SURFACE DEFECT DETECTION WITH MACHINE LEARNING

Prema N S, Prashanth M V, Scinchana S Kumar, Shamitha Raj, Rinla C Mary, Yashwanth S

Department of Information Science Engineering Vidyavardhaka College of Engineering Mysuru,INDIA

ABSTRACT

A product's surface quality is its most important characteristic. Manual defect inspection is a laborious task in an industry. As a result, it is challenging to provide the certainty of an impeccable steel surface. Machine Learning based automatic surface investigation strategies have emerged as a highly effective and popular solution in recent years to suit customer requirements, speed up the inspection process, and increase the industry's overall efficiency. To overcome this issue, we used a conventional machine learning strategy. With the help of image preprocessing techniques, this project aims to improve the model's performance. It then trains and creates a machine learning model to segment and categorize faulty images using the features that were extracted. The NEU provided the input. The NEU surface defect database serves as the source of the input. Six different forms of flaws are included in this database: crazing, inclusion, patches, pitted surfaces, rolled-in scale, and scratches to divide and organize photos of defects.

Keyword : - *Surface Defect, Machine Learning, Classification, Random Forest, Convolution neural network.*

1. INTRODUCTION

Surface defect identification is a crucial step in industrial production that has a big impact on the standard of industrial goods sold. Conventionally, product inspection is mainly conducted by human experts, and this can be time consuming and mistake prone. For example, a human inspector may overlook certain defective products due to long hours of eye strain or other factors. To reduce the burden on human inspectors, we can use a computer to perform product inspection based on methods proposed in the context of pattern recognition. Patterns may include images, sounds, smells, etc. In this paper, we consider image-based product inspection.

The following are compromised since the typical manual detection approach takes a long time and is inaccurate due to the inspector's subjectivity, vigour, and will compromise the following client satisfaction. It is also demonstrated in figure1.

- The cost of production.
- The efficient use of resources.
- Cost of inspection.
- Price-fixing and marketing

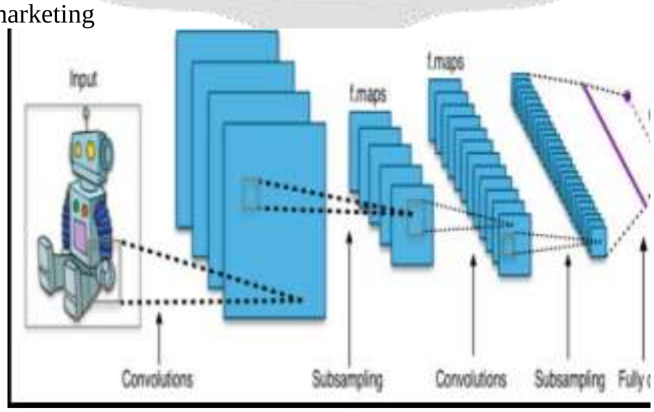

Figure 1: Manual process of metal detection

Thus, it becomes necessary to implement an effective and efficient system of surface defect detection. This objective can be achieved by implementing Surface defect detection using Machine Learning technique.
Currently, flaws and their types are maintained manually using books or ledgers. Humans do quality control, which can occasionally be taxing and imprecise. Cons of the Current System are :

- Manual Method
- Time-
- Expensive
- Inaccurate
- Unsatisfactory consuming to users
- Effective

The classification and identification of flaws in metallic surfaces is offered as an industrial application. It is crucial to continuously improve quality control in all facets of industrial production in order to maximise it within the constraints of possibilities and resources. Applying descriptive or predictive mining to historical data on accidents that have happened along with other crucial information generates an intriguing alternative that could produce beneficial and gratifying outcomes for all parties involved. The suggested solution offers a cutting-edge method of quality control. Accurate classification of metal surface flaws is achieved using machine learning and picture segmentation.
The framework used is streamlit. Web application development, machine learning, and data science are all permitted. We need to have access to the data for machine learning. The database was acquired from the NEU metal surface fault database. We must organise, purge, and categorise our data in order to teach a machine what we need and don't need. Get rid of useless information, material that is missing, and anything ambiguous. We use a machine learning algorithm for data instead of individuals to interpret large amounts of data as is the case with conventional software development. A trained model is used to accomplish the same because the main goal is to recognise and pinpoint the fault in the metal area.

With the knowledge that there is a relationship between the input and the output, we may be presented with a collection of data in the case of supervised learning and already know what the ideal outcome should contain. Supervised learning may be a learning technique that enables us to tackle difficulties even when we have little to no prior knowledge of how our problem may emerge. Unsupervised learning does not produce a response based on a predictable result. Creating tasks that have either a success or failure outcome is a reinforcement learning technique. Although named and unlabelled data are usually used to prepare a limited amount of data, semi-supervised learning sits in the midst between supervised and unsupervised learning.

1.1Objectives

- The proposed system is an application that runs in real time.
- To effectively forecast and categorize the fault, the proposed approach unearths powerful rules concealed in these common item sets.
- The proposed approach predicts and categorizes new images using surface defect images.
- The proposed approach use "machine learning" to accurately forecast metal surface flaws. Proposed system makes use of “machine learning” to correctly predict defects on the metal surface.

2. LITERATURE SURVEY

The quality of manufactured goods is very readily impacted during the industrial production process due to defects and limitations in current technology, bad working conditions, and other issues. One of the simplest methods to determine if a product's quality has been damaged is through surface flaws. In order to guarantee the qualification rate and consistent quality, product surface defects must be detected [1]. A "defect," in general, is an absence, flaw, or area that deviates from the typical sample. The comparison between the normal sample and the faulty sample of industrial goods is shown in the figure. Surface defect detection refers to the detection of scratches, defects, protection against foreign bodies, colour contamination, holes, and other imperfections on the surface of the sample to be tested[2]in order to gather various pertinent details such as the category, contour, location, and size of surface defects from the sample to be tested. In the past, manual defect detection was the norm; however, because results are easily impacted by human subjectivity, it is ineffective and unable to meet real-time detection needs[3].

Businesses can identify metal surface flaws in order to maintain product quality standards and aid in excess production. With this work, we put forth three machine learning (ML) classifiers-Convolutional Neural Network, Random Forest, K Nearest Neighbour to identify, detect and classify the deformity and defect within the dataset. To format photos, information must first be pre-processed. The models are then used to train fault detection classification assignment with the best weights and biases for ML computation at that time. Also, the three models' quality is assessed using a variety of categorization report criteria.
Different strategies have eventually replaced them. Currently, various academics are conducting pertinent study on surface defect identification that considers the most recent approaches, applications, significant issues, and numerous other elements [4]. The literature compares and discusses the advantages and disadvantages of the techniques, and looks at defect detection technology in electronic components, piping, soldering parts, machine parts, and typical quality control applications[5]. It also summarises the state of the art in defect detection techniques, including magnetic particle inspection, liquid penetrant inspection, eddy current inspection, ultrasonic inspection, machine vision, and deep learning [6, 7]. Surface defect detection has been done in the literature using the supervised learning model method [6], the unsupervised learning model method [8], and other ways[6] , including the semi-supervised learning model method and the weakly supervised learning model method. Techniques based on deep learning are described. The following highlights three key concerns about real time, small sample numbers, and comparison with traditional image-based surface defect detection methods. The literature meticulously discussed automated optical (visual) inspection (AOI) technology before describing numerous related procedures and concepts[9].
The literature [10] first discusses the many components of failures before comparing deep learning methods to traditional technology for defect detection. The next section discusses applications of deep learning and ultrasonic detection to defect identification. After researching existing applications, many challenges for defect detection based on the defect detection technique are given. These difficulties include the need for three-dimensional target.
identification, high positioning accuracy, quick target detection, and small targets, among others. In the subject of surface defect identification for industrial items, there is currently limited literature review on machine learning techniques, according to research. Although though various publications list the issues and challenges associated with finding surface flaws in industrial items, their recommendations are not sufficiently systematic. Additionally, the data sets needed to detect surface flaws in industrial items are not yet fully compiled. As a result, this study first covers the research state of surface defect identification of industrial items utilising conventional machine vision methods and deep learning methods in order to address the issues. Then, in the process of identifying industrial surface flaws as well as imbalanced sample issues, it analyses the key issues, such as real-time issues, tiny sample issues, and small target issues.

3. METHODOLOGY

One of the most popular machine learning techniques is classification, which employs a model learned from training data to forecast the class of new samples. Assigning or categorising data samples to class labels in a data set is what classification is generally understood to be. The method of classifying data involves two steps. A classification model is created in the first stage utilising data from a predetermined data set, sometimes referred to as the training set. The model is then used to test or forecast the class labels for newly discovered data in the second step. The data set used to assess the trained model or function's classification capabilities is known as the test set. We employ a Random Forest classifier for picture classification. The data set that classifies the defect as cracks, inclusions, patches, pitted surfaces, rolling flakes, and scratches will be sent to the classifier.

Model 1-Random Forest Classifier:
The supervised learning approach includes the well-known Random Forest machine learning algorithm. It can be used to solve classification and regression problems in ML. It is built on the concept of collaborative learning, a technique for combining various classifiers to handle challenging situations and enhance model performance. As its name implies, Random Forest is a classifier that averages several decision trees applied to various subsets of the supplied data set to improve the predicted accuracy of the data set. Random Forest predicts the outcome based on the votes of most of the predictions rather than depending just on one decision tree. More precision and excessive adjustment are avoided by the forest's larger number of trees, as shown in figure 2.

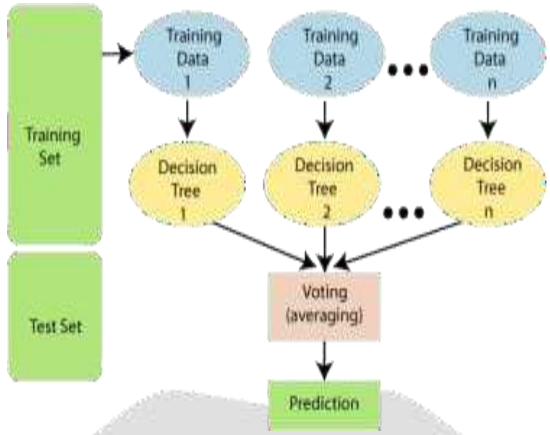

Figure 2: Random forset

How does the Random Forest algorithm work?

Predictions are made for each tree that was created in the first phase after N decision trees are joined to create the random forest. The steps that make up the work process are as follows:

- The first step is to randomly select K data points from the training set.
- Create the decision trees connected to the selected data points in step two (Subsets).
- Step 3: Choose N to represent the size of the decision trees you wish to construct.
- Step 4: Repeat Steps 1 and 2.
- Step 5: Determine each decision tree's predictions for the new data points, and then assign new data points to the category that receives the most votes.

Model 2-CNN Classifier:

A deep learning neural network called a convolutional neural network, or CNN, is made to analyse structured arrays of data, like representations. When it comes to identifying patterns in the input image, such as lines, gradients, circles, or even eyes and faces, CNNs are quite effective. The convolution neural network is extremely strong for machine vision because of this characteristic. CNN does not require any pre-processing and can operate immediately on a raw image. A direct neural network, which seldom reaches 20, is a convolutional neural network. A convolution layer, a particular kind of layer, is what gives a convolutional neural network its strength.

It's crucial to realise that black and white photos are scanned substantially differently from colour images before we go into the specifics of this process. Before examining each of them individually, let's first examine the similarities.

Following are some similarities between the two sorts of images:

- Each pixel contains 8 bits (or 1 byte) of data. Scales ranging from 0 to 255 are used to represent colours. This is because a byte can have any of 256 (28) possible values because bits are binary units and there are 8 of them in a byte. Since we count 0 as the first possible value, we can go up to 255.
- In this approach, 0 represents absolute black and 255 represents absolute white, with different (certainly more than 50!) degrees of grey falling in between.
- Each pixel can hold up to 1 byte (8 bits) of data. On a scale from 0 to 255, colours are represented. This is due to the fact that bits are binary units, and since there are 8 of them in each byte, a byte has a total of 256 (28) possible values. Since we consider 0 to be the first conceivable value, we can go as high as 255.

- Images in colour are three-dimensional, whereas images in black and white are two-dimensional. This results in a different value being assigned to each pixel when it is sent to the neural network. Each pixel's hue in two-dimensional black-and-white graphics is represented by a number between 0 and 255.

Yet, each pixel is represented in a vibrant image at three separate levels. Since each of these colours is a blend of red, green, and blue in varying concentrations, each pixel in a colour image is given a separate value for each of these layers. The consequence is that each of the layers—red, blue, and green—is shown with a value between 0 and 255. Then, they are shown in RGB format. For instance, the neural network would see a "hot pink" pixel as (255, 105, 180). After processing, the image is grouped and put into several categories. In this module, the CNN model's output is shown.

4. RESULT & DISCUSSION

Random Forest predicts the outcome based on the votes of many of the predictions rather than depending just on one decision tree. More precision and excessive adjustment are avoided by the forest's larger number of trees. The model's accuracy was between 67% and 70%. The CNN model assigns the input image to one of the six flaws (Crazing, Inclusion, Patches, Rolling, Pitched Surface, Scratches) The accuracy acquired using CNN is 92-96% .

5. CONCLUSIONS

We can easily see that the machine learning approach performs admirably in both image classification and segmentation. The hot-rolled steel sector can use the classification model developed for this project to identify product flaws in real time. The photos of the output (rolled steel) are continually taken throughout the production line and supplied into the ML model. There is a flaw in the goods if this photograph has been marked as a defect image. Therefore, this product component needs to be removed for further processing. Quality engineers should determine if the product needs to be reworked or destroyed based on the defect levels. However, the initial quality check would have been laboriously expensive and time-consuming if done manually. As described in this project, machine learning can automate this.

6. REFERENCES

1. Artzy, E., G. Frieder, and G.T. Herman, *The theory, design, implementation and evaluation of a three-dimensional surface detection algorithm.* Computer graphics and image processing, 1981. **15**(1): p. 1-24.
2. De Vitis, G.A., P. Foglia, and C.A. Prete, *Row-level algorithm to improve real-time performance of glass tube defect detection in the production phase.* IET Image Processing, 2020. **14**(12): p. 2911-2921.
3. Neogi, N., D.K. Mohanta, and P.K. Dutta, *Review of vision-based steel surface inspection systems.* EURASIP Journal on Image and Video Processing, 2014. **2014**(1): p. 1-19.
4. Rasheed, A., et al., *Fabric defect detection using computer vision techniques: a comprehensive review.* Mathematical Problems in Engineering, 2020. **2020**: p. 1-24.
5. Ma, Y., et al., *Adaptive segmentation algorithm for metal surface defects.* Chinese Journal of Scientific Instrument, 2017. **38**(1): p. 245-251.
6. Tsai, D.-M. and T.-Y. Huang, *Automated surface inspection for statistical textures.* Image and Vision computing, 2003. **21**(4): p. 307-323.
7. Liu, J.-W., Y. Liu, and X.-L. Luo, *Research and development on deep learning.* Application Research of Computers, 2014. **31**(7): p. 1921-1930.
8. Allekotte, I., et al., *The surface detector system of the Pierre Auger Observatory.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2008. **586**(3): p. 409-420.
9. Richter, J., D. Streitferdt, and E. Rozova. *On the development of intelligent optical inspections*. in *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*. 2017. IEEE.
10. Cao, W., Q. Liu, and Z. He, *Review of pavement defect detection methods.* Ieee Access, 2020. **8**: p. 14531-14544.