Journal of Information Technology Management
ISSN-OnLine: 2980-7972
University of Tehran

Ball Screw Surface Defect Impact Level Categorization Using the Collective Constraint Intelligent Prediction Technique

Soorya Prabha Mohan*

*Corresponding author, Research Scholar, Department of Computer Science and Engineering, National Institute of Technology, Trichy, Tamil Nadu, India. E-mail: m.sooriyaprabha@gmail.com

S Jaya Nirmala

Associate Professor, Department of Computer Science and Engineering, National Institute of Technology, Trichy, Tamil Nadu, India. E-mail: sjaya@nitt.edu

Journal of Information Technology Management, 2025, Vol. 18, Issue 1, pp. 34-52
Published by the University of Tehran, College of Management
doi: https://doi.org/10.22059/jitm.2026.106253
Article Type: Research Paper

Received: August 23, 2025
Received in revised form: September 16, 2025
Accepted: December 28, 2025
Published online: January 20, 2026

Abstract

Ball screw failures often exhibit early signs of surface defects, which can arise from contamination and indicate deterioration. Identifying these surface defects is crucial for minimizing repair costs and maximizing machinery uptime. This study presents a prediction algorithm based on transfer learning methods to enhance the accuracy of ball screw surface defect detection throughout its lifecycle. We investigate four transfer learning models: CCIP-V, CCIP-D, VGG16, and DenseNet, utilizing image data mining techniques for defect classification. These models are validated using a specialized dataset of ball screw surface defects, employing Region of Interest (ROI) masking techniques to enhance image classification for each model. Our findings reveal that the proposed hybrid approach, combining CCIP and ROI, demonstrates superior performance, with the best classifier achieving an accuracy of 0.983. Notably, the CCIP classifier, when enhanced with ROI techniques, achieves an impressive accuracy of 0.985, effectively predicting defect impact levels. This research underscores the potential of transfer learning and the integration of CCIP and ROI in improving classifier performance and identifying surface defect severity in ball screws.

Keywords: Predictive Maintenance, Region of Interest, Masking, Transfer Learning

Introduction

Ball screw defect impact level prediction is essential for gaining insights into machine performance and ensuring operational efficiency. Accurately forecasting machine failures presents significant challenges due to a variety of external factors and the inherent uncertainty regarding which components will fail, along with the reasons and timing of such failures. Estimating the probability of imminent failure is a complex task, particularly in industries where the production of tangible goods is involved. If defects go undetected, they can compromise product quality, leading to project delays or even cancellations. Moreover, the manpower required to address these unexpected failures often far exceeds what is necessary for routine machine operations.

Proposing a model that can precisely predict ball screw failure is of absolute importance. This model can offer important insights into the timing, location, and types of ball screw failures that are likely to arise. It has the potential to greatly lower maintenance and replacement expenses, while also decreasing downtime. The capability to foresee ball screw failures is crucial for ensuring project success and producing high-quality products. Implementing early warning systems and proactive strategies is essential for minimizing the impact of failures, as they provide ample time to safeguard and back up data that may be vulnerable during such events. Although many models have been introduced for failure detection, diagnosis, and prediction, additional research is necessary to investigate the limitations of current methods. This paper introduces a hybrid ball screw surface defect impact level prediction model that leverages transfer learning techniques.

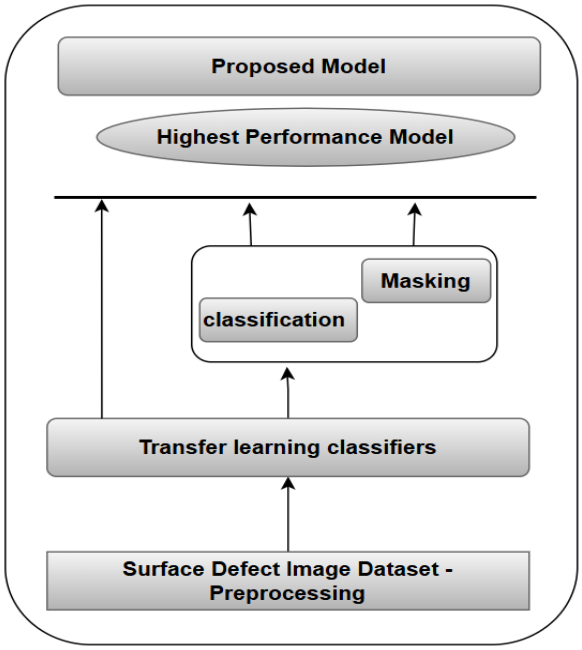

Figure 1. Proposed Model Block Diagram

In this study, we introduce a hybrid prediction model for ball screw failure and its impact level that integrates various transfer learning algorithms and techniques. The model employs four classifiers: CCIP-V, CCIP-D, VGG16, and DenseNet. Additionally, we apply ROI techniques to the outputs of these classifiers and analyze the results. A novel formula has been uniquely developed for the ROI algorithms to highlight the defect area and calculate the percentage of the defect's size. To assess the effectiveness of our proposed approach, we compare the performance of CCIP-V and CCIP-D, which leverage classifier subset evaluations based on VGG16 and DenseNet, respectively. The findings demonstrate that CCIP-V and CCIP-D surpass other methods, exhibiting statistically significant differences. Our proposed approach outperformed others, achieving an accuracy of 97.41%, which exceeds the accuracies reported in Berend et al. (2023), Juan et al. (2018), and Yiming Xu et al. (2021). This improvement is due to our novel combination of a transfer learning model with a customized masking approach.

In the following sections of this paper, we present a review of related research in Section 2 and Section 3 explains the proposed approach and architecture, followed by the results and discussions in Section 4, and summarize the key highlights and contributions of this work in Section 5.

Literature Review

Berend et al. (2023) developed a machine learning application for fleet-based condition monitoring of ball screw systems, aimed at reducing machine downtime and manufacturing costs. Over eight months, data were collected from the automotive sector under varying thermal and positional conditions. Using AutoML tools such as Auto-Sklearn and Auto-Weka, the study achieved accuracies of 82% (Decision Tree), 74% (Naïve Bayes), 80% (K-NN), 88% (SVM), and 89% (Multilayer Perceptron). The investigation also included lubrication degradation affecting long-term performance; however, the approach suffered from high false-alarm rates due to Gaussian and gamma scaling and could not identify defect impact levels.

Chang-Fu et al. (2018) introduced an SVM method grounded in statistical learning theory, outperforming ANN models with fewer samples. The model detects ball screw faults via pattern recognition and identifies machine failures automatically while also monitoring lubrication degradation using frictional torque signals. Its main limitation is the absence of preprocessing or scaling, which lowers accuracy and restricts the detection scope.

Jinwi et al. (2013) focused on fault investigation in ball screws, simulating four failure modes—recirculation system issues, preload loss, ball nut wear, and lubrication starvation. The study recreated preload loss and lubrication starvation and introduced three health-check techniques: Gaussian Mixture Model (GMM), Mahalanobis Distance (MD), and Self-

Organizing Map with Minimum Quantization Error (SOM-MQE). However, the approach relied heavily on labeled data and could not classify unseen samples or assess severity levels.

Juan et al. (2018) proposed a multiple classifier system (MCS) for ball screw degradation detection, employing dynamic classifier selection (DCS) and novel local class accuracy (N-LCA) to enhance results. While a single backpropagation network achieved 80.78% accuracy, the MCS with LCA achieved 93.09%. Although effective, the approach was time-consuming due to multiple classifier layers and was limited to defect identification.

Liang et al. (2019) presented a fault-identification technique using location-based outlier detection and spontaneous rotational frequency estimation. This method automatically identified and localized faults through fault-time estimation and frequency extraction, but was restricted to defect detection without employing machine learning.

Schlagenhauf et al. (2023) visually analyzed wear progression on ball screws, using a dataset of ten spindles to capture all stages from initial condition to failure. A camera system tracked pit growth, and a wear-quantification formula was introduced. However, the technique was limited to defects unsuitable for broader applications.

Yafei et al. (2020) proposed a prognostic study for ball screw health management using an ensembled GRU network embedded in a particle filter framework (GRU-PF). This hybrid data-driven and physical-model approach enabled interval estimation via dropout but increased model complexity and training time.

Our proposed novel approach overcomes all identified drawbacks by enabling classification of new inputs with severity-level interpretation through a masking strategy, reducing training time with optimized epochs, integrating transfer learning models such as VGG16 and DenseNet, lowering false alarms through data balancing, improving performance using effective preprocessing, simplifying model complexity, and expanding applicability to a wider range of defect types.

Proposed Approach

The model suggested in this paper is illustrated as a block diagram in **Error! Reference source not found.**. The model comprises three phases: data preprocessing, implementing separate transfer learning classifiers by applying masking strategies, and identifying the method with the highest performance. The original image is in BGR format, which does not provide sufficient resolution for defect identification. Therefore, it is converted to RGB format to enhance resolution, a key feature of the dataset. This change is necessary, as classifiers need high-resolution images to effectively identify defects.

Mathematical Representation of the Proposed Model

Let X be the dataset containing n instances represented a $X = (x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$, where x_i denotes the image for the i −th instance and yi is the corresponding target class label (defect or no defect). We define the set of classifiers as $C = C_1, C_2, C_3, \ldots, C_m$, Which includes four transfer learning classifiers: VGG-16, CCIP-V, CCIP-D, and DenseNet. Each classifier C_i is trained on the dataset X to predict the surface defect impact level and let F represent the set of features in the dataset. In Stage 1, we train each classifier c in C using the dataset X. In Stage 2, we apply a masking technique with a novel formula to the outputs of each classifier to determine the defect impact level and its area coverage. Finally, in Stage 3, we assess the performance of each classifier based on its f −measure values and select the method that yields the highest performance regarding defect impact level prediction.

The following sections outline each stage of our model

Data Preprocessing

The initial data preprocessing includes several steps: initializing the sample size, reading and properly labeling the images, converting BGR images to RGB format, resizing images to 128x128 for improved quality, and normalizing the images to scale grayscale values between 0 and 1. The target for the classifier is set as P (Pitting) and N (No Pitting).

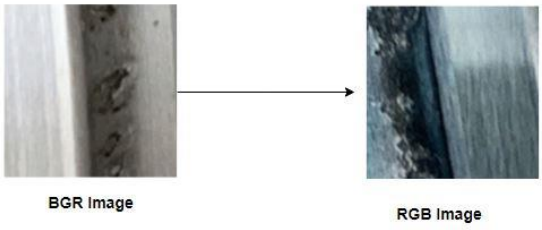

Figure 2. Proposed Model Block Diagram

Applying Transfer Learning Classifiers using the masking approach

The second stage of the proposed model involves implementing the transfer learning technique using the masking approach. In the pre-processed dataset, we applied the VGG-16 and Densenet models without any preprocessing, which overcame the drawback mentioned in Chang-Fu et al. (2018). To overcome the drawback mentioned in the paper Jinwi et al. (2013) and Juan et al. (2018), we developed the CCIP – V, and CCIP-D, in which we combine the transfer learning with masking using a novel region of interest technique and the input as a pre-processed image. For severity examination, we used the masking approach on top of the

Hue Saturation Value (HSV) image. Normally, while reading the image, it's in blue-green-red (BGR) format, especially for our images, it is very difficult to perform the masking, as it is black-shaded. So, we convert it to the HSV format as shown in Figure 3.

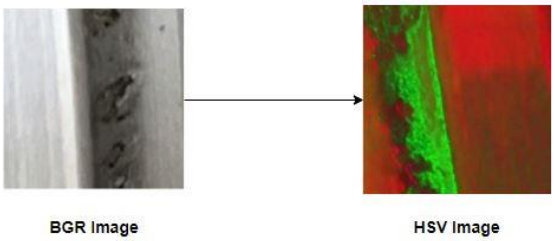

Figure 3. BGR to HSV Image Conversion

Then, perform the masking using the upper and lower boundaries for the pixels to highlight the region of the image having a surface defect. Finally, find the region of the coverage and the percentage of the surface got affected using the formula mentioned below. The count of non-zero pixels from the masked image means the affected pixels, and the count of the total dimension of the input image denotes the total pixel count of the surface-defected image.

$$\text{Rate of Interest (ROI) Percentage} = \frac{Count\ of\ non\ zero\ pixels}{Count\ of\ the\ total\ dimension\ of\ the\ input\ image} * 100$$

Based on the ROI percentage, the severity of the image is classified into different stages: stage 1, stage 2, and stage 3, respectively. Below is the algorithm for the multistage surface defect severity classification framework.

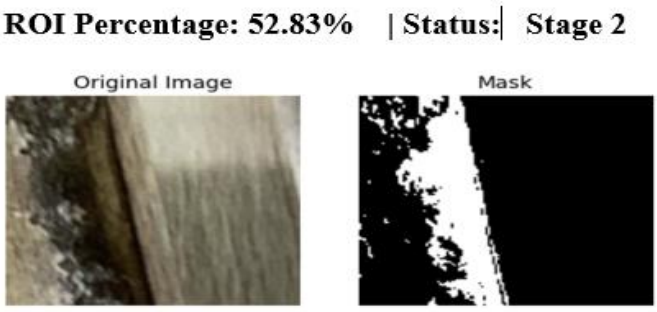

Figure 4. Original to Masked image and its classification stage

CIP – Collective Constraint Intelligent Prediction

The Collective Constraint Intelligent Prediction (CCIP) technique is specifically designed to classify the impact levels of surface defects on ball screws. As illustrated in Figure 1, the first step involves preprocessing the images using a masking approach. This process includes converting images from BGR to RGB format and resizing them to 128 x 128 pixels to enhance quality. Such preprocessing is crucial, as it aids in identifying minor defect levels and determining the area of impact coverage. In the second step, we employed transfer learning models, specifically VGG16 and DenseNet, as base models. We modified these models to accept images sized at 128 x 128 pixels and adjusted the layers accordingly, utilizing a sigmoid function for the activation layer.

Next, we filtered out defective images and applied masking by converting the RGB images to HSV format. This conversion is essential for accurately identifying the defect coverage area. During this step, we established upper- and lower-pixel thresholds and calculated the region of interest by counting the non-zero pixels. This count was then divided by the total pixel count and multiplied by 100 to determine the defect coverage area. Defects were categorized based on their coverage percentage: if less than 10%, it was classified as Level 1; between 10% and 15% as Level 2; and greater than 15% as Level 3, indicating high impact.

In the final step, we compared the performance of the VGG16, DenseNet, CCIP-V, and CCIP-D models based on the evaluation metrics outlined in the experimental section. The algorithm described below explains the CCIP model.

Algorithm: Surface Defect Impact Level Identification using CCIP

Input: *A path to the main folder of the ball screw dataset*
Output: *Degradation level for the given input image*
libraries ← *Import the cv2, matplotlib, tqdm, tensorflow,*
image_count ← *Count of all the images in the ballscrew*
Initialization of variables: *assign zero to variable* ***i***
While(i<image_count) do *// Read all the images from*
    image[i] ← *assign the target label*
    ballscrewlabelset[i] ← *image[**i**], label //store image, label*
end
Initialization of variables: *assign zero to variable* ***j***
While(j<image_count) do *// Read each image one by one*
    *convert the image[**j**] from BGR to RGB using* ***cv2.cvtColor***
    *resize the image[**j**] to the 128*128 using* ***cv2.resize***
    *normalize the image[**j**] using* ***cv2.normalise*** *to bring the*
end
split the ballscrew dataset into training and test sets using
autotune the learning rate using ***ReduceLROnPlateau*** *and set*

*Initialize the model using **Densenet201 for CCIP-D or***
*configure the fully connected layer using activation as **real***
*configure the output layer using activation as **sigmoid** and set*
train the model using vgg_model.fit for CCIP-V and
predict the surface defect using a test set-vgg_model. predict
convert the surface defect image from BGR to HSV using
*resize the surface defect image to 128*128 using **cv2.resize***
***lower** ← assign the numpy array value as [0,88,20]*
***upper** ← assign the numpy array value as [255,255,255]*
***mask** ← the surface defect image using **inRange** //configure*
***count_nonzeropixel** ← count using lower and upper using*
***roi_percentage** ← **(count_nonzeropixel**/ (128*128)) * 100 //*
***if** roi_percentage >=0 and roi_percentage<11*
| ***status** ← 'Level 1'* *status ← 'Level 1'*
***elseif** roi_percentage >=11 and roi_percentage<15*
***status** ← 'Level 2'* *status ← 'Level 2'*
else
***status** ← 'Level 3'* *status ← 'Level 3'*
end

The CCIP-V architecture, in Figure 5, illustrates how the VGG16 model utilizes pre-processed masked images of size 128x128 pixels. Subsequent layers are adjusted to dimensions of 64x64, 32x32, 16x16, 8x8, and 4x4, with the output layer employing a sigmoid activation function. Only images showing surface defects are selected to facilitate surface defect categorization using a Region of Interest (ROI) approach.

The CCIP-D architecture in Figure 6 detailed above demonstrates how the DenseNet model processes preprocessed masked images measuring 128x128 pixels. The following layers are resized to dimensions of 64x64, 32x32, 16x16, 8x8, and 4x4, with the output layer utilizing a sigmoid activation function. Only images exhibiting surface defects are selected to aid in surface defect classification using a Region of Interest (ROI) methodology.

Table 1 below outlines the performance of the CCIP-V model in terms of training and validation accuracy and loss across various epochs. As the number of epochs increases, accuracy also rises, indicating a direct proportionality between epochs and accuracy. However, after epoch 10, there is no further improvement in accuracy.

Table 2 presents the training and validation accuracy and loss of the CCIP-D model over different epochs during the tuning process. Tuning results with various epochs indicate that after epoch 10, there are no significant changes in accuracy and loss.

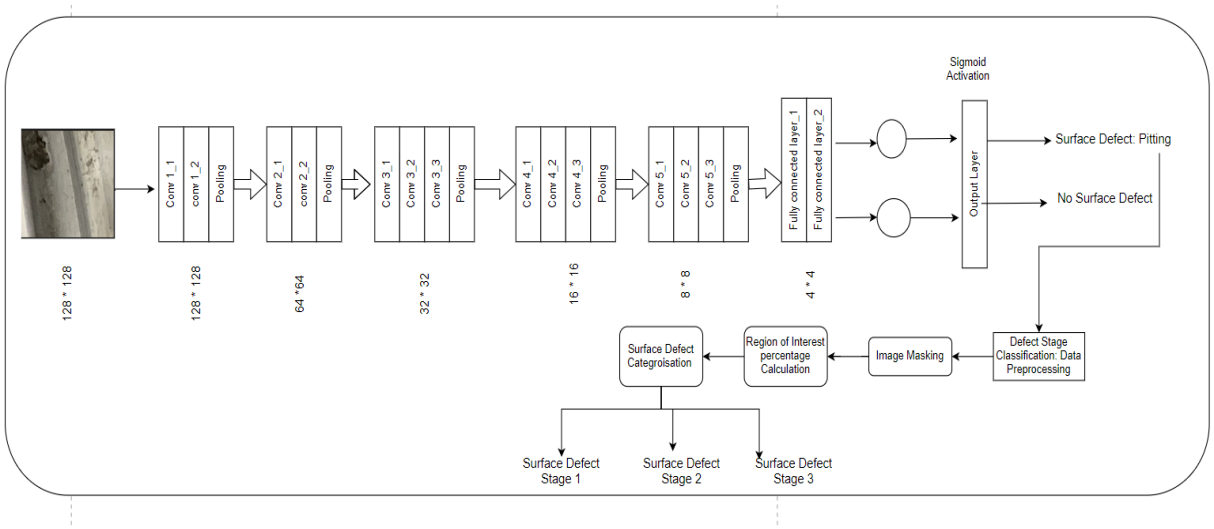

Figure 5. CCIP-V Architecture

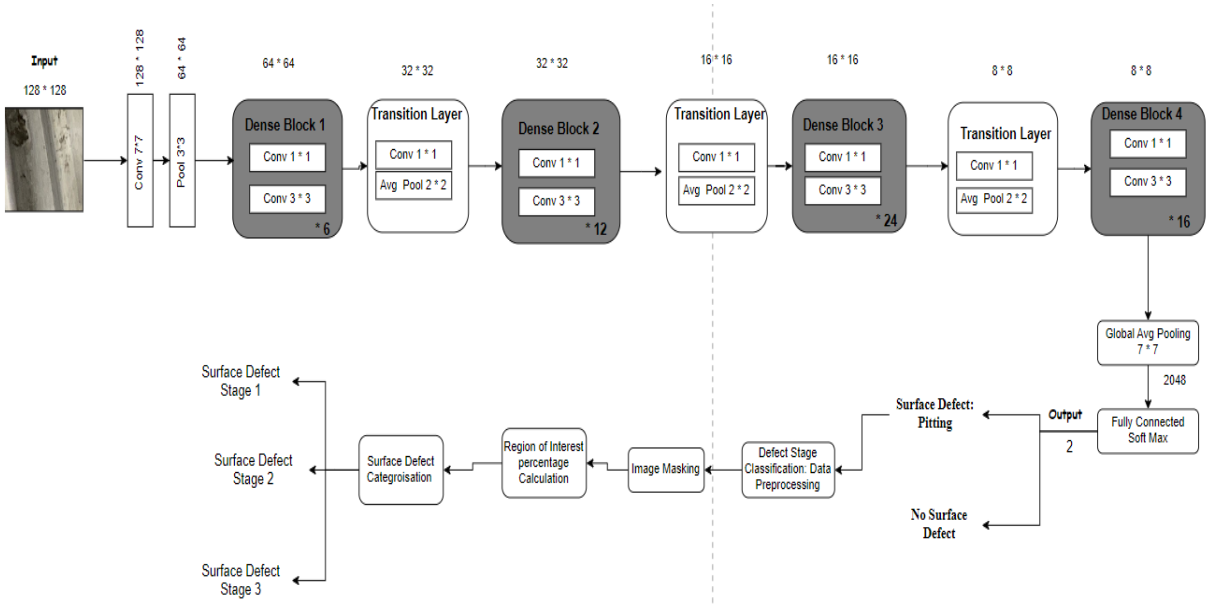

Figure 6. CCIP-D Architecture

Table 1. CCIP – V - Training and Validation accuracy/loss for each epoch

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.856	0.3429	0.9400	0.0010
2	0.936	0.1730	0.9258	0.0010
3	0.9467	0.1390	0.9590	0.0010
4	0.9602	0.1157	0.9633	0.0010
5	0.9669	0.0990	0.9667	0.0010
6	0.9677	0.0923	0.9700	0.0010
7	0.9752	0.0815	0.9733	0.0010
8	0.9752	0.0728	0.9808	0.0010
9	0.9735	0.0783	0.9742	0.0010
10	0.9777	0.0639	0.9742	0.0010

Table 2. CCIP – D - Training and Validation accuracy/loss for each epoch

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.8098	0.2585	0.8708	0.0010
2	0.8725	0.0841	0.8750	0.0010
3	0.8812	0.0565	0.8700	0.0010
4	0.8804	0.0373	0.8792	0.0010
5	0.8831	0.0283	0.8800	0.0010
6	0.8823	0.0243	0.8817	0.0010
7	0.8852	0.0184	0.8883	0.0010
8	0.8871	0.0107	0.8842	0.0010
9	0.8862	0.0110	0.8833	0.0010
10	0.8880	0.0078	0.8858	0.0010

High Performance Method

The final stage of the proposed model involves selecting the optimal method based on the highest performance results. As illustrated in Figure 1, a comparison is made between the outcomes of the previous stages. The method that demonstrates the best performance from each stage is identified as one of the top methods. This comparison is conducted among the four classifiers by combining the recall and precision results, utilizing the f-measure for this evaluation criterion. Figure 8 depicts the comparison of the evaluation metrics among the four classifiers.

Used Dataset

After collaborating with various manufacturers, we found that capturing surface defects from real-time machinery is both challenging and time-consuming. Consequently, we utilized a dataset collected from machinery featuring a range of surface defect images, as presented in the research by Schlagenhauf et al. (2023). This dataset highlights ball screw surface defects that arise during manufacturing due to continuous machinery operation. For our proposed methodology, the ball screw surface defect dataset was sourced from a reputable production plant manufacturer, yielding reliable results. This dataset is well regarded in the research community, with numerous studies conducted on the same operational dataset. It comprises a total of 21,835 images, of which 10,517 images are classified as having no defects (denoted as N), while 10,492 images contain defects (denoted as P).

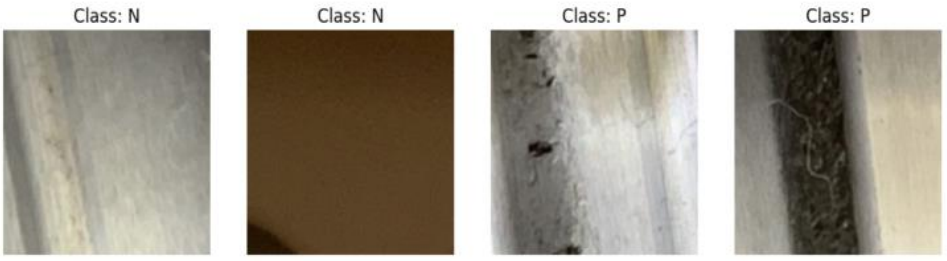

Figure 7. Samples of surface defect images from the dataset

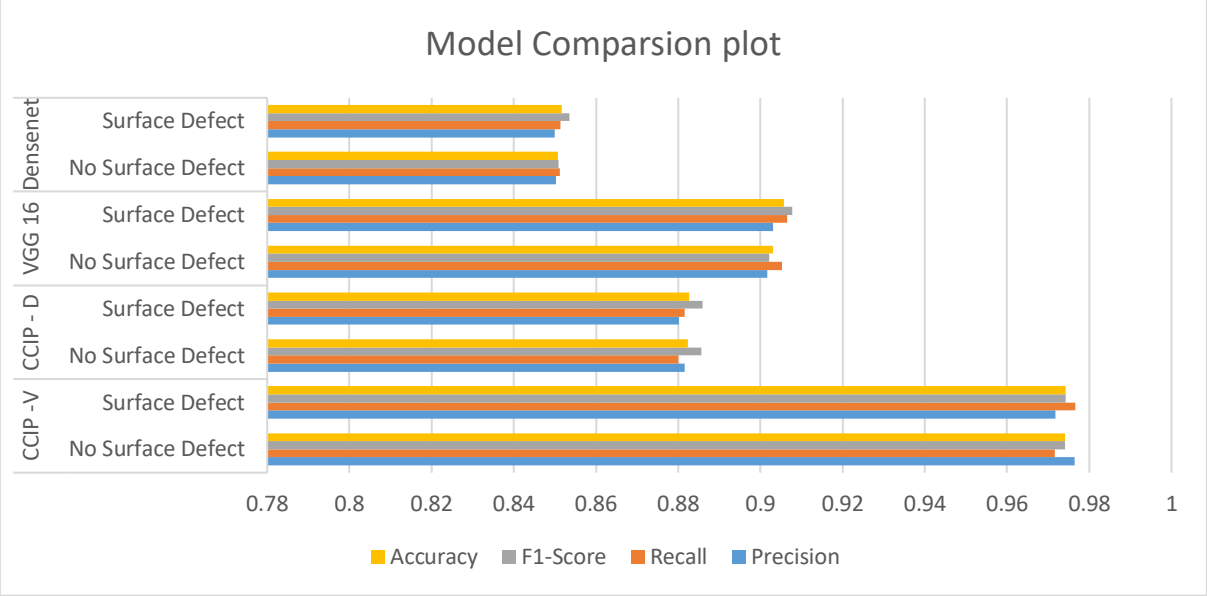

Figure 8. Model Comparison Plot

Proposed Approach

Evaluation Criteria

This section discusses the confusion matrix and the evaluation measures employed. The confusion matrix assesses the results of the proposed model, detailing metrics such as False Negatives, False Positives, True Negatives, and True Positives. The evaluation metrics for our analysis include accuracy, recall, precision, and f-measure, which is used to assess classifier performance. Each of these metrics is defined and calculated as follows:

Accuracy refers to the proportion of correct predictions. A result of 1 indicates perfect accuracy, while 0 signifies the lowest possible accuracy. The calculation of accuracy is done using the following formula:

$$Accuracy = \frac{(TRP+TRN)}{(TRP+TRN+FAP+FAN)} \quad (1)$$

Precision is the ratio of relevant instances to the total number of retrieved instances. A result of 1 indicates optimal precision, whereas a result of 0 reflects the poorest precision. The precision value can be estimated using the following formula:

$$Precision = confidence = \frac{TRP}{TRP+FAP} \quad (2)$$

Recall measures the proportion of true positive instances that are correctly identified. A result of 1 indicates perfect recall, while 0 signifies the lowest recall. It is calculated using the following equation:

$$Recall = \frac{TRP}{TRP+FAN} \quad (3)$$

F-measure is the harmonic mean of precision and recall. It combines both metrics into a single score. The result is calculated using the following formula:

$$F - measure = 2(PrecisionRecall/Precision + Recall) \quad (4)$$

MAE, or Mean Absolute Error, is determined by the total difference between the true and predicted instances. The value of MAE is calculated using the following formula:

$$MAE = Truevalues - Predictedvalues \quad (5)$$

Table 3. Surface Defect Impact Level Classification Results

Model	Label	Precision	Recall	F1-Score	Accuracy	MAE
CCIP -V	No Surface Defect	0.9765	0.9716	0.9741	0.9741	0.0259
	Surface Defect	0.9718	0.9766	0.9742	0.9742	0.0258
CCIP - D	No Surface Defect	0.8815	0.88	0.8857	0.8824	0.1176
	Surface Defect	0.8802	0.8816	0.8859	0.8826	0.1174
VGG 16	No Surface Defect	0.9016	0.9052	0.9022	0.903	0.097
	Surface Defect	0.9031	0.9065	0.9078	0.9058	0.0942
Densenet	No Surface Defect	0.8503	0.8512	0.8509	0.8508	0.1492
	Surface Defect	0.8499	0.8513	0.8536	0.8516	0.1484

Experimental Analysis and Discussion

The comparison is done between the four classifiers using the combination of the recall, precision, f-score, accuracy, and MAE results. Table 3 presents the accuracy, recall, precision, MAE, and F-measure results for the four classifiers. The CCIP-V and VGG16 classifiers achieved the highest prediction accuracy, with values of 0.974 and 0.8825, respectively. Conversely, Densenet and CCIP-D had the lowest accuracy, at 0.851 and 0.883, respectively. The CCIP classifier recorded the lowest MAE value of 0.026, while the other classifiers had MAE values close to or exceeding 0.1. This indicates that CCIP-V excels in both accuracy and MAE prediction. We attribute this performance to CCIP's focus on nominal and enhanced feature data, aided by the image preprocessing techniques we employed for quality enhancement.

Considering precision and F1-Score, the CCIP-V classifier achieved around 0.97, while the other classifiers reported scores below 0.90, as shown in Table 3. This highlights the significant performance of CCIP-V compared to the others. In terms of recall, CCIP-V also excelled with a score of 0.9716, whereas CCIP-D, VGG16, and Densenet reported scores of 0.8857, 0.902, and 0.853, respectively.

Another comparison was conducted among the four classifiers by analyzing the combined training and validation accuracy results. Figure 9 illustrates the training accuracy plots over 10 epochs for VGG16, Densenet, CCIP-V, and CCIP-D, clearly showing that CCIP-V achieves strong accuracy in surface defect classification. Figure 10 presents the validation accuracy plots for the same 10 epochs, further confirming that CCIP-V excels in accuracy for this classification task.

Comparing Figure 9 and Figure 10 for training and validation accuracy indicates that the CCIP-V classifier outperforms the others in both scenarios over the 10 epochs. Another comparison was made among the four classifiers by varying the number of epochs and analyzing the results based on training and validation accuracy. At 10 epochs, the classifiers yielded the following accuracies: CCIP-D at 88%, CCIP-V at 96.4%, VGG16 at 82.1%, and Densenet at 88.7%. However, when the epoch count was increased to 20, the accuracies decreased to 84.3% for CCIP-D, 93.3% for CCIP-V, 80% for VGG16, and 83.5% for Densenet. Therefore, we have fixed the Epoch count at 10, as it provides better accuracy.

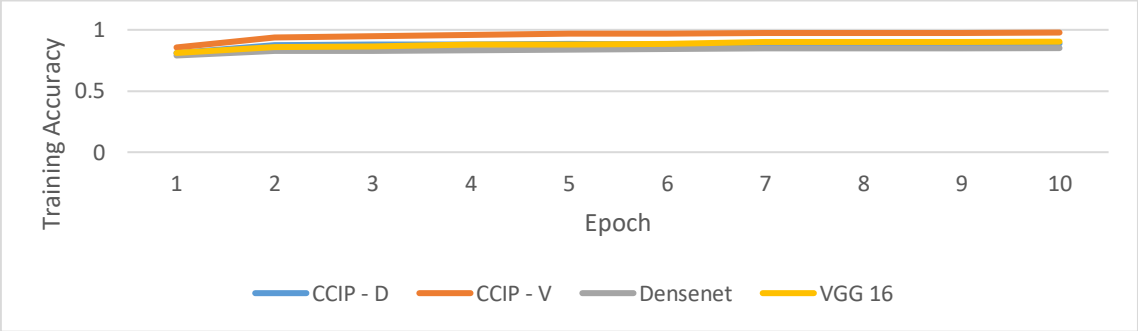

Figure 9. Training Accuracy Plot Graph

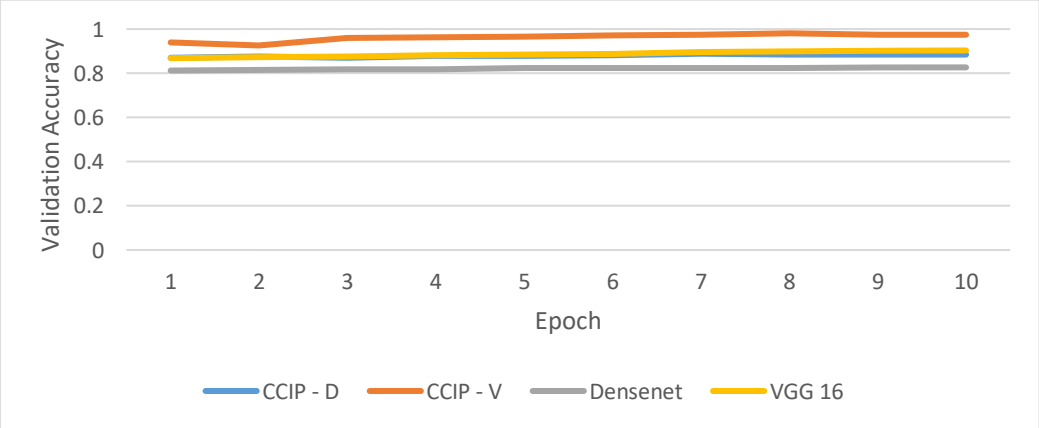

Figure 10. Validity Accuracy Plot Graph

Table 4 presents the comparison of training and validation accuracy for 10 and 20 epochs. The data indicates that increasing the epoch count from 10 to 20 results in a reduction in accuracy. For this dataset and these classifiers, the results show an inverse relationship, where an increase in epoch count leads to decreased accuracy. Figure 11 presents a comparison plot of the classifiers based on varying epoch counts and their corresponding performance. This provides a visual representation of the inverse relationship between epoch count and accuracy among these classifiers.

Table 4. Surface Defect Impact Level Results on various epochs

EPOCH	Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
10	CCIP -D	0.876	0.054	0.88	0.001
	CCIP - V	0.954	0.126	0.963	0.001
	VGG 16	0.836	0.105	0.821	0.003
	Densenet	0.879	0.038	0.887	0.003
20	CCIP -D	0.873	0.083	0.843	0.002
	CCIP - V	0.903	0.167	0.933	0.002
	VGG 16	0.808	0.133	0.8	0.005
	Densenet	0.824	0.063	0.835	0.005

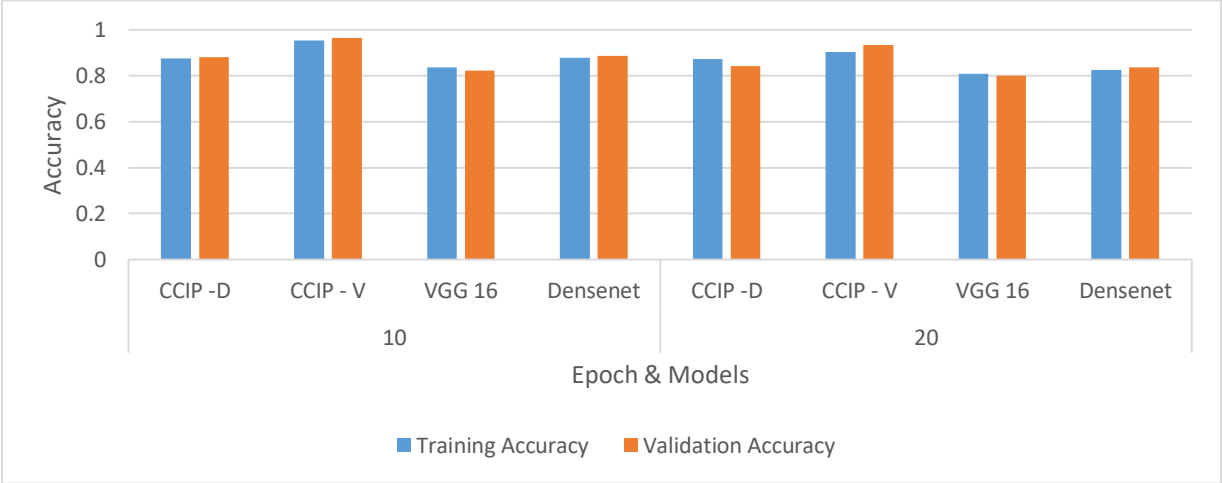

Figure 10. Comparison plot of Models with different epochs

Considering precision and F1-Score, the CCIP-V classifier achieved around 0.97, while the other classifiers reported scores below 0.90, as shown in Table 4. This highlights the significant performance of CCIP-V compared to the others. In terms of recall, CCIP-V also excelled with a score of 0.9716, whereas CCIP-D, VGG16, and Densenet reported scores of 0.8857, 0.902, and 0.853, respectively.

Based on our evaluation using metrics such as precision, recall, accuracy, F1-Score, MAE, and comparisons of training and validation accuracy as well as loss, CCIP-V consistently outperformed the other models. When varying the epochs to identify the optimal model, CCIP-V also showed significant performance across all metrics, achieving an impressive accuracy of 97.41%.

Comparison of the State of the Art

CCIP classification for surface defect impact levels, utilizing various transfer learning models, has shown promising results compared to leading methods in the field. This section will compare our model's performance with a notable existing approach, emphasizing the advancements made. The state-of-the-art method selected for comparison is the work by Juan et al. (2018), which used a dynamic classifier and introduced the LCA and N-LCA models, achieving a maximum accuracy of 96.78%. In contrast, our model attained an accuracy of 97.41%, exceeding the state-of-the-art method by 0.63 points. In Yiming Xu et al. (2021), YOLO and modified YOLO models were enhanced with a feature scaling layer, achieving 79.20% accuracy, but they focused mainly on feature extraction without image preprocessing. In contrast, our approach combined feature selection with preprocessing and masking, resulting in an improved accuracy of 97.41%.

This comparison highlights the superior performance of our surface defect impact level classification. By integrating multiple transfer learning algorithms and modified approaches like CCIP-V and CCIP-D, which include image preprocessing for quality enhancement and feature extraction, we significantly improved prediction accuracy. Notably, the Collective Constraint intelligence (CCIP) classifier demonstrated exceptional effectiveness, surpassing the state-of-the-art method. Additionally, our model utilized masking with Region of Interest (ROI) to specify impact levels and their coverage areas, achieving an accuracy score of 97.41%. This marks a substantial improvement over existing methods, underscoring the value of tailored preprocessing techniques and masking for specific datasets and prediction tasks.

Overall, our comparison with the state-of-the-art method underscores the advancements of our surface defect impact classification model. By leveraging multiple transfer learning algorithms, applying image preprocessing and masking techniques, and utilizing the strengths of the CCIP classifier, we achieved higher accuracy and improved precision. These findings advance surface defect classification and hold practical significance for industries that depend on accurate failure predictions, helping to minimize costs and downtime while ensuring high-quality product delivery. Table 5 presents a comparison of the accuracy results from various researchers' work.

Table 5. Comparison between the accuracy results

Work	Classifier	Dataset	Procedure	Accuracy	Highest Accuracy	Defect level identification
Yiming Xu et al. (2021)	YOLO v5	metal surface defect	additional scale feature layer	75.10 %	79.20%	Not in Scope
	modified YOLO			79.20 %		
Berend et al. (2023)	Decision Tree	ball screw dataset	Gaussian & gamma scaling	82%	88%	Not in Scope
	Naïve Bayes			74%		
	K-NN			80%		
	SVM			88%		
Juan et al. (2018)	BPN	ball screw experiment dataset	Dynamic Classifier System	80.78 %	96.78%	In scope, based on the vibration signal
	Adaboost			86.64 %		
	LCA			93.09 %		
	N-LCA			96.78 %		
This paper	VGG16	ball screw surface defect dataset	NA	90%	97.41%	In scope, % of the defect and its impact level with coverage area using masking with ROI
	Densenet		NA	85.08 %		
	CCIP-V		preprocessing for image quality enhancement, input dimension tuning, transfer learning	97.41 %		
	CCIP-D			88.25 %		

Threats to Validity

This section outlines several issues that could impact the validity of our results. First, the experiment was conducted on a single dataset, which may limit the effectiveness of image preprocessing and masking techniques. Second, relying on only two transfer learning algorithms may not provide a comprehensive view, as different algorithms can exhibit varied performance on the same dataset.

Conclusion

Ball screw surface defect classification is a crucial issue that necessitates proactive measures to reduce associated costs and damages. However, predicting the timing, causes, and locations of defects, as well as their severity, presents significant challenges due to various influencing factors. In this study, we introduced a CCIP model based on transfer learning techniques and applied it to a specific dataset. Our findings yielded promising results, with the CCIP-V classifier delivering the best performance, achieving an accuracy of 0.974 and a mean absolute error (MAE) of 0.025.

Additionally, we examined the effects of classification by comparing results before and after image preprocessing for feature enhancement, based on performance rankings. Both the CCIP-V and CCIP-D models showed significant accuracy improvements following preprocessing. However, the CCIP-V classifier continued to outperform all others, showcasing its robustness and effectiveness in this severity classification task. Remarkably, the CCIP-V method further increased accuracy to 0.974, achieving the highest accuracy in the second phase of our experiment.

In the end, we compared two base transfer learning techniques with the modified CCIP-V and CCIP-D models. The results showed significant improvements in the performance of the CCIP classifiers compared to the individual classifiers.

For future work, we recommend incorporating multiple datasets to validate the model's performance across various scenarios and enhance its generalizability. Given the critical role of image preprocessing and ROI in improving our results, further exploration of different transfer learning algorithms and computer vision techniques is essential to assess their effectiveness on the same dataset. By integrating various algorithms, we can gain deeper insights into feature relevance and identify the most suitable methods for specific ball screw impact level classification tasks.

In conclusion, our ball screw surface defect impact level classification—especially using the CCIP-V classifier and customized masking with ROI—has shown promising results. This model has significant potential to enhance machine life cycle management, reduce defect-related costs, and ensure high-quality production. By tackling the challenges of surface defect classification with advanced transfer learning techniques, our approach improves the efficiency and reliability of predictive maintenance strategies in industries where machine failures can have serious consequences.

Acknowledgments

This study is a part of a PhD thesis. Acknowledgment is bestowed on honorable supervisors and examiners.

Conflict of interest

All authors certify that they have no affiliation with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Amit, S. P., Sunnapwar, B. K., & Oza, A. D. (2022). Effective machining parameter selection through fuzzy AHP–TOPSIS for 3D finish milling of Ti6Al4V. *International Journal on Interactive Design and Manufacturing*.

Anderson, R. N. (2017). Petroleum analytics learning machine for optimizing the Internet of Things of today's digital oil field-to-refinery petroleum system. In *Proceedings of the 2017 IEEE International Conference on Big Data (Big Data)* (pp. 4542–4545).

Berend, N., et al. (2023). Application of machine learning for fleet-based condition monitoring of ball screw drives in machine tools. *The International Journal of Advanced Manufacturing Technology, 127*, 1143–1164.

Bovic, K., Pierre, D., & Xavier, C. (2011). Tool wear monitoring by machine learning techniques and singular spectrum analysis. *Mechanical Systems and Signal Processing, 25*, 400–415.

Chang-Fu, L., et al. (2018). Techniques developed for fault diagnosis of long-range running ball screw drive machines to evaluate lubrication condition. *Measurement*. https://doi.org/10.1016/j.measurement.2018.05.059

Deng, Y., Shichang, D., Shiyao, J., Chen, Z., & Zhiyuan, X. (2020). Prognostic study of ball screws by ensemble data-driven particle filters. *Journal of Manufacturing Systems, 56*, 359-372.

Denkena, B., Dittrich, M.-A., Noske, H., Lange, D., Benjamins, C., & Lindauer, M. (2023). Application of machine learning for fleet-based condition monitoring of ball screw drives in machine tools. *The International Journal of Advanced Manufacturing Technology, 127*, 1143–1164.

Gouarir, A., Garcia, M.-A., Terrazas, R. S., & Benardos, P. (2018). In-process tool wear prediction system based on machine learning techniques and force analysis. *Procedia CIRP, 27*, 501–504.

Han, X., Miller, M., Vogl, G. W., Chen, G., & Jia, X. (2024). Robust feature design for early detection of ball screw preload loss. *Manufacturing Letters, 41*, 1225–1230.

Heng, A., Tan, A. C., Mathew, J., Montgomery, N., Banjevic, D., & Jardine, A. K. (2009). Intelligent condition-based prediction of machinery reliability. *Mechanical Systems and Signal Processing, 23*, 1600–1614.

Jinwi et al. (2013). Methodology For Ball Screw Component Health Assessment And Failure Analysis. *Proceedings of the ASME 2013 International Manufacturing Science and Engineering Conference*

Juan et al. (2018). A new method for identifying the ball screw degradation level based on the multiple classifier systems. *Measurement, 130*, 118–127.

K Venkata, R., M. B.S.N., and N. Rao (2014). Prediction of cutting tool wear, surfac roughness and vibration of work piece in boring of aisi 316 steel with artificial neural network. *Measurement*, 63–70.

Liang et al. (2019). Ball Screw Fault Detection and Location Based on Outlier and Instantaneous Rotational Frequency Estimation. *Shock and Vibration*, DOI 10.1155/2019/7497363.

Lee, W. G., Lee, J. W., Hong, M. S., Nam, S. H., Jeon, Y., & Lee, M. G. (2015). Failure Diagnosis System for a Ball-Screw by Using Vibration Signals. *Shock and Vibration, 2015*(1), 435870.

M.A. Vargas et al. (2019). A mechatronic approach for ball screw drive system: modeling, control, and validation on an FPGA-based architecture. *The International Journal of Advanced Manufacturing Technology*, DOI 10.1007/s00170-019-03945-2

Mo¨hring et al. (2012). Integrated autonomous monitoring of ball screw drives. *CIRP Annals - Manufacturing Technology, 61*, 355–358

Nadeem, I., B.-A. Thorkil, E. N. Finn, and J. Karsten (2022). Outlier detection in sensor data using ensemble learning. *Procedia Computer Science, 176*, 1160–1169.

Nikolai, S., M. Jan, and F. Christoph M. (2018). Big data on the shop-floor: sensor-based decision-support for manual processes. *Journal of Business Economics, 88*, 593–616.

Ringsquandl, M., S. Lamparter, S. Brandt, T. Hubauer, and R. Lepratti (2015). Semantic-guided feature selection for industrial automation systems. *In 14th International Semantic Web Conference*, 225–240.

Schlagenhauf, T., Scheurenbrand, T., Hofmann, D., & Krasnikow, O. (2023). Analysis of the visually detectable wear progress on ball screws. *CIRP Journal of Manufacturing Science and Technology*, *40*, 1-9.

Schlagenhauf, T., Sun, C., & Fleischer, J. (2020). GAN based ball screw drive picture database enlargement for failure classification. arXiv preprint arXiv:2011.10235.

Schlagenhauf, T., & Landwehr, M. (2021). Industrial machine tool component surface defect dataset. *Data in Brief*, *39*, 107643.

Xu, Y., Zhang, K., & Wang, L. (2021). Metal surface defect detection using modified YOLO. *Algorithms*, *14*(9), 257.

Bibliographic information of this paper for citing:

Prabha Mohan, Soorya & Jaya Nirmala, S (2026). Ball Screw Surface Defect Impact Level Categorization Using the Collective Constraint Intelligent Prediction Technique. *Journal of Information Technology Management*, 18 (1), 34-52.
https://doi.org/10.22059/jitm.2026.106253