DOI: 10.21917/ijivp.2022.0380

ANALYZING THE SOFTWARE QUALITY IN IMAGE PROCESSING SOFTWARE IN INDUSTRY USING MACHINE LEARNING

B. Gopinathan[1], R. Kesavan[2], J. Senthil Murugan[3] and M.A. Mukunthan[4]

[1]*Jaya Sakthi Engineering, India*

[2]*Department of Computer Science and Engineering, Jaya Engineering College, India*

[3,4]*Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering, India*

Abstract

The ability of manufacturing organizations to generate defect-free, high-quality products is critical to their long-term success in the marketplace. Despite increased product diversity and complexity, as well as the necessity for cost-effective manufacturing, it is frequently important to conduct a thorough and reliable quality examination. There are bottlenecks in the manufacturing process because there are so many checks done. In this paper, we aim to automate the process of quality control in industries using a machine learning classifier that monitors the manufactured product namely the central processing unit via imaging technique. Development of a model with high quality control improves the productivity and efficacy of production that rejects the malignant and defect pieces from the supply chain. The use of imaging systems or high-speed camera enables the improvement of software quality, where the analysis is built using high clarity input images. The data processed by these imaging systems are transferred to the cyber-physical system for secured access within an organization. The results of classification of input images and process via machine learning improves the efficacy of the model over various machine learning models.

Keywords:

Software quality, Image Processing, Machine Learning, Cyber

1. INTRODUCTION

Throughout history, our reliance on technology has increased in importance dramatically. However, our relationship with information technology (IT) is significant in a number of fields, including healthcare. In response to our increasing reliance on technology, we have witnessed a tremendous shift in the way we communicate with one another. Before, computer systems were designed to run independently of one another, but this is no longer the case. Instead, computer systems are designed to work together. It appears that gadgets are being networked for a wide variety of purposes everywhere you turn. Sensors built into the majority of these devices allow them to communicate with one another and collect data. After the data is made public, people and other devices will be able to get their hands on it at any time [1].

In most workplaces, products are mass-manufactured in large quantities. The challenges of mass manufacturing include issues such as quality, efficiency, prices, and timeliness, all of which must be addressed. To address these difficulties, many firms have turned to automation as a solution. In order to preserve product quality, it is necessary to have a system in place at the end of the production line that scans finished products for imperfections [2]. In order to deal with the aforementioned challenges, the vast majority of industries have turned to automated systems that make use of image processing technology.

When it comes to completely industrial production, a system is absolutely essential. Because of the introduction of high-performance digital cameras and communication interfaces in the last several years, image processing has become both faster and more efficient in recent years. Along with the drop in the cost of image processing software, the quality of the programme has also increased [6].

The automation of the industrial manufacturing system, on the other hand, has its own set of challenges. There has been an increase in the value of time as well as the efficiency of the central processing unit (CPU) [7]. Product quality is extremely important to manufacturers, who place great value on it. In order for parts to be identified during the manufacturing process, it is important to build machines that can use computer vision [8].

In recent years, industrial vision systems have attracted a great deal of interest. Modern industrial machine vision inspection systems outperform and outlast their predecessors in terms of efficiency and resilience. Currently, academics are working on developing industrial vision systems that are more rapid, more efficient, and less expensive [9]. Industrial vision systems can be used to inspect the dimensions, surfaces, structural integrity, and operational quality of a product. In this section, the study look at the structural quality of the items, or, more specifically, how well they're put together, to determine their overall value [10].

The purpose of this article is to discuss quality control in the CPU manufacturing industry. It is possible that errors will arise during the building of CPUs, which will cause a significant number of components to be damaged or destroyed. There needs to be a method in place to ensure that the quality of these things at the end of the assembly line is consistent throughout the process. In this case, we advise building a method that gives the best potential level of production while taking the smallest amount of time possible. It is vital to thoroughly inspect each of the components in a CPU to ensure that they are all in proper working order. All of the devices that are involved must be able to communicate with one another in a reliable manner.

Managing everything that is connected to the system is the responsibility of the Cyber Physical Systems (CPS). Sensors, actuators, processors, and data services are all part of the system architecture and functionality. All operations must be carried out through the use of cloud-based computing resources. The cloud servers allocate bandwidth, storage, and energy in accordance with the needs of the users who are logging on to the servers and using them. Data management, monitoring, and dissemination in real-world contexts are all accomplished through the use of CPCS, which is a very sophisticated system. It is critical that this infrastructure is properly maintained and operated. Because of the large number of devices and organisations that are involved, even a tiny mistake can have a significant influence on the overall system. Industry can benefit from the use of sensors, actuators, and processors in CPSs to help them manage the intricacies of industrial operations.

2. RELATED WORKS

Researchers have investigated machine vision systems for checking the structural device quality in a variety of industries.. A number of automated visual examination approaches and technologies were discussed by Moganti et al. [3]. Template matching technology was first developed by Kim et al. [4] and further developed by Chung et al. [5] for the purpose of detecting irregularities in a car production line and inspecting automobile doors. There are a lot of systems in the literature that can do automated inspections of railroad lines.

Industrial image processing is a technique that is frequently used in inspections. A variety of image processing techniques are employed to enhance desired characteristics, improve image quality, and so on. Identifying patterns in the upgraded images is made feasible by the analysis of these images. Recognition of patterns and similarities in images is concerned with the identification of patterns and similarities in images that may be used to categorise the numerous things or components contained within an image. Over the last few years, analysis methods like fuzzy logic, evolutionary algorithms, and neural networks (NNs) have all been developed (GAs).

In recent years, there has been an increase in the use of machine learning algorithms in inspection systems. This method improves on the popular template matching and design-rule verification [11] methods. In order to accomplish their tasks, automated visual inspection systems rely on supervised learning techniques such as decision trees, statistical classifiers, artificial neural networks, and support vector machines (SVMs) [12].

3. PROPOSED SOFTWARE QUALITY MODELLING

The proposed framework is comprised of three fundamental components that facilitate the following process of predictive models. These components are as follows:

- Data Collection,
- Model Validation,
- Model deployment.

That phase of technological integration into the current IT infrastructure, which includes the predictive model, is not included in this contribution, nor does it constitute a separate contribution. This stage, on the other hand, is far too specific to be considered a component of a strategy that is universally valid and applicable. The layout of the proposed framework is depicted in Fig.1.

3.1 DATA COLLECTION AND PROCESSING

As a result, it is necessary to exclude non-representative data sets from the model database, such as those acquired during manufacturing trials, in order for the model to learn only regular patterns and dependencies from the remaining data. A key component of this step is determining and implementing appropriate methods to preprocess the data, such as identifying and treating outliers with a known unique cause, eliminating redundancies and inconsistencies, and removing missing values.

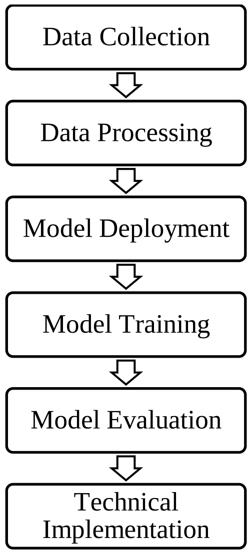

Fig.1. Image Software Quality Testing Architecture

3.2 MODEL TRAINING

A specific set of data is fed into the model training and scoring procedure in order to determine which model is the most successful. Two stages can be used to compare and choose models: training and tests.

To train the models, a tiering structure consisting of inner and outer cross validation as well as hyperparameter optimization is used. It is necessary to train and parameterize a number of supervised learning algorithms at the start of the model-building process in order to select the model with the highest performance. Different methods and algorithms must be evaluated for each unique application because it is not possible to make a broad selection of relevant algorithms for usage in all situations. Selective criteria such as complexity, interpretability, and speed, or the data scientist knowledge as well as insights and results from past efforts, must be considered during the pre-selection process. For the model-based quality inspection application under consideration, prediction time and possible precision, both of which are connected with model complexity, are more relevant considerations than other factors. Although characteristics such as data volume, diversity, and velocity have no effect on algorithm performance, they do have an impact on its efficiency. The mathematical structure of an ensemble technique suggests that a wide range of algorithms be used to look at how well they work with the data set that was given. This will be the main point of this contribution.

It is easy to distinguish between examples that have been successfully classified (true positives and true negatives) and examples that have been incorrectly classified (false positives and false negatives) (false positive, false negative). As a result of this differentiation, a number of statistical quality metrics can be calculated. In the context of classification models, accuracy is defined as the proportion of properly classified examples among the total number of examples. Because of the imbalanced representations of classes that are common in quality-related industrial applications, this measure would result in high accuracy but no added value for class differentiation when used in these applications, and as a result, it should be avoided when used in these applications. As a result, the true-positive-rate and the false-positive-rate are more useful metrics for evaluating models than they were previously. Receiver operating characteristics (ROC) curves are used to show TPR and FPR in a graphical format.

True positives are dictated by the intensity of the inspection process, and false negatives are also dictated by the intensity of the inspection process. Models with their ROC-curves in the upper left corner of the model are thought to be the best models to look at. However, in addition to the technique used, the amount of time it takes to score is dependent on the hardware and software that are being used to execute the game. Most of the time, the reaction time is long enough that the scoring time constraint has little or no effect on the model selection process.

3.3 MODEL DEPLOYMENT AND IMPLEMENTATION

The deployment and execution of models are covered in this section of the document.

We tested the proposed approach after constructing an image dataset of products from the CPU manufacturing line and comparing it to the dataset. Four well-known classifiers were used to categorise the data, and we then compared the results to those obtained by our own method.

We have images of both defective and non-defective products in our database. Following the previous description, the potential faults in an image are retrieved through the use of fuzzy image segmentation and pattern recognition methods, which are described in greater detail below. The data collection involves the use of more than 150 bags (50 positive and 100 negative). There are 271 instances of positivity and 253 instances of negativity in this collection.

- **Sensors**: The sensors for the visual quality control system are provided by cameras.
- **Hardware Specification**: This is a data-intensive technique, parallel computers are the most appropriate hardware configuration. Because of their large data rates, such applications necessitate a significant amount of bandwidth and computational power. PCs with industry-standard parts, on the other hand, can handle most industrial inspection tasks.
- **Communication**: To allow for automation of the production process, image processing must be integrated into the manufacturing process and the results relayed to the necessary authorities. Therefore, additional devices must be able to communicate with one another as a result of this. This system makes use of cloud-based data storage as well as CPCS network connectivity to function. With the use of the cloud, there are a lot of security risks that need to be taken into account.

The model is integrated into the inspection planning process as a result of the organization commitment to the process. As a result, the inspection strategy determines the extent to which the ML model is involved in inspection planning and design. Hybrid techniques, according to the current state of development, have the ability to outperform normal inspection principles, even when they are based only on a prediction model that requires a high level of trust in the model and extremely high precision.

Inspection reliability is ensured through the use of both conventional inspection and quality prediction techniques in tandem. Using quality prediction in quality assurance, it is possible to reduce physical inspection volume while maintaining inspection reliability. This can help to generate more added value. This capability is located upstream of the regular inspection process, and it can be used to dynamically adjust the inspection volume based on the results of the forecast.

There are two approaches to this problem that can be taken, depending on how trustworthy the model is. If the forecast result was non-defective, then only those parts that were physically inspected; if the forecast result was defective, then the reverse is true, and all parts were physically inspected (defective). The increased proportion of alternative prediction classes in the final analysis is responsible for the reduction in inspection volume seen.

The conclusion is that, based on the real requirements and available resources, it is hard to make broad statements about how the technological solution should be designed and implemented. Individual configurations can be based on an overall framework, which can be used as a starting point.

The amount of information that can be accessed is increasing at a rapid rate, and this trend will continue. Because of the networked nature of sensors and devices, data may be accessed at any time and from any location. It should be noted, however, that this new discovery comes with two big drawbacks: high-dimensional and large volumes of data, on the other hand, is that there are a multitude of criteria that go into selecting what kind of equipment you'll need. In this case, it is important to do a thorough analysis before starting the installation and integration process.

In terms of the specifics of how these constraints will be set and enforced, it impossible to make broad statements about the future. Another way to think about how the different parts of the deployment work together is to pay attention to the stated requirements and linkages.

When the processes and elements is taken into consideration, the dimensionality of the data set is considerably reduced, and vice versa. Evaluation of data volume for model training and application must be done separately from the rest of the evaluation. The volume of model applications is governed by the number of classifications performed simultaneously, whereas the volume of training and optimization data is determined by the volume of preceding data sets. The hardware decides what the CPU can do and how much memory and power it needs.

4. RESULTS AND DISCUSSION

Using an imbalanced sample of 4,000 data points, we first performed preliminary training and parameter optimization, and then we ran 5-fold cross validation on the supervised learning algorithms SVM, GB, DT, GBT and LR, respectively.

The findings are tested of accuracy, standard deviation, recall, precision, training time, and scoring time, among other things (Table.1-Table.4). This phase does not include a description of the hardware that was used because the absolute times are not yet essential.

Table.1. Accuracy of various ML on Imaging Software Quality

Model	Accuracy	Recall	Precision
Naïve Bayes	83.5%	94.7%	75.5%
Decision Tree	88.2%	91.9%	84.0%
Logistic Regression	71.9%	77.0%	66.8%
Support Vector Machine	92.9%	96.4%	89.3%
Gradient Boosted Tree	92.6%	89.9%	93.1%

Table.2. Standard Deviation on Accuracy

Model	Standard Deviation
Naïve Bayes	±2.7%
Decision Tree	±1.5%
Logistic Regression	±1.3%
Support Vector Machine	±1.3%
Gradient Boosted Tree	±1.0%

Table.3. Computational time

Model	Training Time (ms)	Testing Time (ms)
Naïve Bayes	4	11
Decision Tree	41	7
Logistic Regression	51	29
Support Vector Machine	302	362
Gradient Boosted Tree	3	41

When evaluated side by side, SVM and GBT performed significantly better in terms of statistical performance. When comparing the scoring times for the GBT with the SVM, the GBT was eight times faster on average. For this reason, GBT was chosen over a system called SVM. It has the ability to grow and can classify a lot of solder joints at the same time.

Table.3. Computational Cost per VM and Scalability

Model	Cost ($)	Scalability Index
Naïve Bayes	21.20	70%
Decision Tree	20.54	72%
Logistic Regression	20.51	76%
Support Vector Machine	18.62	84%
Gradient Boosted Tree	17.03	91%

As part of the hybrid inspection strategy that depends on the prediction model, it is possible to further optimise the GBT model parameters in order to reduce the number of false negatives, hence lowering the inspection volume even further. To determine how much money could be saved by lowering the frequency of false positives, multiple levels of conservativeness were evaluated to see how much could be saved. As a result, false negatives are severely penalised under the high-conservative method, which results in a low percentage of false negatives but a large percentage of false positives, limiting the amount of inspection work that may be saved. Instead of rejecting false negatives in large numbers, the low-conservative model accepts only a small% of false negatives in order to save money on imaging software.

In this phase of the case study, the goal was to come up with a plan for deployment and inspection that was both cost-effective and technologically feasible. The modelling phase showed that there was a strong correlation between parameter values and imaging software.

5. CONCLUSION

In order to be successful in today global marketplace, manufacturing organisations must be able to fully and consistently examine their products before shipping them. The application of machine learning can be used to create inspection procedures that are both efficient and cost-effective. An image processing system examines product photos to determine if there are any flaws. The information gathered as a result of this is then submitted to the appropriate party for further consideration. The most challenging element for our system is extracting the attributes of each instance, which is a time-consuming process. If the incorrect feature extraction processes are not used, the entire visual inspection system is put in danger of malfunctioning.

Another facet is the introduction of new product types and variants to the mix. Because each product variety has its own set of characteristics and defect patterns, it is necessary to train and optimise new models for each one. As a result, a greater amount of research into model management and selection procedures that incorporate some level of automation is necessary as a result. Also, the cloud-based data lake will be linked to data-storage systems, which will allow for the creation of labelled training data for use in training and improving models.

REFERENCES

[1] M.H. Zawawi, F.C. Ng and N.H. Hassan, "Particle Image Velocimetry Analysis on the Liquid-Sediment Model", *Proceedings of International Conference on Climate Change and Water Security*, pp. 201-206, 2022.

[2] B. Guan, W. Kang and H. Lin, "Freshness Identification of Oysters Based on Colorimetric Sensor Array Combined with Image Processing and Visible Near-Infrared Spectroscopy", *Sensors*, Vol. 22, No. 2, pp. 683-694, 2022.

[3] M. Moganti, F. Ercal and C.H. Dagli, "Automatic PCB Inspection Algorithms: A Survey", *Computer Vision and Image Understanding*, Vol. 63, No. 2, pp. 287-313, 1996.

[4] K.H. Kim, Y.W. Kim and S.W. Suh, "Automatic Visual Inspections System to Detect Wrongly Attached Components", *Proceedings of International Conference on Signal Processing Applications and Technology*, pp. 13-16, 1998.

[5] S. Sledz and M.W. Ewertowski, "Evaluation of the Influence of Processing Parameters in Structure-from-Motion Software on the Quality of Digital Elevation Models and Orthomosaics in the Context of Studies on Earth Surface Dynamics", *Remote Sensing*, Vol. 14, No. 6, pp. 1312-1324, 2022.

[6] B. Vineetha and R.B. Madhumala, "Providing Security and Managing Quality Through Machine Learning Techniques

for an Image Processing Model in the Industrial Internet of Things", *Proceedings of International Conference on Smart IoT for Research and Industry*, pp. 161-177, 2022.

[7] G.U. Nneji, J. Cai, J. Deng and C.C. Ukwuoma, "Multi-Channel Based Image Processing Scheme for Pneumonia Identification", *Diagnostics*, Vol. 12, No. 2, pp. 325-334, 2022.

[8] J.L. Salgueiro, L. Perez and C. Miguez, "Microalgal Biomass Quantification from the Non-Invasive Technique of Image Processing through Red–Green–Blue (RGB) Analysis". *Journal of Applied Phycology*, Vol. 32, pp. 1-11, 2022.

[9] X. Tian, Y. Li, D. Ma and L. Xia, "Strand width Uniformly Control for Silicone Extrusion Additive Manufacturing based on Image Processing", *The International Journal of Advanced Manufacturing Technology*, Vol. 11, No. 2, pp. 1-14, 2022.

[10] J. Wang, Y. Yang and Y. Hua, "Image Quality Enhancement using Hybrid Attention Networks", *IET Image Processing*, Vol. 16, No. 2, pp. 521-534, 2022.

[11] B. An and Y. Kim, "Image Link Through Adaptive Encoding Data Base and Optimized GPU Algorithm for Real-time Image Processing of Artificial Intelligence", *Journal of Web Engineering*, Vol. 65, No. 1, pp. 459-496, 2022.

[12] V. Laghi, V. Ricci, F. De Santa and A. Torcinaro, "A User-Friendly Approach for Routine Histopathological and Morphometric Analysis of Skeletal Muscle using Cell Profiler Software", *Diagnostics*, Vol. 12, No. 3, pp. 561-568, 2022.