1. Introduction

Galvanising is one of the most effective methods for protection of steel against corrosion. Every year more than three million tonnes of zinc, which is 65–70% of the world zinc production, is being consumed worldwide in galvanising of steel sheets.1) Hot dip galvanisation is the most widely used process for galvanisation of steel sheets, which involves hot dipping of steel sheets in a galvanising bath to produce a thin protective coating of zinc-rich alloy on steel. The molten zinc bath often contains a small amount of lead and/or antimony and the coating surface is characterised with flowery structure consisting of bright and dull phases called ‘spangles’. On the basis of macroscopic surface appearance, spangles may be classified into three types2,3): shiny or mirror-like which are highly reflective with prominent hexagonal dendritic structures; feathery spangles also have high reflectivity with a lath-like pattern of alternate shiny dendritic and dull regions; and dull spangles, the appearance (low reflectivity) of which is principally caused by precipitated Pb particles. A lead-free coating may appear entirely shiny. Microscopic observations lead to three different forms of dull areas: (i) dimpled structure with spherical Pb particles randomly spread across the uneven zinc surface; (ii) ridged structure with precipitated Pb particles arranged in lines, and (iii) orthogonal-dendritic structure with secondary dendritic arms, lying at 90 degrees to its originating trunk, and precipitated Pb particles in between the dendritic arms.

Although the quality of spangles improves the product’s aesthetics and acceptability among low end users and commands some premium in the domestic market, spangles are undesirable in products with prospective painting potentials such as in automobile industries or metal buildings. The uneven spangle grains sometimes emerge through the paint, impairing the appearance.

In this study, an effort was made to establish a correlation between the spangle size in galvanised cold rolled sheets and process parameters at one of the continuous galvanising lines at Tata Steel. All the process related data were collected from the CRM database, while the information on spangle size was generated through actual measurements. Statistical (factor analysis) and mining (neural classification mining) analyses were carried out. The most significant input variables with respect to spangle size were extracted. The artificial neural network classification model was developed using 849 records for training with a prediction accuracy of 57%. Strip thickness appears to be most sensitive on the spangle formation; whereas lead and antimony concentration in zinc bath, and the pressure difference between the top and bottom air knives seem to be more sensitive amongst the other eight significant parameters. The classification model can be used for prediction of spangle size given the process parameters. It can also be used as an important tool to set and adjust the process parameters to produce a given spangle size.

KEY WORDS: galvanising; spangle; data mining; factor analysis; neural classification.
annealed, cooled nearly to the bath temperature and then immersed in a molten zinc bath. As the sheet is withdrawn from zinc bath, excess liquid zinc on the surface is removed by an air jet wiping (as shown in Fig. 1 as a line diagram), and solidification of sheet surface starts. Development of spangles, like any other crystallographic grain structures, occurs through a two-step process e.g. nucleation and growth. Nucleation starts at the interface of steel sheet with a passivating layer of Fe₂Al₅ and liquid zinc. During any solidification process, undercooling plays an important role in nucleation. Although high thermal undercooling (10–20°C) in the galvanised layer, which is required for homogeneous nucleation, has been reported by some researchers, contradictory reports of very low undercooling (<1°C) based on actual measurements are also available. Hence, heterogeneous nucleation of zinc dendrites with preferred nucleation sites on the rough steel–zinc interface has been proposed. No significant influence of bath composition (Al addition: 0.2 to 0.4% and Pb up to 0.05%) on undercooling was observed by Strutzenberger et al., which is again a contradiction that spangle formation is directly related to high melt undercooling, caused by Pb addition, as proposed by Spittle and Brown.

Once the heterogeneous nucleation takes place at some preferred sites on the steel/zinc interface, dendritic development of these zinc crystals forms within a small region of slightly undercooled layer (Fig. 2(a)), as proposed by Strutzenberger and Faderl. Heat is transferred from steel sheet towards the coating surface by conduction. The newly formed dendritic layer of zinc crystal makes the heat transfer slower from sheet to the liquid zinc surface. But heat can be transferred without any hindrance through the areas that are yet to be covered by flat dendrites and the undercooling of these areas increases. Thus a rapid lateral expansion of the nucleated zinc crystals parallel to steel–zinc interface takes place until the entire interface is covered (Fig. 2(b)). This is the first stage of solidification, which happens within fractions of a second. Final shape, size and appearance of a growing spangle are principally determined during this stage, controlled by an interaction of thermal conditions in the undercooled layer and crystal orientation of the nucleated zinc grain.

The second stage of solidification starts after the entire steel surface is covered with a layer of dendritic zinc grains. Further growth of these grains with slow thickening takes place, which is now controlled by transfer of heat from the coated sheet to atmosphere through radiation and convection, and consumes the major part of the total solidification time. The surface solidification starts when the thickest part of any uneven growing grain having an inclined basal plane, emerges with shiny surface through the thin film of liquid zinc, as shown in Fig. 2(c). With the progress in thickening of zinc grain layer, the residual liquid becomes continuously enriched in Pb and other alloying elements already added to the bath. Finally, the eutectic Zn–Pb composition is reached towards the end of solidification, depending on the initial Pb concentration, when the zinc phase crystallizes at the primary zinc dendrite and lead particles precipitate on the coating surface. Thus a grain with inclined basal plane may finally generate a typical shiny-dull divided spangle with the shiny part at the area that emerges first through the liquid, and dull area where solidification ends with eutectic precipitation of dimpled lead particles (Fig. 2(d)). According to Biber, different insoluble particles, mainly compounds/intermetallics of aluminium, other than pure lead particles, precipitate at the coating surface during the solidification process; the size of particles being...
larger in rough (dull) than in smooth (bright) spangles. The spangle structure is fully developed with the completion of eutectic reaction.

2.2. Parameters Determining Spangle Formation

Factors responsible for spangle formation can broadly be classified into three categories: (i) bath chemistry, (ii) cooling rate and undercooling and (iii) substrate topography and cleanliness.

2.2.1. Bath Chemistry

The effects of alloying additions on spangle formation have been studied by many researchers. Most of them found a direct influence of addition of at least one element i.e. Pb in the bath. Fasoyinu et al. found that Pb, Sb and Bi, when added in the zinc bath at concentrations more than 0.04% (weight), produce spangles on coated surface and the size of spangles increases with solute concentration. Similar additions of Sn, Cd and Mg do not produce spangle. Similar results with Pb addition were achieved by several researchers. Pb is supposed to increase the undercooling necessary for nucleation and decrease the density of active nucleation sites. It was observed that spangles would reach to a maximum size at about 0.07% Pb in the bath and remain constant with further lead addition, whereas an interaction of Al and Pb contents on spangle size was found by Faderl et al. According to Seré et al., addition of Pb influenced a crystallographic orientation with pyramidal planes (less reflection) parallel to the surface, and thus the amount of bright crystals decreased with the increase in lead content. But an increase in antimony content did not change the crystal texture or the brightness. Operational experience suggests optimum spangle size with a combined percentage of 0.16% of Pb and Sb. Also, the nature of spangles produced by preponderance of Pb differs from that produced by Sb. Very high lead content creates deep relief of spangles with etched boundaries.

2.2.2. Cooling Rate and Undercooling

Strutzenberger and Faderl have shown that a slow cooling rate of 2 K/s (natural convective cooling in air) or a higher one (15 K/s) by jet cooling with a gas flow rate of 10 m³/h of 5% H₂/N₂ does not make much difference in temperature gradient, negligible in both cases, in the thin film of liquid zinc prior to solidification. Consequently, sufficient undercooling is not produced by either a slow or fast cooling to start a homogeneous nucleation. However, during lateral expansion of zinc dendrites in the first stage of solidification, cooling rate seems to play an important role to determine final spangle sizes.

2.2.3. Substrate Topography and Cleanliness

A spangle appears to develop dendritically from a single nucleus. Spangle size therefore inversely varies with the number of nuclei, which is determined by an interaction of preferential nucleation rate and the rate of lateral expansion of already nucleated dendritic zinc crystals. In view of very low undercooling, a heterogeneous nucleation at the steel–zinc interface has been proposed. In such heterogeneous nucleation, the condition of substrate surface is therefore very important for preferential nucleation sites. Faderl et al. observed 40 to 100% larger spangles in a smoother material (IF steel) in comparison to a less smooth material (AK steel). Different precleaning treatments also caused wide variations (up to 400%) in spangle size, particularly at low concentration of Pb in the bath. The difference was hardly recognisable in presence of higher Pb concentration (0.05% and above).

2.2.4. Other Factors

Other factors like substrate thickness, dipping time controlled by the line speed, sheet temperature before entry into bath, zinc bath temperature, air knife (wiping) pressure and position etc. may also affect the final spangle size. But most of them alter either the cooling rate and undercooling or the substrate topography, which are the actual reasons for variation in spangle size. However, no significant influence of the amount of wiping gas and hence of the coating thickness and of sheet temperature on spangle size was found by Faderl et al. Since the coating thickness is generally much less than strip thickness, the latter is one of the prime factors that determine the solidification time and so the final spangle size.

3. Data Mining Process

Generally, four steps are followed in a data mining process: data selection, data transformation, data mining (or the analysis part), and data interpretation (results and discussion). The first three are described in this section and the fourth one will be elaborated in the next section.

3.1. Data Selection

Based on the available information and availability of process related data responsible for spangle formation, pertinent parameters were first selected. The description of important process parameters, bath chemistry and information on spangle size, which were used in this study, is given in Table 1.

3.1.1. Collection of Process Data

The relevant process parameters are recorded at the galvanising line and stored in CRM database. These records were restored with the help of a structured query language (SQL) program. The air knife pressure during air jet wiping of liquid zinc as the sheet steel comes out of the zinc bath is important for spangle formation. The air knife pressures for both top and bottom surfaces of the sheet (as shown in Fig. 1) were collected. Since spangle size on the top surface (SPNLG_T) is the target parameter here, the air knife pressure for the top surface (AIR_PR_T) and the difference in air knife pressure for the top and bottom surfaces (AIR_PR_DIFF) were considered in this study. Amongst the other available process parameter data, the line speed (SPEED), zinc bath temperature (TE_BATH), temperature of strip before entry to the zinc bath (TE_STR), the tension level of strip inside the zinc pot (TEN_BATH), strip thickness (THK_STR) and the zinc coating weight on the top surface of sheet (WT_CT_T) were chosen to be important.

3.1.2. Collection of Bath Analysis Data

The shift-wise zinc bath analysis information (weight percent of Al, Pb, Sb and Fe as AL_WT, PB_WT, SB_WT and FE_WT) was collected from CRM. These data were appended to the process parameter data according to the day and shift of operation. Coils processed in any one shift,
thus, are assumed to have the same bath chemistry.

3.1.3 Collection of Information on Spangle Size

No information on spangle size is available at the galvanising line. Therefore the spangle sizes for all coils were manually determined. Determination of spangle size involves measurement of macro grains. As per ASTM (E112), there are three basic procedures for grain size estimation12):

1. Comparison Procedure – the grain structure is compared with a series of graded images, either in the form of a wall chart, clear plastic overlays, or an eyepiece reticle. This is applied to completely recrystallised or cast materials with equiaxed grains.

2. Planimetric Procedure – number of grains within a known area (circular or rectangular of usually 5 000 mm²) is actually counted. This method is reasonably precise, free of bias and reproducible.

3. Intercept Procedure – an actual count of the number of grains intercepted or the number of grain boundary intersections by a test line is done. Intercept procedures are also precise, free of bias and reproducible, but more convenient than Planimetric procedure and are recommended particularly for non-equiaxed grain structures.

Many authors have referred the spangle sizes in millimetre but most of them have not mentioned of any particular procedure that they followed for spangle size measurement. Only in an article by H.-B. Chen,9) it was mentioned that spangle size was calculated by counting grain numbers in a given area and then converting that given area into circles of equal diameter.

In the present study, the spangle size has been correlated with various process parameters by applying an Artificial Neural Network (ANN) technique. Since there were many input parameters, huge sets of data were required to achieve a good correlation by this ANN technique (minimum 10 times the number of input variables). Spangle size was determined only for those coils for which the entire process variable data were available in the CRM database. After collecting those data and coil IDs, the corresponding samples were traced in the CRM Laboratory and spangle sizes were measured. It was decided to identify the spangles in different groups on the basis of their size to make the size determination process faster with sufficient accuracy so that enough data could be generated. Four groups were identified as described in Table 2.

The Intercept Procedure was followed. Total number of spangles, both dull and bright, intercepted by a fixed length (100 mm) of a line was counted based on the following principle:

\[
\text{Total No. of spangles} = \frac{\text{No. of spangles completely intercepted}}{\text{No. of spangles partly intercepted}} + \frac{1}{2}
\]

The spangle size was determined by dividing the fixed length by the total number of spangles. This was repeated at three different places of sheet sample and the group was determined based on average of the three. Since the groups have distinctly wide size-ranges, after having an initial experience of size measurement of few tens of samples and a feel about the groups, it was easy to identify the group of a

---

Table 1. Details of parameters used in this study.

| Variable ID | Variable description                                      | Unit       | Variable type |
|-------------|----------------------------------------------------------|------------|---------------|
| X0          | AIR_PR_DIFF                                              | milli-bar  | Integer       |
| X1          | AIR_PR_T                                                 | milli-bar  | Integer       |
| X2          | AL_WT                                                   | %          | Decimal       |
| X3          | FE_WT                                                   | %          | Decimal       |
| X4          | PB_WT                                                   | %          | Decimal       |
| X5          | SB_WT                                                   | %          | Decimal       |
| X6          | SPEED                                                   | m/min      | Integer       |
| X7          | SPNGLE_T                                                | -          | Categorical   |
| X8          | TE_BATH                                                  | deg C      | Integer       |
| X9          | TE_STR                                                  | deg C      | Integer       |
| X10         | TEN_BATH                                                | kg/mm²     | Integer       |
| X11         | THK_STR                                                  | mm         | Decimal       |
| X12         | WT_CT_T                                                 | g/m²       | Integer       |

Table 2. Different groups of measured spangles and their size ranges.

| Group | Size range       |
|-------|------------------|
| 1     | Up to 5 mm       |
| 2     | >5 – 10 mm       |
| 3     | >10 – 15 mm      |
| 4     | >15 mm           |
sample quite accurately just by a thorough visual observation of the sample.

### 3.2. Data Transformation

Having collected the raw process data, the missing value records were discarded. Only those coils, for which all fields were available in the database, were considered for analysis. All information regarding the process parameters, zinc bath chemistry and spangle size, was collated into one file. Two such datasets were prepared—one with 849 records for training of the model in subsequent analysis and another with 100 records for the prediction.

The information on spangle size was initially generated in numeric form i.e. 1 to 4, as described earlier. Since our objective is to classify the spangle size into four categories, characters such as A, B, C and D were used to replace 1, 2, 3 and 4 respectively.

### 3.3. Data Mining

Data mining analysis was done in three steps: descriptive statistical analysis, factor analysis and classification mining. Information from various references13–18 was helpful during this mining analysis.

#### 3.3.1. Descriptive Statistical Analysis

At first, a descriptive statistical study was carried out for the whole dataset. To summarise the data, distribution patterns for all input parameters along with the minima, maxima, mean values and standard deviations, were determined. The outliers were discarded during the subsequent analysis.

#### 3.3.2. Factor Analysis

The number of input parameters was large. Factor analysis, a common technique used in multi-variate data analysis, was initially done with the help of a data mining software.13) It is a data reduction technique and is generally used to reduce a large number of variables in the dataset to a manageable size. First, it examines the interrelations amongst large number of variables and then tries to explain these relationships using a smaller number of variables. Factor analysis uses the overall correlation matrix of the variables and establishes the parameters that are having an underlying dimension. Factors or the underlying dimensions are linear combinations of original variables. A number of factors are mathematically identified, which best represent the observed correlations between the initial set of variables. Generally, the parameters that are highly correlated with each other, are represented by a single factor or dimension. The number of significant factors, which can represent a major part of the variance in the dataset, is much less than the total number of variables. Once the factors or dimensions are identified, the next step is to determine the parameters that fall within their respective dimensions. This is done with the help of a mathematical procedure called rotation. The influence of the parameters on any particular factor is thus determined. After the factor analysis, one can finally decide on a fewer number of variables which are responsible for the maximum variance of the dataset.

#### 3.3.3. Classification Mining

Intelligent Miner provides two classification algorithms: tree induction, and neural induction. Both are based on supervised learning technique, i.e., the model is created from a training set of input records, already containing the predefined classes. After the model is induced, it can be used for prediction of classes for unclassified records. The neural induction technique, which was chosen for the model development, is based on a network-architecture of interconnected nodes and weights between the input parameters and output parameters; thus it is based on ANN. In order to minimise the errors during training run, a generalised supervised learning algorithm, called back propagation is used to adjust the weights by propagating the errors backwards from the output.

**Figure 3** shows the 9–7–4 architecture used in developing the neural classification model. The network consists of an input layer with 9 input nodes, a hidden layer with 7 processing nodes, and an output layer with 4 output nodes. The number of input nodes is same as the number of important input parameters after the factor analysis. The number of output nodes is same as the number of different classes pertaining to the output field in the training dataset. The network architecture has been created in a way such that every node in the input layer is connected to every node in the hidden layer, and every node in the hidden layer is connected to every node in the output layer. The connections between nodes, called weights, are initialised before training, and adjusted during the training phase when each record (pattern) in the input dataset has been fed to the network for learning. Before feeding in the network, all the input data fields (for both continuous and categorical variables) were normalised between 0 and 1 by the data mining software. An input dataset having 849 patterns or records was taken for training. During training, when a pattern is fed to the input layer, the input travels through the network until it reaches the output layer. This forward pass produces the predicted output pattern for a given input pattern. Since this is a supervised learning algorithm, the desired output pattern is also given in the input dataset. The difference between the actual and the predicted output for a given pattern produces an error signal, which is used as the basis of backpropagation that propagates the error backwards in order to adjust the weights to produce the desired output. The weights are corrected by using an error-minimizing techni-
nique called gradient-descent. The learning rate or step size ($\eta$) of the back-propagation gradient descent was chosen to be 0.2. The momentum ($\alpha$), which helps in dampening the oscillation during training, was taken as 0.5.

4. Results and Discussion

It was seen from the statistical analysis that the distribution was mostly normal for all data fields. However, the outliers were discarded during bucketing in the mining analysis. Since 13 data fields are fairly large, a data reduction technique called factor analysis was carried out so as to reduce the number of data fields to a manageable figure. The multi-variate correlation analysis shows the inter-relation-ship of variables in the form of correlation coefficients in a 13×13 correlation matrix, as shown in Table 3. As the correlation matrix is symmetric about its diagonal, only bottom left triangle is shown. The value at the junction of a given row and a given column shows the correlation between the row and the column variables. The values in the matrix, close to $\pm 1$ mean positive correlation, while that close to zero $(0.1$ to $0.1)$ mean negative correlation. The values close to zero $(-0.1$ to $0.1)$ suggest that no correlation exists between the variables and even the positive or negative signs are insignificant. It can be seen from the Table that there was a significant positive effect of strip thickness and strip temperature before zinc bath on spangle size; whereas line speed had a negative effect. The heat content in the strip increased with its thickness and temperature, which resulted in a longer solidification process of liquid zinc on the strip, thus allowing the spangles to grow in size. Generally, the line speed depends on the strip thickness: higher the thickness lower is the speed (the correlation was $-0.85$, as shown in the table). Therefore, an opposite effect is reflected by the line speed. Since this Table gives a one-to-one relationship between any two variables and many of these variables are inter-related as shown in the Table, no conclusion should be directly made based on these data.

The complete correlation matrix is used to extract the factors or the underlying dimensions of variables. The analysis results produce a matrix showing the factor loadings by all variables to the significant factors, as shown in Table 4 for unrotated factor matrix. The factor loadings are the correlation between each variable and the factors; higher the value, more representative is the variable. The significant factors are extracted based on a commonly used technique, the ‘latent root criterion’, which defines a factor to be a significant one, if its eigenvalue is greater than one. The extent of variance in dataset covered by a given factor is explained by the eigenvalue. It is calculated from the factor loadings data, by the column sum of squared loadings for a particular factor. The factor extraction criterion, along with the eigenvalues is depicted in Fig. 4. This shows that there are only five significant factors (out of a total of 13) having eigenvalues greater than one and they are extracted in the factor analysis. The total amount of variance on which the factor analysis is based, or the total ‘trace’ is obtained by summing up squares of the diagonal values of the correlation matrix, which is 13 here. The percentage of trace is calculated by dividing each factor’s sum of squares or, the eigenvalues by total trace, and is given in Table 4. This suggests how well a particular factor accounts for what all the variables together represent; a low value means all the variables are very different from one another. The
total trace shows that over 76% of total variance is explained by these five factors. Table 4 also shows the communality of all variables for the significant factors, which is obtained by the row sum of squared factor loadings for a given variable. This explains the extent of variance in a given variable covered by the five factors taken together. Table 4 shows that these five factors extract maximum variance in most of the variables. The extraction of these five factors out of 13 should not produce much error during further analysis.

However, from the ‘unrotated factor matrix’, it is difficult to select the important variables since there are significant loadings by some variables to more than one factor (as indicated in bold in Table 4). The first factor, for example, accounts for largest amount of variance (24.23%) with eight variables having moderate to high loadings. A graphical representation of the loadings of all variables on (any) two factors, Fact_2 and Fact_3 is shown in Fig. 5(a). Thirteen variables labelled in numbers are shown in a two dimensional factor diagram, with ‘0’ as origin and axes spanning between −1 and +1. The horizontal and vertical axes are taken as Fact_2 and Fact_3 respectively. It can be seen that variables 3, 4 and 5 (overlapped with 4) load highly in positive direction, variable 6 loads moderately also in positive direction, while variables 7 and 11 load moderate values in negative direction. All other variables have significantly lower loadings on Fact_2. However, on Fact_3, except variable 9, all other variables have moderate to low loadings. It is not obvious from the unrotated factor matrix, which factor is loaded most by any given variable. Therefore, a matrix rotational technique, called ‘quartimax rotation’ was used to rotate the factor matrix, as given in Table 5. It is to be noted here that although the loading pattern of five factors is different in the ‘rotated factor matrix’, the total trace and the communalities remain unchanged. By rotation, the variances associated with factors are more evenly distributed for the first three factors. It is now apparent from Table 5 which variable has contributed the maximum to which factor. This is more evident when a comparison is made in the graphical representation of the contribution of all variables on Fact_2 and Fact_3, between unrotated and rotated matrix (Fig. 5). As given in Table 5, there are 11 out of 13 parameters had values greater than 0.60 (indicated in bold) and they were considered as significant. However, since FE_WT showed significant contribution in more than one factor, it was not considered. As SPNGL_T was the target parameter, the remaining nine variables were considered as the input parameters to the ANN classification mining.

The result of the ANN model after the training with 849 records, as explained earlier, is shown in terms of a confusion matrix (Table 6). The confusion matrix is a table with the numbers of actual versus predicted classes. It can be seen that the model had learnt 74 out of 140 ‘A’ s correctly. Hence, the prediction accuracy of class A is 52.9%. Similarly, the accuracy of class B is 78.3%, class C is
28.9% and class D is 59.9%. An overall accuracy of 57% was obtained from the training model. As described earlier, cleanliness and topography of the substrate surface plays an important role in spangle formation. Since no pertinent data are captured in the plant, information on strip surface condition could not be included in the present dataset. Moreover, shift-wise steady state zinc bath chemistry was assumed in this study, which is far from the reality, considering the dynamics within and drag rate from the bath for individual strips. These two reasons may be responsible for a moderate accuracy of 57% after training. The ANN model so trained was then used for prediction of 100 records. The model prediction is shown in Table 7 in terms of a ‘confusion matrix’. It can be seen that the accuracy of overall model prediction is 57%. The prediction for class A and B and D is good (73.6% accuracy). In real process, this is important in the sense that the model can predict the smaller (up to 10 mm) and very large (>15 mm) spangles efficiently. However, the prediction for the class C seems to be skewed towards left as also observed in the training results. A sensitivity analysis of nine input variables on the target parameter i.e. spangle size, as learnt by the ANN classification mining model, is given in Fig. 6. It shows that the strip thickness has got the maximum sensitivity (20.8%) on spangle size, as expected from the heat transfer point of view, explained earlier. Even if the working range...
of lead and antimony in zinc bath has been small and the mean value is 0.08 for both, they appear to be quite sensitive on spangle size. It is also evident that the pressure difference between the top and bottom air knives, coating weight and the tension level of strip inside liquid zinc bath play important roles in determining the spangle size. There are references of the strip thickness, coating weight and bath composition as important parameters from previous thermodynamic and other studies as discussed earlier. But there is no reference of the difference of air pressure and tension level as significant parameters for spangle size. It was possible to extract this information with the application of data mining technique into such analysis for the first time.

Table 5. Factor loadings of rotated factor matrix (after quartimax rotation).

| Variable ID | Fact_1 | Fact_2 | Fact_3 | Fact_4 | Fact_5 | Commu.
|-------------|--------|--------|--------|--------|--------|--------
| AIR_PR_DIFF | -0.05  | -0.02  | 0.02   | -0.05  | -0.87  | 0.76   |
| AIR_PR_T    | 0.5    | -0.02  | 0.59   | -0.15  | -0.1   | 0.63   |
| AL_WT       | 0.07   | -0.03  | -0.04  | 0.98   | 0.02   | 0.96   |
| FE_WT       | 0.02   | 0.74   | -0.11  | 0.62   | 0.01   | 0.94   |
| PB_WT       | -0.01  | 0.96   | -0.05  | -0.04  | -0.02  | 0.93   |
| SB_WT       | -0.04  | 0.97   | -0.07  | -0.07  | -0.02  | 0.96   |
| SPEED       | 0.89   | -0.02  | 0.13   | 0.02   | -0.06  | 0.81   |
| SPNGE_T     | -0.77  | -0.01  | 0.29   | -0.06  | 0.02   | 0.68   |
| TE_BATH     | -0.18  | -0.05  | 0.04   | -0.02  | 0.51   | 0.30   |
| TE_STR      | -0.28  | -0.02  | 0.78   | -0.13  | 0.13   | 0.72   |
| TEN_BATH    | -0.22  | 0.05   | -0.78  | -0.06  | 0.23   | 0.72   |
| THK_STR     | -0.87  | 0.03   | -0.34  | -0.04  | 0.09   | 0.89   |
| WT_CT_T     | 0.08   | -0.07  | 0.76   | 0.06   | 0.14   | 0.62   |

Eigenvalue 2.58 2.43 2.38 1.41 1.14 0.14 9.94
% of Trace 19.85 18.69 18.31 10.85 8.77 8.14 76.47

Table 6. Confusion matrix of actual vs. predicted values after the training of neural classification model.

| Actual | Predicted values after training | Total |
|--------|---------------------------------|-------|
|        | A     | B   | C   | D   | Unknown |
| A      | 74    | 58  | 2   | 0   | 6       | 140    |
| B      | 24    | 253 | 23  | 5   | 18      | 323    |
| C      | 2     | 119 | 69  | 18  | 31      | 239    |
| D      | 0     | 28  | 88  | 88  | 23      | 147    |

Table 7. Confusion matrix for prediction.

| Actual | Predicted values | Total |
|--------|------------------|-------|
|        | A    | B   | C   | D   | Unknown |
| A      | 12   | 5   | 0   | 0   | -       | 17     |
| B      | 2    | 29  | 5   | 0   | 4       | 38     |
| C      | 1    | 16  | 4   | 5   | 2       | 28     |
| D      | 0    | 2   | 2   | 12  | 1       | 17     |

Fig. 6. Sensitivity of input variables for the target parameter.
5. Conclusions

An effort has been made in the present work to establish a correlation between spangle size during hot dip galvanising and the process parameter. A large dataset was created, and statistical and data mining analysis was carried out with the help of a data mining software, for the first time in the area of galvanising. The significant variables were first extracted using Factor Analysis technique and an ANN classification mining model was then developed. The model predicts with 57% accuracy. The prediction was more accurate for smaller and very large spangles. The sensitivity of input variables towards the spangle size was found out. Strip thickness appears to be most sensitive, whereas lead and antimony concentration in zinc bath, the difference in air pressure between the top and bottom air knives, coating weight and the strip tension are also important amongst the nine significant parameters. This information will be useful in producing galvanised products with different spangle sizes of customers’ choice. The neural classification model can be used for prediction of spangle size for given process parameters. It can also be used as an important tool in adjusting the process parameters to produce a given spangle size.

Acknowledgements

The authors would like to thank the management of The Tata Iron and Steel Co. Ltd., India for giving the permission to publish this work. The kind help by Mr. B. Sinha of Information Technology Services Division in collecting the process data from the CRM database is also thankfully acknowledged.

REFERENCES

1) X. G. Zhang and F. E. Goodwin: Proc. of 5th Int. Conf. on Zinc and Zinc Alloy Coated Steel Sheet (GALVATECH 2001), Verlagstahleisen GmbH, Düsseldorf, (2001), 311.
2) D. Jaffrey, J. D. Browne and T. J. Howard: Metall. Trans., 11B (1980), 631.
3) J. Strutenberger and J. Faderl: Metall. Mater. Trans., 29A (1998), 631.
4) F. A. Fasoyinu and F. Weinberg: Metall. Trans., 21B (1990), 549.
5) J. A. Spittle and S. G. R. Brown: Proc. of 3rd Int. Conf. on Zinc Coated Sheet (INTERGALVA 91), European General Galvanisers Association, London, (1991), 54K/1.
6) H. E. Biber: Metall. Trans., 19A (1988), 1603.
7) S. Chang and J. C. Shin: Proc. of Use and Manufacture of Zinc and Zinc Alloy Coated Steel Sheet Products (GALVATECH ‘95), ISS, Warrendale, PA, (1995), 783.
8) J. Faderl, W. Maschek, J. Strutenberger: Proc. of Use and Manufacture of Zinc and Zinc Alloy Coated Steel Sheet Products (GALVATECH ‘95), ISS, Warrendale, PA, (1995), 675.
9) H.-B. Chem. China Steel Tech. Rep., 5 (1991), 52.
10) T.-W. Kim and R. S. Patil: Proc. of 1st Int. Conf. on Zinc Coated Steel Sheet, Zinc Development Association, London, (1983), D1.
11) P. R. Sere, J. D. Culcas, C. I. Elnor, A. R. Di Sarli: Surf. Coat. Technol., 122 (1999), 143.
12) ASTM, Designation: E112–96, ASTM, West Conshohocken, PA, (1996).
13) IBM Intelligent Miner for Data, Version 2, Release 1, 2nd ed., IBM Corporation, USA, (1998).
14) P. Cabena, P. Hadjimour, R. Stadler, J. Verhees, A. Zanasi: Discovering Data Mining – from Concept to Implementation, IBM, Prentice Hall, PTR, New Jersey, (1998), 1.
15) J. F. Hair, Jr., R. E. Anderson, R. L. Tatham and W. C. Black: Multivariate Data Analysis with Readings, 4th ed., Prentice Hall, PTR, New Jersey, (1984), 1.
16) R. L. Kennedy, Y. Lee, B. V. Roy, C. D. Reed, R. P. Lippman: Solving Data Mining Problems through Pattern Recognition, Data Warehousing Institute Series, Prentice Hall, PTR, New Jersey, (1997), 1.
17) M. J. A. Berry, G. Linoff: Data Mining Techniques – for Marketing, Sales and Customer Support, John Wiley & Sons Inc., (1998), 1.
18) B. E. Hayes: Measuring Customer Satisfaction – Survey Design, Use, and Statistical Analysis Methods, 2nd ed., ASQC Quality Press, Wisconsin, (1997), 153.