Proceedings Paper:
Rubio-Solis, A. and Panoutsos, G. (2017) An ensemble data-driven fuzzy network for laser welding quality prediction. In: 2017 IEEE International Conference on Fuzzy Systems. 2017 IEEE International Conference on Fuzzy Systems , 09-12 Jul 2017, Naples, Italy. IEEE . ISBN 9781509060344

https://doi.org/10.1109/FUZZ-IEEE.2017.8015496

An Ensemble Data-Driven Fuzzy Network for Laser Welding Quality Prediction

Adrian Rubio-Solis and George Panoutsos
Automatic Control and Systems Engineering Department (ACSE), The University of Sheffield
Email: ; {a.rubiosolis,g.panoutsos}@sheffield.ac.uk

***Abstract*—This paper presents an Ensemble Data-Driven Fuzzy Network (EDDFN) for laser welding quality prediction that is composed of a number of strategically selected Data-Driven Fuzzy Models (DDFMs). Each model is trained by an Adaptive Negative Correlation Learning approach (ANCL). A monitoring system provides quality-relevant information of the laser beam spectrum and the geometry of the melt pool. This information is used by the proposed ensemble model to asist in the prediction of the welding quality. Each DDFM is based on three conceptual components, i.e. a selection procedure of the most representative welding information, a granular comprehesion process of data and the construction of a fuzzy reasoning mechanism as a series of Radial Basis Function Neural Networks (RBF-NNs). The proposed model aims at providing a fuzzy reasoning engine that is able to preserve a good balance between transparency and accuracy while improving its prediction properties. We apply the EDDFN to a real case study in manufacturing industry for the prediction of welding quality. The corresponding results confirm that the EDDFN provides better prediction properties compared to a single DDFM with an overal prediction performance** $> 78\%$.

***Keywords*—Data-Driven Fuzzy Models (DDFMs), RBF Neural Networks, Granular Computing, ensemble networks, Adaptive Negative Correlation Learning (ANCL).**

I. Introduction

VIA the design and implementation of data-driven fuzzy methods for complex systems modelling usually one can gain a deeper insight of the system. This is translated into a better understanding of the process dynamics via the construction of a transparent and interpretable fuzzy reasoning mechanism and kowlegde representation as a series of linguistic rules [1].

In approximate reasoning, it is also known the combination of a number of different predictors usually improves results in prediction [2]. Particularly, an ensemble network groups the ability of individual learners to improve its generalisation properties in classification and then in prediction when single neural networks frequently disagree. In that sense, a number of different learning methodologies have been reported to construct an ensemble network by encouraging its individual members to learn different parts of a data set and then combine them to generalise better [3-7].

In automotive industry, especially those processes that involve the joint of multiple pieces by laser welding, the understanding of the dynamics of the melt pool during the application of the laser beam has recently gained a lot of attention. This is due to the advantages that it offers over traditional welding techniques such as resistance spot welding. Improvements by using laser welding results mainly from the laser spot size, the penetration depth of the weld into the material and its flexibility to be applied on complicated geometries at low thermal distortion and hight speed for free contact assembling [8]. As for any other welding technique, imperfections may occur as a consequence of the small features of the laser welding. Usually the laser welding's quality relies on a number of parameters associated to laser beam such as laser power, beam size and shape on the workpiece, beam divergence and wavelength of the laser beam [8]. As a consequence, advanced monitoring systems for laser welding applications have resulted to mantain a high product quality consistency, particularly to control the melt pool geometry and then the microstructure and surface properties of the material.

In order to assist the monitoring system and gain a better understanding of the laser welding process, in this paper an ensemble network that aggregates the ability of single DDFM's is proposed. Each DDFM consists of a systematic construction of a neural fuzzy inference that is based on the Radial Basis Function Neural Networks (RBFNNs), Granular Computing [9] and a Fast Correlation-Based filter [10] for the selection of the most relevant features of the welding process. With the objective to speed up the network learning and to find the optimal performance we combine an Adaptive Back Error Propagation approach (ABEP) and the Negative Correlation Learning that we call Adaptive Negative Correlation Learning (ANCL) [7].

The rest of this article is organized as follows: Section II provides an overview of the laser welding process and the monitoring system. Section III, describes the ensemble data-driven fuzzy model while section IV shows a comparison of the prediction results obtained by the proposed model and a single DDFM. Finally, in section V the conclusions and recommendations for future work are drawn.

II. Overview of the Monitoring System for the Laser Welding Process

A based-industry laser welding process is considered in this article. Such a process consists of a monitoring system combined with an offline Non-Destructive Test (NDT) which is used to verify the associated weld quality. The monitoring process involves the integration of two subsystems for monitoring the laser power signal spectrum and the geometry of the melt pool respectively.

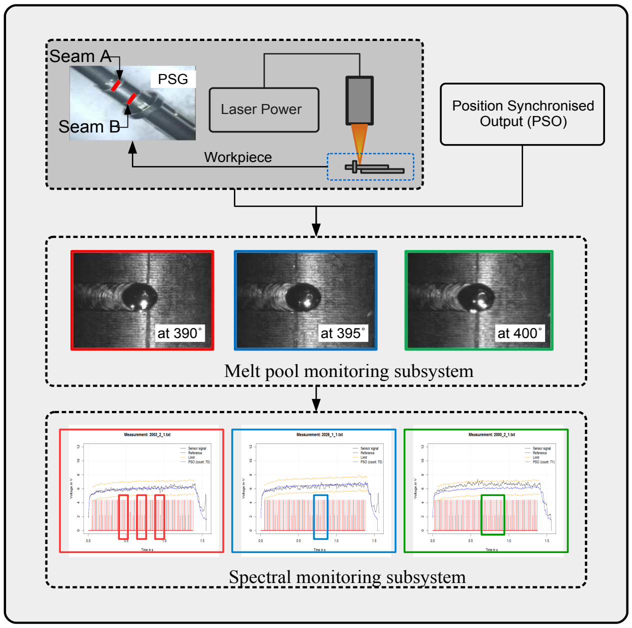

Fig. 1. Monitoring System for laser welding (Image provided by HIDRIA).

On the one hand, the spectral monitoring subsystem conducts a high-speed, high-resolution processing search for spectral isolation of the laser power signal. To properly perform the first monitoring task, an axis unit is employed to rotate the workpiece and evaluate the performance of the laser welding process at different angles. On the other hand, the melt pool monitoring subsystem is an illumination laser technology that employs a high-speed camera that provides an online quality-relevant information of the seam position. That is, with the objective of correlating the output from the spectral monitor and the NDT, at each $5°$ a Position Synchronised Output (PSO) from the axis unit is obtained triggering the high-speed camera while the PSO signal is recorded at each seam position by a spectral sensor. In other words, the high-speed images and the signal spectrum are synchronised and then link to each seam position during the welding process. Finally, an offline leaking NDT was carried out in order to verify the laser weld quality of each manufactured workpiece (PSG). As indicated in Fig. 1 every welding process involves two separate seams (”A” and ”B”) along the surface of the workpiece.

A. Overview of the Monitoring System Data

81 experiments per each seam (”A” and ”B”) with a different focal length of the welding head ranging from 100 mm up to 200 mm, an initial activation laser power of 100 W with a constant welding speed between one seam and another were carried out. A high-speed camera is used to provide 69 frames per experiments in relation to the geometry of the melt pool where 9 signals were extracted. Several metrics were calculated for the monitoring and NDT data. Table I summarises the signal sets and the associated estimate metrics. Finally, to identify the welding quality per PSG workpiece, an offline leaking NDT was effectuated to verify the corresponding welding quality of each seam (”A” or ”B”).

TABLE I
SIGNALS EXTRACTED BY MS.

Melt Pool Monitoring Subsystem (Per frame)
1. Average of the weld width (mean-W)
2. Standard deviation of the weld width (stddev-W)
3. % coeficient of variation of the weldth width (Varco-W)
4. Average of the upper pixel position (Mean-U)
5. Standard deviation of the upper pixel position (Stddev-U)
6. % coefficient of variation of the upper pixel position (varco-U)
7. Average of the lower pixel position (mean-L)
8. Standard deviation of the lower pixel position (stddev-L)
9. % coefficient of variation of the lower pixel position (varco-L)
Spectral Monitoring Subsystem
10. Sensor Signal
11. Reference
12. Position Synchronised Output (PSO)
13. Signal too low/high
Leaking Non Destructive Test (NDT), PSG total fault
Seam A: 49 Ok samples/32 Non-Ok samples
Seam B: 49 Ok samples/32 Non-Ok samples

III. ENSEMBLE DATA-DRIVEN FUZZY NETWORK

This section describes the proposed Ensemble Data-Driven Fuzzy Network illustrated in Fig. 2. According to Fig. 4, a systematic parameter identification for each DDFM follows three major conceptual components: a selection process of the most relevant features, a granulation step that groups similar data and a reasoning mechanism based on the RBF Neural Network (RBF-NN). The process of feature selection obtains the smallest set of features that better represents the process dynamics. The idea behind granulation is a clustering mechanism to create a number of semantic fuzzy rules based on the concept of Granular Computing (GrC). In other words, granulation aims at providing an interpretable fuzzy model within a unified concept based on functionality and data compatibility. Thus, the RBF-NN exploits such a granular signature in order to discriminate the role of each fuzzy set and the input variables while preserving a balance between transparency and interpretability [1]. The parameter identification of the EDDFN follows a Negative Correlation Learning (NCL) [7] and an Adaptive Back Error Propagation (ABEP) approach [11] that we call for short Adaptive Negative Correlation Learning (ANCL). The NCL introduces a penalty term in the cost function of each individual DDFM minimising its Mean Square Error (MSE) together with the correlation of the ensemble network so that every DDFM is finally trained by the ANCL. [7]. The ensemble model is viewed as a multi-input-single-output (MISO) FLS $f : U \subset R^n \rightarrow R$ having n inputs $x_k \in [x_1, ..., x_n]^T \in U_1 \times U_2 \times .. \times U_k .. \times U_n \triangleq U$, where U is the universe of discourse, the training set $\{\vec{x}_p, d_p\}_{p=1}^{P}$ such as $\vec{x}_p = \{x_1, \ldots, x_n\}$ and the ensemble output is:

$$f_{ens}(\vec{x}_p) = \frac{1}{N} \sum_{j=1}^{N} y_j(\vec{x}_p) \quad (1)$$

where the cost function of each DDFM is computed as:

$$e_j = \sum_{p=1}^{P} (y_j(\vec{x}_p) - d_p)^2 + \lambda p_j \tag{2}$$

and λ and p_j is a weighting and regularised term respectively.

$$p_j = -\sum_{p=1}^{P} (y_j(\vec{x}_p) - f_{ens}(\vec{x}_p)) \tag{3}$$

And the ensemble error function is:

$$E_p = \frac{1}{p} \sum_{p=1}^{P} (f_{ens} - d_p)^2 \tag{4}$$

Due to its functional equivelence to Fuzzy Logic Systems, every ith fuzzy rule in the RBF-NN can be stated as [12]:

$$\begin{aligned} R^i : IF\ x_1\ is\ F_1^i\ and \dots\ x_k\ is\ F_k^i\ and \dots \\ and\ x_n\ is\ F_n^i\ THEN\ y\ is\ G^i;\ i = 1, \dots, M \end{aligned} \tag{5}$$

And $F_1^i \times ... \times F_n^i = A^i$, hence Eq. (5) can be expressed as:

$$R^+ : F_1^i \times ... \times F_n^i \rightarrow G^i = A^i \rightarrow G^i; i = 1, ..., M \tag{6}$$

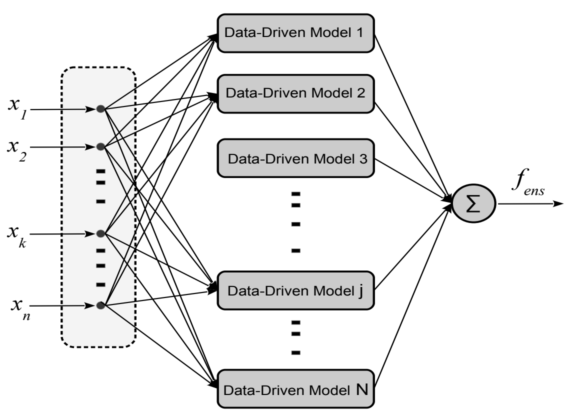

Fig. 2. Ensemble Data-Driven Fuzzy Network (EDDFN) structure.

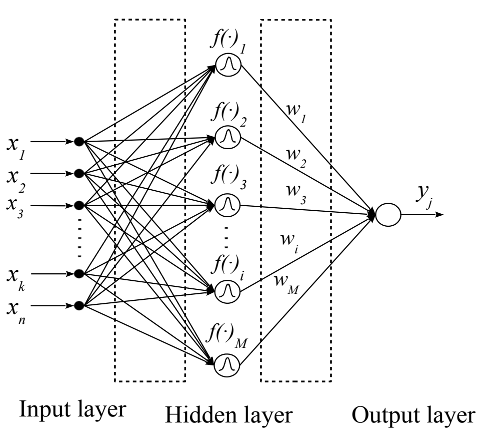

Fig. 3. RBF-NN Structure used by a single DDFM (Taken from [11]).

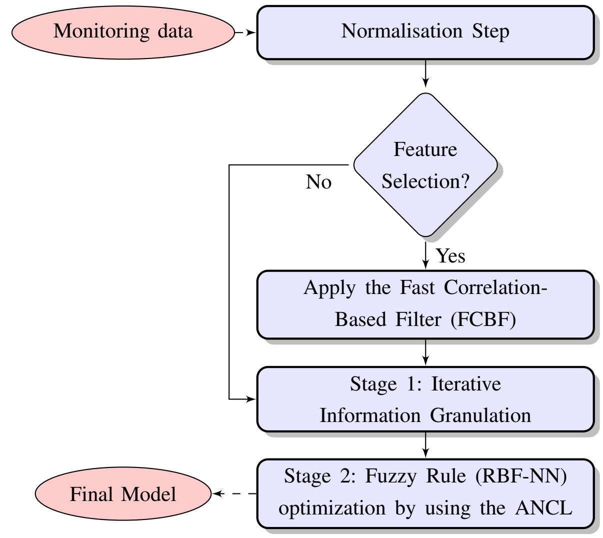

Fig. 4. Parameter identification applied to each Data-Driven Fuzzy Model (DDFM) (Taken from [13]).

A rule R^i is described by $\mu_{R^i}(\vec{x}_p, y) = \mu_{R^i}[x_1, ..., x_n, y]$, where a Mamdani implication is used as:

$$\mu_{R^i}(\vec{x}_p, y) = \mu_{A^i \rightarrow G^i}(\vec{x}_p, y) = \left[T_{k=1}^{n} \mu_{F_k^i}(x_k) \star \mu_{G^i}(y)\right] \tag{7}$$

Where each firing strength f_i is defined as

$$\mu_{R^i}(\vec{x}_p, y) = \mu_{A^i \rightarrow G^i}(\vec{x}_p, y) = f_i \left(exp \left[-\frac{\|\vec{x}_p - \vec{x}\|^2}{\sigma_i^2} \right] \right) \tag{8}$$

A. Adaptive Negative Correlation Learning (ANCL)

Although a conventional BEP leads the MSE to a good global minimum, it usually does not represent the optimal performance [14]. For that reason, we apply an Adaptive BEP to the NCL that follows the update rules:

$$\Delta w_i(t+1) = -\eta \frac{\partial E_p}{\partial w_i} + \gamma \Delta w_i(t) \tag{9}$$

$$\Delta \sigma_i(t+1) = -\eta \frac{\partial E_p}{\partial \sigma_i} + \gamma \Delta \sigma_i(t) \tag{10}$$

$$\Delta m_k^i(t+1) = -\eta \frac{\partial E_p}{\partial m_k^i} + \gamma \Delta m_k^i(t) \tag{11}$$

At iteration 't', a performance index $P_i(t+1) = \frac{1}{P}\sum_{p=1}^{P} E_p^2$ is monitored by the NCL:

- if $Pi(t+1) \geq Pi(t)$ Then

$$\eta(t+1) = h_d \alpha(t),\ \gamma(t+1) = 0$$

- if $Pi(t+1) < Pi(t)$ and $\left|\frac{\Delta Pi}{Pi(t)}\right| < \delta$ Then

$$\eta(t+1) = h_i \alpha(t),\ \gamma(t+1) = \gamma_0 \tag{12}$$

- if $Pi(t+1) < Pi(t)$ and $\left|\frac{\Delta Pi}{Pi(t)}\right| \geq \delta$ Then

$$\eta(t+1) = \eta(t),\ \gamma(t+1) = \gamma(t)$$

Algorithm 1 Fast Correlation-Based Filter Strategy [10].

Input:
  $S(F_1, F_2, \ldots, F_N, C)$ ▷ a training data set
  δ ▷ a predefined threshold
Output: S_{best}
procedure :
  for $i = 1\ to\ N$ **do**
    Calculate $SU_{i,c}$ for F_i;
    if $(SU_{i,c} \geq \delta)$ **then**
      append F_i to S'_{list};
    end if
  end for
  order S'_{list} in descending $SU_{i,c}$ value;
  $F_p = getFirstElement(S'_{list})$;
  do
    $F_q = getNextElement(S'_{list}, F_p)$;
    if $(F_q <> NULL)$ **then**
      do
        $F'_q = F_q$;
        if $(SU_{p,q} \geq SU_{q,c})$ **then**
          remove F_q from (S'_{list});
          $F_q = getNextElement(S'_{list}, F'_q)$;
        else $F_q = getNextElement(S'_{list}, F_q)$;
        end if
      while $(F_q <> NULL)$
    end if
    $F_p = getNextElement(S'_{list}, F_p)$;
  while $(F_p <> NULL)$
  $S_{best} = S'_{list}$
end procedure

where $h_d, (0 < h_d < 1)$ and $h_i, (1 < h_i)$ are the decreasing and increasing factors, respectively - δ is a threshold rate for the MSE. Thus, the partial derivatives for each DDFM are:

$$\frac{\partial E_p}{\partial w_i} = \left((y_j(\vec{x}_p) - d_p) - \lambda(y_j(\vec{x}_p) - f_{ens}(\vec{x}_p))\right) A_i \quad (13)$$

$$\frac{\partial E_p}{\partial s_i} = 2\left((y_j(\vec{x}_p) - d_p) - \lambda(y_j(\vec{x}_p) - f_{ens}(\vec{x}_p))\right) \times (w_i - y_j(\vec{x}_p))\, A_i \left(\frac{\sum_{k=1}^{n}(x_k - m_{ki})^2}{s_i^3}\right) \quad (14)$$

$$\frac{\partial E_p}{\partial m_{ki}} = 2\left((y_j(\vec{x}_p) - d_p) - \lambda(y_j(\vec{x}_p) - f_{ens}(\vec{x}_p))\right) \times (w_i - y_j(\vec{x}_p))\, A_i \left(\frac{(x_k - m_{ki})}{s_i^2}\right) \quad (15)$$

where $A_i = 2f_i / \sum_{i=1}^{M} fi$

B. Fast Correlation-Based Filter (FCBF) for Feature Selection

In this article we apply a Fast Correlation-Based Filter (FCBF) to find the most representative attributes/features [10]. The FCBF is aimed to find the optimal trade-off between fitness and complexity [1] while contributing to the DDFM accuracy. According to the **Algorithm 1**, the FCBF estimates the uncertainty of a random variable based on the information-theoretical concept of entropy of an attribute X_I in relation to the attribute X_J as [[10], [15]]:

$$H(X_I|X_J) = -\sum_{j=1}^{n_1} P(x_j) \sum_{i=1}^{n_1} P(x_i|x_j) log_2(P(x_i|x_j)) \quad (16)$$

where $X_I = \{x_1, \ldots\ldots, x_{n_1}\}$, $n = \{n_1, n_2, \ldots, n_m\}$ is the cardinality of the dimension I, $P(x_i)$ and $P(x_i|x_j)$ is the prior and posterior probability of X_I, respectively. The *"Information Gain"* is computed by Eq. (17) which is interpreted as the decrease of entropy of X_I given X_J.

$$IG(X_I|X_J) = H(X_I) - H(X_I|X_J) \quad (17)$$

When $IG(X_I|X_J) < IG(X_K|X_J)$, X_J is more correlated to X_K than to X_I. That is, the value of $'IG'$ is biased to those attributes/features whose cardinality is higher. To compensate for information gain's bias toward features with more values, a Symmetrical Uncertainty measure SU that is normalized to the range $[0, 1]$ is used to determine the correlation between the feature F_i and the label class C, where 0 and 1 represent the lowest and highest respectively [16].

$$SU(X_I, X_J) = 2 * \frac{IG(X_I|X_J)}{H(X_I) + H(X_J)} \quad (18)$$

SU is used as a goodness measure that evaluates the relevance of each feature and discriminates those that are redundant. A subset S' of relevant attributes can be extracted from S based on a predefined threshold δ, such that $\forall F_i \in S'; 1 \leq i \leq N, SU_{i,C} \geq \delta$. In a like manner to [10], we use the predominant correlation to extract the less redundant features that most contribute with information, i.e. *iff*, there exist a $SU_{i,C} \geq \delta$, and $\forall F_j \in S'(j \neq i)$ such that there is no F_j where $SU_{j,i} \geq SU_{i,C}$.

C. Iterative Information Granulation

Iterative information granulation is a data mining approach for data grouping based on a compatibility index $compat(\cdot, \cdot)$ that evaluates the similarity in the data [9], [14], [17]. In this article, the data provided by the monitoring system is granulated to describe the system dynamics as a series of linguistic fuzzy rules and used as the initial parameters of each DDFM. Hence, the process for iterative information granulation involves two main steps:

- Find the two most 'compatible' information granules and merge them together as a new information granule containing both original granules.
- Repeat the process of finding the two most compatible granules until a satisfactory data abstraction level is achieved.

The compatibility between two any granules A and B is:

$$compat(A, B) = D_{MAX} - d_{A,B} e^{(-\alpha R)} \quad (19)$$

where R and $d_{A,B}$ are:

$$R = \frac{card_{A,B} / Cardinality_{MAX}}{L_{A,B} / Length_{MAX}} \quad (20)$$

$$d_{A,B} = \sum_{k=1}^{n} (w_k/n)(max(u_{Ak}, u_{Bk}) - min(l_{Ak}, l_{Bk})) \quad (21)$$

The metrics associated to each resulting granule are the multidimensional average distance $d_{A,B}$ and length L_{AB}, while D_{MAX}, $Length_{MAX}$ and $Cardinality_{MAX}$ are the distance and length of the largest granule and the total number of granules in the data set respectively. With w_k playing the role of importance weight for the dimension k, $k = 1, \dots, n$. In Eq. (19) and (21), α is a weighting term for the rate cardinality/length and l_{Ak} and u_{Ak} are the lower and upper limits (corners) of the granule A respectively.

IV. Experiment Simulations

The experimental setup for laser welding prediction is carried out with an ensemble model of 7 DDFM units. Each DDFM uses the ANCL for its parameter identification (See Fig. 4) with a different correlation threshold, $\delta_T < \delta \leq \delta_{max}$. For this case study, the threshold for the third ensemble hidden unit is $\delta = \delta_T(\delta_{max} + 3/7)$. Similarly, the number of fuzzy rules per DDFM is defined by $M = M_o + \Delta_T(j-1)$, where $M_o = 2$ and $\Delta_T = 2$. A granular weight $\alpha = 0.3$ for the iterative information granulation was used. As indicated in section II, two welding seams per each PSG were produced in one experiment. Thus, a set of 41×26 dimensional feature vectors per seam were extracted by the monitoring system. Hence, we perform a set of 10×2 experiments by randomly selecting a training (60%) and a testing data set (40%) at each experiment. To compare the efectiveness of the proposed ensemble model to some existing techniques, we performed an identical number of experiments by using a single DDFM introduced in [13]. Such a model follows the methodology described by a single DDFM unit that uses a different number of hidden units.

It was found that the highest performance for the single DDFM is obtained with **five fuzzy rules**. In table II, the average number of the most relevant features and their associated correlation are presented. Fig. 5 and Fig. 6, the band witdh for the fuzzy decision surface that corresponds to the DDFM unit 1 and N, and the Average Ensemble Surface (**AES**) are depicted. In order to construct the fuzzy rule base, we extracted from the ensemble model the most frequent features selected at each single DDFM (ensemble neuron). Although, the set of consequents G^i for the fuzzy model are singleton, we use the average of each DDFM output and its corresponding standard deviation to construct the set of Gaussian functions. Consequently in Fig. 7, the corresponding fuzzy rule for the Average Ensemble Surface (AES) is presented for both seams "A" and "B".

TABLE II
INPUT RANKING USING THE FCBF FOR FEATURE SELECTION.

List of the most relevant and less redundant features.					
No.	Feature	SU	No.	Feature	SU
1	Mean of Mean-W	0.85	**5**	Mean of Varco-U	0.76
2	Mean of Varco-W	0.76	**6**	Mean of PSO	0.54
3	Mean of stddev-U	0.76	**7**	Mean of Sensor Signal	0.51
4	Mean of stddev-L	0.51	**8**	Mean of reference	0.51

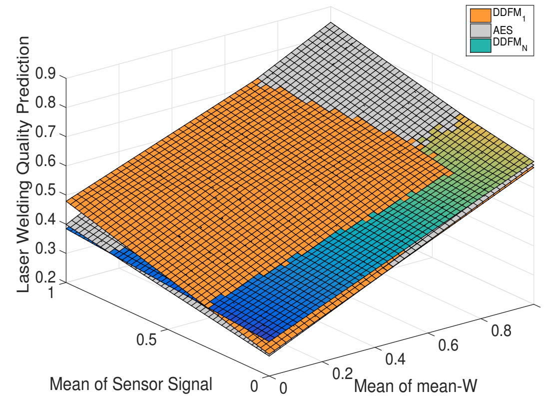

Fig. 5. Band width for the Fuzzy model surface that correspond to seam 'A'

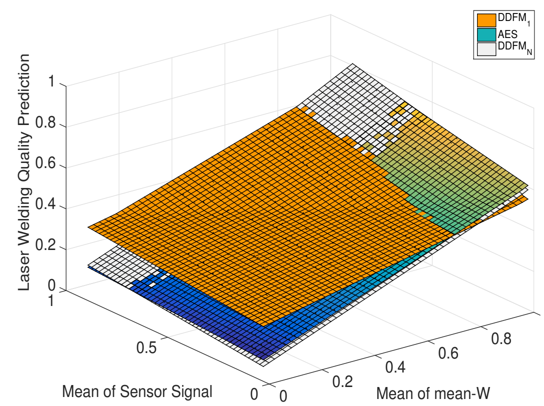

Fig. 6. Band width for the Fuzzy model surface that correspond to seam 'B'

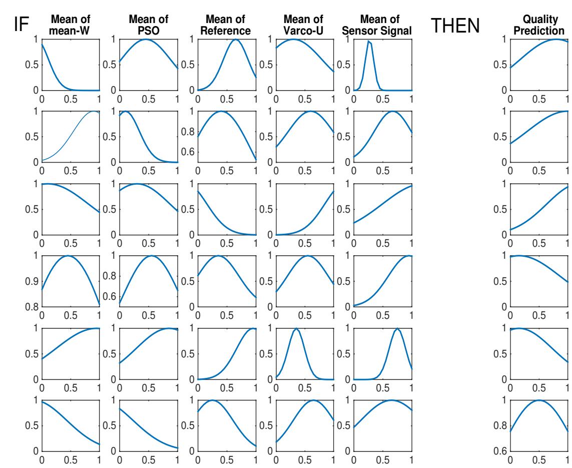

Fig. 7. AES Fuzzy model for laser welding quality prediction.

For diagnostic test evaluation, sensitivity and specificity are calculated in order to quantify the proportion of 'Ok' and 'Non-Ok' samples that were correctly classified respectively, while accuracy is the overall percentage of both measures [18]. Such metrics are computed by:

$$Specificity = \frac{TN}{TN + FP} \tag{22}$$

TABLE III
PERFORMANCE: LASER WELDING QUALITY PREDICTION

Model	Specificity[%]	Sensitivity[%]	Accuracy[%]
- Training -			
- Seam A -			
Single DDFM	71.23	74.25	72.74
Ensemble Network	73.68	82.76	78.22
- Seam B -			
Single DDFM	76.56	83.21	79.88
Ensemble Network	75.86	68.72	72.29
- Testing -			
- Seam A -			
Single DDFM	60.00	65.00	62.50
Ensemble Network	65.76	75.00	70.38
- Seam B -			
Single DDFM	61.54	77.31	69.42
Ensemble Network	74.67	82.0	78.33

$$Sensitivity = \frac{TP}{TP+FN} \tag{23}$$

$$Accuracy = \frac{TN+TP}{TP+TN+FP+FN} \tag{24}$$

While *sensitivity* measures the proportion of 'OK' seams (TP) that are identified correctly by the EDDFN, *specificity* quantifies the proportion of worpieces with a low-quality (failed-Non-OK) welding (TN). *FP* and *FN* represent the good seams-'OK' predicted as 'Non-OK' and the 'Non-OK' seams predicted as good seams. Although both models exhibited a similar performance during training, the EDDFN showed an improvement of approximately 10% for predicting new data. Particularly, for seam B the predictions results achieved by the EDDFM observed an accuracy of around 80%.

V. CONCLUSION

In this article, an Ensemble Data-Driven Fuzzy Network (EDDFM) is proposed. Each hidden unit of the EDDFN consists of a systematic construction of a Data-Driven Fuzzy Model that is beased on three conceptual components, i.e.: information theory for feature selection, granular cmputing and neural fuzzy systems. The overall framework is designed to specifically improve the generalisation prediction properties of a single DDFM. We employ the proposed ensemble model to not only model the process of laser welding, but also to predict the associated welding quality.

Results show that an ensemble model can have significant impact on the performance of single data-driven fuzzy models, particularly in the ability of the model to recognise (predict) new data (data that were not used in the learning/training process) with an overall performance of $> 78\%$ accuracy.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support from the EU H2020 programme (European Union) under the grant agreement 636902 (project COMBILASER), and HIDRIA AET and 4D LORTEK Spectrum(Germany) for providing the manufacturing case study and associated data.

REFERENCES

[1] Z. Shang-Ming, and J. Q. Gan. "Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling." Fuzzy Sets and Systems 159.23 (2008): 3091-3131.

[2] K. Anders, and J. Vedelsby. "Neural network ensembles, cross validation, and active learning." Advances in neural information processing systems 7 (1995): 231-238.

[3] Y. Liu , and X. Yao. "Simultaneous training of negatively correlated neural networks in an ensemble." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 29.6 (1999): 716-725.

[4] M. Alhamdoosh and D. Wang. "Fast decorrelated neural network ensembles with random weights." Information Sciences 264 (2014): 104-117.

[5] Y. Y. Yang, M. Mahfouf and G. Panoutsos, *"Development of a parsimonious GA-NN ensemble model with a case study for Charpy impact energy prediction"*, Advances in Engineering Software, vol. 42, no. 7, pp. 435-443, 2011.

[6] A. Mozaffari, and N. L. Azad. "Optimally pruned extreme learning machine with ensemble of regularization techniques and negative correlation penalty applied to automotive engine coldstart hydrocarbon emission identification." Neurocomputing 131 (2014): 143-156.

[7] H. Chen and X. Yao. (2009). "Regularized negative correlation learning for neural network ensembles". IEEE Transactions on Neural Networks, 20(12), 1962-1979.

[8] C. A. Walsh, "Laser weldingliterature review." Materials Science and Metallurgy Department, University of Cambridge, England (2002).

[9] W. Pedrycz and B. Andrzej, *"Granular clustering: a granular signature of data"*, *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 32, no. 2, pp. 212-224, 2002.

[10] L. Yu and H. Liu, *"Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution"*, *Twentieth International Conference in Machine Learning, Washington DC*, vol. 3, pp. 856-863, 2003.

[11] A. Rubio-Solis and G. Panoutsos, *"Interval Type-2 Radial Basis Function Neural Network: A Modelling Framework"*, *Fuzzy Systems, IEEE Transactions on*, vol. 23, no. 2, pp. 457-473, 2015.

[12] J.SR. Jang and C.T. Sun, *"Functional equivalence between radial basis function networks and fuzzy inference systems"*, *Neural Networks, IEEE Transactions on*, vol. 4, no. 1, pp. 156-159, 1993.

[13] A. Rubio-Solis, G. Panoutsos, and Steve Thornton. "A Data-driven fuzzy modelling framework for the classification of imbalanced data." In Intelligent Systems (IS), 2016 IEEE 8th International Conference on, pp. 302-307. IEEE, 2016.

[14] A. Rubio-Solis and G. Panoutsos, *"Granular computing neural-fuzzy modelling: A neutrosophic approach"*, *Applied Soft Computing, Elsevier*, vol. 13, no. 9, pp. 4010-4021, 2013.

[15] U. Martinez-Hernandez, T. J. Dodd, M. H. Evans, T. J. Prescott, and N. F. Lepora, "Active sensorimotor control for tactile exploration", Robotics and Autonomous Systems", vol. 87, pp. 1527, 2017.

[16] U. Martinez-Hernandez, A. Damianou, D. Camilleri, L. Boorman, N. Lawrence, and A. Prescott, "An integrated probabilistic framework for robot perception, learning and memory", in 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2016.

[17] A. Rubio-Solis and G. Panoutsos, *"Iterative Information Granulation for novelty Detection in Complex Datasets"*, IEEE World Congress on Computational Intelligence, Vancouver Canada, 2016, in press.

[18] A.J. Saah and D.R. Hoover, *"Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings.*, Annals of internal medicine, vol. 126, no. 1, pp. 91-94, 1997.