# Development of Edge-based Deep Learning Prediction Model for Defect Prediction in Manufacturing Process

Kyung-Taek Lee
Korea Electronics Technology Institute
Seoul, Korea
ktechlee@keti.re.kr

Youn-Sung Lee
Korea Electronics Technology Institute
Seoul, Korea
yslee@keti.re.kr

Hyoseok Yoon
Division of Computer Engineering
Hanshin University
Osan-si, Korea
hyoon@hs.ac.kr

***Abstract*— In a series of lens module manufacturing process for mobile cameras, many process parameters affect the final quality of the manufactured lens module. Defect prediction of lens module in manufacturing process using many of these process parameters can help design efficient manufacturing parameters at an early stage of production. In addition, at an early stage of the manufacturing process, predicted defect products can be discarded without going to the next stage, thereby avoiding unnecessary additional costs. Many existing approaches use shallow architectures in their prediction models that cannot learn features in multi-parameters sufficiently. In this paper, we propose a method to improve productivity and reduce the manufacturing cost by predicting the product quality using deep learning. We designed prediction models using SVM, DNN and CNN approaches for quality prediction where CNN prediction model showed the best performance. Furthermore, to enable real-time defect prediction on-device to improve productivity, we propose a low-cost and edge-based solution that does not rely on expensive server or cloud solution.**

***Keywords—on-device AI, deep learning, defect prediction, manufacturing process***

## I. INTRODUCTION

Recent advances in Industrial Internet of Things (IIoT) [1–3] and the vision of Industry 4.0 [4, 5], enable efficient development and scalable deployment of productivity improvement, care services, and predictive maintenance [6, 7] on many different facets of Human, Machine, and Environment (HME) data for cyber-physical systems (CPS) [8]. Such intelligent systems benefit from well-orchestrated system integrations and often wireless connectivity technologies that catalyze massive data collection across interconnected components. In the lens manufacturing process for mobile cameras, the quality of the lens is affected by many process parameters. Defect prediction of lens in manufacturing process using these process parameters can offer the possibility of designing better manufacturing parameters at an early stage of production. In addition, predicted defective products at an early stage of the manufacturing process can be discarded without going to the next stage, thereby avoiding unnecessary additional costs. Many existing approaches fail at providing favorable results due to shallow architecture in their prediction models that cannot learn multi-parameters features sufficiently. In this paper, we propose a method to improve productivity and reduce the manufacturing cost by predicting the product quality using deep learning. Moreover, thanks to the huge advance in processor technology, recent mobile platforms are capable of over one Tera-flops [9]. In edge computing, mobile devices now serve a few core functions that were used to be handled only by data center servers. Edge computing is beneficial for many real-time applications such as object detection because it can provide a timely solution without needing to communicate with remote servers through network. We deployed our trained model on NVIDIA Jetson TX2 without any prediction accuracy loss.

## II. MANUFACTURING SYSTEM DESCRIPTION

### A. System description

Fig. 1 shows the manufacturing process for a smartphone camera lens module.

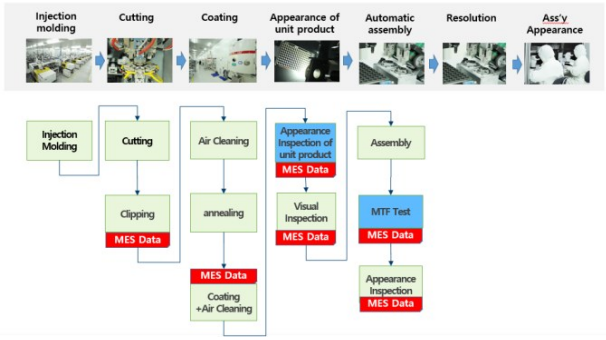
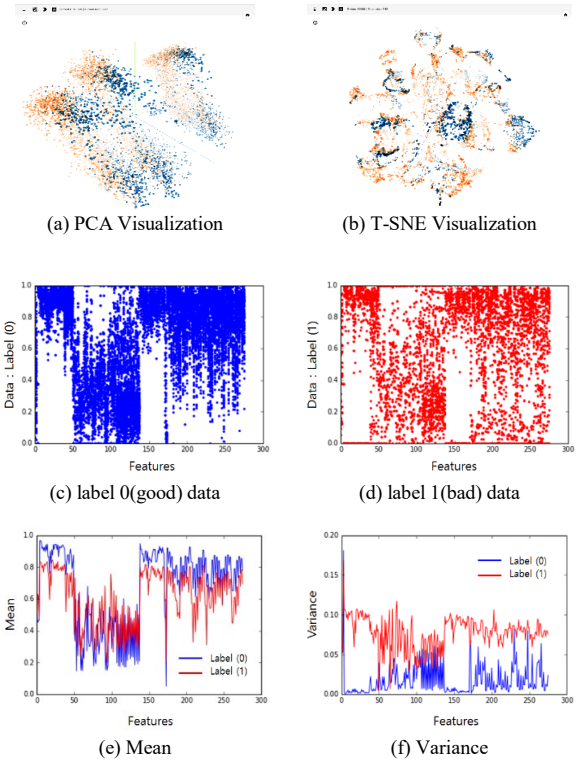



Fig 1. Manufacturing Process for Mobile Lens Module

The manufacturing process of lens modules consists of seven steps: injection molding, cutting, coating, appearance inspection of unit product, automatic assemble, resistance and assay appearance inspection. At the end of the manufacturing process, the MTF(Modulation Transfer Tunction) test equipment is used to check for a defect in lens modules. The defect rate of the lens module is a very important indicator directly related to the profit of the products. Whenever a change is made to the lens specification, an immediate defect rate becomes high and it is necessary to adjust and optimize the process parameters. At each step, the manufacturing equipment provides MES(Manufacturing Execution Wystem) data as process parameters and this MES data are collected by an IoT Gateway. Then the IoT gateway sends the collected data to the edge device that performs defect prediction.

### B. Data preparation

In the MES data collected from the manufacturing equipment, the number of process parameters for a lens module is 343. These parameters were converted into 272 dimensioned tensor by preprocessing. The following preprocessing was performed on the 343 parameters. First, the normal lens modules were labeled 0 and the defect lens modules were labeled 1. Missing parameters were found in the input data and the missing parameters were replaced with mean values of that parameters. The process parameters were normalized so that each parameter value ranges from 0 to 1. Fig. 2 shows statistics of the process parameters where Fig. 2(a) and Fig. 2(b) show the visualized process parameters through the PCA(Principal Component Analysis) and T-SNE(Stochastic Neighbor Embedding) algorithms. Fig. 2(c) and Fig. 2(d) represent the values of the parameters for good and defect lens module, and Fig. 2(e) and Fig. 2(f) represent the mean and variance of that 272 parameter values. The number of data prepared for training and testing are 200,000 and 39,884 respectively as shown in Table 1.




(a) PCA Visualization (b) T-SNE Visualization

(c) label 0(good) data (d) label 1(bad) data

(e) Mean (f) Variance

Fig 2. Statistics of manufacturing data

TABLE I. TABLE TYPE STYLES

| **Data** | ***No. of data*** |
|---|---|
| Parameters | 272 |
| Training data | 200,000 |
| Testing data | 39,884 |
| Total data | 239,884 |

## III. IMPLEMENTATION

Fig. 3 shows the implemented functional system diagram for defect prediction in manufacturing process of mobile lens modules. The MES data from the manufacturing equipment are collected by the IoT Gateway. Then the IoT gateway sends the collected data to the edge device that performs preprocessing and defect prediction. The preprocessing block processes synchronization between each stage data, missing parameter handling, normalization of data, and unnecessary parameter elimination, resulting 272 dimensioned tensor output.

### A. Prediction Model

The 272 dimensioned tensor data from the preprocessing block are inputs into the defect prediction model. If the model predicts "defect" at each step, the parameters of the process can be analyzed and changed to reduce the defect rate and the products that are expected to be defective can be stopped without going to the next stage, thereby avoiding unnecessary additional costs. We designed the prediction model to predict defects on the input data using SVM(Support Vector Machine), DNN(Deep Neural Network), and CNN (Convolutional Neural Network) approaches and optimized the model parameters for the best performance.

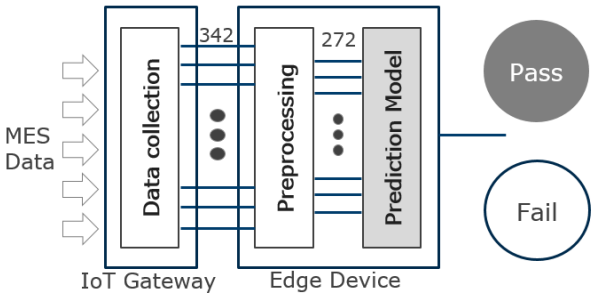



Fig 3. Functional System Diagram for defect prediction in manufacturing process

### B. Edge decvie platform

A commercially available Nvidia Jetson TX2 platform was selected as a target platform for the embedded optimization of the presented prediction model. The Nvidia Jetson TX2 development board features a low-power Nvidia Tegra X2 system-on-chip (SOC). The TegraX2 features an ARM-based CPU cluster and a powerful GPU (Pascal architecture) with 256 CUDA cores operating at up to 1,465 MHz. It can achieve a throughput of up to 0.75 TFLOPS in FP32 (1.5 TFLOPS in FP16). The CPU consists of a Denver processor (ARM-based) with two-cores operating at up to 2.0 GHz and an ARM Cortex-A57 with four cores running at up to 2.0 GHz. The SOC contains 8 GB LPDDR4 RAM that is shared between the CPU and the GPU. The maximum power consumption is 15 W. The standard operating system for the Jetson TX2 is a Nvidia variant of Ubuntu 16.04 (Nvidia Jetpack (Version3.2)) with Cuda 9.0 and CuDNN 7.0 support. Our inference model, data collection and preprocessing blocks were ported to this platform.

### C. Network optimization

One way to improve the performance of CNN model on embedded devices without losing accuracy is the removal

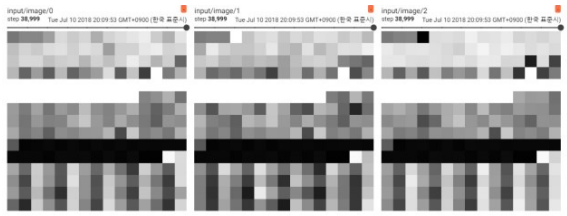
(pruning) of redundant operations/layers/information in the network. Most reduction approaches achieve this in an off-line, post-training approach. For this, the main approaches are kernel pruning and weight quantization. In coefficient/kernel pruning, coefficients/kernels that do not contribute to the network response in a significant way (for most inputs) are removed completely from the network. For weight quantization, the floating-point coefficients of a network are replaced by fixed point representations to which the input data and output responses are mapped [10]. Pruning and weight quantization can be utilized separately or in combination and typically lead to improved inference speed and reduced power consumption.

## IV. Results and Conclusions

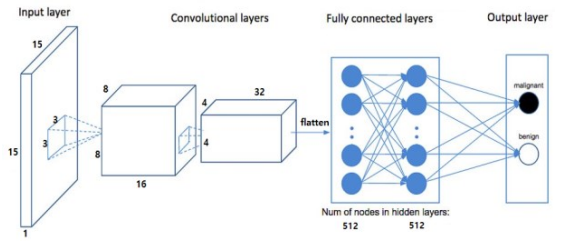
We optimized the parameters for the SVM, DNN, and CNN models and measured the prediction performance of each model. Table 2 shows the prediction performance for the SVM, DNN and CNN model. The CNN model shows the best prediction performance. Fig. 4 shows (a) the image of input data, (b) CNN model structure, (c) loss function and (d) prediction accuracy of CNN model. We deployed the trained CNN prediction model on NVIDIA Jetson TX2 without any prediction accuracy loss for real time defect prediction.

TABLE II. Prediction Performance of Models

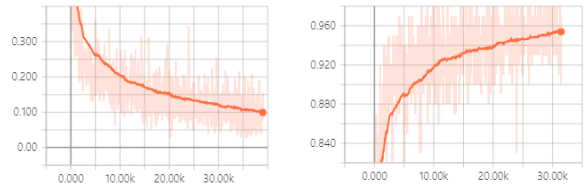
| Performance | SVM | DNN | CNN |
|---|---|---|---|
| Accuracy | 82 | 92.34 | 95.1 |



(a) Input data for CNN




(b) CNN model structure




(c)Loss function (d) Prediction Accuracy

Fig 4. CNN model and performance


## Acknowledgment

This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D program (N045400001, Development of Human and Machine Predictive Maintenance and Care Services based on Industrial IoT).


## References


[1] L.D. Xu, W. He, S. Li, IEEE Trans. Ind. Informat. 10, 2233 (2014)

[2] J. Wan, S. Tang, Z. Shu, D. Li, S. Wang, M. Imran, A.V. Vasilakos, IEEE Sensors J. 16, 7373 (2016)

[3] M.H.u. Rehman, E. Ahmed, I. Yaqoob, I.A.T. Hashem, M. Imran, S. Ahmad, IEEE Commun. Mag. 56, 37 (2018)

[4] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, B. Yin, IEEE Access 6, 6505 (2018)

[5] P. Zheng, H. wang, Z. Sang, R.Y. Zhong, Y. Liu, C. Liu, K. Mubarok, S. Yu, X. Xu, Front. Mech. Eng. 13, 137 (2018)

[6] J. Yan, Y. Meng, L. Lu, L. Li, IEEE Access 5, 23484 (2017)

[7] H. Yan, J. Wan, C. Zhang, S. Tang, Q. Hua, Z. Wang, 6, 17190 (2018)

[8] Yoon, H.; Lee, Y.S.; Lee, K.T. Human-Machine-Environment Data Preparation Using Cooperative Manufacturing Process Triggers. In Proceedings of the 2018 International Conference on Information Technology, Engineering, Science & Its Applications, Yogyakarta, Indonesia, 1–2 August 2018.

[9] Nvidia.com, NVIDIA Jetson Modules and Developer Kits for Embedded Systems Development, 2017.

[10] D. D. Lin et al., "Fixed Point Quantization of Deep Convolutional Networks," CoRR, vol. abs/1511.0, 2015