TOWARDS AN AI-Driven Smart Manufacturing of Diecastings

F. Liu, S. Wang, X. Liu, T. Zhang, B. Yang, and Q. Han

Purdue University, West Lafayette, Indiana

D. Yang,

University of Missouri, Columbia, Missouri

Corey Vian

Fiat Chrysler Automobiles, Kokomo, Indiana

# abstract

This paper describes our initial effort in applying supervised machine learning approaches to analyze data collected at a FCA casting plant during the production of crossmember castings. The data contain results of X-ray inspection on castings and processing variables such as alloy compositions, cooling conditions of the shot tooling and die, and the operation parameters of the diecasting machine. In addition to the conventional statistical machine learning approaches, such as polynomial regressions, logistic regressions and decision trees, deep neural networks were also used to investigate the relationship between the scrap rates of the crossmember casting and the operation variables of the casting process. Various data science techniques, such data cleaning, data normalization and data augmentation were also applied to further boost the validity and accuracy of the results. The finding of this work shows that machine learning approaches have an excellent potential to reduce scrap rates of the castings. The work also exemplifies how state-of-art artificial intelligence techniques can be applied in die casting to enable smart manufacturing.

# introduction

More and more data are collected, often automatically, to register how a part is produced in a machine, an assembly, a production line and a factory. With the advance of data sciences and artificial intelligence, it makes sense to apply machine learning techniques to analyze the collected data and to pursue smart manufacturing. In this paper, we examine data collected from a FCA casting plan during the production of crossmember castings. Datasets collected includes the following three groups of information: 1) Alloy compositions; 2) Casting process data including vacuum levels, water and lube flowrates, temperatures, cycle time, shot velocity, intensification pressure, etc.; and 3) X-Ray data for all 46 X-Ray views. Data collected are all in structured format, with no images or voices, and are stored in CSV format. Each row in the CSV file represents a part and is labeled with PASS or FAIL. There are total of 91 factors (columns) and 345,465 parts (rows). Among them, 310,221 parts are labeled as pass and 35,244 parts are labeled as fail.

Prevailing machine learning algorithms, such as neural networks, have become increasing accurate at predictions. But the black-box approach makes the model difficult to explain and hard to infer cause and effect relationship. To this end, two types of studies are explored in this paper. The first study leverages the traditional statistical analyses and aims at discovering factors that caused higher scrape rate. The second study applies the current machine learning algorithms and intends to predict if a part can pass the test. Put it differently, the first study will help practitioners understand how to reduce the scrap rate while the second study will help us understand, given data in its current form, is it sufficient to skip the manual inspection process and declare if a part is good or bad?

# EXPERIMENT 1: STATISTICAL METHODS

First, we introduce the statistical study. The goal is to find a best model from a candidate pool and find important explanatory variables for passing probability. The study also provides optimal values of explanatory variables to potentially improve the quality of manufactured parts.

## Generalized Lineara Model

We applied the generalized linear model (GLM) approach to the front crossmember casting data. GLMs are popular in analyzing PASS/FAIL data [1]. They are extended from ordinary linear models for continuous response variables to a class of models for either continuous or discrete response variables. Basically, a GLM for PASS/FAIL data can be expressed as

(1)

where is the link function, is the probability of PASS, and is a deterministic function of explanatory variables. Three kinds of GLMs are extremely important in the analysis of PASS/FAIL data. They are the logistic, the probit, and the cloglog models. A logistic model is derived if one chooses . A probit model is derived if one chooses , where is the cumulative distribution function (CDF) of the standard normal distribution. A cloglog model is derived if one chooses Note that is between 0 and 1. The purpose of these choices is to make the left hand side of equation (1) to cover all of the real numbers such that the model is always well-defined for any kind of . The choice of is flexible. A common way is to formulate it by linear and quadratic effects of explanatory variables. The primary issue is to determine the best choice of such that the model can provide the optimal relationship between the probability of PASS and explanatory variables.

At the beginning, we assumed that was composed of quadratic functions of all of the explanatory variables such that it was expressed as

(2)

where  represented the *j*th explanatory variable, *p* was the total number of explanatory variables,  and were the coefficients of the quadratic and linear effects of the *j*th explanatory variable, respectively, and *c* was formulated by constant terms of the quadratic functions of explanatory variables as well as the impact of machines and cavities. We classified coefficients contained by *c* as nuisance parameters since they were not major interests.

We fitted the logistic, probit, and cloglog models by the standard GLM package contained in the **R** software. It provided the estimates of coefficients, their standard errors, and estimates of nuisance parameters contained by *c*. A Wald statistic for each coefficient was calculated by the ratio of the coefficient and its standard error. The *p*-value of the coefficient was calculated by comparing the value of the Wald statistic with the standard normal distribution. The significance of the coefficient was concluded if the *p*-value was less than 0.05. All of them were calculated by the output of the standard GLM package.

To determine the optimal relationship between the probability of PASS and explanatory variables, we evaluated the importance of the coefficients contained in for each *j*. It was determined by the significance of and . If neither of them was significant, then we set both and as zero, indicating that the explanatory variable could be removed from (2) such that we set . If was significant but was not, then we set as a linear function of . If was significant, no matter whether was significant or not, we kept both of them. Then, we had with a nonzero , and a quadratic effect of was derived. After excluding unimportant coefficients, we obtained the final result of . It was composed by many linear functions and many quadratic functions of the explanatory variables. The number of explanatory variables contained in the final result of was lower than *p*.

## RESULTS

The individual effects of explanatory variables were completely determined by properties of . It contained three important scenarios when the quadratic effect was present. The first scenario was given by a negative estimated value of (Table 1). The function had a maximum value at , where and were the estimates of and . Assuming that the values of all of the rest explanatory variables did not change, the maximum probability of PASS was obtained at , indicating that it was a maximizer of equation (1). It was the best value for us to set up the explanatory variable. As all of them were within the range (i.e., between minimum and maximum) of the corresponding variables, the maximizers given by Table 1 could be treated as nice approximations to their true values. The second scenario was given by a positive value of (Table 2). The function had a minimum value at . Assuming that values of the rest explanatory variables did not change, the minimum probability of PASS was obtained at , which was a minimizer of equation (1). It was the worst value that we needed to avoid. For the same reason, the minimizers given by Table 2 could also be treated as nice approximations to their true values. The third scenario was induced by the absence of (result not shown), where was a linear function of . Since it was either strictly increasing (when ) or strictly decreasing (when ) in , the optimal (i.e., maximum or minimum) value of did not exist. Therefore, we should pay more attention to those explanatory variables involved in the first and the second scenarios.

Table 1- Coefficients of quadratic (i.e.,) and linear (i.e.,) effects as well as the p-value of the quadratic effects of explanatory variables with negative values in the logistic, probit, and cloglog models, where the best choice (i.e., maximizer) is calculated given by .

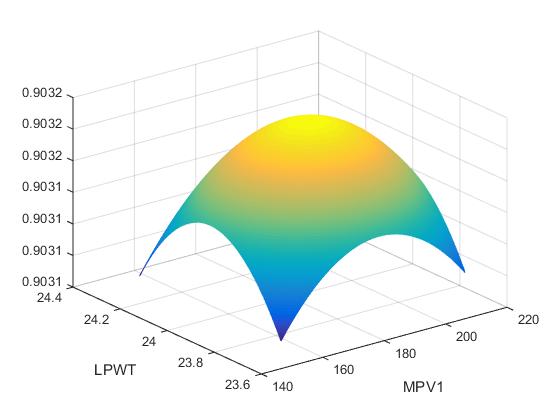
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Range** | **Model** | **Quadratic** | **Linear** | **Maximizer** | ***p*-value** |
| Cu | [0.003,0.06] | Logistic  Probit  Cloglog | -1312  -689.9  -535.5 | 53.62  27.29  20.71 | 0.0204  0.0198  0.0193 | 1.24e-5  8.48e-6  6.90e-6 |
| Ni | [0.001,0.017] | Logistic  Probit  Cloglog | -171600  -91540  -71550 | 640.2  337.1  259.8 | 0.0019  0.0018  0.0018 | 0.0005  0.0003  0.0002 |
| Sr | [0.0058,0.38] | Logistic  Probit  Cloglog | -103.2  -49.02  -33.99 | 26.14  12.57  8.789 | 0.1266  0.1282  0.1293 | 7.19e-8  4.72e-7  4.08e-6 |
| Al | [85.44,89.62] | Logistic  Probit  Cloglog | -1.273  -0.6009  -0.4088 | 227.2  107.3  73.07 | 89.24  89.28  89.37 | 0.0001  0.0003  0.0013 |
| Imp | [0,0.97] | Logistic  Probit  Cloglog | -6.177  -3.181  -2.504 | 8.859  4.364  3.190 | 0.7171  0.6859  0.6369 | 0.0033  0.0034  0.0036 |
| IS | [1,75] | Logistic  Probit  Cloglog | -1.424e-4  -7.376e-5  -5.572e-5 | 0.0102  0.0053  0.0041 | 35.81  35.92  36.79 | 0.0091  0.0120  0.0180 |
| HC5 | [0,155] | Logistic  Probit  Cloglog | -7.014e-5  -3.725e-5  -2.940e-5 | 0.0150  0.0081  0.0065 | 107.1  108.7  110.5 | 9.31e-9  8.69e-9  9.80e-9 |
| HC7 | [0,176] | Logistic  Probit  Cloglog | -3.358e-5  -1.831e-5  -1.474e-5 | 9.203e-3  4.757e-3  3.596e-3 | 137.0  129.9  122.0 | 5.79e-7  6.22e-8  9.70e-8 |
| MPV1 | [48,1599] | Logistic  Probit  Cloglog | -5.689e-7  -3.325e-7  -2.950e-7 | 2.212e-4  1.467e-4  1.493e-4 | 194.4  220.6  253.1 | 0.0012  0.0004  5.84e-5 |
| WF2 | [0,26.24] | Logistic  Probit  Cloglog | -1.561e-3  -7.684e-4  -5.608e-4 | 0.0180  0.0077  0.0047 | 5.765  5.010  4.190 | 0.0002  0.0007  0.0018 |
| WT4 | [0,74.25] | Logistic  Probit  Cloglog | -2.488e-4  -1.253e-4  -9.378e-5 | 0.0152  0.0077  0.0059 | 30.39  30.72  31.46 | 1.38e-8  2.76e-8  5.47e-8 |
| WT13 | [0,47.99] | Logistic  Probit  Cloglog | -1.563e-3  -8.205e-4  -6.433e-4 | 0.0395  0.0212  0.0170 | 12.65  12.92  13.21 | 1.63e-7  4.27e-8  1.01e-8 |
| LPWT | [0,41.78] | Logistic  Probit  Cloglog | -6.723e-3  -3.719e-3  -3.009e-3 | 0.3224  0.1786  0.1442 | 23.98  24.01  23.96 | 0.0010  0.0001  0.0001 |

## Discussions

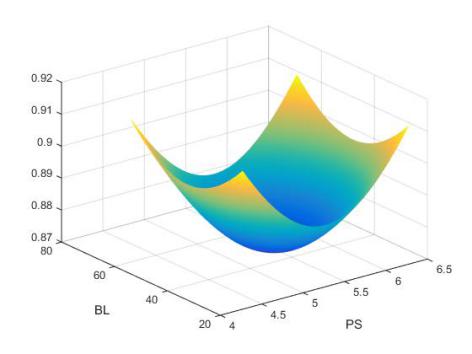
As *p* is determined by all of the explanatory variables contained in the final expression of , it is necessary to study their joint effects. The relationship between *p* and explanatory variables becomes complicated. It may have local maximizers, local minimizers, and saddle points. Local maximizers appear in the case when only variables in Table 1 vary. An example is given by Figure 1. As the coefficients of quadratic effects of LPWT (Low Pressure Water Temperature) and MPV1 (Minimum Pressure Vacuum 1) are both negative, a local maximizer appears. It provides the best pass rate based on LPWT and MPV1. Local minimizers appear in the case when only variables in Table 2 vary. An example is given by Figure 2. As the coefficients of quadratic effects of BL (Biscuit Length) and PS (Plunger Speed) are positive, a local minimizer appears. It provides the worst pass rate based on BL and PS. If variables in Tables 1 and 2 vary together, then saddle points may appear. An example is given by Figure 3. The point is treated as a maximizer in one direction or a minimizer in another direction. An optimal value (local maximizer, minimizer, or saddle point) is not present when the quadratic effect is absent. An example is given by Figure 4. As only the linear term of IP (Intensification Pressure) is significant, *p* is strictly decreasing in IP (because its coefficient is negative), indicating that we need to choose the value of IP as large as possible. As the relationship approximated by a statistical model may be imprecise when values of explanatory variables are outside of their observed windows, IP should be studied between its minimum and maximum values given by the data set. Overall, the statistical model defined by equations (1) and (2) can be used to interpret various relationships between *p* and explanatory variables. It is important to study the joint effects of three categories of parameters (or variables): the alloy composition, the process related to the shot sleeve, and the process related to die/mold). The classification of the three categories has been previously discussed [2].

***Table 2-Coefficiens of quadratic (i.e.,) and linear (i.e., ) effects as well as the p-value of the quadratic effects of explanatory variables with positive values in the logistic, probit, and cloglog models, where the worst choice (i.e., minimizer) is calculated by***

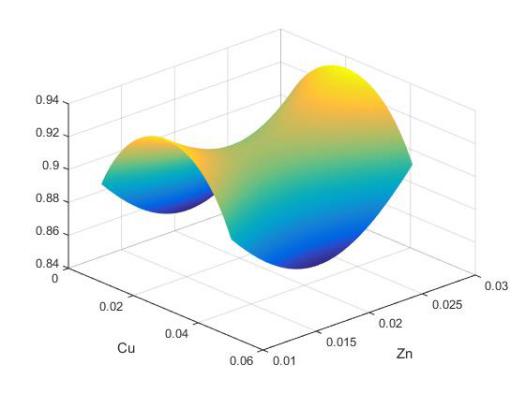
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Range** | **Model** | **Quadratic** | **Linear** | **Minimizer** | ***p*-value** |
| Mn | [0.506,0.686] | Logistic  Probit  Cloglog | 67.29  33.35  24.78 | -77.50  -38.39  -28.50 | 0.5610  0.5756  0.5751 | 1.51e-9  3.65e-9  5.58e-9 |
| Si | [9.288,11.42] | Logistic  Probit  Cloglog | 1.185  0.5464  0.3570 | -22.53  -10.32  -6.665 | 9.506  9.444  9.335 | 0.0014  0.0035  0.0115 |
| Zn | [0.013,0.113] | Logistic  Probit  Cloglog | 4592  2418  1855 | -192.7  -100.7  -76.35 | 0.0210  0.0208  0.0206 | 0.0001  0.0004  0.0004 |
| Ti | [0.045,0.131] | Logistic  Probit  Cloglog | 1332  648.4  432.0 | -177.7  -86.59  -57.79 | 0.0667  0.0668  0.0669 | <1e-16  <1e-16  <1e-16 |
| Ca | [0.0002,0.0132] | Logistic  Probit  Cloglog | 67770  33700  24820 | -508.1  -254.3  -188.2 | 0.0037  0.0038  0.0038 | 5.42e-9  5.58e-9  7.63e-9 |
| Al4 | [59.67,113.7] | Logistic  Probit  Cloglog | 1.329e-3  6.273e-4  4.268e-4 | -0.2877  -0.1357  -0.0923 | 108.2  108.2  108.1 | 0.0095  0.0165  0.0321 |
| BL | [12,147] | Logistic  Probit  Cloglog | 3.260e-4  1.572e-4  1.101e-4 | -0.0271  -0.0127  -0.0085 | 41.58  40.39  38.60 | 2.48e-14  3.30e-12  5.65e-10 |
| PS | [1.589.7.719] | Logistic  Probit  Cloglog | 0.2699  0.1411  0.1105 | -2.855  -1.483  -1.152 | 5.289  5.255  5.213 | <1e-16  <1e-16  <1e-16 |
| HC4 | [0,177] | Logistic  Probit  Cloglog | 3.721e-5  2.007e-5  1.604e-5 | -7.561e-3  -4.199e-3  -3.453e-3 | 101.6  104.6  107.6 | 1.05e-7  8.66e-8  9.79e-8 |
| MPV2 | [44,1599] | Logistic  Probit  Cloglog | 5.364e-7  2.831e-7  2.285e-7 | -8.388e-4  -4.316e-4  -3.369e-4 | 781.9  762.3  737.2 | 0.0011  0.0011  0.0007 |
| WF6 | [0,24.33] | Logistic  Probit  Cloglog | 1.436e-3  6.786e-4  4.782e-4 | -0.0180  -0.0086  -0.0064 | 6.254  6.337  6.692 | 0.0010  0.0025  0.0052 |
| WF7 | [0,24.99] | Logistic  Probit  Cloglog | 9.739e-4  5.159e-4  3.972e-4 | -0.0403  -0.0210  -0.0159 | 20.71  20.35  20.01 | 1.71e-5  5.49e-6  2.69e-6 |
| WF10 | [0.,29.77] | Logistic  Probit  Cloglog | 1.132e-3  5.648e-4  4.188e-4 | -0.0145  -0.0067  -0.0046 | 6.40  5.93  5.49 | 1.03e-5  2.52e-5  5.96e-5 |
| WF13 | [0,42.14] | Logistic  Probit  Cloglog | 8.431e-4  4.398e-4  3.448e-4 | -0.0351  -0.0181  -0.0140 | 20.80  20.58  20.30 | 3.07e-6  2.02e-6  1.01e-6 |
| WT3 | [0,90.07] | Logistic  Probit  Cloglog | 1.951e-4  1.064e-4  8.642e-5 | -5.921e-3  -3.179e-3  -2.547e-3 | 15.17  14.94  14.74 | 0.0005  0.0002  0.0001 |



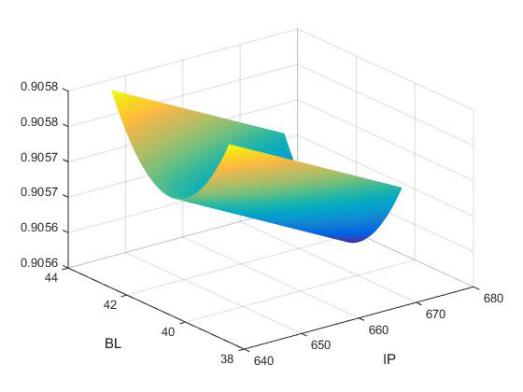
***Figure 1- Probability of PASS as functions of two explanatory variables when both quadratic terms are negative, where a local maximizer appears.***



***Figure 2- Probability of PASS as functions of two explanatory variables when both quadratic terms are positive, where a local minimizer appears.***



***Figure 3- Probability of PASS as functions of two explanatory variables when one quadratic term is negative and the other is positive, where a local saddle point appears.***



***Figure 4-Probability of PASS as functions of two explanatory variables when only one quadratic term is significant.***

## Alloy Compositions

The solidification range of the aluminum-silicon based alloys decreases with increasing silicon content for hypoeutectic allows. The smaller the freezing range, the smaller the shrinkage porosity during solidification. Shrinkage of liquid on cooling can be compensated by liquid feeding. The local minimizer means that many bubbles are likely to occur when Si is around 9.5. It recommends changing composition of Si to other values. One reason is that FCA has a strict control on the silicon content so the effect of silicon from melt to melt on the pass rate is minimized. By studying the quadratic effect, however, one can still find a way to adjust Silicon content to improve the pass rate. The signs of the estimates quadratic coefficients of Cu and Zn are distinct. A good result may be obtained by adjusting Cu only, but it may become worst if Zn also changes. It recommends that Cu should be adjusted with Zn together. The phenomenon can be more complicated if other alloy compositions parameters, such as Sr, Mn, and Mg [2], are involved. As neither the quadratic nor the linear effect of Mg is significant, it can be ignored in the study of the pass rate.

During solidification, the segregation of an alloying element is controlled by the partition coefficient of the element [3]. The partition coefficient is defined as the concentration of the element in solid divided by the concentration of the element in the liquid during solidification. The smaller the partition coefficient is, the severer the element segregating to the region last to solidify if no secondary phase is formed to consume that element. Si, Cu, and Zn are elements that have partition coefficients less than one. Cu and Zn form intermetallic phases at temperatures lower than that of the silicon phase. As a result, they tend to segregate at the liquid pools last to solidify. Solidification of the liquid in these liquid pools cannot be compensated, leading to the formation of isolated micro porosity which cannot be detected by X-rays. It might be of interest to adjust the contents of Cu and Zn in the alloy to disperse some of the dissolved hydrogen porosity into isolates small bubbles. Small sized pores are not as detrimental as large sized pores in terms of fatigue life of the casting. As the quadratic term of Fe is not significant, its coefficients are not given in Tables 1 and 2. The linear coefficient of Fe was -7.5 in the logistic model. Since Fe is very low already, further decreasing in iron can lead to significant increase in the costs of the alloy. Mn is added to enhance the soldering resistance of the alloy so it cannot be reduced, especially when the Fe concentration is low. Mg is an element that is added to strengthen the alloy by forming Mg2Si nanoparticles during T5 and T6 heat treatment. Note that Mg is not significant. Slightly increasing Mg can improve the quality of the alloy without reducing the pass rate in solidification. Thus, less sensitive variables to bubbles are still important if they can affect physical properties of alloys.

## PRoCESS PArameters (SHot SLEEVE)

Among the estimated coefficients for the process parameters that are related to the actions in the shot sleeve, the linear coefficient of CY (Cycle Time) is negatively significant and its quadratic effect is not significant, indicating that the pass rate increases if CY is shorter. Coefficients of quadratic effects of BL (Biscuit Length) and PS (Plunger Speed) are positive, but coefficients of IS (Intensification Stroke) are negative, indicating that saddle points of process parameters are present. More bubbles may appear if BL and PS are close to their minimizers. Less bubbles may appear if we set IS close to its maximizer. It is necessary to study their joint effects to determine the best combination of these values. Although quadratic effects of IP (Intensification Pressure) and FT (Fill Time) are not significant, the estimates of their linear effects (not shown) are all negative. Slightly decreasing IP and FT can improve the pass rate.

BL is a measure of the fill of the shot sleeve during the pouring stage. The longer the biscuit size, the high percentage the shot sleeve is filled, and the less change that air are entrapped when the plunger pushes the liquid metal into cavity. Slightly increasing or decreasing the BL value can enhance the pass rate. The increase of BL induces the decrease the CY and the increase of the erosion of shot sleeve. Note that the coefficient of CY is negative. A solution is to slightly decrease BL and simultaneously increase CY. It can improve the pass rate without increase of the erosion of shot sleeve. PS is proportional to the velocity of molten metal at the in-gate. It is well-known that the internal quality of a casting improves with increasing velocity at the gate. Our finding indicates that the quality becomes better if the speed is either higher or lower than 5.25 (the minimizer given by Table 1). The velocity at the in-gate has to be high enough to achieve atomization or to form a large number of liquid droplets to ensure high quality of a casting. When the velocity is lower than 5.25, then more bubbles may be present if we slightly increase the PS value. It is important to set up PS higher than its minimizer. After that, the pass rate can be enhanced by increasing PS values. The quadratic effect of FT is not significant. The linear coefficient of FT (-13.43) is negative. It is not large relative to variations of FT values. A decrease of FT can improve the pass rate but the impact is limited. The die cavity is filled with a fraction of a second at fast shot velocity. FT refers to the pouring of the molten metal to the shot sleeve and the time spent at slow shot velocities. An interpretation could be related to wave formation in the shot sleeve during slow shot profile. The formation of waves entraps air bubbles which may affect porosity in the casting.

## PROcess Parameters (DIE/Mold)

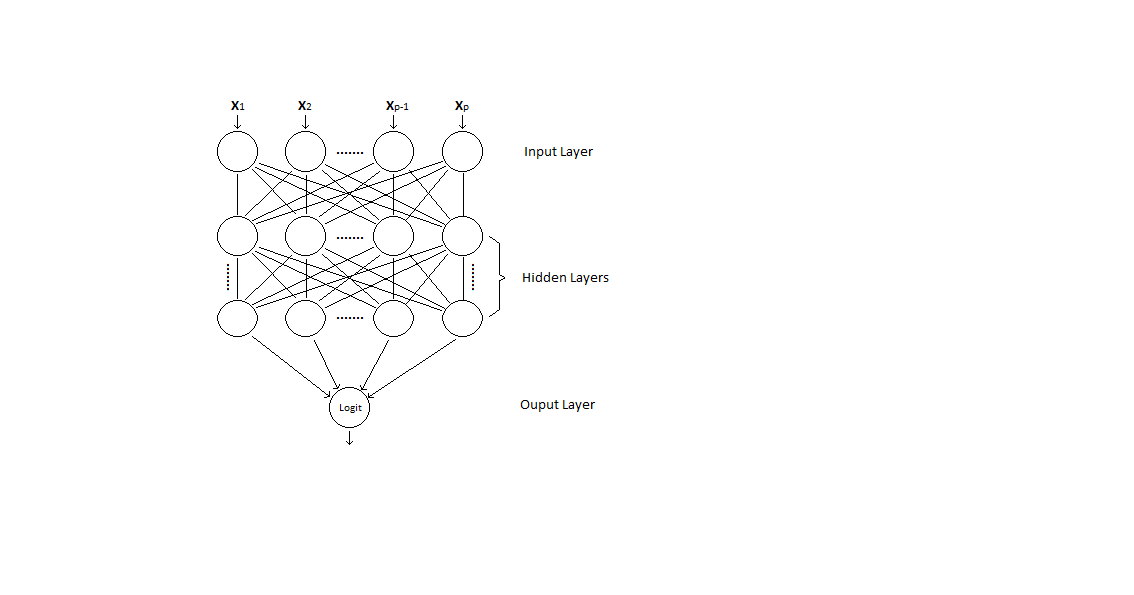
Among the process parameters that are related to the actions in the die/mold, quadratic coefficients of MPV1 (Minimum Pressure Vacuum 1) and LPWT (Low Pressure Water Temperature) are negative but the quadratic coefficient of MPV2 (Minimum Pressure Vacuum 2) is positive. Although their linear effects are significant, quadratic effects of IWT (Incoming Water Temperature), IWP (Incoming Water Pressure), HPWT (High Pressure Water Pressure), HPWF (High Pressure Water Flow), and Air Pressure (AIRP) are not. Saddle points are present as MPV1 and MPV2 vary. The determination of vacuum pressures can improve pass rate according to one measurement but not to both. The study of the joint effect is critical. Water is used for cooling the die/mold. Cooling lines are embedded in the dies at FCA casting plants. Changes of water pressure can cause changes of change of mold temperatures, which can reduce porosity and increase the soldering tendency if the core pins and dies.

# EXPERIMENT 2: Predict Pass/Fail using deep Neural Networks (DNN)

In this experiment, we used the popular Deep Neural Network (DNN) to construct and tune models to predict the quality of a part. This section begins with a short review of the DNN, followed by the construction of a DNN for the diecasting plan data. Results are examined and discussed at the end of this section.

## Introduction of Multi-layer perception

Neural Network technology is a powerful and flexible tool. It is widely accepted by many disciplines and is promised to become one of the foundational techniques for intelligent manufacturing [4][5]. The first basic brain-like formal model was presented by McCulloch and Pitts in 1943 [6]. Among different neural network tools, the most accessible applied network is the Multi-layer Perceptron (MLP). It usually has an input layer, one or more hidden layers and an output layer [7], as shown in Figure 5. Each layer is often connected using mesh architecture and the goal is to figure out the fittest parameters (weights of each link) to minimize the lose function [6]. In experiment 2, we constructed a few Deep Neural Networks to solve the binary classification problem using the data on crossmember castings collected at a FCA casting plant.



***Figure 5- The basic architecture of Multi-Layer Perceptron***

## Methodology of Deep Neural Network

The characteristic of FCA crossmember data shares similarity with semiconductor manufacturing data in that it is also highly imbalanced [8]. The widely accepted methods for solving extra imbalanced data classification problem include sampling methods, active learning methods, and cost-sensitive learning methods, and etc. [9][10]. For this extra imbalanced manufacturing data, we implemented an over-sampling technology (SMOTE) to enhance the accuracy of classification [11]. In addition, Exponential Linear Unit (ELU) is adopted as the activation function. Batch normalization was also employed in this DNN experiment to improve learning characteristics [12]. For the output of neural network, moving the threshold toward success-prone result so that the accuracy of prediction of failure-prone part classification could improve dramatically [13].

## Result and Discussion

The performance of deep neural network is shown in Figures 6 and 7. In Figure 6, the training loss of initial model was higher than 1 and it dropped very quickly. When the value of training loss was below 0.4, the rate of variance began to slow down. The value of testing loss was higher than the training loss initially and the reducing rate was slower than the training loss. Nevertheless, the reducing rate increased when the value was below 0.9. After reaching 0.4, the rate slowed down again. In the entire process of iterations, the testing loss was always higher than the training loss and the absolute value of their differences changed from enlarging to reducing.

In Figure 7, it shows initial training accuracy was only about 70% but it increased quickly. After reaching 0.96, this value begins to fluctuate. This may suggest the model was still not complex enough to fit the data. The testing accuracy started below 70% and it kept going down in the first 12 iterations (epochs). It then started to increase until it hit 90%. There was no noticeable improvement from epoch 70 till epoch 200. The final value of the test accuracy stops at 90.8%.

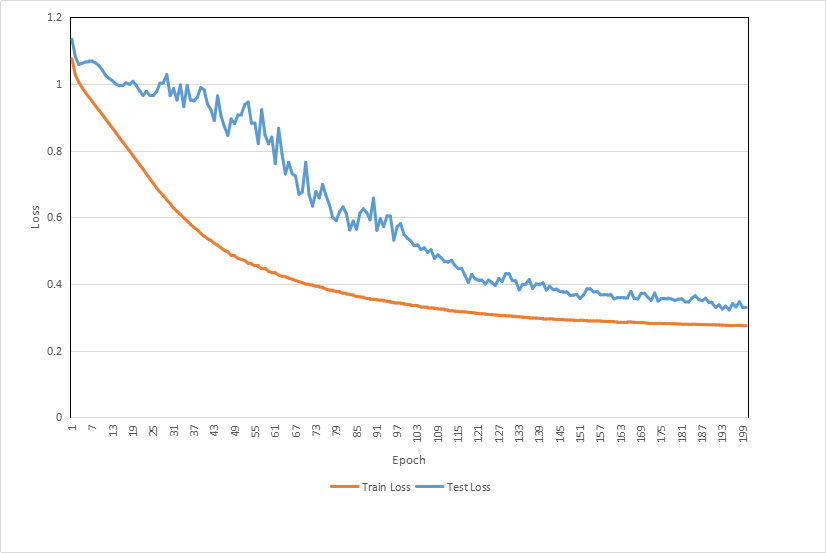


Figure 6- The training loss and testing loss of deep neural network

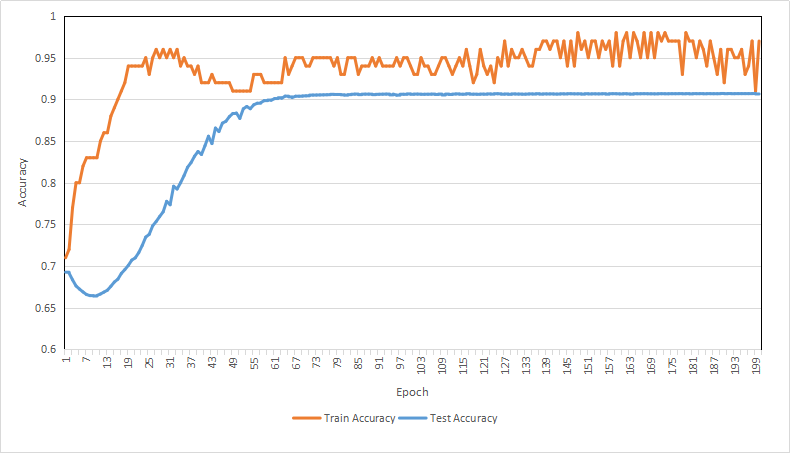


Figure 7- The training accuracy and testing accuracy of DNN

When the depth of Deep Neural Network is fixed, the more neuron nodes per layer, the stronger flexibility this DNN model would have [14]. For example, the 400-400-400-200 model has more variance and less bias than the 250-250-250-125 model [15]. This is also the case in our experiment. As shown in Figure 8, the lowest point of the testing accuracy of the 400-neuron model is higher than the models with less neurons. In addition, its rate of restoring is slower than others because it’s harder to adjust the model due to its complicity. Although these models reached the highest testing accuracy at diffident speed, their best testing accuracy were all limited to 90.8%.

Due to the time and resource constraints, we did not try more powerful models with hundreds or thoughts of layers. Finding and tuning DNN are certainly no easy task and the fittest models are often specific to each dataset. While disappointed by the mediocre results from the a few self-designed DNNs, we turned our attention to other models, as described in the next section.

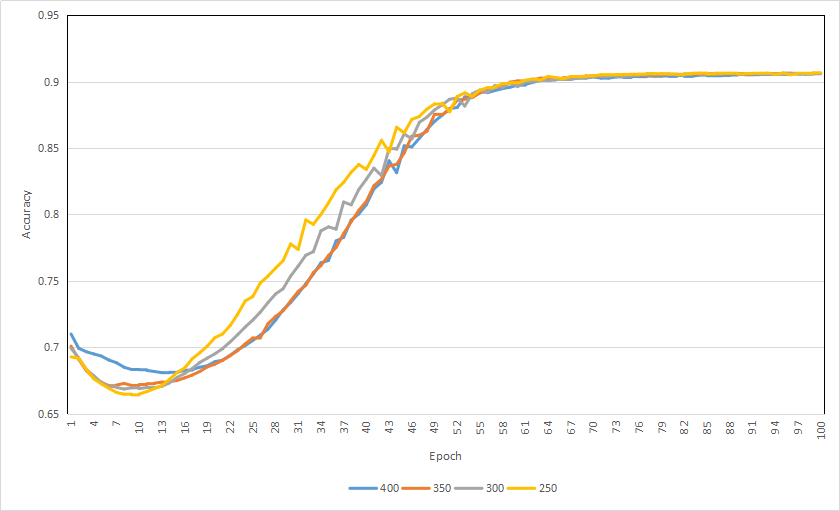


Figure 8- The test accuracy of DNNs with different neuron nodes

# EXPERIMENT 3: Predict Pass/Fail using XGBoot

From experiment 2, it is clear that tuning DNN is a daunting process of its own and the results may not be exciting until a fit model is discovered. In the current stage of machine learning, the advices has been try more models and find the fittest one. This is because the most complex model does not always produce the best results. In this section, we turn our attention to explore the ensemble statistical machine learning approach, XGBoost.

## Introduction of XGBoost

XGBoost is one of the most popular machine learning techniques among data scientists [16]. It is a parallel tree boosting system, following the principle of gradient boosting. It produces a strong prediction model from multiple weak learners, as shown in Figure 9. To do this, it optimizes the objective function of a sequence of decision trees by taking the advantage of gradient descent. XGBoost is designed to be highly effective and flexible as the algorithm is cache-aware and could be run on multiple cores.

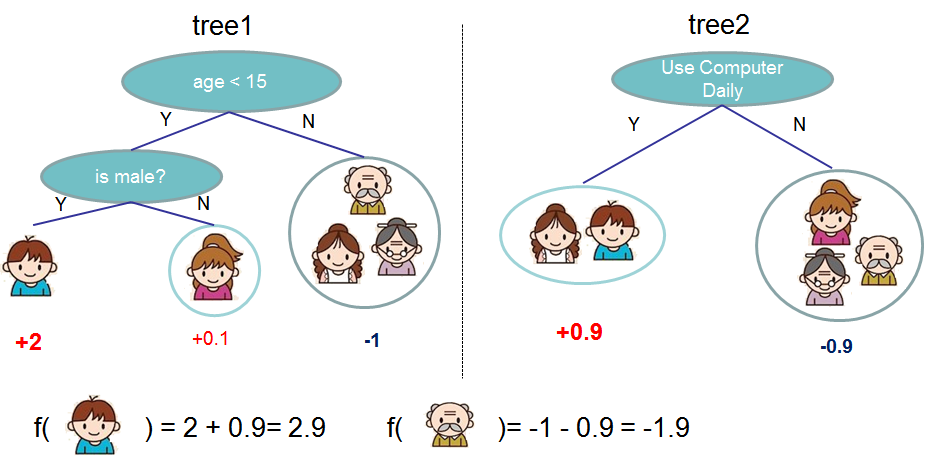


Figure 9– Weak trees form a strong model

## METHODALOGY of EXPERIMENT 3

Prior to training phase, we extracted desired features from the dataset. We first extracted “year”, “month”, “day”, “hour” features from the timestamp for each record. Then, we extracted “part of the day” feature (morning, afternoon, night, etc.) from the hour feature. After the time-related feature processing, we used one-hot encoding for all the categorical features. We got 96 features after feature extraction. Note again, the FCA crossmember diecasting data is imbalanced, the proportions of “PASS” records and “FAIL” records are roughly 0.9 and 0.1, respectively. Thus, SMOTE, an over-sampling technique, is applied to balance data for training phase. The test data part is not subjected to the over-sampling transformation to ensure the prediction resembles the real environment.

## Results and discussion

We split the dataset into training and test set in the ratio of 4:1. We tuned multiple parameters of XGBoost, e.g. “*learning rate*”, the step for gradient descent; “*n\_estimators*”, the number of trees used in the model; “*max\_depth*”, the depth of every tree model; “*max\_depth*” , the minimum sum of instance weigh needed in a child; “*colsample*”, subsample ratio of columns; and “*subsample*”, the proportion of train data used for training. All except the last two parameters are tuned sequentially. The last two parameters, “*colsample*” and “*subsample*”, are tuned jointly. And during the training phase, 5-folder cross validation (cv) is applied for every tuning. At last, we achieved an accuracy of 94% with 1,000 trees.

Figure 10 shows the training/cv accuracy on the dataset for “*n\_estimators*”. From the figure, we can find that both accuracy of both the training and the cv increase as the number of trees increases. After the number of trees exceeds 150, the accuracy increased very slowly, even though it’s still growing. When the tree number exceeds 350, the training accuracy still grows but the cv accuracy stop growing. This increase a bigger and bigger gap between the training accuracy and the cv accuracy, a sign of overfitting. To avoid overfitting, we set “*n\_estimators*” to 300.

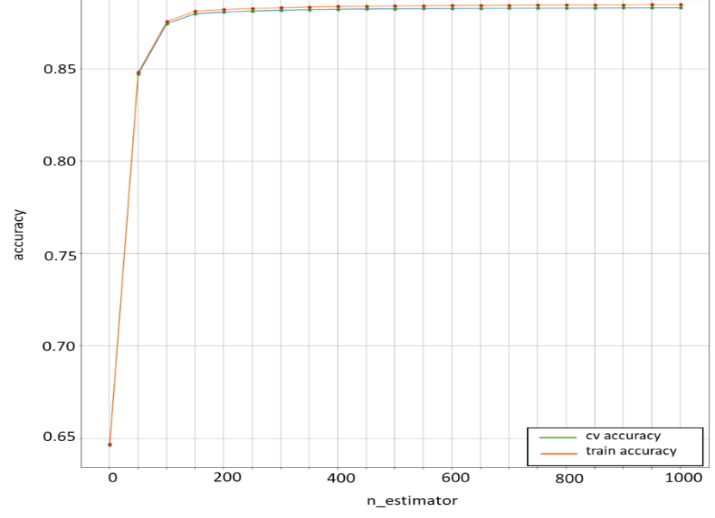


Figure- 10 Training/cv accuracy for n\_estimators

Figure 10 shows the accuracy plot for “*colsample*” and “*subsample*” tuning. The x-axis is the combination tuples of them. We chose 0.89, 0.79 for “*colsample*” and “*subsample*” as both training/cv accuracy are highest.

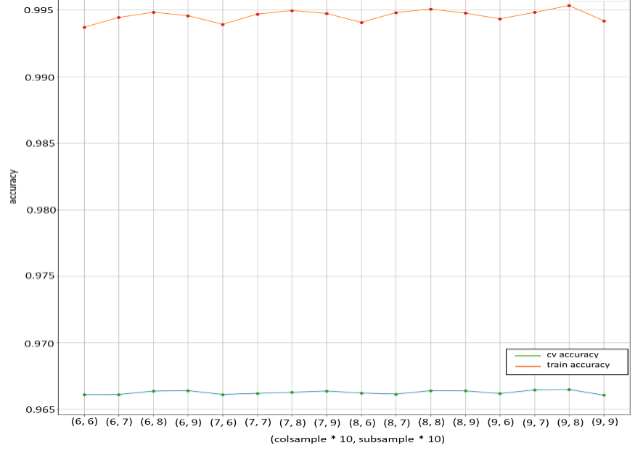


Figure- 11 Training/cv accuracy for colsample and subsample

Our final parameters are as follows:

* *n\_estimators*: 300
* *max\_depth*: 13
* *min\_child\_weight*: 5
* *subsample*: 0.89
* *subsample*: 0.79

It’s evident from this experiment that XGBoost has more predicting power than the limited DNN models we have explored in experiment 2. The results from XGBoost can be further improved if the data is cleaned by the domain experts who knows which row may have bad data.

# Concusions and Future work

This work examined a few prevalent data science techniques to better understand how it can lead to smart manufacturing for a diecasting plan. Our experiment suggests the traditional statistical analysis, if used correctly, can effectively and efficiently explain the cause effect in the production environment. If some of the variables can be controlled by the workers, then the statistic model should be used to improve the quality of a product. As an uprising star in the AI, DNN didn’t work well in our experiments. This is because tuning DNN is often long and a tedious job. In industry, it is not uncommon for an AI to spend over six months to find the best DNN structure. Given the time constraints of this project, our efforts were not able to reveal the magic power of DNN. On the other hand, XGBoost works very well on the structured data and should be considered by the practitioners for classification problems in diecasting plants.

Our study shows that pass rates of casting can be influenced by alloy composition and process parameters (either shot sleeve or die/mold). Some parameters have significant effects on air bubbles. These parameters must be carefully treated before casting. A few parameters may not have significant effects on air bubbles. If they have significant effects on physical quality of alloys, then adjustment of these parameters can significantly improve the quality without change the risk caused by air bubbles. It is important to have a nice way to classify parameters into bubble-related or quality-related in a casting process.

The results can be greatly improved if more data are collected, such as the humidity and temperature of the production environment. And missing data and outliers need to be purified by the professionals with the knowledge of diecasting manufacturing. In addition, one of the strengths in today’s AI is in computer visions. If the X-ray images of each part can be collected, then it is highly likely that the advanced Convolutional Neural Networks (CNN) can be leveraged to automatically ‘inspect’ a part.

# acknowledgments

Financial support by Fiat Chrysler Automobiles is acknowledged.

# REFERENCES

1. Faraway, J.J. *Extending the Linear Model with R*. CRC Press, Boca Raton, Florid, (2016)
2. Han, Q., McClure, D., Wood, D., and Yang, D. Statistical Analysis of the Effect of Operational Parameters on the Scrap Rates of Crossmember Casting, *North American Die Casting Engineer*, 38-43, November 2017.
3. Felmings, M.C., Solidification Processing, McGraw-Hill, USA (1974).
4. Joarder Kamruzzaman,Rezaul K.Begg, Ruhul and A.Sarker, *[Artificial neural networks in finance and manufacturing](https://books.google.com/books?hl=en&lr=&id=fUy9AQAAQBAJ&oi=fnd&pg=PR1&dq=artificial+neural+network+manufacturing&ots=Ukq6rCnkBt&sig=IqncTGVhJ27vmbTRrWWV2q_O3dQ)*, Idea Group Publishing, Hershey, PA, (2006)
5. Cihan H.Dagli, *[Artificial neural networks for intelligent manufacturing](https://books.google.com/books?hl=en&lr=&id=K4ftCAAAQBAJ&oi=fnd&pg=PP12&dq=artificial+neural+network+manufacturing&ots=XzUoqrCuN2&sig=hoSOwl6-uFs_znE1wpsLTuzw-Oc)*, 1st ed., Springer Science+Business Media Dordrecht, (1994)
6. Huang, S.H., Zhang, H.C., “Artificial Neural Networks in Manufacturing: Concepts, Applications, and Perspectives,” *IEEE Transactions on components, Packaging and Manufacturing Technology*-Part A 17(2), 212-227 (1994)
7. Kuo, R.J., Wang, Y.C., Tien, F.C., “Integration of artificial neural network and MADA methods for green supplier selection,” *Journal of Cleaner Production*, 18(12), 1161-1170 (2010)
8. Adam, A., Chew, L.C., Shapiai, M.I., Jau, L.W., Ibrahim, Z., & Khalid, M., “A Hybrid Artificial Neural Network-Naive Bayes for Solving Imbalanced Dataset Problems in Semiconductor Manufacturing Test Process,” *Hybrid Intelligent Systems (HIS), 2011 11th International Conference on*, pp. 133-138 (2011)
9. Haibo He, Edwardo A.Garcia, “Learning from Imbalanced Data,” *IEEE Transactions on Knowledge and Data Engineering*-Page 1263-1284, 1041-4347(2009)
10. Chawla, N.V., Japkowicz, N., Kolcz, A., “Editorial:Special Issue on Learning from Imbalanced Data Sets,” *SIGKDD Explorations Special Issue on Learning from Imbalanced Datasets* 6(1), 1-6(2004)
11. Jeatrakul, P., Wong, K.W., Fung, C.C., “Classification of Imbalanced Data by Combining the Complementary Neural Network and SMOTE Algorithm”, *ICONIP 2010.LNCS*, vol.6444, pp.152-159.Springer, Heidelberg(2010)
12. Clevert, D.-A., Unterthiner, T., Hochreiter, S., “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)”, *arXiv preprint* arXiv:1511.07289(2015)
13. Zheng J, “Cost-sensitive boosting neural networks for software defect prediction”, *Expert Syst Appl* 37(6):4537-4543, (2010)
14. [Aurélien Géron](https://www.google.com/search?tbo=p&tbm=bks&q=inauthor:), Hands-On.Machine.Learning.with.Scikit-Learn.and.TensorFlow, 1st ed., Published by O’Reilly Media, Inc., 1005 Cravenstein Highway North, Sebastopol, CA 95472(2017)
15. Gareth James,Daniela Witten,Trevor Hastie, Robert Tibshirani, An Introduction of Statistical Learning, Springer Science+Business Media New York, Vol 112, NY(2013)
16. Tianqi Chen, Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”. In *22nd SIGKDD Conference on Knowledge Discovery and Data Mining*, 2016.