REVISTA SOLDAGEM & INSPEÇÃO

ISSN 1980-6973 (Online) | ISSN 0104-9224 (Print)

ORIGINAL PAPER

Development of Machine Learning Models to Predict the Weld Defect Using Resistance Spot Welding Experimental Data

Santhosh Mathi[1], Pedro Bamberg[1], Alexander Schiebahn[1], Uwe Reisgen[1]

[1] RWTH Aachen University, Aachen, Germany.

How to cite: Mathi S, Bamberg P, Schiebahn A, Reisgen U. Development of machine learning models to predict the weld defect using resistance spot welding experimental data. Soldagem & Inspeção. 2023;28:e2810. https://doi.org/10.1590/0104-9224/SI28.10

Abstract: Advanced materials and automated processes in manufacturing pose a challenge in terms of adaptability. Introduction of 3rd-generation advanced high strength (3rd-gen AHSS) steels aimed for weight reduction in the automotive without compromising its strength and efficient fuel consumption. Nevertheless, welding 3rd-gen AHSS using resistance spot welding (RSW) is often affected by liquid metal embrittlement (LME) or other quality matters. Identifying the process window to control and produce defect-free welds, requires huge experimental work with an enormous time. Therefore, this paper aimed to use machine learning (ML) to identify the process window by computing the relationship between the input parameters and output weld defect categories like 'Splash', 'LME', 'Insufficient nugget size' and 'Good weld'. Classification-based algorithms, K-nearest-neighbour (KNN) and Naive Bayes algorithms were used. Among these, Naive Bayes exhibits better prediction efficiency of 71% and KNN has 63%. Using these models, predicting the incidence of weld spot defects with the fore-mentioned predictability is possible. Therefore, this work supports the industry experts and researchers to study and predict the process window for the welding process to produce defect-free welds and this idea could implement in different manufacturing processes.

Keywords: Machine learning model; Resistance spot welding; AHSS.

1. Introduction

Application of the resistance spot welding (RSW) process is more predominant in the automotive industry in a way that approximately 3000-6000 weld spots were to be made to fabricate a single car [1,2]. RSW is a process of joining two or more metal sheets by passing a focused electric current at a weld spot. A high electrical resistance created due to the concentrated electric current between the metal sheets, creates a weld spot [3]. Nevertheless, the weld defects are more often created during RSW due to the inappropriate welding parameters and the type of weld materials due to the effect of its alloying elements (i.e.: its protective Zn layer) [4]. Especially, advanced high strength steels (AHSS) steels are coated with zinc on its surface for corrosion protection, which infiltrates into grain boundaries at critical process conditions, causes liquid metal embrittlement (LME) cracks. There are several factors which could influence the cause of weld defects like material related parameters (e.g., sheet thickness, steel grade, type of coating), process related parameters (e.g., welding current, welding time, electrode force) and electrode parameters (e.g., electrode diameter, shape). To guarantee a proper welding window, free of defects and discontinuities, a suitable range of welding parameters must be validated. It needs a deep understanding of the theory of the process, properties of materials and mechanisms of the defect formation. Therefore, to overcome this issue, most of the studies developed weld lobe curves for a robust welding process. The weld lobe curve shows the boundaries of certain working range of parameters for defect-free welding [5]. Last few decades, an advance in the artificial intelligence and the machine learning plays an important role in the automation of manufacturing process. Especially, in terms of recording the data of the process and its parameters at each moment generates a huge data from the machines that helps creating a better understand of the respective welding process [6]. Computing the historical data from the machines with the help of machine learning can help obtaining information about the relationship between input and its corresponding outputs [7]. Machine learning (ML) is a subfield of artificial intelligence that helps to predict or identify the useful information from the available data collected from the machine during a certain process. Using a different mathematical algorithm, a machine learning could develop a model using empirical machine data to analyze the pattern of the process. Compared to weld lobe curve, the ML model helps to predict the output of the process for a large range of parameters without utilizing many resources and time in the industry, which is useful for decision-making purposes [8]. However, implementing the machine learning methods in the manufacturing process, poses several challenges in terms of data quality and process complexity.

Received: 21 Mar., 2023. Accepted: 13 Nov., 2023.
E-mails: santhosh.mathi@rwth-aachen.de (SM), bamberg@isf.rwth-aachen.de (PB), schiebahn@isf.rwth-aachen.de (AS), office@isf.rwth-aachen.de (UR)

Therefore, the idea of this paper is to develop a machine learning model that helps to predict the incidence of weld defect/crack using available experimental data. As per the study of Bhattacharya [5], the influence of defect in the RSW process mainly depends on factors like higher electric current, lower electrode force, higher welding time, type of electrode cap, angle of electrode, thickness of the sheet metal and types of surface coating, etc. Especially, some parameters at critical conditions create a crack more often during welding. The focus on highly influential parameters combinations and its boundaries could help predicting the incidence of cracks. However, to identify the combination of parameters which produces defect in the weld metal, a large number of experiments is required. The dataset from a previous related work Mathi et al. [9] containing the most influencing process parameters to produce cracks on 3rd gen AHSS type CR850Y1180T steel, with a range of parametric values to know its boundaries was used in this work. By using the experimental data and the ML algorithms, a model was developed that computes mathematically, an occurrence of the defect which shows the influence of parameters in a crack formation. The behavioral pattern of the occurrence of the defects helps to compute/predict the output for many other different combinations of the parameters and its values. The developed ML models were evaluated by comparing the prediction accuracy. At the end, the model could predict the output category of the weld metal for the corresponding process parameters whether it produced a defect or not. The open-source algorithm developed in this work can support industrial use on selecting properly welding parameters to deliver weld spots with proper quality, process prognosis based on the monitoring of welding parameters during the process and further developments of ML models to increase the predictability quality prognosis of resistance spot welding processes.

2. Methodology

As mentioned, the resistance spot welding dataset from Mathi et al. [9]. was used and it contains the input columns with the different combination of varied process parameters and its corresponding weld output (i.e., defect or defect-free weld). As per the literature study [10-12], the most influencing process parameters for defect formation like welding current, welding time and electrode force were considered as an input. Welding defects such as LME, Splash and Insufficient nugget diameter were considered as an output. Totally, a list of 202 datapoints with its input and output were listed and the welded samples were categorized based on the type of defect as an output in which the data set was obtained. A dataset with its input and output was imported with the help of python libraries using web-based programming language ‘jupyter notebook’. Since the output in our data is categorical, classification-based supervised learning algorithms such as ‘K nearest neighbour’ and ‘Naive Bayes were used and implemented due to the suitability of multi class classification [13]. Initially the data was preprocessed to improve the data quality and then splits 80% of data as a training dataset and 20% of data as a testing dataset [14]. Then, the ‘training data’ was trained using the algorithms by mathematical computations to identify the relationship between the input parameters (weld current and weld time, electrode force and temperature) and an output category (LME, good weld (i.e., without any defect), splash and insufficient nugget diameter). The trained model was evaluated by assessing the prediction accuracy using data allotted for testing (i.e., testing data). Finally, the models are compared, and a better model was identified based on the higher prediction accuracy score. Figure 1 shows the steps to develop a machine learning model to predict the defects for the type of dataset used in this study.

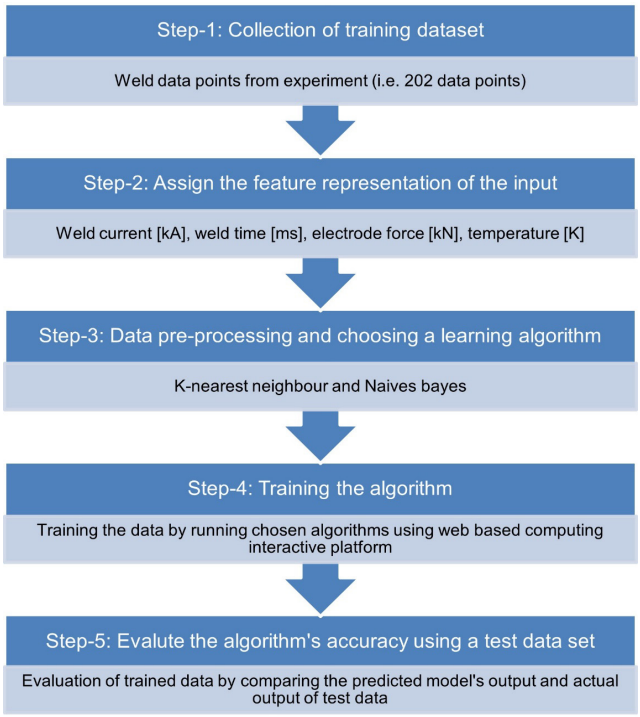

Figure 1. Steps in developing a machine learning model.

3. Implementation of the Machine Learning Model

3.1. Collection of the dataset

A dataset with a list of 202 data points from the experimental work of Mathi et al. [9], was imported and saved in an excel (XLSX) format. The welding experiments were performed using a C-type welding gun with 1000Hz mid-frequency direct current transformer (MFDC). To avoid overheating, water-cooled electrodes were attached with a flow of 4 liters/minute and the electrode type F1 (dome type) with size of 16×6 mm was used in the experimental work. A 45×45×1.2 mm zinc coated CR850Y1180T type 3rd gen AHSS sheet along with 45×45×0.8 mm CR4 steel as a counterpart were used in the experimental work. A single central part weld without shunt was made to join the above mentioned two sheets with an overlapping of 40×40 mm. As per the literature study of [5], the more sensitive parameters and its values for weld defects were focused. Then based on the full factorial design of experiments, the number of experiments were identified by different combination of three parameters by varying its values. To avoid experimental error, three samples were welded for each parameter and its values. Welding current, welding time and the electrode force are the chosen parameters with its values varied in each sample to understand the influence of the defect formation. The sample data from the dataset that was used to develop a model has shown in the Table 1. The data contains a multiple output category (i.e., more than two types of output).

Table 1. A sample data from excel format used for model development with input and output variables.

Welding Current [A]	Welding Time [seconds]	Electrode Force [kN]	Output
7800	0.1	2.0	Spatter
7800	0.2	2.0	Spatter
7800	0.3	2.0	LME
7800	0.4	2.0	LME
7800	0.5	2.0	LME

3.2. Data pre-processing

The data collected from the experimental work cannot be used directly to develop a ML model. The data might contain missing rows, outliers (i.e., irrelevant/wrong output due to experimental errors) and that leads to misinterpretation of information from it. Therefore, the data pre-processing is used to convert raw data into understandable labelled format [15]. Basic python libraries like NumPy, pandas and matplotlib were used for data pre-processing in the 'Jupyter Notebook' platform. The pandas library was used to import the dataset that is saved in the computer. Once the data was imported, the training of a model has started by assigning an independent and dependent variable from the dataset [16]. From the reference of Table 1, the first three columns like welding current, welding time and the electrode force are assigned as input/independent variables (x) in this study. The output column represents the dependent variable (y) which is the outcome of the identified defect category. Since, the dataset contains categorical output, encoding (i.e., converting variables into numeric form for machine readable format) is not needed. For the dataset, 80% of the rows were splatted and assigned as a training data and 20% of the data was assigned as a testing data. The model can mis-predict the output of the category, due to different scales of values in the dataset. Therefore, the standardization method is used to scale a data to fit in a standard normal distribution [17].

3.3. Training and evaluation

Once data has been pre-processed to establish a connection between input and output parameters, it can undergo training. The data was trained using two distinct supervised learning algorithms: 'K-nearest neighbour' and 'Naive Bayes.' This was done to compare the results and mitigate the risk of incorrect predictions caused by coding errors. The selection of these two algorithms for implementation in the project was based on their superior suitability for multi-class classification compared to other available algorithms [13]. KNN method does not rely on underlying data distribution. Therefore, it is beneficial while dealing with categorical data that might not follow a particular statistical distribution [18]. Similarly, naive bayes method is conditionally independent to the given parameters where the presence or absence of one category doesn't belongs to the other category. It treats each category as a separate feature and thus these methods are computationally efficient for categorical data [19]. Firstly, the data was trained by running the algorithm. The algorithm consists of mathematical equations (see Figure 2 that compute and correlate the inputs and outputs of the given data. Then, it develops automatically a model based on the trained data by understanding its working pattern. Subsequently, the model predicts the output of test data based on working pattern of the process by considering its input parameters alone. Then, the algorithm evaluates the model by comparing the predicted output with the actual output of test data. Finally, based on the percentage of correct predictions, the accuracy score of the model gets generated [20].

3.3.1. K Nearest Neighbour (KNN) model

The flow chart for training and validation of the dataset using the KNN algorithm is shown in Figure 2. KNN algorithm trains the data by calculating the Euclidean distance iteratively between the new input data point and the nearest neighbours (K) data points [21]. In our case, 20% of the data, which means 40 number of data points are the total neighbours (K). The primary task is to determine the number of neighbour data points (i.e., Optimal K value) to be considered. The optimal K value is identified based on the model's prediction accuracy or error rate. The model is assigned with different K values starting from 1 to 40, and the corresponding accuracy or error rate is calculated. Finally, based on the lowest error rate, the corresponding K value is identified as an optimal K. Steps to determine the optimal K value are shown in the flowchart (see Figure 3). The error rate curve for K values 1 to 40 is drawn as shown in Figure 4.

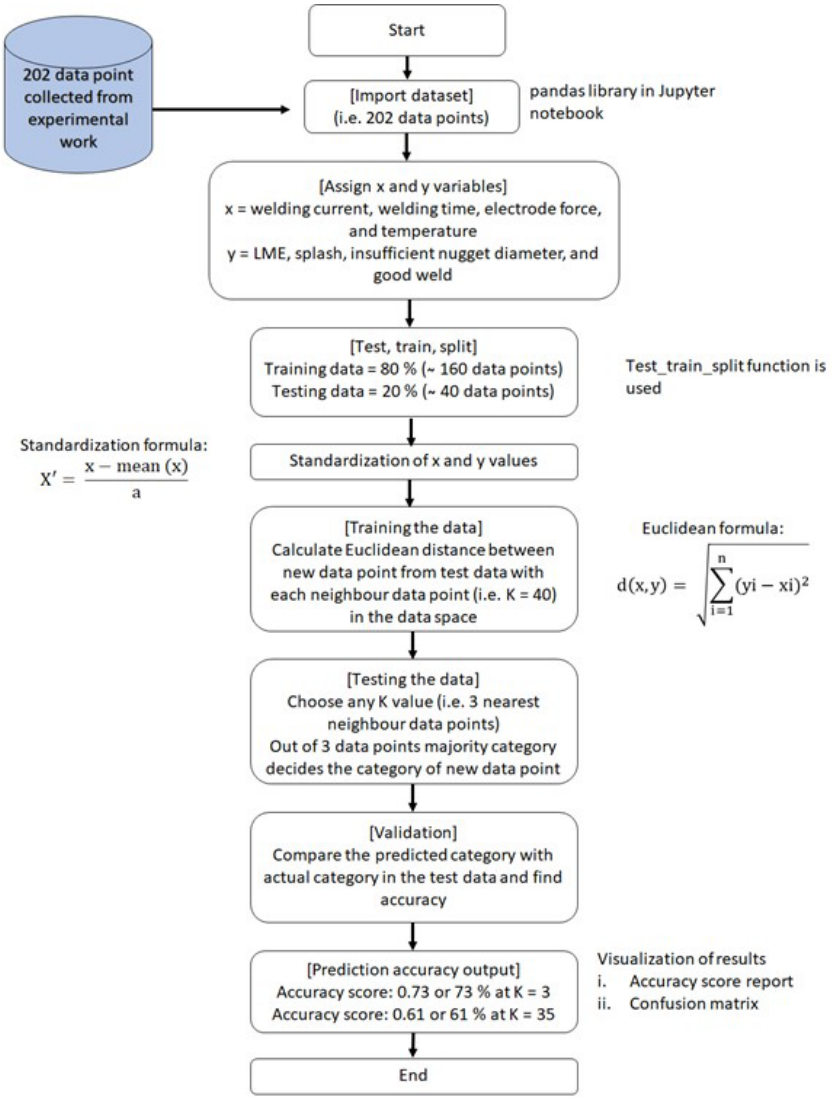

Figure 2. Flow chart for implementing KNN model.

According to the generated curve, the K value of 33 and 35 has the lowest error rate, which means considering 33 or 35 nearest neighbours has the highest prediction accuracy. Therefore, in this work, K=35 is considered as an optimal K value, and the model was developed.

Figure 5 shows an overview of the KNN model generated. Figure 5A represents the data points collected from the experimental work with three categories (i.e., LME, Splash and Good weld). Where x1 and x2 are the independent variables in X and Y axis respectively, let us say welding current and welding time in our case. The red point is the new data point whose category is unknown and to be determined. The Euclidean distance between the red point and all other data points are calculated which is shown in the Figure 5B. Based on the Euclidean distance, three blue points are the nearest neighbours with the majority. Therefore, the red colored data point is considered to be belong to the blue category.

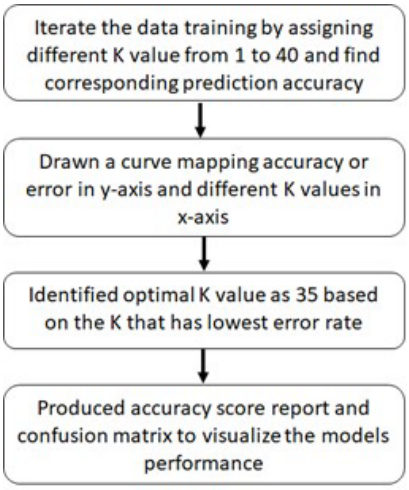

Figure 3. Steps used to determine optimal 'K' value.

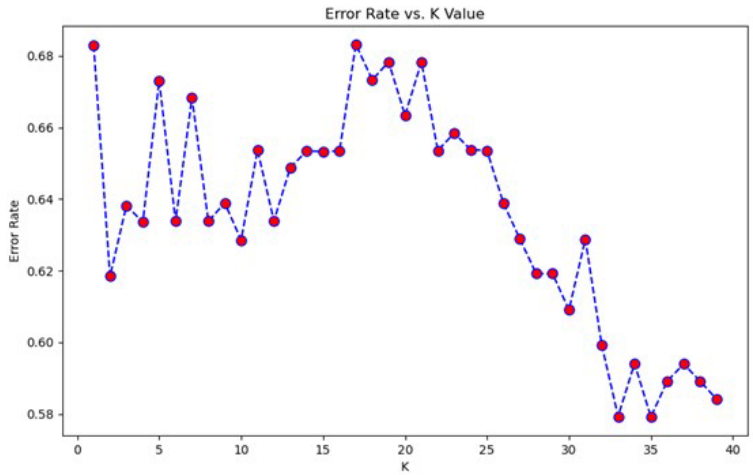

Figure 4. Error rate of prediction for different 'K' values.

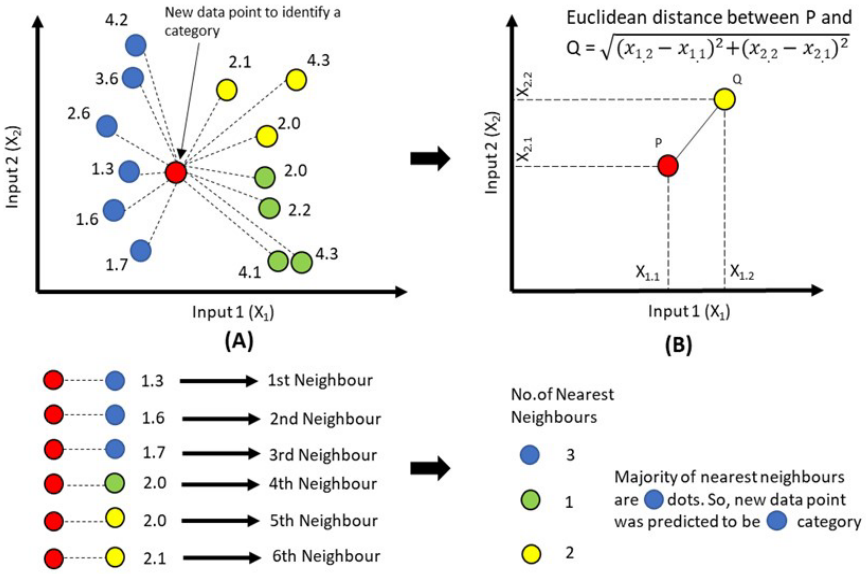

Figure 5. (A) KNN prediction of new data point using the plotted 202 experimental data points using Euclidean distance formula; (B) Euclidean distance calculation and its formula [20].

The predicted output is compared with the actual output and visualized by the confusion matrix shown in the Figure 6. The confusion matrix contains 'Actual label' on the y-axis and 'Predicted label' on the x-axis. The actual label is the output category obtained from the experimental work and the predicted label is obtained from the machine learning model. The diagonal (from top left to bottom right) of the matrix shows the number of data points predicted correctly and the other boxes show the wrong predictions. The dark-coloured box shows a higher number of samples belongs to the corresponding box and the light-coloured box denotes a smaller number of the output in the corresponding category.

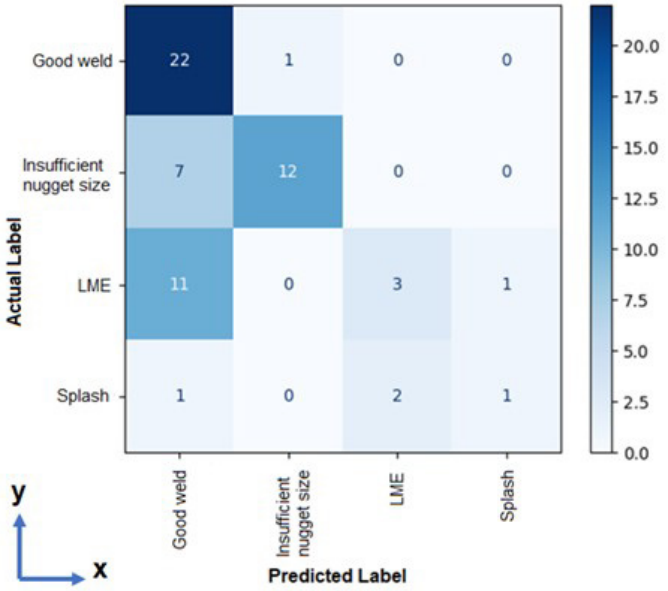

Figure 6. Confusion matrix of 'K Nearest Neighbour' trained model.

Direction (x+, y+): Starting from the left bottom cell (x-Good weld and y-Splash), shows number 1. It means the model predicted 1 'Splash' category sample as a 'Good weld' sample which is an incorrect prediction. The remaining cells in the diagonal have the value of '0', which means the model doesn't predict any category in the corresponding diagonal between the x and y-axis. These diagonal cell values represent the number of samples that are predicted wrongly.

Direction (x-, y+): The left topmost corner cell (x-Good weld and y-Good weld) shows a value of 22, which means the model predicted 22 samples as 'Good weld' that is actually belongs to 'Good weld' sample category. Similarly, all other cells denote the same category in x (Predicted label) and y (actual label) axis shows correct prediction. Simply, the values in the diagonal cell denote the number of samples predicted correctly. The dark shaded in the cell indicates higher values, and the white cell indicates zero value. The higher or zero value denotes the number of samples considered in that particular cell.

Horizontal (only x+): The dataset contains 23 samples of the 'Good weld' category which lies in top horizontal boxes. In that, 22 samples predicted correctly and only 1 sample is predicted wrongly as 'Insufficient nugget size'. Likewise, out of 19 samples, 12 samples are predicted correctly that actually belongs to the 'Insufficient nugget size' category. In the LME category, out of 15 samples, only '3' samples are predicted correct, which shows the worst accuracy. The model doesn't predict the LME samples efficiently. Similarly, out of 4 samples, only '1' is predicted correctly in the 'Splash' category.

Based on the number of correctly and wrongly predicted samples, the accuracy will be calculated. The prediction report is shown in the Table 2 with precision, recall, f1 score and support which is crucial in performance evaluation, decision threshold tuning, interpretability, and quality of model. Precision defines the accuracy of positive predictions made by the model (i.e. the ratio of correctly predicted positive instances to the total positive predictions). Recall quantifies the model's ability to correctly identify all relevant instances in the dataset (i.e. the ratio of correctly predicted positive instances to the total actual positive instances). The 'f1' score is a harmonic mean of precision and recall. Support is the number of actual occurrences of the category in the given dataset. It doesn't affect the performance of the models but instead diagnoses the evaluation process [18]. In other words, it could be defined as:

$$Precision = (True\ Positives) / (True\ Positives + False\ Positives)$$

$$Recall = (True\ Positives) / (True\ Positives + False\ Negatives)$$

$$F1-Score = 2 * (Precision * Recall) / (Precision + Recall)$$

For instance, 71% of the 'Good weld' is predicted correctly from the actual output in test data. Similarly, the model's performance is shown in the 'f1 score' column and the last row shows the overall accuracy of the model including the score of all the categories is 63%.

Table 2. Prediction accuracy report for K = 35.

Category	Precision	Recall	f1-score	support
Good weld	0.57	0.94	0.71	17
Insufficient nugget size	0.86	0.67	0.75	9
LME	0.60	0.38	0.46	8
Splash	1.00	0.14	0.25	7
Accuracy			**0.63**	**41.00**

3.3.2. Naive Bayes model

The naive bayes supervised learning methods uses a statistical method for classification. It considers independent assumptions between the features and calculates the probability of the outcomes between each feature. This model mainly helps to solve predictive problems based on probability calculation [22]. The probability of forming LME, splash, insufficient nugget diameter and good weld category concerning each data point contains features like welding current, welding time, electrode force and temperature are calculated using the conditional probability formula. For instance, the probability of forming LME with relation to welding current should be calculated for all the test data points. Similarly, the probability is calculated for all other features [23]. Then, the new data point with a category with the highest probability value is considered its category. Finally, the predicted category of new test data is compared with the actual category from the experimental observation. Based on the correct prediction, the accuracy score was calculated. The flow chart for the implementation of naive bayes is shown in the Figure 7. The naive bayes model has a 71% accuracy score, as shown in the classification report in Table 3.

Table 3. Prediction accuracy report of naive bayes algorithm.

Category	Precision	Recall	f1-score	support
Good weld	0.63	0.80	0.71	15
Insufficient nugget size	0.83	0.83	0.83	12
LME	1.00	0.50	0.67	8
Splash	0.50	0.50	0.50	6
Accuracy			**0.71**	**41.00**

The model's prediction accuracy is visualized by generating a confusion matrix shown in Figure 8. The confusion matrix and its prediction performance were explained through its cell direction.

Direction (x+, y+): Bottommost left corner cell indicates the model has predicted the 'Splash' category sample as a 'Good weld' sample which was a wrong prediction. The remaining cell in the diagonal has the value of zero, which means there is no other wrong prediction of category in the diagonal cells.

Direction (x-, y+): The topmost left corner cell denotes 21 'Good weld' category samples are predicted as 'Good weld', which was a correct prediction. Ten sample that belongs to the 'Insufficient nugget size' category is also predicted correctly. Similarly, 9 samples in the LME category and 4 samples in the 'Splash' category are predicted correctly.

Horizontal (only x+): Top horizontal row cells show the number of actual 'Good weld' samples, and it contains a total of 22 samples. Out of 22 samples, only 1 sample was mis-predicted, which shows good efficiency in overall. Similarly, out of 11 samples, the model predicted 10 samples from 'Insufficient nugget size' samples correctly. Compared to the KNN model, the Naive Bayes model predicted 'LME' category samples more efficiently. Out of 12 actual 'LME' samples, 9 samples are predicted correctly.

Figure 9 shows a comparison between the efficiency of the developed models in terms of prediction accuracy using a bar chart. The naive bayes model had predicted more accurately compared to the KNN model for the given dataset. Both the model predicted the 'insufficient nugget diameter' category more accurately and that shows the obvious process parameters boundaries. The difference in prediction accuracy between both the models is about 8%. The Naive Bayes model significantly predicted the 'LME' category sample more efficiently with a higher accuracy score.

On the other hand, the category 'splash' shows less accuracy in both models. It shows that the model cannot predict the 'splash' category more accurately based on the trained data and that means the data has no pattern in the splash formation or maybe insufficient training data which has the category of the splash.

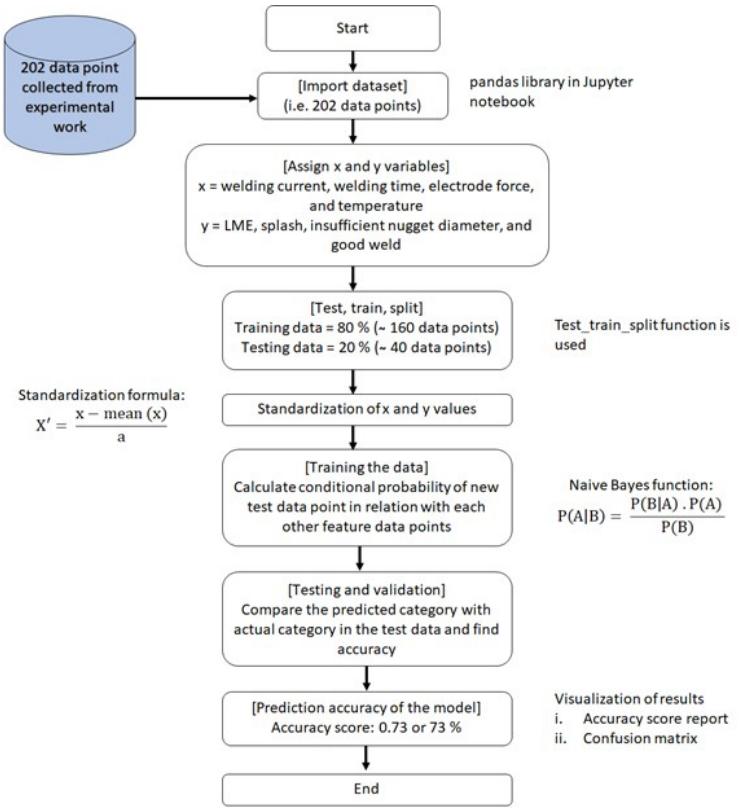

Figure 7. Implementation of Naive Bayes model.

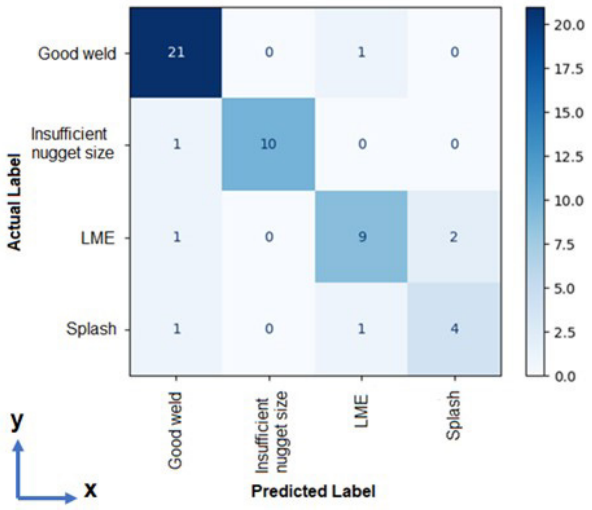

Figure 8. Confusion matrix of Naive bayes model.

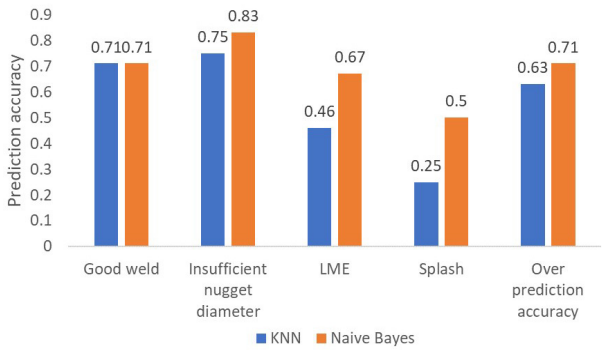

Figure 9. Comparison bar chart between KNN and Naive bayes model developed for LME crack prediction.

4. Conclusion

Deploying a newly developed or advanced materials for light weight vehicle manufacturing in the industry requires a lot of resources and time to achieve a defect-free fabrication. In our case, the data collected from the experimental work in Mathi et al. [9] is limited. It requires material preparation, welding a sample with different parameters, microscopic analysis and chisel test for nugget measurement which consumes lot of time and energy. Moreover, finding the phenomenon of LME formation in relation between material parameters and its range of values with the available data may not be so accurate. In comparison with the result of Mathi et al. [9], the resulting weld lobe contains a shorter boundary of values in a process window where defect-free welds could be produced. Whereas, using machine learning models, a large range of values can be tested and identified with higher accuracy. In this work, using the machine learning, LME cracks are predicted with 67% accuracy and overall prediction accuracy with 71%, using the available data. Therefore, using a machine learning model to predict the LME crack for new materials could be advantageous and less effort in the industrial applications.

The collection of data and the way of handling it majorly influences the model's accuracy score. The generation of the machine learning model in this project work is a strategical and faster approach that helps to adapt a new materials like 3rd generation AHSS for the RSW process by predict/detect LME crack issues in the fabrication industries. Moreover, the approach doesn't need long term research and many resources. This project work approach gives a guidance for an industrial specialist to develop a machine learning model adapting new materials like 3rd generation group AHSS with RSW process in their industries. Moreover, it could be a better reference for the researchers who works in the formation of the LME on 3rd generation AHSS steels and RSW process.

5. Outlook

Developing a machine learning model to predict LME crack within the available data will be a solution for at least the same generation of steel as done in this project work. Otherwise, a considerable amount of research should be required for LME crack prediction for every recently developed material. However, a more generalized model is required to adapt a large group of materials in the RSW process without LME crack issues.

A theoretical understanding of the phenomenon of LME crack formation under various constraints should be studied. Even to develop a machine learning model, data collection plays a significant role which helps to develop a more generalized model. In the study of the reference [9], the measurement of transition resistance induced during welding is less informative. It decides the heat generation in the weld spot and directly affects LME formation during the welding. The proper setup for accurate measurement of transition resistance needs to be developed. It helps to understand the resistance induced or heat generation or temperature in the weld zone much precisely. The relationship between all the parameters which influences the LME crack should be identified. By doing this, a general theory for the formation of LME crack could be developed possibly. This approach could help to produce a generalized model that could predict LME crack for larger group of metals in RSW process. Once the data is more sensible and contains all influencing input parameters with its accurate output, a generalized model could be developed with higher accuracy to predict LME cracks using machine learning.

Moreover, many other material factors like the material's microstructure, tensile strength, type of electrode cap, coating type, the different material thickness should also be included in the experimental data collection to categorize the output. Then, the large amount of data is needed to generate a more precise model as well. Therefore, more experimental work should be carried out by varying an extensive range of parameters and their values to generate enough data. Alternatively, simulations are an appropriate tool that could help to produce a large amount of data based on the theoretical approach to generate a precise and more generalized machine learning model to predict LME crack.

Authors' contributions

SM: literature study of the topic to achieve the aim of the project; firsthand implementation of project with the different ideas; collection of data from the machine; statistical data analysis and programming; report writing. PB: guiding the complete project until the end of the result with ideas; aligning and issuing the resource for the project, closely following the implementation of the project; proofreading the complete work of the report. AS: identified the scope of the project for industrial benefits; mentoring the project; reviewing the ideas and achievements of the project. UR: allotting the budget and resources of the project; proofreading the summary of the project.

Acknowledgements

This work was supported by the Institute of Welding and Joining Technology (ISF) at RWTH Aachen University in Germany with providing experimental and computational facilities.

References

[1] Dhawale P, Kulkarni M. Electric resistance spot welding: a state of art. International Journal of Engineering Research & Technology. 2017;5(1):1-6.

[2] Zhang X, Chen G, Zhang Y, Lai X. Improvement of resistance spot weldability for dual-phase (DP600) steels using servo gun. Journal of Materials Processing Technology. 2009;209(5):2671-2675. http://dx.doi.org/10.1016/j.jmatprotec.2008.06.008.

[3] British Standards Institution. Welding and allied processes: definitions of metal welding processes. London: BSI; 2004.

[4] Kong J-P, Kang C-Y. Effect of alloying elements on expulsion in electric resistance spot welding of advanced high strength steels. Science and Technology of Welding and Joining. 2016;21(1):32-42. http://dx.doi.org/10.1179/1362171815Y.0000000057.

[5] Bhattacharya D. Liquid metal embrittlement during resistance spot welding of Zn-coated high-strength steels. Materials Science and Technology. 2018;34(15):1809-1829. http://dx.doi.org/10.1080/02670836.2018.1461595.

[6] Zhou B, Svetashova Y, Byeon S, Pychynski T, Mikut R, Kharlamov E. Predicting quality of automated welding with machine learning and semantics: a Bosch case study. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management; 2020; Ireland. New York: Association for Computing Machinery; 2020. http://dx.doi.org/10.1145/3340531.3412737.

[7] Wuest T, Weimer D, Irgens C, Thoben K-D. Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research. 2016;4(1):23-45. http://dx.doi.org/10.1080/21693277.2016.1192517.

[8] Hu F, Hao Q, editors. Intelligent sensor networks. New York: Auerbach Publications; 2008.

[9] Mathi S, Bamberg P, Schiebahn A, Reisgen U. Estimation of the incidence of liquid metal embrittlement crack in 3rd generation advanced high strength steel during resistance spot welding using 2D and 3D weld lobe curve. International Journal of Engineering Research & Technology. 2022;11(3):345-354.

[10] Emre HE, Kacar R. Development of weld lobe for resistance spot-welded TRIP800 steel and evaluation of fracture mode of its weldment. International Journal of Advanced Manufacturing Technology. 2016;83(9-12):1737-1747. http://dx.doi.org/10.1007/s00170-015-7605-1.

[11] Jung G. Liquid metal embrittlement of high Mn TWIP steel. Pohang: Graduate Institute of Ferrous Technology, POSTECH; 2015.

[12] Sierlinger R, Gruber M. A cracking good story about liquid metal embrittlement during spot welding of advanced high strength steels. Linz, Austria: voestalpine Stahl GmbH; 2016. p. 1-6. White Paper.

[13] Mehra N, Gupta S. Survey on multiclass classification methods. International Journal of Computer Science and Information Technologies. 2013;4(4):572-576.

[14] Lipovetsky S. Pareto 80/20 law: derivation via random partitioning. International Journal of Mathematical Education in Science and Technology. 2009;40(2):271-277. http://dx.doi.org/10.1080/00207390802213609.

[15] García S, Ramirez-Gallego S, Luengo J, Benitez JM, Herrera F. Big data preprocessing: methods and prospects. Big Data Analytics. 2016;1(1):9. http://dx.doi.org/10.1186/s41044-016-0014-0.

[16] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. Journal of Machine Learning Research. 2011;12:2825-2830.

[17] Brownlee J. Data preparation for machine learning, v1.1 Hrsg. San Juan: Machine Learning Mastery; 2022.

[18] Hastie T, Tibshirani R, Friedman J. The elements of stastical learning, learning: data mining, inference, and prediction. New York: Springer; 2009. http://dx.doi.org/10.1007/978-0-387-84858-7.

[19] Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. Cambridge: Cambridge University Press; 2008. http://dx.doi.org/10.1017/CBO9780511809071.

[20] Wang S, Chen S. Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. Journal of Petroleum Science Engineering. 2019;174:682-695. http://dx.doi.org/10.1016/j.petrol.2018.11.076.

[21] Learned-Miller EG. Introduction to supervised learning. Amherst: Department of Computer Science, University of Massachusetts; 2014. p. 3.

[22] Nasteski V. An overview of the supervised machine learning methods. Horizons B. 2017;4:51-62. http://dx.doi.org/10.20544/HORIZONS.B.04.1.17.P05.

[23] Rish I. An empirical study of the naive Bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence; 2001; Seattle, Washington, USA. Washington, DC: American Association for Artificial Intelligence; 2001. p. 41-46.