Prediction of inclusion state in molten steel by morphology and appearance of inclusions in liquid steel samples

Tuomas Alatarvas, Tero Vuolio, Eetu-Pekka Heikkinen, Qifeng Shu, Timo Fabritius*

Dr. T. Alatarvas, T. Vuolio, Doc. E.-P. Heikkinen, Assoc. Prof. Q. Shu, Prof. T. Fabritius
Process Metallurgy Research Unit, P.O. Box 4300, FI-90014, University of Oulu, Finland
E-mail: tuomas.alatarvas@oulu.fi

Keywords: inclusions, inclusion analyses, calcium treatment, classification

Inclusions are unwanted but to some extent inevitable in molten and solid steel. Usually solid inclusions are considered to be the most harmful. Inclusions can be converted into a less detrimental form with calcium treatment. The success of the calcium treatment can be evaluated by analyzing the state of the inclusion population. The state of inclusions is usually determined by computational thermodynamics making use of the chemical composition of the inclusions and system conditions. In this process, the liquid and solid inclusions are usually distinguished. This study presents a classification procedure which combines computational thermodynamics and data-driven reasoning. The objective of this work is to study the predictability of the inclusion state based on its appearance and morphological properties. As a result, $\text{Al}_2\text{O}_3\text{--CaO--MgO--CaS}$ inclusions were classified as liquid and solid ones based on their aspect ratio, equivalent circle diameter and mean grey value with a recall of 82.7% and precision of 84.9% by making use of a logistic regression based classifier.

1. Introduction

In aluminum killed steels, solid alumina (Al_2O_3) and spinel ($\text{MgO}\cdot\text{Al}_2\text{O}_3$) inclusions are transformed into less detrimental liquid inclusions with calcium treatment. After the addition of calcium to molten steel, alumina and spinel inclusions react with the added calcium, and their compositions move towards low-melting calcium aluminates. The ratio between Al_2O_3 and CaO in calcium aluminates must be within the correct range to obtain liquid inclusions at steelmaking temperatures.^[1] Inadequate calcium additions or a low yield may lead to remaining Al_2O_3 or $\text{MgO}\cdot\text{Al}_2\text{O}_3$ solid phases in the inclusions, whereas excess calcium leads to the formation of solid CaO or CaS, depending on the steel composition.^[2]

During steel melt shop processes, clear benefits are observed when the inclusions are liquid. They lead to better castability by preventing nozzle clogging during continuous casting. It has been reported that for satisfactory casting performance, inclusions do not need to be fully liquid.^[1,3] Fuhr et al.^[4] observed a decreasing risk of nozzle clogging when the liquid fraction of inclusions exceeded 30–50%. After continuous casting, inclusions, especially in large quantities or sizes, weaken the mechanical properties of the steel.^[5]

The liquid fraction and consequently the roundness of inclusions can be considered as an indication of the success of calcium treatment. The chemical composition of inclusions acquired by scanning electron microscopy combined with energy-dispersive X-ray spectroscopy (SEM–EDS) from polished cross sections of samples is frequently used to assess the inclusion state, such as the liquid fraction, via the oxide composition in the Al_2O_3 –CaO–MgO ternary diagram. Using computational thermodynamics computer software, such as FactSage, liquid fractions in oxide systems at steelmaking temperatures can be readily calculated with arbitrary axes. For instance, Pistorius et al.^[6] presented 100% and 50% liquid areas in Ca/O vs. Mg/O graphs. In addition, Yang et al.^[7] presented 50%, 75%, and 100% liquid areas in Al_2O_3 –CaO–MgO and Al_2O_3 –CaO–CaS diagrams. A similar approach was presented by Verma et al.,^[8] with 50%

liquid areas in Al–Ca–Mg and Al–Ca–S ternary diagrams. In their work, the effect of Mg on the liquid fraction of inclusions with varying Ca/O levels was also illustrated.

In the literature, practical approaches for rating the inclusion composition have also been presented. Story et al.^[9,10] plotted the inclusion data to Al–Ca–S ternary diagrams and linked the composition to process performance indicators of steelmaking such as casting stability and the tendency for nozzle clogging and ladle gate erosion. Bi et al.^[11] investigated nozzle clogging during the continuous casting of stainless steel and noted the frequent occurrence of irregular and clustered inclusions in clogged heats, whereas non-clogged heats expectedly showed mainly spherical inclusions. Blais et al.^[12] investigated the dependence between Ca/S ratios of heats and the average shape factors of inclusions containing (Ca,Mn)S phases. This approach was extended by Lis^[13] using Ca/Al ratios, taking note of oxide inclusions as well. Abdulsalam et al.^[14] utilized principal component analysis and clustering tools for inclusion grouping by chemistry. The researchers successfully differentiated MnS and complex inclusions and presented deviations in their average areas and aspect ratios.

Despite the extensive research into the inclusion compositions and shapes, the correlation between inclusion morphology and liquid fractions has not been widely reported in the literature. Some results were presented by the authors,^[15] with differences between the use of aspect ratio and shape factor thresholds for assessing the inclusion composition in Al₂O₃–CaO–CaS ternary diagrams.

The objective of this study was to develop a fast method for inclusion state classification based on morphological features. The reference data for the classification was built based on elemental analyses and morphological data of inclusions, acquired by SEM-EDS. The thermodynamic calculations to assess the liquid fractions in the Al₂O₃–CaO–MgO(–CaS) system were carried out using the FactSage 7.1 computer software package.

Based on the reference data, binary logistic regression models were trained to classify inclusions found in calcium treated and aluminum killed steel samples based on their morphological properties. The flowchart of the classification procedure is presented in **Figure 1**.

2. Methodology

2.1. Samples and available data

19 heats each of aluminum killed, calcium treated steels were investigated. Liquid steel samples were taken from the ladle after calcium treatment and purging of approximately 5 minutes. The mold samples were taken during continuous casting. The composition ranges of the steel samples concerning the most relevant elements to non-metallic inclusions are presented in **Table 1**.

The samples were analyzed using a Jeol JSM-7000F field emission scanning electron microscope (FESEM) with an acceleration voltage of 15 kV using backscattered electrons imaging mode. The Oxford Inca Feature software was used to analyze the inclusions, recording composition and extracting features for each inclusion detected, as depicted in **Table 2**. The inclusions were analyzed over the whole inclusion area with a live time of 1.0 s using energy-dispersive X-ray spectroscopy (EDS). The analyzed areas in each sample were approximately $6\text{ mm} \times 6\text{ mm}$, totaling 36 mm^2 .

The data was treated so that it covered the inclusions which had the largest diameter above $3\text{ }\mu\text{m}$. The reasoning behind the pre-treatment was firstly related to the presumed accuracy of the elemental analysis, as Pistorius & Patadia illustrated how the EDS analyses of small inclusions might be erroneous because the steel matrix affects the analyzed (Ca-S)/Al ratios.^[16] Secondly, a sufficient number of pixels in imaged inclusions for morphological assessments were ensured by omitting the smallest inclusions from the dataset.

Manganese sulfides and titanium nitrides are mostly formed only during the solidification of the steel sample. Thus, they are not significant in the liquid steel samples studied, and inclusions containing those components were not considered. In order to study only the oxides and oxide–sulfides, inclusions with calculated contents of MnS or TiN exceeding 5 mol% were discarded from the dataset. As a result, a total of 2,041 oxide/oxide–sulfide inclusions in the size range of 3–25 μm were taken into account. The average compositions of the studied inclusions in each heat are presented in **Figure 2a and 2b**.

2.2. Particle shape analysis

Non-metallic inclusions in steel are found in a variety of shapes, heavily depending on their composition. Inclusions that are observed in a spherical shape are considered to have been liquid in the molten steel before the solidification of the steel sample.^[17] That is, roundness is an indication of an inclusion being liquid in molten steel, and that the calcium treatment has been successful. However, Beskow et al.^[18] also noted composition homogeneity as a criterion for an inclusion being liquid.

In particle analysis, a shape can be presented with numerous factors, as depicted by Beddow & Meloy.^[19] In the Inca Feature software analyses, which are widely used in industry, the aspect ratio (AR), and shape factor (SF) are recorded with each detected inclusion, presented in **Equation 1 and 2**. To be exact, the shape factor utilized in Inca Feature is the inverse of a factor known as high sensitivity circularity.^[20] Both the aspect ratio and shape factor are close to unity with spherical inclusions and increase with elongation or an increasing perimeter in complex shaped inclusions, respectively.

$$AR = \frac{L}{B} \quad (1)$$

$$SF = \frac{P^2}{4\pi A} \quad (2)$$

where L is the longest Feret diameter (length), B is the shortest Feret diameter (breadth), P is the perimeter of the inclusion, and A is the observed inclusion area. In **Figure 3**, typical shapes for non-metallic inclusions are presented, with varying aspect ratios and shape factors. The effect of imaging resolution is clearly visible in Figures 3a and 3b with low and high resolutions, respectively. The observed shape factor in Figure 3b is high, partly caused by the voids imaged, which contribute to the reported perimeter.

2.3. Estimation of liquid phase fraction of inclusions by making use of EDS data and computational thermodynamics

The inclusions in the samples are assumed to be a mixture of Al_2O_3 – CaO – MgO oxides, CaS , MnS and TiN components. The component fractions in each inclusion were estimated with **Equation 3**, based on the elemental analyses of each inclusion. As a result of the set of equations, molar amounts of Al_2O_3 , CaO , CaS , MgO , MnS , and TiN were obtained.^[21] Further, the composition of the oxide part of the inclusion in the Al_2O_3 – CaO – MgO system was calculated with the ratios of $n_{\text{Al}_2\text{O}_3}$, n_{CaO} , and n_{MgO} .

$$\begin{bmatrix} 2M_{\text{Al}} & 0 & 0 & 0 & 0 & 0 \\ 0 & M_{\text{Ca}} & M_{\text{Ca}} & 0 & 0 & 0 \\ 0 & 0 & 0 & M_{\text{Mg}} & 0 & 0 \\ 0 & 0 & 0 & 0 & M_{\text{Mn}} & 0 \\ 0 & 0 & M_{\text{S}} & 0 & M_{\text{S}} & 0 \\ 0 & 0 & 0 & 0 & 0 & M_{\text{Ti}} \end{bmatrix} \begin{bmatrix} n_{\text{Al}_2\text{O}_3} \\ n_{\text{CaO}} \\ n_{\text{CaS}} \\ n_{\text{MgO}} \\ n_{\text{MnS}} \\ n_{\text{TiN}} \end{bmatrix} = \begin{bmatrix} m_{\text{Al}} \\ m_{\text{Ca}} \\ m_{\text{Mg}} \\ m_{\text{Mn}} \\ m_{\text{S}} \\ m_{\text{Ti}} \end{bmatrix} \quad (3)$$

where M_i is the molar mass of element or component i, n_i is the molar amount of compound i in the inclusion, and m_i is the mass of element i in the inclusion. The six unknown n_i values are readily solved with the set of six linear equations, with a constraint $n_i \geq 0$. In practice, sulfur is bound with all manganese available, and excess sulfur is bound with calcium. The remaining

calcium, if any, forms oxides along with aluminum and magnesium, whereas titanium acts only as a nitride former.

The construction of the reference data for the $\text{Al}_2\text{O}_3\text{--CaO--MgO}$ system was carried out by making use of a full factorial experimental design matrix, which is considered an optimal experimental design matrix. The properties of the design matrix can be found in the works of Fang et al.^[22] The step size for the experimental design was 0.5 wt%, which consequently resulted as 20,301 data rows, with all the possible composition combinations of the three components. The liquid fraction was calculated for each composition in the $\text{Al}_2\text{O}_3\text{--CaO--MgO}$ experimental design matrix. The computations were carried out for 1823 K, and 1 atm with FactSage 7.1 using FactPS and FToxid databases.

For each $\text{Al}_2\text{O}_3\text{--CaO--MgO--CaS}$ inclusion in the dataset, the calculated oxide composition was rounded to 0.5 wt% accuracy, and the corresponding liquid fraction of the FactSage calculated results was selected for the oxide fraction, taking CaS into account and regarding its fraction as solid. This way, the liquid fraction could be estimated for any composition in the $\text{Al}_2\text{O}_3\text{--CaO--MgO--CaS}$ system without further use of FactSage.

For comparison, the liquid fraction was also calculated with FactSage for each inclusion in the dataset, taking the calculated Al_2O_3 , CaO, MgO and CaS contents as inputs. This allowed the solubility of CaS in the liquid phase, in addition to the presence of pure CaS.

2.4. Classification model

The binary classification (liquid/solid = 1/0) was carried out by making use of a logistic regression model. The logistic regression is a popular choice for classification due to its simplistic formulation and thus a more interpretable model structure. In the logistic regression, the probability of an inclusion of being liquid ($Y = 1$) is given as:^[23]

$$p(Y = 1 | X) = \frac{1}{1 + e^{-(b_0 + \sum_{i=1}^J x_i b_i)}} \quad (4)$$

where Y is the dependent variable, X is the set of independent variables, b_i is a regression coefficient for a corresponding variable i , and b_0 is the bias-term. The inclusion state classes are assigned based on the following conditions:

$$Y = \begin{cases} 1, & p(Y = 1 | X) \geq r \\ 0, & p(Y = 1 | X) < r \end{cases} \quad (5)$$

where r is the decision threshold value. The threshold value is optimized so that it maximizes the precision (Positive Predictive Rate, PPV) and recall (True Positive Rate, TPR) of the classifier. The logistic regression model coefficient answers how much the selected independent variable contributes to the probability of a dependent variable of belonging to class 1. The regression coefficients can be solved by making use of the maximum-likelihood method, in which the given likelihood function is maximized with respect to regression coefficients. The log-likelihood function for the formulated classification problem can be given as:^[23]

$$\arg \max J(\hat{b}, x) = \arg \max \sum_{i=1}^n y_i \ln p(Y = 1 | X) + (1 - y_i) \ln(1 - p(Y = 1 | X)) \quad (6)$$

As the given likelihood function is non-linear in nature, the regression coefficients need to be solved iteratively. In this study, a genetic algorithm was applied for the parameter identification task. Genetic algorithms belong to a family of metaheuristic optimization methods, in which a population of individuals is used to explore the search domain. The genetic algorithm makes use of operations that are familiar from evolution theory, i.e. the survival of the fittest. The basic operations of genetic algorithm are selection, crossover and mutation. In genetic algorithm, there are some parameters of which selection needs to be considered in terms of reliability of

the computational results. The usual considered criteria are for example the accuracy and repeatability of the results. These parameters include number of individuals (n_{pop}), number of generations (n_{gen}), crossover probability (p_c) and mutation probability (p_M). A more detailed description of the method as well as the computational parameters is given for example in refs.^[24,25] The computational parameters that were observed to yield realistic and reliable classification results were $n_{\text{pop}} = 200$, $n_{\text{gen}} = 500$, $p_c = 0.8$ and $p_M = 0.05$. Due to large differences in the domain values of the feature candidates, prior to training the classifier, the classification features were scaled so that $X = \{0, 1\}$. The performance of the classifier was evaluated with some standard figures of merit, which were precision (Positive Predictive Value, PPV), recall (True Positive Rate, TPR), and the F1-score. The definitions of these are given as:^[26]

$$TPR (\%) = \frac{TP}{TP + FN} \cdot 100, \quad (7)$$

$$PPV (\%) = \frac{TP}{TP + FP} \cdot 100, \quad (8)$$

$$F1 - \text{score} (\%) = 2 \frac{TPR(\%) \cdot PPV(\%)}{TPR(\%) + PPV(\%)} \quad (9)$$

where TP is true positives, FP is false positives and FN is false negatives. In the equations, true positives denote inclusions that were correctly identified as liquid. False positives stand for inclusions that were identified as liquid, when they were solid in reality. Similarly, false negatives were inclusions identified as solid, when they were actually liquid.

3. Results and Discussion

3.1. Liquid fraction of inclusions

The ternary contours of the calculated liquid fractions of the Al_2O_3 – CaO – MgO oxide system at 1,823 K are presented in **Figure 4**. The 50% liquid fraction contour is highlighted, showing the preferred composition area for oxides from the point of view of castability.

The liquid fraction was calculated for each inclusion using the reference dataset of calculated liquid fractions in the Al_2O_3 – CaO – MgO system taking CaS as solid, and with FactSage in the Al_2O_3 – CaO – MgO – CaS system, allowing the solubility of CaS in the liquid phase. From a total of 2041 inclusions in the dataset, the liquid fraction was over 50 wt% in 1256 inclusions, denoted as liquid. The remaining 785 inclusions were considered solid. A comparative illustration of the computational results is presented in **Figure 5**. It can be seen that there is only a slight difference between the results, with CaS assumed solid or possibly dissolved in a liquid oxide phase. A larger variance between the calculated results is observed in the higher liquid fractions, whereas in the inclusions with lower liquid fractions the results are consistent. However, this does not have an effect on the classification results, as the liquid fraction limit between the inclusion state classes was set to 50 wt%. The computational result indicates that CaS may be observable as a separate solid phase in the inclusion images, and only a minor amount of CaS is dissolved in the liquid phase. Kim et al.^[27] measured the solubility of CaS in solid CaO – Al_2O_3 oxides and found that sulfur was soluble only in $12\text{CaO}\cdot 7\text{Al}_2\text{O}_3$ at approximately 1.2 mass%, while the sulfur content in the other solid calcium aluminate compounds was very low. The computational results are in good agreement with their experimental results.

As can be seen in **Figure 6**, there is no clear correlation between the observed aspect ratios of inclusions and their liquid fraction. It can be concluded that only using aspect ratio values, liquid fraction of inclusions cannot be evaluated. A direct observation is that inclusions with aspect ratios larger than 4, all contain none liquid phase in this case. On the other hand, using

only these values it is not possible to evaluate the minimum liquid fraction to achieve the sphericity of inclusions. Therefore, for the assessment of inclusions being adequately liquid or not, other features must be taken into account as well.

3.2. Classification of inclusions with logistic regression

In this stage of the analysis, in the feature selection and parameter identification, the inclusion data was split so that 85% was used for training and classifier tuning and 15% was used for testing the final classifier performance. The feature selection for the classifier was carried out using forward selection, so that the selection was continued until the classification accuracy no longer significantly improved. Consequently, the selected feature subset for inclusion classification consists of the sampling location (0/1 = ladle/mold), AR, equivalent circle diameter (ECD) and mean grey value (MGV). The selected feature set as well as the corresponding model parameters are given in **Table 3**, where negative values indicate increasing probability of inclusions being liquid, when the corresponding feature value increases. From the values of the model parameters it can be deduced that when the AR approaches a value of 1, the probability of an inclusion being liquid approaches 1. Under the conditions of this study, a higher ECD value indicates a liquid inclusion, whereas a lighter shade of the inclusion is an indicator of a solid inclusion, caused by the presence of CaS phases. It can also be seen that the inclusions in the ladle samples the inclusions are more likely liquid, also attributed to higher CaS contents in the mold samples.

The tuning of the threshold value r was carried out by making use of the ROC-graph, maximizing the sum of $TNR = 100 - FPR$ and TPR . The chosen threshold value is $r \approx 0.64$. The ROC-graph and the classifier performance as a function of the threshold value r are shown in **Figure 7a and 7b**. The area that is left under the ROC-curve indicates that the classification result is satisfactory, as the classification result is significantly better than the random

classification, which is presented as a diagonal in **Figure 7a**. All in all, the classification can be considered good, regardless of the uncertainties that relate to industrial data.

The confusion matrix that is constructed based on the selected classifier is presented in **Table 4**. In the confusion matrix, the number of each observed class are compared to corresponding predicted ones. From the table it can be seen that the classifier is able to distinguish solid and liquid inclusions reasonably well. This can be attributed to reasonably good precision and recall values for the independent test set, which were 84.9% and 82.7%, respectively. From this, it can be said that the classification result is realistic. A further analysis of the compositional dependence of the classification performance indicates that the logistic regression classifier struggles to classify inclusions for which the liquid phase fraction is well above 0.5. However, a more complex classifier, a support vector machine with a fine Gaussian kernel function (SVM-FGK), did not yield any observable improvements in the classification performance based on the same set of features. The quantitative figures of merit of the SVM-FGK determined for the test set were TPR = 80%, PPV = 80% and F1-score = 80%. The results indicate that improvements in the classification performance could be acquired by clustering the inclusions into smaller subclasses, which would on the other hand require a more detailed feature extraction from the inclusion images. This matter is discussed in more detail in Section 3.3.

3.3. Implications of the classification results

The classification results emphasized the need for a more detailed feature extraction procedure. It can be seen in **Figure 8a** that inclusions with high CaS contents have higher mean grey values on average compared to inclusions containing only oxides. Because of their lower density, calcium sulfides can be observed with a lighter shade in the backscattered electron images used in the inclusion analyses. However, there is no clear trend in the grey shades of the oxide inclusions, as shown in the ternary contours in **Figure 8b**. In fact, there is a marked variance in the grey values of liquid calcium aluminates. Based on the results, the mean grey value can be

used to distinguish calcium sulfides from oxides to some extent, but its applicability to differentiate oxide phases is rather low.

The resolution used for the imaging during the scanning electron microscopy may affect the aspect ratio and shape factor values, especially at low magnifications. In this study, inclusions smaller than 3 μm were omitted for ensuring accurate morphological assessment of inclusions.

With the incomplete modification of alumina or spinel inclusions into liquid calcium aluminates, solid phases may be fully surrounded by liquid oxide phases. In such cases, only the liquid oxide composition affects the observed roundness and the morphological values of the inclusion, whereas the solid phases affect the actual elemental analysis of the whole inclusion. Besides of alumina and spinel, calcium sulfide is another solid phase found in liquid steel samples of aluminum-killed, calcium treated steels. Depending on the formation mechanism of the CaS phase, it can be found in several shapes. Zhao et al.^[28] proposes that CaS can be observed in oval shape if it is formed from the collision of oxide and sulfide inclusions. On the other hand, CaS may take a crescent shape if precipitated on the surface of oxide inclusions. With higher total sulfur contents in steel, CaS may form around an inclusion, observed as a full ring on the cross section.^[29] Therefore, CaS-bearing inclusions in liquid steel samples are often observed in round shapes if the oxide part in inclusions is molten. It is evident that this will lead to false positives in classification.

For comparison, two same-sized inclusions with differing compositions are presented. In the inclusion in **Figure 9a**, the CaS phase forms the majority, accompanied by C3A and C12A7 calcium aluminates. The oxide part is almost fully liquid, with a 94.9% liquid fraction. However, the liquid fraction of the whole inclusion is only 38.4% because of the presence of CaS. In **Figure 9b**, solid MgO phases are observed mainly inside the inclusion, surrounded by calcium aluminate phases close to the CA composition. As the inclusion comprises almost entirely of oxides, the liquid fractions calculated both ways are very close to each other, with 40.4 and

41.2%. Both inclusions are round in shape, and these types are not expected to lead to instability issues during continuous casting. However, solid fractions increasing beyond 60% could lead to nozzle clogging, as reported by Fuhr et al.^[4] Based on these observations, considering the liquid fraction of the whole inclusion is recommended for the classification.

According to the results, there was no large deviation in the calculated liquid fractions of inclusions, whether the calculated CaS content was considered fully solid or possibly solute in a liquid phase. Calcium sulfide was, however, observed in lighter shades in backscattered electron images, compared to oxides. This emphasizes the need for more accurate feature extraction methods – such as morphological data not provided by Inca Feature software or using image analysis methods.

The classification procedure presented here is not limited for use with Inca Feature. Instead, a similar approach can be applied in light optical or laser scanning microscopy, given that morphological data of inclusions is obtained. However, the accuracy of the classifier could be enhanced with other morphological data not obtained using the Inca Feature software, such as the number of corners detected in the inclusions. Also, a more detailed analysis of the pixel intensity histogram could provide potential alternative features for classification. An example of such a feature would be the dimensionless area ratio that is defined as a ratio between lighter and darker areas, calculated based on the pixels, in the inclusion. By ensuring constant SEM imaging conditions between the samples, grey level variations due to the differences in phase densities within inclusions would be observable.

With additional morphological data available, other applications, in addition to the estimation of liquid fraction of inclusions, could be developed. For instance, particle analysis data could be pre-treated, by recognizing scratches, defects, voids, and differentiating true inclusions from the dataset. Fully making use of the morphological data of observed inclusions could potentially

lead to reduced EDS analysis times, without losing accuracy in the inclusion classification into relevant classes.

4. Conclusion

The objective of this work was to study the predictability of the state of inclusions – liquid or solid – based on their appearance and morphological properties. The main findings of this study can be summarized as follows:

1. Under the conditions of this study, it is possible to distinguish between liquid and solid $\text{Al}_2\text{O}_3\text{--CaO--MgO--CaS}$ inclusions based on their aspect ratio, ECD and mean grey value with a recall of 82.7% and precision of 84.9% by making use of a logistic regression based classifier. The classification accuracy given in terms of F1-score was 83.8%. The results suggests that a regular circle shape indicates a liquid inclusion under the conditions of this study. However, it was also observed that a more detailed feature set would produce a more accurate classification result.
2. The training of the classification model needs to be carried out by making use of reference data concerning the inclusion state. A pre-trained classification model can be used for rapid classification of inclusions based on their morphological features. Under the conditions of this study, the proper feature set constituted of the aspect ratio, ECD and mean grey value of the image of the inclusion.
3. The results of this study are applicable in light optical or laser scanning microscopy, which lack the ability to obtain elemental analyses of inclusions. By investigating morphological values, the success of calcium treatment may be determined for individual heats in steel production. The method proposed in this study is straightforward to implement using image analysis software.

Acknowledgements

This research was carried out under the Academy of Finland project no. 311934 (Genome of Steel). SSAB Europe Oy is acknowledged for sharing and allowing the use of data. Authors would like to thank Mrs. Leena Määttä and Mr. Tuomas Antola (SSAB Europe Oy) as well as Mr. Tuomo Saarinen (Sandvik AB) for the discussions on the topic.

Received: ((will be filled in by the editorial staff))

Revised: ((will be filled in by the editorial staff))

Published online: ((will be filled in by the editorial staff))

References

- [1] J. M. A. Geldenhuis, P. C. Pistorius, *Ironmaking Steelmaking* **2000**, 27, 442.
- [2] X. Zhang, L. Zhang, W. Yang, Y. Wang, Y. Liu, Y. Dong, *Metall. Mater. Trans. B* **2017**, 48, 701.
- [3] S. K. Choudhary, A. Ghosh, *ISIJ Int.* **2008**, 48, 1552.
- [4] F. Fuhr, G. Cicutti, G. Walter, G. Torga, *Iron Steelmaker* **2003**, 30, 53.
- [5] P. Kaushik, H. Piolet, H. Yin, *Ironmaking Steelmaking* **2009**, 36, 561.
- [6] P. C. Pistorius, P. Presoly, K. G. Tshilombo, 2006 TMS Fall Extraction and Processing Division: Sohn International Symposium, **2006**.
- [7] W. Yang, C. Guo, C. Li, L. Zhang, *Metall. Mater. Trans. B* **2017**, 48, 2267.
- [8] N. Verma, P. C. Pistorius, R. J. Fruehan, M. S. Potter, H. G. Oltmann, E. B. Pretorius, *Metall. Mater. Trans. B* **2012**, 43, 830.
- [9] S. R. Story, T. J. Piccone, R. J. Fruehan, M. Potter, ISSTech Conf. Proc (2003).
- [10] S. R. Story, Q. E. Liu, C. R. Cathcart, M. Molnar, Proc. 9th Int. Conf. on Clean Steel, Budapest, **2015**.
- [11] Y. Bi, E. Roos, A. Karasev, P. G. Jönsson, Proc. 8th Int. Conf. on Clean Steel, Budapest, **2012**.
- [12] C. Blais, G. L'Espérance, H. LeHuy, C. Forget, *Mater. Charact.* **1997**, 38, 25.
- [13] T. Lis, *Metalurgija* **2009**, 48, 95.
- [14] M. Abdulsalam, T. Zhang, J. Tan, B. A. Webler, *Metall. Mater. Trans. B* **2018**, 49, 1568.
- [15] T. Alatarvas, E.-P. Heikkinen, T. Antola, L. Määttä, T. Fabritius, Proc. 10th Int. Conf. on Clean Steel, Budapest, **2018**.
- [16] P. C. Pistorius, A. Patadia, Proc. 8th Int. Conf. on Clean Steel, Budapest, **2012**.

- [17] S. K. Choudhary, A. Ghosh, *ISIJ Int.* **2009**, *49*, 1819.
- [18] K. Beskow, J. Jia, C. H. P. Lupis, D. Sichen, *Ironmaking Steelmaking* **2002**, *29*, 427.
- [19] J. K. Beddow, T. P. Meloy, *Testing and characterization of powders and fine particles*, Heyden, London **1980**.
- [20] G. E. Christidis (ed.), *Advances in the Characterization of Industrial Minerals*, The European Mineralogical Union and the Mineralogical Society of Great Britain & Ireland, London **2011**.
- [21] T. Alatarvas, *PhD Thesis*, University of Oulu, **2018**,
<http://urn.fi/urn:isbn:9789526220987>.
- [22] K. Fang, R. Li, A. Sudjianto, *Design and modeling for computer experiments*, Chapman & Hall/CRC, Boca Raton **2006**.
- [23] F. E. Harrell Jr, *Regression modeling strategies : with applications to linear models, logistic regression and survival analysis*, Springer, New York **2001**.
- [24] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, Reading, Mass. **1989**.
- [25] K. Deb, *Sadhana* **1999**, *24*, 293.
- [26] T. Fawcett, *Pattern Recogn. Lett.* **2006**, *27*, 861.
- [27] S. Kim, M. Kageyama, X. Gao, S. Ueda, S. Kitamura, *ISIJ Int.* **2019**, Article ID ISIJINT-2019-127, <https://doi.org/10.2355/isijinternational.ISIJINT-2019-127>
- [28] D. Zhao, H. Li, C. Bao, J. Yang, *ISIJ Int.* **2015**, *55*, 2115.
- [29] L. Holappa, 84th Steelmaking Conference, Baltimore, Maryland **2001**.

Figure 1. Flowchart for classification model identification for inclusion state detection.

Figure 2. Average inclusion composition in each heat on Al_2O_3 –CaO–MgO and Al_2O_3 –CaO–CaS ternary diagrams.

Figure 3. Typical inclusion shapes with varying aspect ratio and shape factor values.

Figure 4. Contours of liquid fractions in the Al_2O_3 - CaO - MgO oxide system at 1,823 K, calculated by FactSage 7.1.

Figure 5. Calculated liquid fraction of inclusions with and without the solubility of CaS in liquid phase.

Figure 6. Aspect ratio of inclusions as a function of calculated liquid fraction.

Figure 7. a) ROC-graph, b) TPR, TNR and classification error as a function of the threshold value.

Figure 8. a) Observed mean grey values of inclusions as a function of the CaS content in inclusions, b) Contour plot of the observed mean grey values in the oxide inclusions.

Figure 9. Scanning electron images and outlines of a) CaS+CAx and b) CAx+MgO inclusions. CAx: varying calcium aluminates, M: MgO.

Table 1. Composition ranges of the studied steel samples.

	Aluminum	Calcium	Oxygen	Sulphur	Magnesium
Concentration (ppm)	260 – 430	13 – 23	8 – 17	4 – 21	< 10
Method	OES	OES	LECO	OES	OES

Table 2. Data acquired with Inca Feature runs for each inclusion.

Feature	Elemental analysis	Length	Breadth	Area	Aspect Ratio	Shape factor	Perimeter	Equivalent circle diameter (ECD)	Mean grey value
Unit	wt%	μm	μm	μm ²	-	-	μm	μm	-
Range	0-100	3-25	0.7-25	1.2-498	1-8	0.97-13.4	6.5-204	1.2-26	50-146

Table 3. Selected classification features and their corresponding model parameters.

Feature	Bias	Sampling location	AR	ECD	MGV
Parameter	3.7	-0.8	-23.0	3.4	-3.2

Table 4. Confusion matrix and the quantitative figures of merit for the logistic regression classifier.

Training				Test			
True class	Predicted class			True class	Predicted class		
		Liquid	Solid			Liquid	Solid
	Liquid	850	209		Liquid	163	34
	Solid	143	533		Solid	29	80
	TPR	PPV	F1-score		TPR	PPV	F1-score
	80.3%	85.6%	82.9%		82.7%	84.9%	83.8%

An inclusion classification procedure is proposed, combining computational thermodynamics and data-driven reasoning. The state of inclusions in molten steel is predicted based on their appearance and morphological properties. Al_2O_3 – CaO – MgO – CaS inclusions are classified as liquid and solid ones based on their aspect ratio, equivalent circle diameter and mean grey value by making use of a logistic regression based classifier.

Non-metallic inclusions

T. Alatarvas*, T. Vuolio, E.-P. Heikkinen, Q. Shu, T. Fabritius

Prediction of inclusion state in molten steel by morphology and appearance of inclusions in liquid steel samples