingot to predict if it will end in structure loss. At least the analysis will indicate if it is possible to predict structure loss from production data, and the features of the signal that is needed for good prediction can be investigated by controlling the input

data. The investigated data include ingot diameter, pull speed, heater power and a pyrometer reading. For each of these measurements there is a sequence of up to thousands of datapoints, for which long short-term memory (LSTM) [3] is selected for classifying ingots as having structure loss or not. Data for a portion of the ingots are left to test the trained classifier against the actual occurrence of structure loss.

Methods

The first strategy before turning to machine learning directly on the timeseries is to calculate features from the sensor timeseries and assess if they have separate range of values from ingots with structure loss than ingots without structure loss. Such a feature may be used for prediction and classification of ingots with and without structure loss. As a second approach, long short-term memory is applied to aggregated and cleaned sensor readings to classify ingots as ending in structure loss or not. Predictions are then compared with the actual occurrence of structure loss in test set ingots. The workflow of the LSTM analysis is shown in Figure 2.

A. Data

Silicon production industry is highly digitalized, and a lot of information is continuously monitored. In this work, focus is given to diameter, pull speed, heater power and a pyrometer reading. This choice is motivated by the fact that these parameters are related to the growth process and changes in correspondence to growth conditions.

The equivalent of one-year worth of data has been collected. In total, the dataset is made of 13589 ingots. Roughly 66% of them without structure loss and 33% are with structure loss. In Figure 3, frequencies of structure loss by crystal length are given. Almost 50 % of all structure losses happen before 100 mm of body. The first bin with negative lengths in Figure 3 represents structure loss in crown or shoulder.

For each ingot, the data are collected from start of neck until ingot growth is stopped in crown, shoulder or body, or until the end of body for ingots without structure loss. In order to reduce the data size, the sensor readings have been aggregated as averages over each 30 seconds. By visual inspection, this is found to resolve the fluctuations in all signals.

Occasional sensor readings are erroneous (zeros) and they are replaced by linear interpolation. For the diameter reading, also spikes occurs. Diameters differing more than 2.5 mm from a median filter are corrected to the median filter value. The median is calculated for a window of 65 points, as this is the highest value that still tracked the signal trend (particularly in the neck). For correct median filter calculation in start and end of the signal, the signal is padded with 65 points equal to the first or last sensor reading, respectively. This is also done as the signal is split at the start of body in order for the median filter to capture a top in diameter here. The padding is subsequently removed and the sections merged to represent the whole ingot. The median filter, corrected diameter and original diameter readings are shown for the start of an ingot in Figure 4. The diameter is calculated from tracking the meniscus in a camera feed. The meniscus may be difficult to track in crown growth and the diameter signal may be erroneous over long periods in this stage.

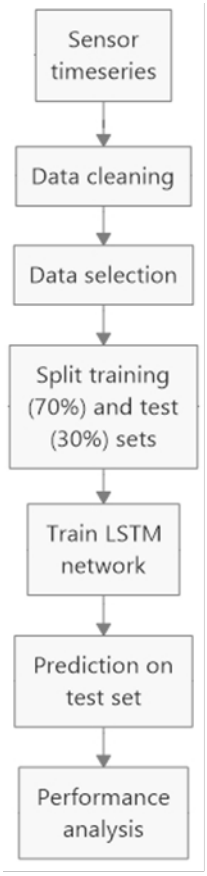

Figure 2: Schematic of the workflow.

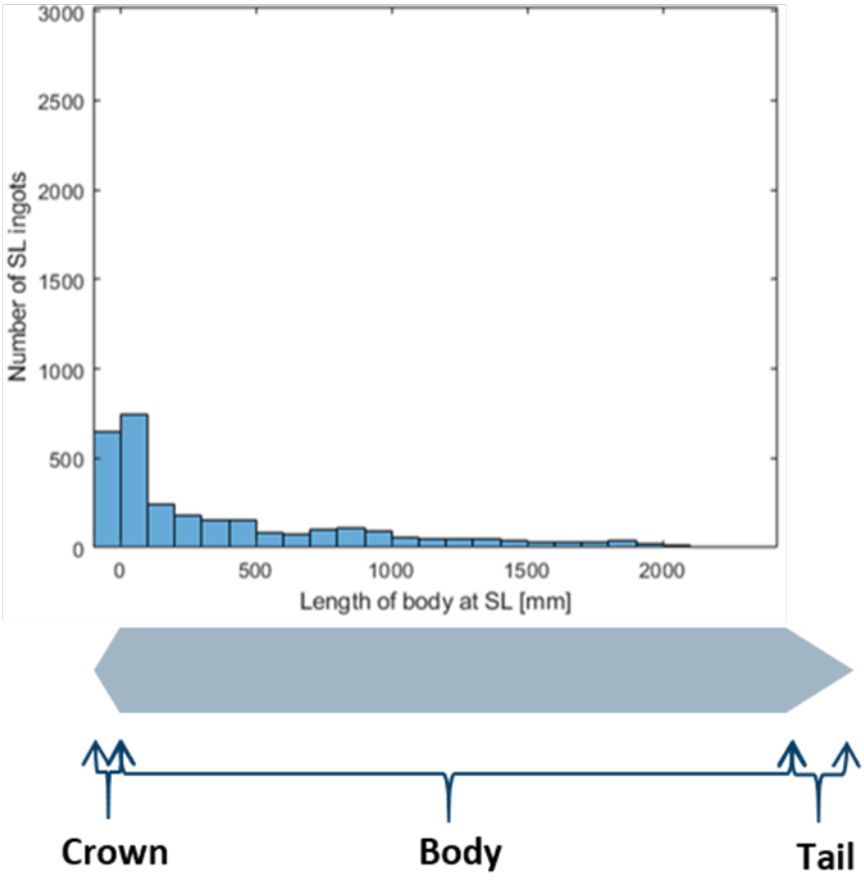

Figure 3: Frequencies of structure loss occurrence by crystal

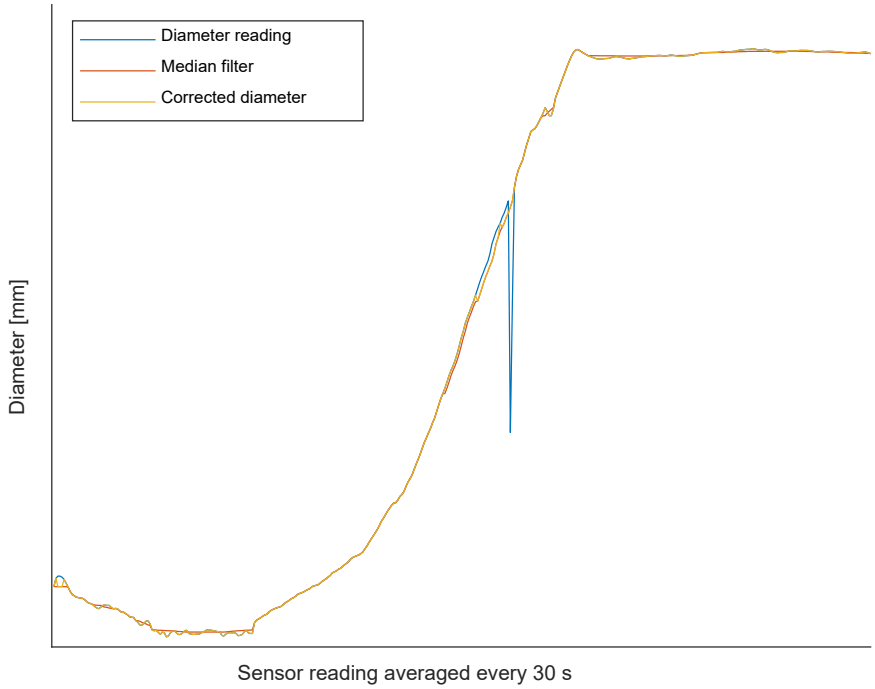

Figure 4: Diameter reading, median filter and correction in neck, crown and start of body.

B. Calculation of features

A range of features is calculated from the timeseries of different sensor readings and their value is compare between ingots with structure loss and ingots without structure loss, to visually assess whether each feature can be used to classify ingots as having structure loss or not. Table I summarizes the assessment of features inspected visually. Table II show features which have their mean for ingots with structure loss and for ingots without structure loss compared by t-tests with a 5% significance level. The ingots used for Table II have the most typical production type (6.7 inch ingots), crucible and feedstock material. The analyses differentiate between first (A-ingot) and second (B-ingot) ingot of a run.

Table I: Features calculated from sensor readings and used to visually assess if they can classify ingots with structure loss (SL ingots) and ingots without structure loss (non-SL ingots). The column “Diff.” indicates whether SL and non-SL ingots could be visually separated by the feature.

Feature	Diff.	Comment
Slope of pyrometer reading in body after minimum	Yes	Higher slope towards full-length ingots. Lower slope of SL ingots likely an effect of short ingots.
Slope of pyrometer reading in 1 h intervals between 10-25 h after start neck	No	
Diameter at 30 mm crown length	No	
Crown angle (to vertical) after 30 mm diameter	No	
Mean diameter fluctuation frequency weighted by amplitude	No	
Prominence of power valley in crown	No	
Prominence of power peak following the valley in crown	No	
Change between adjacent pull speed points (15s interval)	No	
Mean of pull speed in neck	No	
Duration of neck	No	
Length of neck	No	
Pyrometer reading at end neck	No	
Minimum pyrometer reading	No	
Time of minimum pyrometer reading after start neck	No	
Prominence of largest power valley in ingot	Yes	The largest power valley extends to the end of the ingot, and the separation is due to short SL ingots.
Power at neck end	No	
Minimum power	No	
Power drop from neck to minimum in early body	No	

Table II: Features calculated from sensor readings and used assess if they can classify ingots with structure loss (SL ingots) and ingots without structure loss (non-SL ingots). The column “Diff.” indicates whether the mean of the feature for SL ingots is significantly different from the mean of non-SL ingots at 5% significance level.

Feature	Diff.	Comment
Number of power peaks over 10 standard deviations after 4 h of body	Yes	Higher number of peaks for non-SL ingots than SL ingots, due to manual adjustments which cause peaks reduce risk of structure loss [4].
Stabilization time from end of melting to start neck	No	
Mean power between 100-375 mm of body	No	
Mean pyrometer reading between 100-375 mm of body		Significantly higher for SL ingots than non-SL ingots for B-ingots, but not A-ingots.
Minimum pull speed after 6 hours of body	Yes	Significantly higher for SL ingots than non-SL ingots, because pull speed is reduced towards the end of full-length ingots, so this is an effect of short SL ingots.
Standard deviation of diameter fluctuation, in last hour of SL ingots and in body for non-SL		Significantly higher for non-SL ingots than SL ingots for B-ingots, but not A-ingots.
Power at start neck	No	

C. LSTM network training and testing

A LSTM network is set up according to an example in MathWorks Documentation [5]. The architecture has five layers: Sequence input layer, bidirectional LSTM layer with 75 hidden units, fully connected layer, softmax layer and classification layer. High accuracy of prediction can be achieved with 50, 75 and 100 hidden units in the LSTM layer, while 300 hidden units yields lower accuracy.

70% of ingots from one or two pullers are used for training, in order to complete training within a couple of hours, and the remaining 30% are used for testing. For training on ingots only from one puller, all ingots from a second puller is used for testing.

The number of training ingots is at most 270. With such low number of training ingots, best results are achieved when all ingots are included in each iteration, meaning the minibatch size is up to 256 (the minibatch size is a lower number of ingots included in each iteration). The training algorithm pads ingots with zeros to the signal length of the longest ingot. The training ingots are sorted according to signal length to minimize padding.

The trainings are run for 100 epochs, with all ingots being processed in each epoch, and the network is saved after each epoch. To avoid underfitting and overfitting, each network is used for classification of test ingots. The accuracy and confusion matrix are assessed for each network and the network with highest accuracy and lowest number of ingots without structure loss classified as ingots with structure loss (false positives) is selected. False positives are considered a more serious error than false negatives (classifying structure loss ingots as not having structure loss). In case an algorithm to predict structure loss is implemented in production, a false positive could trigger termination of an ingot that would continue to grow with high quality, while a false negative would not

initiate any action and the structure loss will be handled as before implementation the algorithm. The performance of the selected network is also assessed with Receiver operating characteristics (ROC) curve.

Results and discussion

First, 70% of ingots from one puller is used to train LSTM networks. After each epoch during training a network is saved. The performance of each network is tested on the remaining 30% of the data from the puller. Figure 5 shows accuracies and number of false classifications for each network. The networks after epoch 44 and 55 reach 100% accuracy with no false classification on this test set.

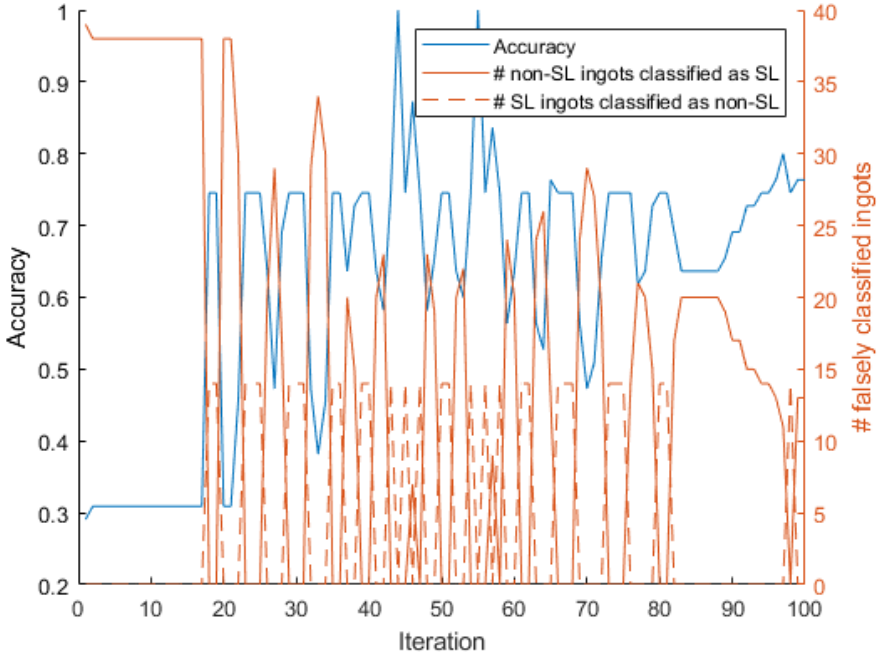

Figure 5: Accuracies and number of falsely classified ingots from networks saved after each epoch (iteration) during training.

The network after epoch 55 is selected for testing on all ingots from a second puller. It provides an accuracy of 99% with two ingots without structure loss classifies as having structure loss (false positives), as shown in the confusion matrix in Table III. Figure 6 shows the ROC curve for the prediction, having an area under the curve close to one.

Table III: Confusion matrix for best network trained on one 70% of ingots from puller and tested on all ingots of a second puller.

Actual \ Prediction	Non-SL	SL
Non-SL	75	2
SL	0	132

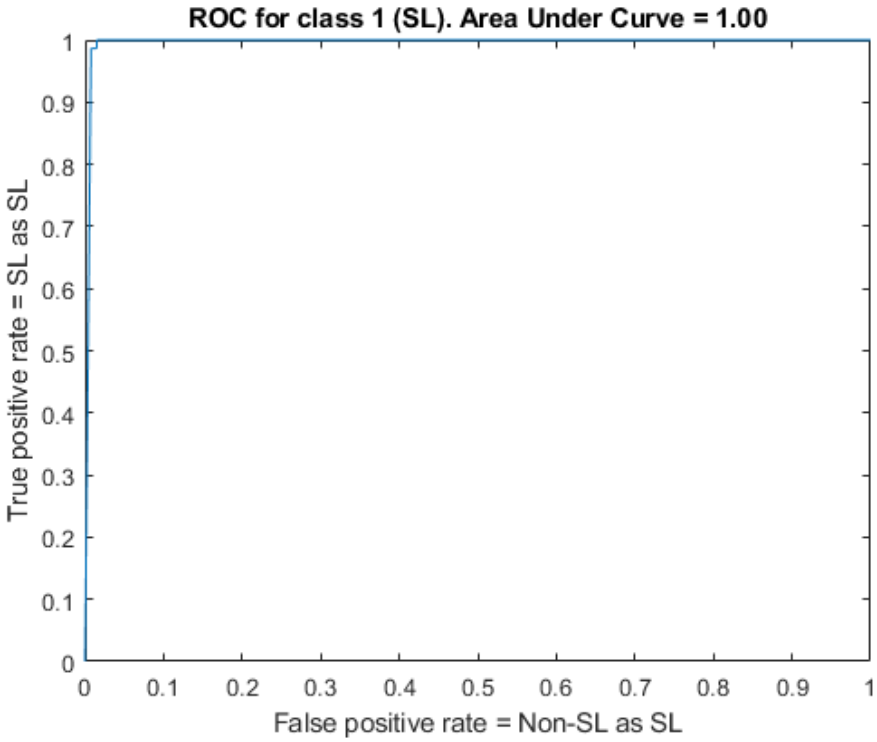

Figure 6: Receiver operating characteristics for best network trained on 70% of ingots from one puller and tested on all ingots of a second puller.

Best accuracy is achieved when the network is trained and tested on ingots of similar diameter. A network trained on a puller with 6.5-6.7 inch diameter ingots achieves 91% accuracy when tested on ingots with the same diameters from a second puller. The confusion matrix is shown in Table IV and 19 ingots without structure loss is classified as having structure loss (false positive). When tested on 8.4-8.5 inch diameter ingots from a second puller, the accuracy is 82% and the confusion matrix is shown in Table V with 35 false positives. Training a network on 8.4-8.5 inch diameter ingots form a third puller achieves a higher accuracy of 95% with 5 ingots without structure loss falsely classified to have structure loss in the confusion matrix in Table VI.

Table IV: Confusion matrix for ingots trained and tested on 6.5-6.7 inch diameter ingots.

Actual \ Prediction	Non-SL	SL
Non-SL	115	19
SL	0	75

Table V: Confusion matrix for ingots trained on 6.5-6.7 inch and tested on 8.4-8.5 inch diameter ingots.

Actual \ Prediction	Non-SL	SL
Non-SL	96	35
SL	2	71

Table VI: Confusion matrix for ingots trained and tested on 8.4-8.5 inch diameter ingots.

Actual \ Prediction	Non-SL	SL
Non-SL	106	5
SL	3	60

To investigate which sensors provide information best used to predict structure loss, models are trained and tested for all combinations of four sensors, namely diameter (D), Main Heater power (MHP), a pyrometer reading (T) and pull speed (PS). Several combinations achieve 99.0% accuracy of prediction with one non-SL ingot classified as SL as shown in Table VII. Many of these combinations include the diameter reading. Furthermore, the lowest number of sensors achieving this high accuracy is diameter alone. It is thus decided to only use the diameter reading for further investigations. It is noted in Table VII that including all the sensors provides the worst prediction, possibly due to increasing noise from sensors not effective in predicting structure loss. This seems different than the high accuracy achieved for training and testing on the same data presented in Figures 5-6 and Table III.

Table VII: Comparison of prediction accuracy and number of false positives (Non-SL as SL) and false negatives (SL as Non-SL) for different combinations of sensor signals: Diameter (D), Main Heater power (MHP), a pyrometer reading (T) and pull speed (PS).

Sensors	Accu-racy	Non-SL as SL	SL as Non-SL
D	99.0 %	1	1
MHP,D	99.0 %	1	1
MHP,PS	99.0 %	1	1
MHP,PS,D	99.0 %	1	1
MHP,T,D	99.0 %	1	1
T,PS	99.0 %	1	1
T,PS,D	99.0 %	1	1
MHP,T	99.0 %	2	0
MHP,T,PS	99.0 %	2	0
T,D	99.0 %	2	0
PS,D	99.0 %	2	1
MHP	98.6 %	3	0
T	97.1 %	1	5
PS	92.3 %	16	0
MHP,T,PS,D	90.4 %	20	0

Different sections of the signal along the length of the ingots are tested to assess if only part of the signal can be used for prediction of structure loss. If structure loss could be predicted in advance, an algorithm could collect data during growth and the process could be stopped before structure loss would occur, saving production time. Thus it is tested to load an increasing length of the signal of test ingots to use for prediction. Structure loss can not be predicted until the whole signal is included in the classification algorithm. Actually, structure loss is first predicted when the length of the padding after a structure loss extended for a few hours. This indicate that the length of the padding may be used by the network to predict structure loss. Using only the part of the diameter signal in the body which is relatively constant, structure loss can not be predicted and also the first transient at least from the top in diameter at the transition from shoulder to body needs to be included in order to predict structure loss.

As an alternative to padding, the diameter signal is stretched to the length of the longest ingot so that the signal length of all ingots is equal to the signal length of the longest ingot. This is achieved using an algorithm for resizing images with bilinear interpolation in which the output value is a weighted average of the nearest two values. A network trained on this data is able to predict structure loss in 11 out of 30 ingots with structure loss in the test set, with the remaining 19 ingots with structure loss falsely classified as not having structure loss. All 67 ingots without structure loss is correctly classified. Although this data transformation avoids padding of the signal with zeros, the length of the ingots still has effect on the signal as the duration of neck and crown spans an increasing number of points for shorter ingots with structure loss. A second approach extrapolated the diameter signal including noise that resembles the fluctuations in the diameter signal. The network is then not able to predict structure loss and all ingots are predicted as not having structure loss, indicating that the padding of the signal plays a role in the algorithm for predicting structure loss.

Ingots with structure loss and similar length is extracted from all pullers and pooled together with a random selection of ingots without structure loss with cut-off of the diameter signal to the same length. The 5-minute window with the largest number of structure loss ingots is 159.5-164.5 min (after start neck), encompassing 85 ingots with structure loss in this time. Within this window, the shorter ingots are padded with extrapolation of the diameter and adding noise. 85 ingots without structure loss was added and the signals were cut at 164.5 min. Training on 70% of the ingots and testing on 30% of the ingot provided the confusion matrix in Table VIII, with a mix of correct and both types of errors. The accuracy for predictions on the test set is 61.5%, not much better than random.

Table VIII: Confusion matrix for 85 structure loss ingots lasting 159.5-164.5 min and 85 non-structure loss ingots with signal cut at 164.5 min.

Actual \ Prediction	Non-SL	SL
Non-SL	19	6
SL	14	13

Figure 7 are results from training and testing on only part of the diameter signal from start neck to 250 minutes of body, after the diameter has stabilized around a constant value. The foresight axis counts the number of hours the ingot ends after this window. Ingots at zero foresight ends within this window used for training and prediction. It is observed that structure loss is only predicted for ingots which ends due to structure loss within the window used for prediction, meaning the signal is padded after the structure loss occurred. Ingots that ended in structure loss after the window up to 250 minutes of body are not predicted to have structure loss, but are predicted as non-structure loss ingots like all ingots for which the signal completed the window and is not padded.

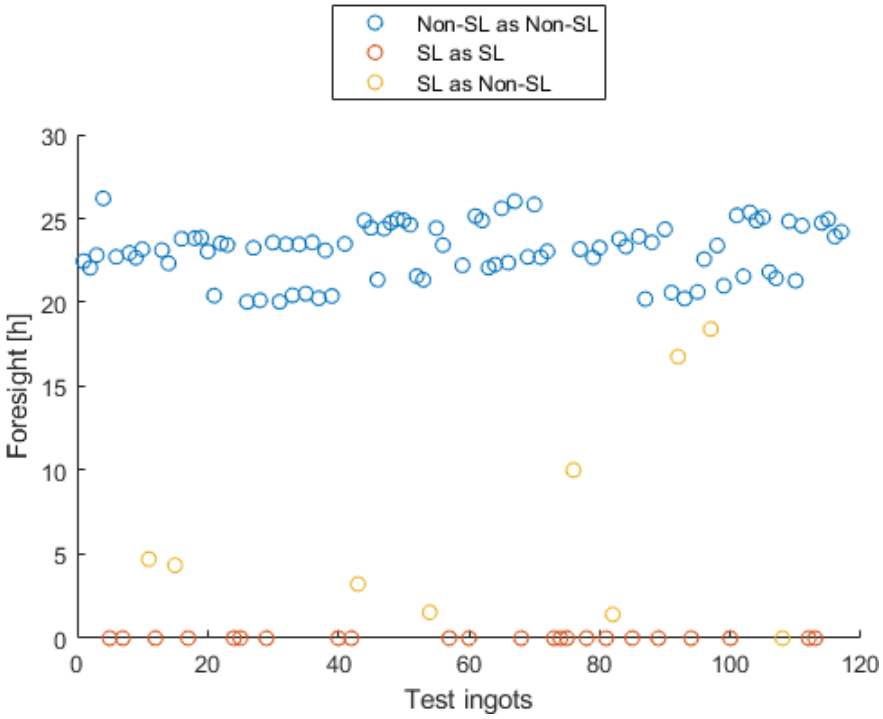

Figure 7: Predictions for test ingots from neck to 250 minutes of body, showing number of hours ingots end after this signal length used for training and classification.

For implementation to a puller, a trained network would predict structure loss for a growing ingot. The network should thus be able to predict structure loss for only one ingot. Test with only one ingot do not provide any prediction of structure loss, possibly because there is no padding of the signal for only one ingot in the minibatch. Prediction

of structure loss for one specific ingot as a part of a larger minibatch with dummy ingots is also not successful.

Bidirectional LSTM is able to use the context from both past and future points at any point in the sequence, while unidirectional LSTM only utilize the context from past points. The necessity of bidirectional LSTM is assessed by comparing to unidirectional LSTM, by replacing the bidirectional LSTM layer in the network architecture. The unidirectional LSTM network failed to predict structure loss for training and test sets for which bidirectional LSTM predicted 30 correct structure losses, with an accuracy of 100%.

To investigate what features of the diameter signal the network use to predict structure loss, the signal is smoothed. A network is trained and tested for the smoothed signal and for only the fluctuation remaining when subtracting the smoothed signal from the original diameter reading. The smoothing is done with a moving average of 50 points (25 minutes) window. The smoothed and original signal are shown in Figure 8.

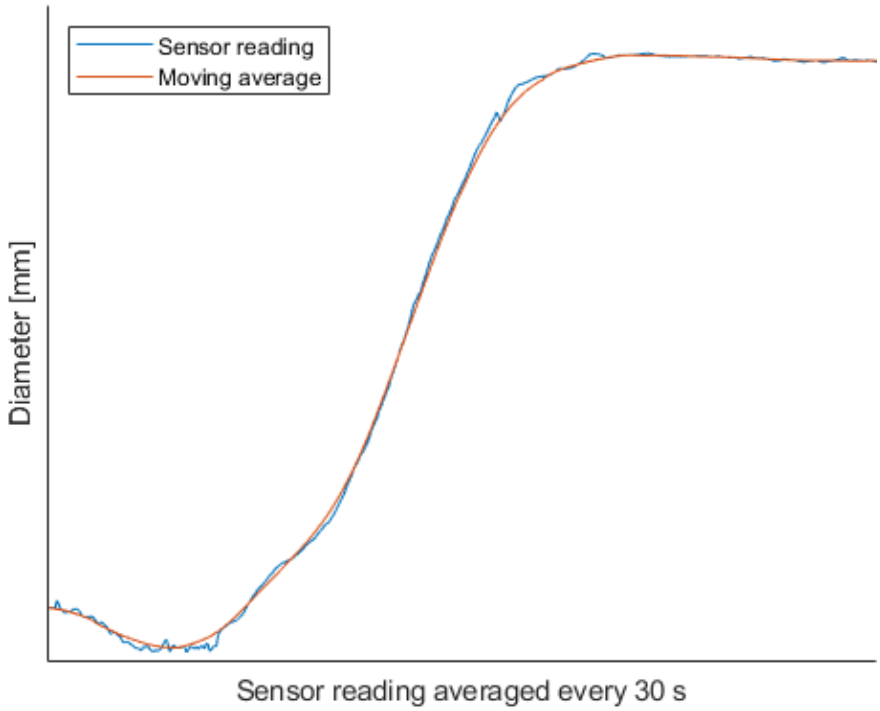

Figure 8: Moving average with 25 min window size and original diameter reading of points every 30 s with average sensor reading.

Using the moving average to train and test a LSTM network results in no structure loss ingots being predicted, even though also this signal is padded with zeros after the end of ingots. The same result is obtained when using the fluctuation of the signal (subtracted by the moving average) for training and prediction.
Power, pull speed, pyrometer and diameter readings averaged over 30 s does not seem to contain information that can be used to predict if an ingot will have structure loss. Large fluctuations of the crystallization rate and temporal remelting at local points on the solidification front are considered as a cause for structure loss [6], and such fluctuations are not captured in the signal of overall pull speed. A more detailed estimation or measurement of local crystallization rate and remelting could prove useful to predict structure loss and add more information on the effect of crystallization rate fluctuations and remelting on the occurrence of structure loss.

Conclusion

Feature engineering and long short-term memory (LSTM) is assessed for predicting structure loss in industrial Czochralski silicon ingots. Some features have different values for ingots with and without structure loss, but except features reported in [4], these were due to ingots with structure loss being shorter than ingots without structure loss. LSTM can provide high accuracy of prediction using only the diameter signal. However, the length of the signal and padding with zeros in the training algorithm could not be excluded as the feature used for prediction. Structure loss could only be accurately predicted if the signal ends and is padded.

Acknowledgements

This work was performed within the Research Centre for Sustainable Solar Cell Technology (FME SuSolTech, project number 257639), co-sponsored by the Norwegian Research Council and research and industry partners. NorSun AS is gratefully acknowledged for providing industrial data investigated in this work, and Hendrik Schön is acknowledged for fruitful discussions.

References

[1] G. Dhanaraj, K. Byrappa, V. Prasad, and M. Dudley (Eds.), *Springer Handbook of Crystal Growth.* New York: Springer-Verlag, 2010, pp. 217, 1338-1339.
[2] J. Friedrich, "Methods for Bulk Growth of Inorganic Crystals: Crystal Growth", *Encyclopedia of Condensed Matter Physics,* 2005, pp. 262-274
[3] S. Hochreiter, and J. Schmidhuber. "Long short-term memory." *Neural computation.* Vol. 9, Number 8, 1997, pp.1735–1780.
[4] Ø. S. Sortland, M. Jomâa, M. M'hamdi, E. J. Øvrelid, M Di Sabatino. "Statistical Analysis of Structure Loss in Czochralski Silicon Growth." In: *SiliconPV 2019, THE 9TH INTERNATIONAL CONFERENCE ON CRYSTALLINE SILICON PHOTOVOLTAICS.* American Institute of Physics (AIP), 2019, ISBN 978-0-7354-1892-9. pp. 100002-1-100002-7.
[5] MathWorks: Documentation, https://se.mathworks.com/help/deeplearning/examples/classify-sequence-data-using-lstm-networks.html, 21 October 2019.
[6] O. V. Smirnova, N. V. Durnev, K. E. Shadrakova, E. L. Mizitov, V. D. Soklakov. "Optimization of furnace design and growth parameters for Si Cz growth, using numerical simulation", *J. Cryst. Growth.* Vol. 310, 2008, pp. 2185-2191.

NTNU

Norwegian University of
Science and Technology