CHAPTER 5

DESIGN

Systems design is the process of defining the architecture, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering.

5.1 Corrosion Detection Architecture

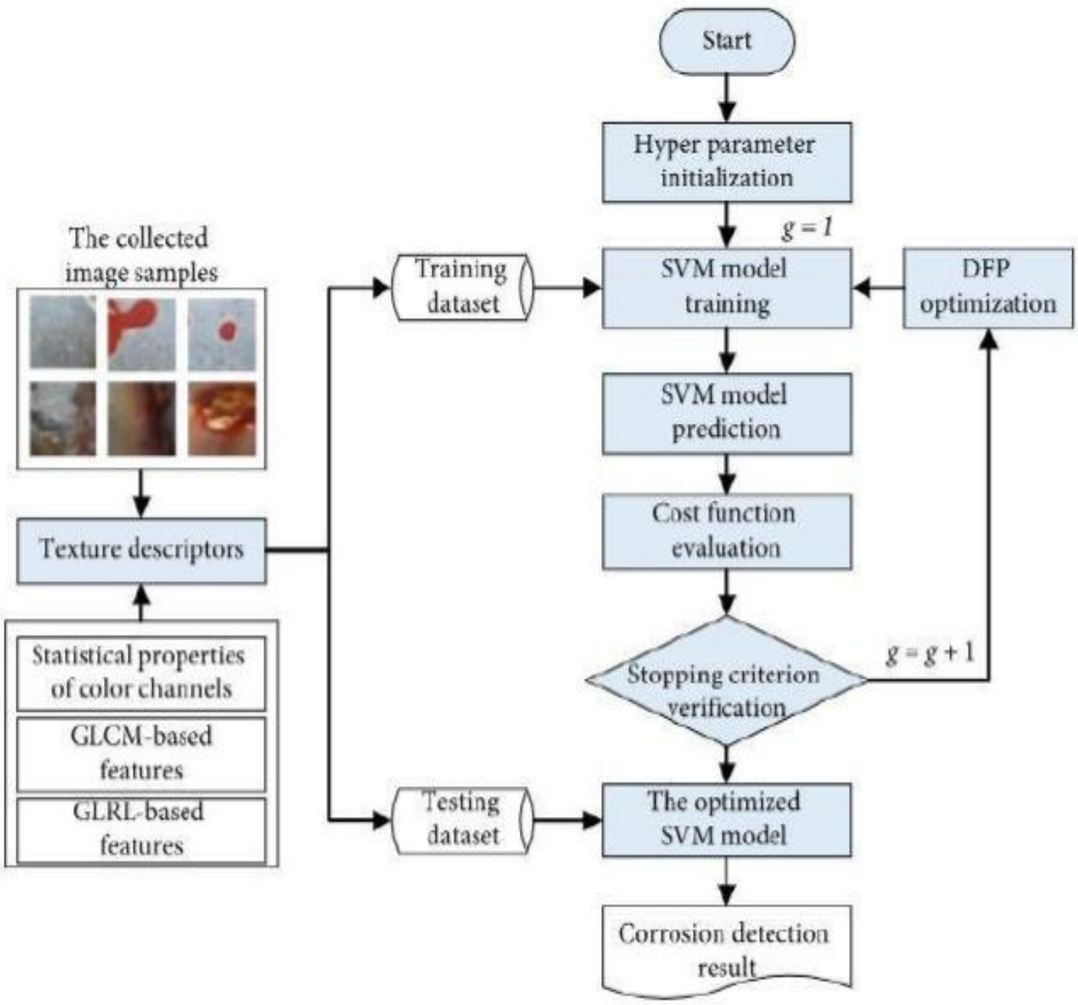

Figure 5.1: Corrosion Detection Architecture

The architecture describes the structure of the newly developed hybrid model of image processing and metaheuristic-optimized SVM for pipe corrosion detection. The proposed model, named as MO-SVM-PCD, is a combination of image texture analysis and a metaheuristic-optimized machine learning approach. As mentioned earlier, the statistical measurements of color channels, GLCM, and GLRL are used to extract texture-based features from image samples. These image datasets are then divided into training and testing dataset which is used to train the SVM model and then to test the trained SVM model to get the accuracy.

5.2 MODULE DESCRIPTION

A module description provides detailed information about the module and its supported components, which is accessible in different manners.

The Modules in this system are:

• Filtration of image.

• Identifying text descriptors of the image.

• Pre-processing by the Z-score data normalization method.

• Pattern recognition and classification of the image.

1. **Filtration of image:**

- Removal of noise by image filtering techniques:Noises present in an image is caused by sharp and sudden disturbances in the image signal. Median filtering technique is used to remove salt and pepper noise present in an image.
- Transformation of color space from RGB to HIS: Image is converted to HIS color space because identification of corrosion in the image in RGB color space becomes computationally expensive and cumbersome.

2. **Identifying Text descriptors of the image:**

- Statistical Properties of color Channels: Mean, standard deviation, skewness, kurtosis, entropy and range are calculated for particular image.This paper reflects the statistical characteristics of the three channel colors of a sample of pictures (red, green and blue). This is why an image in a RGB format is defined. In addition to RGB, other color spaces such as HSV can help detect corrosion. However, in this study, we are based on the original digital camera RGB color model. In this paper, we extract six features in three different colors (Red, Green, and Blue). The six statistical properties extracted from the images are Mean, Standard Deviation, Skewness, Kurtosis, Entropy, and Range. All these six properties are extracted at three different colorchannels, thereby extracting (3*6)18features.

- Grey-Level Co-Occurrence Matrix (GLCM): This is a widely used procedure for studying an image. The first step is to convert a color image to a grayscale image. Let $\delta = (r, \theta)$ be a variable quantity represented in the coordinate. In this paper, we will be extracting 4 features at 4 different angles. The 4 features extracted are Angular Second Moment, Contrast, Correlation and entropy, each of these features are extracted from the image for different angles namely {0°, 45°, 90°, 135°}. Therefore, a total of 4*4=16 features are extracted from theimage.

- Grey-Level Run Lengths (GLRL): This procedure is proposed by Galloway for texture classification. This approach is very effective for differentiating textures of various fineness and was successfully applied in different fields of research. A run-length p (i*j) matrix is calculated in a certain direction. This can be used to measure short-term stress (SRE), long-term concentration (LRE), Gray-level non-homogeneity (GLN), run-time non-homogeneity (RLN), and run-percentage (RP). Chu also proposed the low-gray running emphasis (LGRE) markers and high-gray running accent (HGRE). Dasarathy and Holder suggested that low gray short- term accent (SRLGE), high-gray short-term emphasis (SRHGE), low-gray long-term emphasis (LRLGE) and long-gray high-gray long-term focus (LRHGE) be calculated. Table 1 summarizes the above measures. Consequently, 11 features for 4 different angles {0 °, 45 °, 90 °, 135 °} which gives a total of 44 features are obtained.

3. **Preprocessing by Z-score data normalization method:**

- It converts all indicators to a common scale with an average of zero and standard deviation of one. Handles the outliers.

4. **Pattern recognition and classification of image:**

- SVM: SVM is a robust method for recognition of patterns for statistical education theory. Since the input function is classified in to two categories: −1 (non-corrosion) and +1 (corrosion), an SVM model is responsible for constructing a decision-making spatial area by using a hyperplane it divides the input into two different regions. The goal of the SVM algorithm is to define a decision boundary in order to achieve the greatest possible distance between groups. SVM also uses the kernel to turn a nonlinear classification function into a linear classification function. First, an SVM model maps input data to a hyperplane which can separate the data from the original room. After extracting all the 78 features using the image processing technique, we train our SVM model with the 78 features. As explained above SVM model will map the input data and using hyperplane creates two different regions (corrosive and non-corrosive). The model will try to identify a pattern from the above features extracted and will use it fortesting.

5.3 COMPONENT DESIGN

5.3.1 Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. A DFD is often used as a preliminary step to create an overview of the system without going into great detail, which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design).

A DFD shows what kind of information will be input to and output from the system, how the data will advance through the system, and where the data will be stored. It does not show information about process timing or whether processes will operate in sequence or in parallel, unlike a traditional structured flowchart which focuses on control flow, or a UML activity workflow diagram, which presents both control and data, flows as a unified model.

Data flow diagram (DFD) graphically representing the functions, or processes, which capture, manipulate, store, and distribute data between a system and its environment and between components of a system. The visual representation makes it a good communication tool between User and System designer. Structure of DFD allows starting from a broad overview and expands it to a hierarchy of detailed diagrams. DFD has often been used due to the following reasons:

- Logical information flow of the system
- Determination of physical system construction requirements
- Simplicity of notation
- Establishment of manual and automated systems requirements

DATA FLOW DIAGRAM LEVEL 1

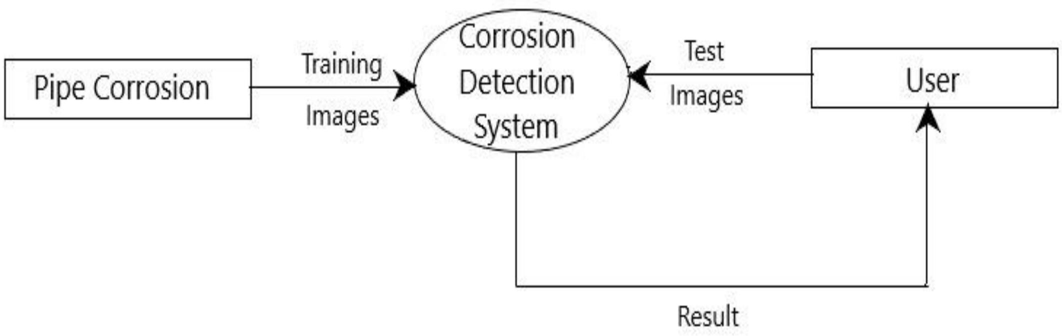

Figure 5.2: Data flow diagram for corrosion detection of pipelevel 1

First when the idea of proposing the model was taken, The main aim was to build a Support Vector Machine classification model which could classify the images, The model was built using the pipe images taken as training set and finally the user here gives test images for testing the model So as to see whether if the model correctly classifies the test pipe images.

DATA FLOW DIAGRAM LEVEL 2

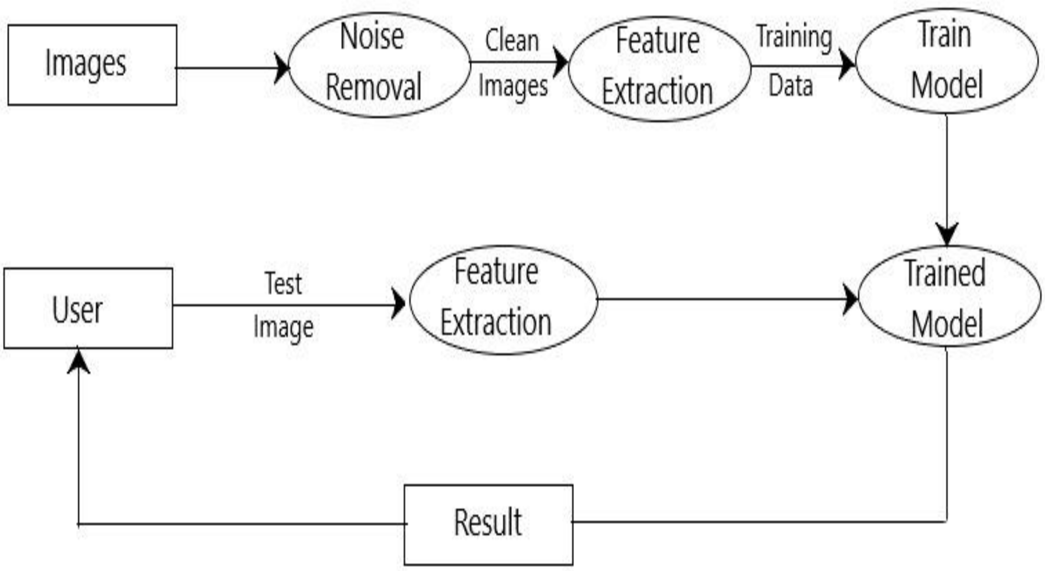

Figure 5.3: Data flow diagram for corrosion detection of pipelevel 2

The pipe images collected from publically available datasets are Raw images,These raw images are then preprocessed i.e scaled to particular size and noise removal/Normalization and then these processed images are converted to an Pixel array form , and these arrays are further converted to compatible form which are used for training the model along with the label. Using these training set a SVM model is constructed and the model is saved , Now the user gives the test images which are also preprocessed and transformed but without the label to the model, Thus the model is able to classify the image of pipe as corroded or not corroded.

DATA FLOW DIAGRAM LEVEL 3

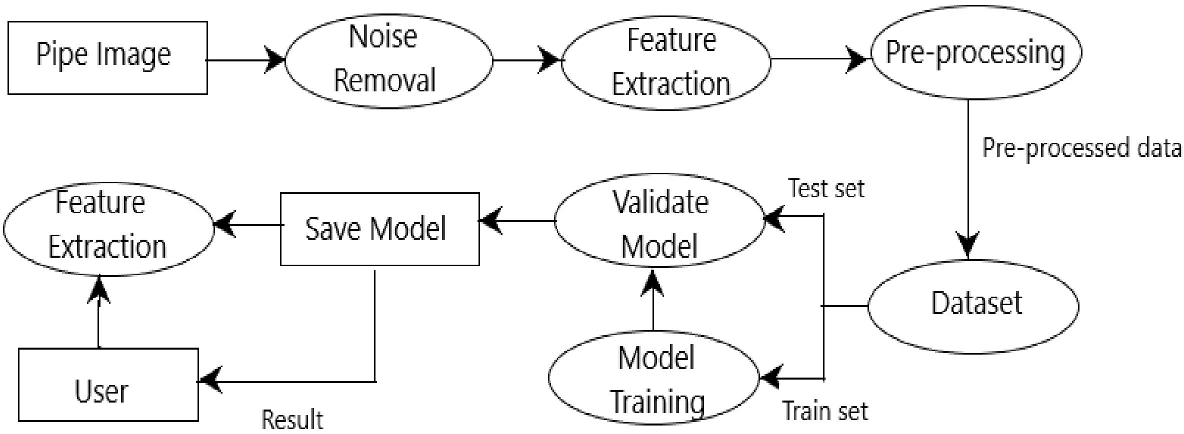

Figure 5.4: Data flow diagram for corrosion detection of pipelevel 3

Images are collected from the public available datasets, Then the datasets are pre-processed i.e scaling etc followed by pre-processing is the feature Extraction from the Images by statistical properties of color channels, GLRL, GLCM.The available Dataset is split into 2 parts namely Training set (80%) and testing set (20%).

Using the training set and the learning algorithm a SVM model is trained, Using this Model, a validation model is constructed which obtains the test data as input for checking the validity of the model, Finally result analysis is performed and if the accuracy conditions are met the model is saved.

5.3.2 Sequence Diagram

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function.

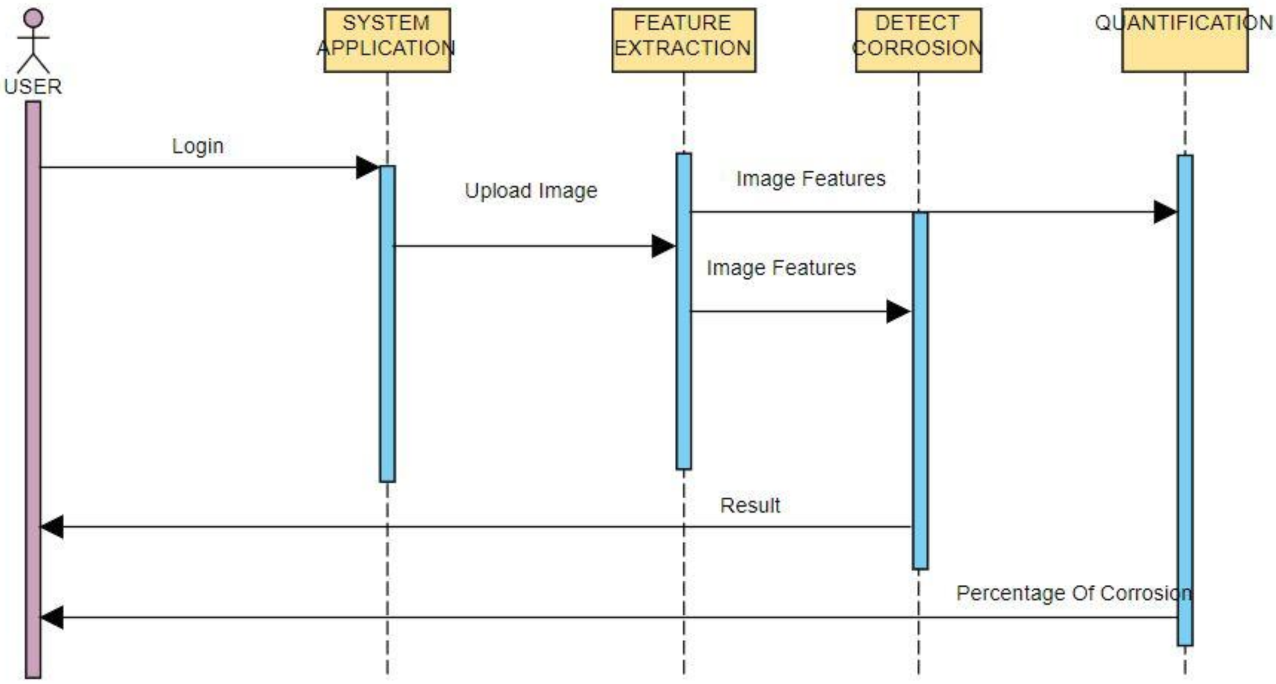

Figure 5.5: Sequence Diagram