IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Defect Prediction in CWDM Optical Modules Using Multimodal Learning

KYU-JEONG CHOI[1, 2], JIA YANG[1], BOTAMBU COLLINS[1], SUNG-GEUN KIM[2], DO-JIN LIM[2], and JIN-TAEK SEONG[1],

[1]Graduate School of Data Science, Chonnam National University, Gwangju, Republic of Korea
[2]Opticis Co., Ltd., Republic of Korea

Corresponding author: Jin-Taek Seong (e-mail: jtseong@jnu.ac.kr).

"This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2023-00242528)."

ABSTRACT Reliable defect detection in coarse-wavelength division multiplexing (CWDM) optical modules is critical for ensuring stable high-speed optical communication and minimizing network disruptions. Traditional inspection methods, such as manual video screening and optical testing, are labor-intensive, prone to human error, and often fail to identify underlying defect patterns, thereby necessitating multiple rounds of testing. To overcome these limitations, this paper proposes a deep learning–based multimodal learning framework for defect prediction that integrates manufacturing process data (tabular features) and optical eye diagram images from the module's receiver. Leveraging the complementary nature of these data sources enhances defect detection performance, enabling the proposed model to surpass conventional single modal alternatives. The model was comprehensively evaluated on a real CWDM module dataset through cross-validation, based on performance metrics such as accuracy and F1-score. The multimodal learning schemes achieved an accuracy of 91.0% and an F1-score exceeding 90%, significantly outperforming both traditional single modal approaches and manual inspection techniques. These results demonstrate the effectiveness of multimodal learning in improving defect prediction accuracy, reducing reliance on repetitive testing, and lowering overall inspection costs. The proposed approach represents a scalable and efficient solution for automated quality control in optical module manufacturing, with potential applications in optical network maintenance and defect detection for other photonic components.

INDEX TERMS CWDM OSA, optical link, deep learning, multimodal learning, tabular data, binary classification, imbalanced dataset.

I. INTRODUCTION

An optical module [1] in Coarse Wavelength Division Multiplexing (CWDM) [2] is a critical electronic device that facilitates high-speed, high-capacity data transmission over optical fibers by converting electrical signals into optical signals and vice versa. It comprises a transmitter, which utilizes a laser or LED to convert electrical signals into optical signals, and a receiver, which employs a photodetector to revert optical signals into electrical form. Optical modules are essential for long-distance, high-bandwidth communication and play a pivotal role in enhancing data transmission efficiency, particularly in data centers and high-performance networking environments.

The optical module used in this study is designed for transmitting digital audio/video (A/V) signals between input and output devices using optical communication technology, commonly referred to as an optical link (Figure 1) [3]. This technology enables high-fidelity optical signal transmission across various types of A/V equipment, including PCs, DVDs, TVs, projectors, and speakers. It is used in applications that demand high-resolution video transmission, such as medical diagnostic systems, control rooms, and video walls in monitoring centers, where signal integrity and reliability are critical.

The inspection quality of an optical module is primarily determined by its inherent characteristics and manufacturing precision. Common symptoms of defects that can degrade performance include display failure, noise, flickering, and low resolution, all of which can significantly impact the reliability and visual fidelity of transmitted signals.

To ensure reliable defect detection, various inspection methods are employed, including screen inspection [4],

FIGURE 1: Optical modules for digital A/V applications.

flicker detection using neural networks [5], signal analysis using eye diagrams [6], and optical power measurement [7], [8]. Screen inspection involves connecting the optical module to a system and continuously monitoring for display anomalies. However, since defects occur randomly, accurately identifying failures under real-world conditions remains a significant challenge. To address this problem, extended-duration screen inspection tests can be conducted (Figure 2). These tests are intended to detect defects such as flickering over an extended period. Units that exhibit no abnormal behavior during the extended-duration period are classified as "pass", while those displaying defects such as noise or flickering (Figure 3) are labeled "fail". The final dataset is constructed by assigning a label of "0" for pass and "1" for fail, which provides a structured basis for defect prediction.

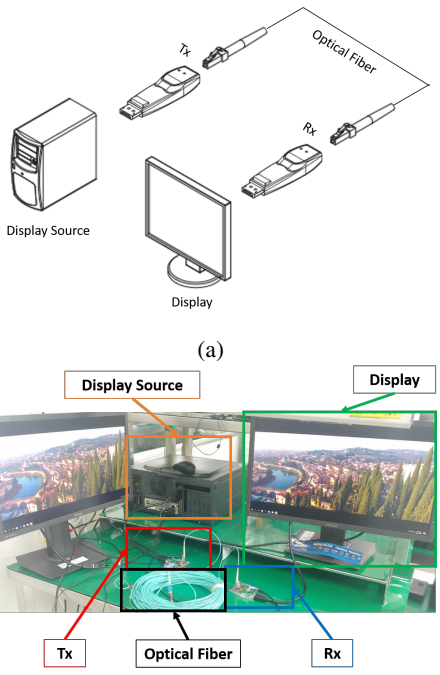


FIGURE 2: Screen inspection setup and process: (a) schematics and (b) actual implementation.

Eye diagram analysis is a widely used technique for detecting display defects; however, its effectiveness primarily lies in identifying severe issues, such as low resolution or

FIGURE 3: Examples of screen inspection results: (a) normal display, (b) no display, and (c) noise.

complete display failure. This process involves measuring jitter–which represents the variation in signal timing during transmission–to evaluate signal quality [9]. Measurements are obtained from the eye diagram using an oscilloscope, and the corresponding images are captured and stored for further analysis.

Measuring optical power is essential for evaluating the output power of an optical module as the optical power directly impacts signal integrity and transmission reliability. Because a higher optical power input to the receiver optical subassembly (ROSA) leads to a lower bit error rate (BER) [10], maintaining an adequate power level is critical. To ensure proper functionality, optical products must be operated at specified minimum optical power levels. As shown in Figure 4, optical power is measured using an evaluation board (EV B/D), where a testing current is injected into the optical module. The transmitted optical signal propagates through an optical fiber and is subsequently measured by a power meter connected to a photodetector (PD); this ensures accurate assessment of the module's performance.

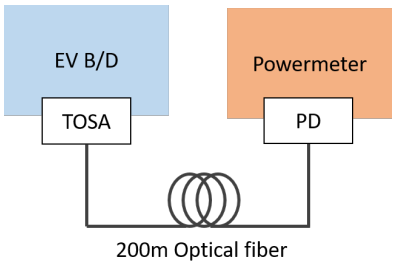


FIGURE 4: Schematic of optical module output inspection setup.

Conventional defect detection methods, including eye diagram analysis, optical power measurement, and short-duration screen inspections, have shown limited effectiveness in identifying defects that emerge during extended-duration screen inspections, achieving only 87.5% detection accuracy. These short-term tests often fail to capture randomly occurring defects over prolonged periods, which leads to a loss of valuable data. Moreover, existing approaches do not effectively exploit the subtle pattern variations in ROSA eye diagrams, nor do they fully leverage critical process data generated during manufacturing. To overcome these limitations, we propose a deep learning–based approach that integrates a broader range of informative features for defect prediction.

Specifically, our approach is designed to enhance the accuracy and robustness of defect detection by leveraging advanced deep learning techniques to capture fine-grained pattern variations in eye diagram images. The proposed multimodal deep learning framework for defect detection in digital A/V optical modules integrates both image-based and tabular data to improve defect classification accuracy. Our model is structured into four key stages: data preprocessing, feature extraction and fusion, classification, and result analysis. To extract meaningful representations effectively, we utilize a class of ResNet-50 models for image-based feature extraction, while tabular data are processed using the proposed deep learning architecture. Furthermore, we employ multimodal learning techniques to fuse heterogeneous data sources and leverage data augmentation strategies to address class imbalance, which ensures robust defect detection.

One of the key advantages of our approach is its ability to integrate inspection data from the manufacturing process with extracted data from production equipment, which maximizes the utilization of available information. The core contributions of this study are as follows:

- **First application of deep learning to defect detection in digital A/V optical modules.** Unlike traditional inspection methods, our approach leverages both image-based and process-driven data to enhance defect prediction accuracy.
- **Significant improvement in defect detection accuracy.** By incorporating multimodal learning techniques, our method increases detection accuracy from **87.5%** to **91.0%**, surpassing conventional techniques.
- **Reduction in inspection time through automated defect detection.** The proposed approach streamlines the manufacturing process by eliminating the need for extended-duration inspections, thereby improving production efficiency.

By integrating deep learning with multimodal data fusion, this work provides a more robust and efficient solution for defect detection in digital A/V optical modules. Our findings highlight the potential of artificial intelligence in enhancing quality control processes within the optical communication industry. Based on an extensive review of existing literature, we reaffirm that this study is the first to apply multimodal deep learning to the inspection of digital A/V optical modules. Unlike conventional inspection techniques that rely solely on signal metrics or manual evaluation, our approach maximizes the utility of information by integrating both inspection data generated during the manufacturing process and signal data extracted from production equipment through a multimodal deep learning framework.

II. RELATED WORK

A. VARIOUS APPROACHES OF EVALUATION FOR CWDM OPTICAL MODULES

Various analytical methods have been employed to improve and evaluate the performance of CWDM optical modules. To analyze the signal quality of optical modules, jitter measurements can be performed based on eye diagrams [11], which provide an intuitive assessment of signal integrity. However, since eye diagrams alone do not allow for precise numerical analysis, quantitative signal quality evaluation is typically conducted based on the BER and Q-factor [12]– [14].

Additionally, simulation-based analysis is widely utilized during the design and development stages of optical modules to identify defects and mitigate potential performance degradation before actual hardware implementation [15]. In simulations, Q-factor and BER are also employed as key performance evaluation metrics [16]. Furthermore, in video transmission systems, the video and audio quality of optical modules is frequently tested using the 881/882-HDMI Video Generator [18] to ensure compliance with performance standards. Additionally, visual inspection and BER testing are often replaced by the R&S VTC/VTE video tester [19] for A/V quality performance testing. In this test method, the HDMI port is used instead of directly connecting the monitor to the video source to detect transmission errors.

Although direct human observation is generally required during the hardware inspection of display screens, this method is inefficient for long-duration testing. To address this issue, a technique that utilizes an optical detector to detect LCD screen flickering has been developed, which reduces operator fatigue and improves inspection efficiency [4].

B. MULTIMODAL LEARNING

Multimodal learning has been explored to leverage the complementary nature of multiple data types and enhance predictive performance [20]– [22]. Generally, multimodal learning can be categorized into early, intermediate, and late fusion approaches [23], [24]. Early fusion involves integrating features at the input level, thereby concatenating modality-specific data directly. However, this approach may reduce interactions between individual modalities [25]. Intermediate fusion combines features learned from different modalities, which enables effective modeling of cross-modal interactions [26]– [29]. Late fusion aggregates predictions from independently trained modality-specific models for classification [30]– [32]. Since each modality is learned separately, this approach offers greater flexibility in incorporating new modalities. However, it may fail to fully capture interactions between modalities [33].

To address the aforementioned limitations and enhance fusion strategies, alternative approaches that analyze and integrate modality dynamics have been introduced [34], [35]. Our proposed model adopts a dual-modality input structure comprising tabular and image data. Specifically, due to the unique structure of CWDM, an additional intermediate fusion is performed using four image features extracted from each channel. To capture the interactions between the tabular and image modalities effectively, the model compensates for modality imbalance and enhances representational synergy by incorporating modality dynamics.

TABLE 1: Summary of A/V Optical Module Inspection Methods.

Author & Year	Technique / Model	Device / Signal / Data Type	Dataset / Purpose	Limitation	Performance (Accuracy / F1-score)	Ref
Chen et al., 2016	Flicker auto-inspection	LCD module (A/V)	Detect flicker via backlight	Not for optical modules; no ML-based classification	–	[4]
Patel et al., 2017	ANN + Frame Difference + DCT	Video frames	YouTube video clips	Small dataset (60 frames); no real industrial context	0.933 / –	[5]
Our Work, 2025	Multimodal visual inspection (DL-based)	CWDM OSA (Eye diagram (RGBC) + Tabular)	Predict defects using eye jitter pattern	First DL-based defect detection on A/V OSA; no public benchmark dataset	**0.910 / 0.904**	–

TABLE 2: Optical Module Evaluation Techniques Using Eye Diagram and CWDM Components.

Author & Year	Method / Approach	Evaluated Metric	Evaluation Target	Limitation	Ref
Hancock, 2004, Stephens, 2004	Eye diagram inspection	Jitter, Q-factor	Optical lab testbench	Visual/manual only; no automated detection	[9], [11]
Priyadarshi et al., 2004	BER analysis	Bit Error Rate	CWDM test module	One-point based; lacks image classification	[15]
Sialm et al., 2005	Optical power + coupling	Optical output	Multimode VCSEL	Alignment focus only; not for final screen or DL-based diagnosis	[8]
Park et al., 2010	Optical Power Testing	Output optical power	CWDM TOSA	No classification or feature-level analysis	[7]
Tizikara et al., 2022	ML-aided Optical Monitoring	Eye diagram, BER, OSNR	Optical coherent system testbed	Not A/V-specific; no CWDM or visual inspection	[6]
Huang et al., 2023	Optical Sampling + Eye Monitoring	Q-value, BER, Time Jitter	Optical Tx (2.5 Gbps, 40–80 km)	No DL classification; focused on signal quality only	[17]
Our Work, 2025	**Multimodal DL classification**	**Accuracy, Recall, F1-score**	**CWDM OSA (Tabular + Eye Diagram)**	**First DL-based visual classification for CWDM A/V module**	–

TABLE 3: Comparison of Deep Learning-based Defect Detection Methods Using Tabular and/or Image Data.

Author & Year	Model / Approach	Data Type	Dataset	Limitation	Performance (Accuracy / F1-score)	Ref
Kong & Ni, 2020	Semi-supervised Ladder Net & VAE	Image (Wafer bin maps)	Two real-world wafer map datasets	Requires pre-cleaned maps, single GFA only	0.951 / 0.948	[71]
Xu et al., 2022	ResNet + CBAM + Cosine Norm	Image (Wafer maps: WM-811K)	Public WM-811K	Class imbalance affects rare patterns	0.939 / 0.923	[72]
Jiang et al., 2023	RAR-SSD (multi-scale + attention-enhanced SSD)	Image (PCB AOI images)	Internal PCB defect benchmark	Less effective for very small defects	0.967 / 0.908	[70]
Kumbhar et al., 2023	CNN + RNN + GAN (DeepInspect framework)	Image (Manufacturing images)	Internal annotated dataset	GAN instability, no public benchmark	0.973 / 0.908	[73]
Cho et al., 2023	ResNet + FNN + Logit Adjust.	Image + tabular (semiconductor)	Chip defect dataset	Not A/V; low image resolution; domain-specific tuning needed	0.961 / 0.950	[46]
Our Work, 2025	**ResNet-50 + FNN fusion**	**Eye diagram (4-ch) + tabular**	**CWDM OSA (real production)**	**First on CWDM A/V OSA; real data but no public dataset**	**0.910 / 0.904**	–

As shown in Tables 1–3, previous studies have either focused solely on conventional visual inspection or utilized performance evaluation methods not tailored to A/V optical modules. In contrast, our work introduces a novel approach that integrates both eye diagram images and manufacturing process data commonly used in performance evaluation, specifically targeting defect detection in A/V optical modules.

C. METHODS FOR IMBALANCED DATASET

Imbalanced datasets are handled using primarily two types of techniques [36]. The first approach involves resampling-based methods, which include oversampling and undersampling techniques. Oversampling increases the number of samples in the minority class to balance the dataset [37]– [39]. However, this method may lead to overfitting for the minority class, resulting in high training accuracy but poor generalization on the test set [40]. Conversely, undersampling reduces the number of samples from the majority class to match the minority class size [41], [42]. While this approach mitigates class imbalance, it risks the loss of critical information from the majority class, which potentially degrades model performance.

The second approach is based on cost-sensitive methods [43], which instead of adjusting class sample distributions assign different misclassification costs to different classes [44]. Also referred to as cost-adjusted or weighted loss functions, these techniques can suffer from a lack of Fisher consistency, which prevents convergence to the true probability distribution. To address this issue, logit-adjusted loss functions have been introduced [45], which provide adequate Fisher consistency while improving performance across varying imbalance ratios [46].

III. MULTIMODAL LEARNING-BASED OPTICAL MODULE DEFECT DETECTION MODEL

A. DATASET

The image dataset was compiled using the setup illustrated in Figure 5, where TOSA and ROSA pair was mounted on an EV B/D, which facilitated stable signal transmission and reception. The TOSA was connected to a PC via an optical fiber. After being received by the ROSA, the transmitted signal was captured by an oscilloscope, which enabled further analysis of transmission integrity and system performance. To measure the eye diagram, the oscilloscope was directly connected to the ROSA, which captured the transmitted signal for evaluation. The acquired signals were stored as images for further analysis. Each TOSA–ROSA pair consisted of four channels (4ch), and the dataset included four images per pair. The original dataset (Figure 6) comprised a total of 3,760 subsets and contained 15,040 images (3,760 × 4 channels), with each image having a resolution of 1026 × 770 pixels.

Additionally, the PC and monitor were connected to the TOSA and ROSA to facilitate screen defect testing. Based on the screen inspection results, the data were systematically recorded in a tabular format, where normal (pass) and defective (fail) samples were labeled "0" and "1", respectively. Each image set was annotated with the corresponding TOSA serial number and wavelength information. As shown in Figure 7, the normal images typically displayed minimal noise, with the cross-point centrally aligned. In contrast, the defective images(containing defect Types 1 and 2) often exhibited noticeable noise, with the cross-point shifted either above or below the center. All images were captured using a 10 GHz oscilloscope with a sampling rate of 40 GSa/s under controlled experimental conditions to ensure consistent and reliable data collection.

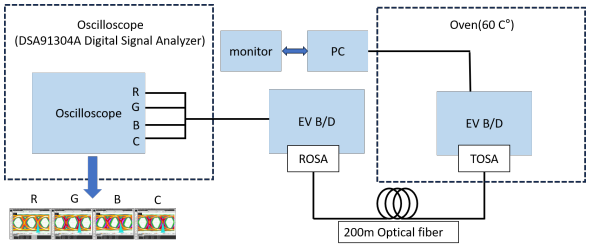


FIGURE 5: Configuration settings for acquiring image data acquisition.

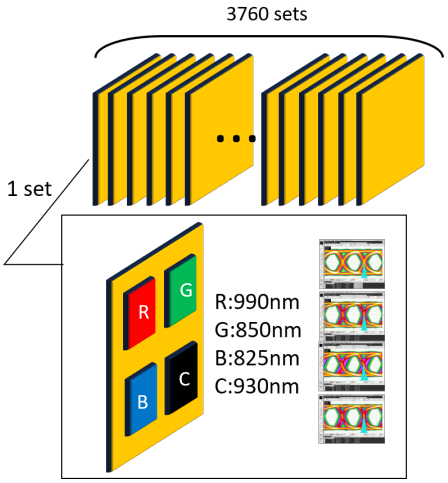


FIGURE 6: Structure of the image dataset.

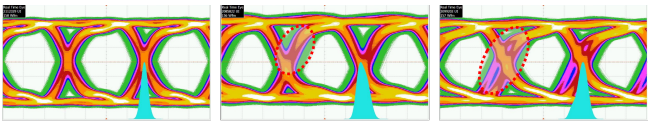


FIGURE 7: Eye diagram analysis: (a) Normal, (b) Defect Type 1, (c) Defect Type 2.

The tabular dataset consisted of numerical values recorded during the optical module manufacturing process, encompassing key parameters such as equipment position details, optical power at varying temperatures, jitter measurements, and screen inspection results as shown in Figure 8. This

dataset comprised 35 distinct features, which can be categorized into the following four groups(Table 4):

- Process data (optical aligner), represented by 16 features that describe the physical alignment of the optical module.
- Optical power data, recorded at 0°C, 25°C, and 60°C to evaluate temperature-dependent variations in optical performance.
- Jitter data, collected from four optical sources during eye-diagram analysis to evaluate signal integrity.
- Final inspection data (pass/fail), represented by binary labels, where "0" indicates a functional module (pass), and "1" denotes a defective module (fail).

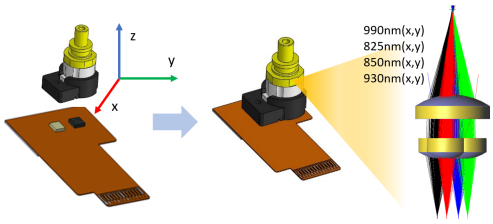


FIGURE 8: Schematic illustrating optical alignment process.

B. DATA PREPROCESSING

1) Image Downscaling

The raw images(Figure 5) were captured directly from the oscilloscope, representing the signals received by the ROSA. The dataset included images corresponding to four distinct wavelengths: 825 nm, 850 nm, 930 nm, and 990 nm. As part of the preprocessing step, the original 1026 × 770 pixel images were resized to 360 × 490 pixels, as shown in Figure 9. This downscaling reduced the image size to approximately one-fourth of the original, significantly lowering memory consumption. Moreover, the resizing helped suppress irrelevant noise, thereby enhancing the accuracy of subsequent analyses.

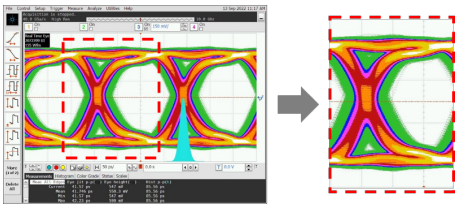


FIGURE 9: Image downscaling.

2) Data Augmentation

The dataset exhibits a class imbalance with a pass-to-fail ratio of 8:2, with the minority class containing fewer than 500 sets. Such imbalance complicates model training as the dominance of the majority class causes the model to underperform on minority-class samples, ultimately degrading overall performance [47]. When trained on the original dataset, the model achieved an accuracy of 87.2%; however, it suffered from overfitting to the majority class, which limited its ability to generalize effectively to unseen data.

To enhance model performance, oversampling techniques [48] based on data augmentation were applied. Given the structural characteristics of signal images, preserving their original shape during the augmentation process is crucial. Therefore, the preprocessed images from Figure 9 were augmented using random erasure [49] and Gaussian noise injection [50], [51] to generate additional synthetic samples while maintaining the inherent properties of the images. Furthermore, pixel shifting along the X- and Y-axes (ranging from 1 to 50 pixels) [52] was performed to introduce controlled variations into the dataset. For tabular data augmentation, Conditional Tabular GAN (CTGAN) [53] was utilized to generate synthetic data belonging to both the normal and defective classes, which ensured a more balanced dataset. Table 5 presents the final dataset distribution after image and tabular data augmentation.

1) Random image erasure: The total erasure area was randomly set to 2–40% of the entire image area, while the aspect ratio of the erased region was randomly set to 0.3–3.33. The actual size of the erased region was determined by these two parameters, which ensured variability in the augmentation process. The erasure location was randomly assigned relative to the top-left corner of the image, which added further randomness to the dataset. Figure 10 illustrates the results of the random erasure process.

2) Noise injection: Gaussian noise was generated based on a normal distribution and applied directly to the original images. The noise was defined with a mean (μ) of 0 and a standard deviation (σ) of 30, whereby controlled variation was maintained within the augmented dataset. Figure 10 presents the noise injection results.

3) Pixel shift: Pixel shifting was performed randomly along both axes, with horizontal and vertical displacements ranging between -50 and 50 pixels. Any blank regions introduced by the shift were filled with black pixels exhibiting red, green, and blue colors (255, 255, 255) to preserve image consistency. Figure 10 displays the results of the pixel shift–based augmentation.

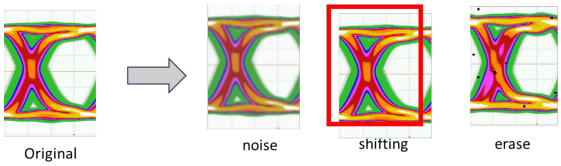


FIGURE 10: Three data augmentation methods: noise injection, pixel shifting, and random erasure.

4) CTGAN: The CTGAN method was utilized to generate synthetic tabular data and thus ensure a balanced dataset. The preprocessed tabular dataset was loaded from a CSV file, and

TABLE 4: Description of Tabular Dataset with Four Feature Groups.

Feature Group	Description	Key features (examples)
Process data (optical aligner)	Position alignment and optical power during align	Position coordinates (X, Y, Z), aligned optical power
Optical power data	Optical power variation across different temperatures and channels	Temperature (°C), optical power (per channel)
Jitter data for each channel	Signal quality and jitter measurement data	Jitter magnitude, Jitter average(total channel)
Final inspection data (Pass/Fail Judgment)	Final pass/fail classification for product quality assessment	Inspection result (normal/defective)

the CTGAN model was trained for 300 epochs to produce new synthetic samples. As shown in Figure 11, the generated tabular data closely followed the original data distribution while maintaining its statistical properties.

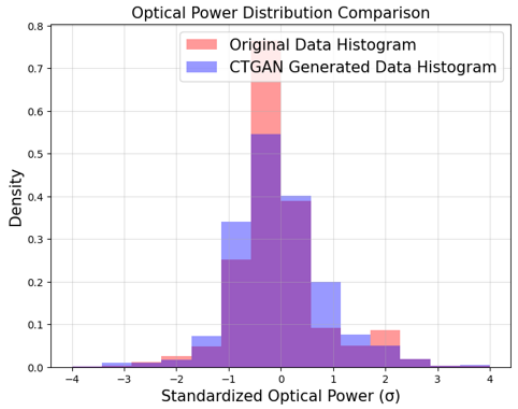


FIGURE 11: Comparison of data histogram before and after CTGAN application.

TABLE 5: Summary of Data Augmentation.

(a) Augmentation for Image Data

Label	Base	Random Erasing	Noise	Shift	Total (set)
Pass (0)	3167	3167	500	-	6834
Fail (1)	458	1779	2290	2290	6817

(b) Augmentation for Tabular Data

Label	Base	CTGAN	Total (set)
Pass (0)	3167	3667	6834
Fail (1)	458	6359	6817

C. DESIGN AND ARCHITECTURE OF MULTIMODAL LEARNING

1) Single Modal Structure

The ResNet-50 model [54] mmitigates the vanishing-gradient problem through residual learning, which incorporates shortcut connections in the following form:

$$\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{x}) + \mathbf{x}. \tag{1}$$

The ResNet-50 architecture has 50 layers and demonstrates superior performance in image classification tasks compared to Visual Geometry Group (VGG) networks. However, the computational overhead introduced by residual learning necessitates substantial memory and processing power. In this work, ResNet-50 is utilized as the backbone feature extractor as shown in Figures 12 and 13.

The preprocessed eye diagram images were processed through an extraction function (E) to extract features on a per-channel basis. TThe features extracted from all four channels (4ch) were then concatenated into a single-column vector and passed through a fully connected layer for final classification [55]. The model's binary classification performance was evaluated using 11 classifiers, namely Support Vector Machine (SVM) [56], Random Forest [57], k-Nearest Neighbors (KNN) [58], Logistic Regression [59], Gradient Boosting [60], Naïve Bayes [61], Decision Tree [62], AdaBoost [63], XGBoost [64], LightGBM [65], and CatBoost [66].

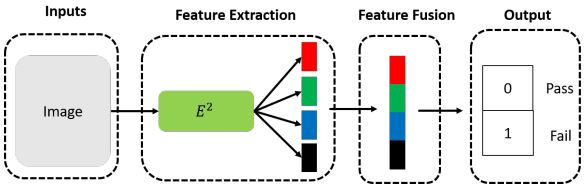


FIGURE 12: Structure of single modal model.

2) Binary Classification Model based on Multimodal Learning

Apart from image data, tabular data were incorporated as input. Feature vectors were extracted from the tabular data using a feedforward neural network [67] and from the image data using ResNet-50. The extracted feature vectors were then fused, with class-imbalance adjustments applied to the pass-fail ratio to improve classification accuracy. As shown in Figures 14 and 15, the proposed model could handle imbalanced datasets effectively [68], [69]. Figure 16 shows the multimodal feature extractor architecture of the proposed network structure, where B denotes the batch size.

Herein, i denotes the ith individual OSA, while n denotes the number of modalities; x corresponds to the input data for the classifier f, which can be either tabular or image data; y represents the ground truth, indicating whether the sample is classified as normal or defective; and the vector h represents the fusion of features extracted from different modalities, serving as the combined representation for final classification.

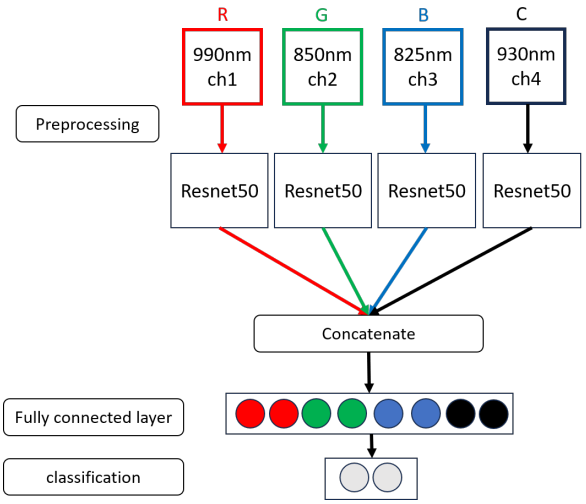


FIGURE 13: Binary classification model based on ResNet-50.

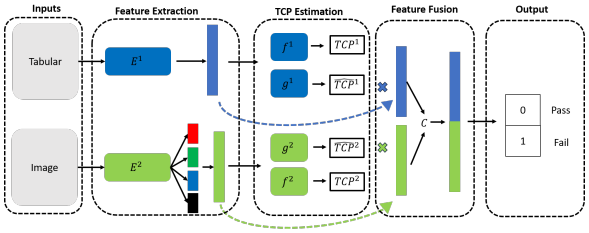


FIGURE 14: Structure of multimodal learning framework.

The probability $p^n(y \mid x_i^n)$ represents the logit-based classification probability obtained from the softmax function in classifier f. It quantifies the likelihood of correctly classifying a given OSA i for modality n.

$$y_i = f^n(x_i^n). \tag{2}$$

The concept of true class probability (TCP) [69] was utilized to weight feature vectors based on the relative importance of each modality. Specifically, if image data were deemed more reliable, the TCP value increased, whereas if tabular data appeared less reliable, the TCP value decreased; the final prediction was influenced accordingly. By assigning higher weights to more informative modalities, the TCP-based approach enhanced the model's robustness, particularly when handling imbalanced datasets.

$$TCP_i^n = p^n(y \mid x_i^n) = \text{Softmax}\left(f^n(x_i^n)\right). \tag{3}$$

One limitation of TCP is its reliance on ground-truth labels for computation, which restricts its applicability to the training phase. To enable the use of TCP during the testing phase, a TCP surrogate model must be trained.

This limitation was addressed by introducing a confidence regression network (CRN) [69], denoted as g^n, is introduced. The neural network g^n was specifically designed to predict TCP values. It incorporated a sigmoid activation function in the output layer to ensure that the TCP values remain within

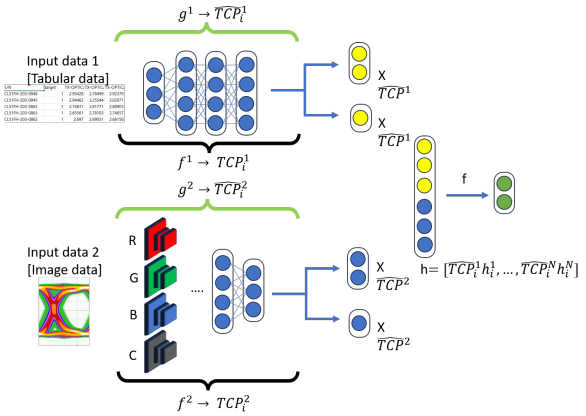


FIGURE 15: Binary classification model with multimodal learning framework.

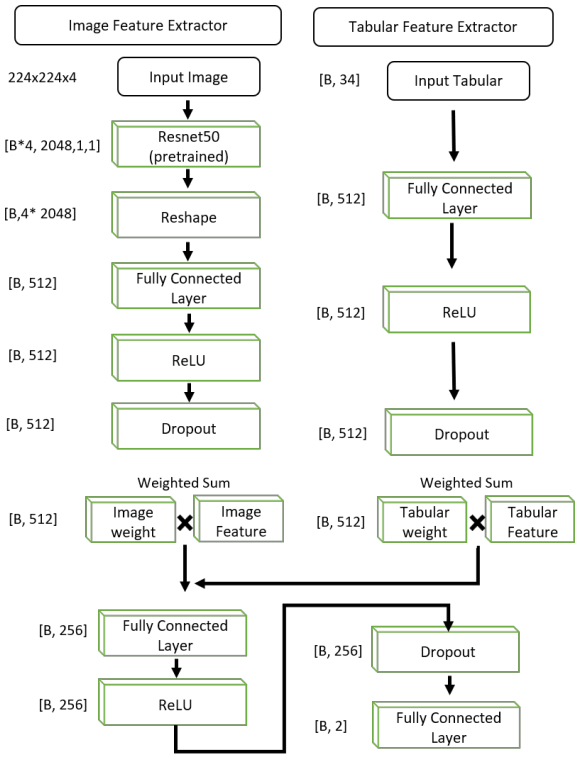


FIGURE 16: Multimodal Feature Extractor Architecture.

the range $(0, 1)$. This approach enabled the model to maintain the same architecture. As the individual modality classifiers, with only the output layer being trained separately.

$$\widehat{TCP}_i^n = g^n(x_i^n) \approx TCP_i^n. \tag{4}$$

The feature vector extracted from each modality, h_i^n, was combined with its corresponding approximated TCP value, $\widehat{TCP}_i^n$, to incorporate the reliability of each modality into the final representation. This process can be mathematically expressed as follows:

$$h_i = \left[\widehat{TCP}_i^1 h_i^1, \ldots, \widehat{TCP}_i^n h_i^n\right]. \tag{5}$$

TABLE 6: Correlation Coefficients between Features and Target Variable.

Key Feature	Correlation Coefficient
TX-OPTICAL-ALIGN-EPOXY-990	-0.23
TX-60D-850	0.20
TX-0D-825	0.28
TX-OPTICAL-ALIGN-DISTANCE-930	0.24
TX-0D-930	0.24
TX-60D-930	0.23
JITTER-930	0.22

To address class imbalance and prevent underfitting for the minority class, logit adjustment loss [45] was employed as the loss function.

IV. EXPERIMENTS AND RESULTS

A. DATASET CHARACTERISTICS

TThe correlation between the tabular features and the target variable (screen inspection results) was relatively weak, as shown in Table 6. This table presents the top seven features that exhibit the highest correlation with the target variable. In the feature names in Table 6, the notations 990, 850, 825, and 930 represent wavelengths. The key features included TX-60D (TOSA optical power at 60°C), TX-0D (TOSA optical power at 0°C), JITTER (eye diagram jitter), TX-OPTICAL-ALIGN-EPOXY (optical power from the optical aligner dataset), and TX-OPTICAL-ALIGN-DISTANCE (distance measurement from the optical aligner dataset).

A portion of the receiver signal was captured using an oscilloscope, which generated images for analysis. Each dataset (one set) consisted of four channels: R (990 nm), G (850 nm), B (825 nm), and C (930 nm). The shape of the eye diagram and the presence of signal noise were found to be strongly associated with product defects. If any of the four channels exhibited a signal defect, the likelihood of the entire sample being classified as defective increased.

As shown in Figure 7, the image dataset contained a mixture of image types classified as normal or defective, including Types 1, 2, 3, 4, and 5. For example, Type 1 corresponded to images where all four channels were normal, while Types 2 and 3 contained only one or two normal images, respectively. However, all images in a set being classified as normal did not guarantee that the final sample would be classified as a pass. Similarly, a set consisting entirely of defective images would not always be classified as a fail.

Table 7 categorizes the compositions of pass and fail images within a single set (R, G, B, and C) into five types. The probability of a pass classification was determined based on the composition of normal and defective images in the set. The likelihood of a set passing the final screen inspection increased with the number of channels containing normal images, following the expected order: Type 1 > Type 2 > Type 3 > Type 4 > Type 5.

However, the final classification was not determined solely by image data. For instance, even if a set exhibited Type 1 composition, it could still be classified as defective if the optical power was low. Conversely, a Type 5 set could still pass if the image structure and optical power indicated acceptable conditions. This suggests that the image data alone were not directly used to classify defects but rather constituted an indirect factor influencing defect classification.

TABLE 7: Example of Pass Probability based on Image Dataset.

Type	R	G	B	C	Pass probability
1	O	O	O	O	a
2	X	O	O	O	b
3	X	X	O	O	c
4	X	X	X	O	d
5	X	X	X	X	e

O: normal image (Pass), X: defective image (Fail).

Each row in the tabular dataset represented a single set (one set) and comprised 35 features associated with the optical module manufacturing process. These features include optical alignment data, optical measurement data (X-axis and Y-axis alignment information and optical power levels), eye-diagram jitter, and screen inspection results.

Model performance was evaluated based on the confusion matrix presented in Table 8with accuracy, precision, recall, and F1-score being the key metrics. To ensure reliability, five-fold cross-validation was performed. Table 9 details the experimental setup used for training.

TABLE 8: Confusion Matrix.

		Predicted Class	
		Fail	Pass
Actual Class	Fail	True Positive (TP)	False Negative (FN)
	Pass	False Positive (FP)	True Negative (TN)

TABLE 9: Experimental Setup and Training Environment.

OS	Windows 11
CPU	13th Gen Intel(R) Core(TM) i7-1360P
RAM	32GB
Library	Python 3.8 Scikit-learn 1.3.2 torch 2.4.0+cu118 numpy 1.24.3

B. RESULTS AND DISCUSSION

The impact of the balanced dataset, multimodal learning, and hyperparameter optimization on model performance was evaluated. Among the tested classifiers, only those that demonstrated superior empirical performance were selected for the final evaluation to ensure a fair and meaningful comparison across model types. The baseline models, i.e., SVM, fandom forest, KNN, logistic regression, gradient boosting, and decision tree, exhibited significantly lower recall and F1-scores on the imbalanced dataset than on the balanced dataset. As shown in Table 10, the dataset augmentation process performed to mitigate class imbalance led to performance improvements across all models. However, despite these enhancements, the baseline models struggled to generalize effectively for defect detection.

TABLE 10: Dataset with Data Augmentation.

Dataset	Data Augmentation (Image) pass(0) set				Data Augmentation (Image) fail(1) set			
	Base	Random Erasing	Noise	Shifting	Base	Random Erasing	Noise	Shifting
Imbalanced-1 (6834:2237)	3167	3167	500	0	458	1779	0	0
Imbalanced-2 (4500:916)	3167	0	1333	0	458	0	458	0
Imbalanced-3 (4586:458)	3167	1419	0	0	458	0	0	0
Balanced-1 (6834:6817)	3167	3167	500	0	458	1779	2290	2290
Balanced-2 (4586:4527)	3167	1419	0	0	458	1779	0	2290
Balanced-3 (2212:2237)	2212	0	0	0	458	1779	0	0

TABLE 11: Description of Model Parameters.

Parameter	Description
`num_classes`	Number of output classes in classification
`epochs`	Number of training iterations over the dataset
`batch_size`	Number of samples per training batch
`learning_rate`	Step size for optimizer updates
`tau`	Scaling factor or threshold parameter (specific to some algorithms)
`class_weights`	Weighting factors assigned to different classes for imbalanced datasets
`C`	Regularization parameter in models like SVM and Logistic Regression
`solver`	Optimization algorithm used in Logistic Regression (e.g., 'lbfgs')
`max_iter`	Maximum number of iterations for convergence in optimization
`n_estimators`	Number of trees in ensemble models like Random Forest and Gradient Boosting
`max_depth`	Maximum depth of decision trees
`min_samples_split`	Minimum number of samples required to split a node in decision trees
`n_neighbors`	Number of nearest neighbors considered in KNN
`weights`	Weighting function for nearest neighbors (e.g., 'uniform')
`p`	Power parameter for Minkowski distance in KNN ($p = 2$ corresponds to Euclidean distance)
`kernel`	Type of kernel function used in SVM (e.g., 'rbf')

TABLE 12: Model Parameters for Different Datasets.

Model	Dataset	Parameters
Multimodal Learning	Balanced-1	`num_classes=2, epochs=300, batch_size=16, learning_rate=1e-4, tau=1.0, class_weights=[1.0, 1.0]`
	Balanced-2	`num_classes=2, epochs=300, batch_size=16, learning_rate=1e-4, tau=1.0, class_weights=[1.0, 1.0]`
	Balanced-3	`num_classes=2, epochs=300, batch_size=16, learning_rate=1e-4, tau=1.0, class_weights=[1.0, 1.0]`
	Imbalanced-1	`num_classes=2, epochs=300, batch_size=16, learning_rate=1e-4, tau=1.0, class_weights=[1.0, 3.0]`
	Imbalanced-2	`num_classes=2, epochs=300, batch_size=16, learning_rate=1e-4, tau=1.0, class_weights=[1.0, 5.0]`
	Imbalanced-3	`num_classes=2, epochs=300, batch_size=16, learning_rate=1e-4, tau=2.0, class_weights=[1.0, 10.0]`
SVM	Balanced-1 / Imbalanced-1	`SVC(C=1.0, kernel='rbf')`
Random Forest	Balanced-1 / Imbalanced-1	`n_estimators=100, max_depth=None, min_samples_split=2`
KNN	Balanced-1 / Imbalanced-1	`n_neighbors=5, weights='uniform', p=2`
Logistic Regression	Balanced-1 / Imbalanced-1	`C=1.0, solver='lbfgs', max_iter=100`
Gradient Boosting	Balanced-1 / Imbalanced-1	`n_estimators=100, learning_rate=0.1, max_depth=3`
Decision Tree	Balanced-1 / Imbalanced-1	`max_depth=None, min_samples_split=2`
Multimodal Learing (optimization)	Balanced-1	`num_classes=2, epochs=300, batch_size=64, learning_rate=0.001, tau=0.5, class_weights=[1.0, 1.0]`
	Imbalanced-1	`num_classes=2, epochs=300, batch_size=16, learning_rate=0.0005, tau=0.3, class_weights=[1.0, 3.0]`

For the imbalanced dataset, KNN achieved the highest F1-score (0.619) among the baseline models, although its recall remained low (0.572). The other baseline models performed even worse, exhibiting recall values mostly below 0.5; this highlights their limitations in handling class imbalance. On the balanced dataset, the baseline models performed better, with KNN achieving an F1-score of 0.853 and a recall of 0.834, the highest values among the baseline models. However, even with balanced data, the baseline models were outperformed by the multimodal learning model, as shown in Table 13, which provides a detailed comparison of the models across datasets.

The multimodal learning model consistently outperformed all baseline models in terms of both accuracy and robustness. When trained on the balanced dataset, this model achieved an F1-score of 0.885, a recall of 0.880, and an accuracy of 0.872, surpassing all the baseline models. On the imbalanced dataset, its F1-score (0.635) and recall (0.551) were still higher than those of the best-performing baseline model; this reinforces the advantage of incorporating multimodal information. The parameter configurations used for these models are detailed in Table 11, while their specific values across

TABLE 13: Performance of Different Models on Balanced and Imbalanced Datasets.

Model	Dataset	Accuracy	F1-score	Precision	Recall
SVM	Balanced-1	0.815±0.003	0.789±0.004	0.920±0.003	0.690±0.005
	Imbalanced-1	0.772±0.002	0.290±0.010	0.842±0.015	0.175±0.007
Random Forest	Balanced-1	0.833±0.002	0.817±0.003	0.903±0.003	0.746±0.003
	Imbalanced-1	0.772±0.001	0.252±0.006	0.987±0.003	0.145±0.004
KNN	Balanced-1	0.856±0.002	0.853±0.003	0.873±0.004	0.834±0.007
	Imbalanced-1	0.813±0.002	0.619±0.006	0.674±0.004	0.572±0.011
Logistic Regression	Balanced-1	0.779±0.003	0.761±0.004	0.829±0.003	0.704±0.005
	Imbalanced-1	0.735±0.001	0.050±0.004	0.505±0.045	0.026±0.002
Gradient Boosting	Balanced-1	0.788±0.003	0.764±0.004	0.867±0.004	0.683±0.004
	Imbalanced-1	0.738±0.001	0.036±0.006	0.743±0.049	0.018±0.003
Decision Tree	Balanced-1	0.745±0.004	0.747±0.005	0.741±0.004	0.753±0.007
	Imbalanced-1	0.677±0.006	0.413±0.011	0.399±0.010	0.428±0.012
Multimodal Learning	Balanced-1	**0.872±0.010**	0.885±0.009	**0.891±0.017**	0.880±0.015
	Balanced-2	0.869±0.014	**0.887±0.011**	0.880±0.012	0.880±0.012
	Balanced-3	0.834±0.021	0.828±0.022	0.858±0.023	0.801±0.022
	Imbalanced-1	0.781±0.016	0.635±0.015	0.750±0.019	0.551±0.028
	Imbalanced-2	0.699±0.027	0.440±0.014	0.318±0.015	0.722±0.017
	Imbalanced-3	0.464±0.023	0.236±0.008	0.136±0.005	**0.896±0.024**
Multimodal Learning (Optimization)	Balanced-1	**0.910±0.012**	**0.904±0.014**	**0.968±0.004**	**0.859±0.029**
	Imbalanced-1	0.832±0.037	0.692±0.031	0.670±0.097	0.735±0.066

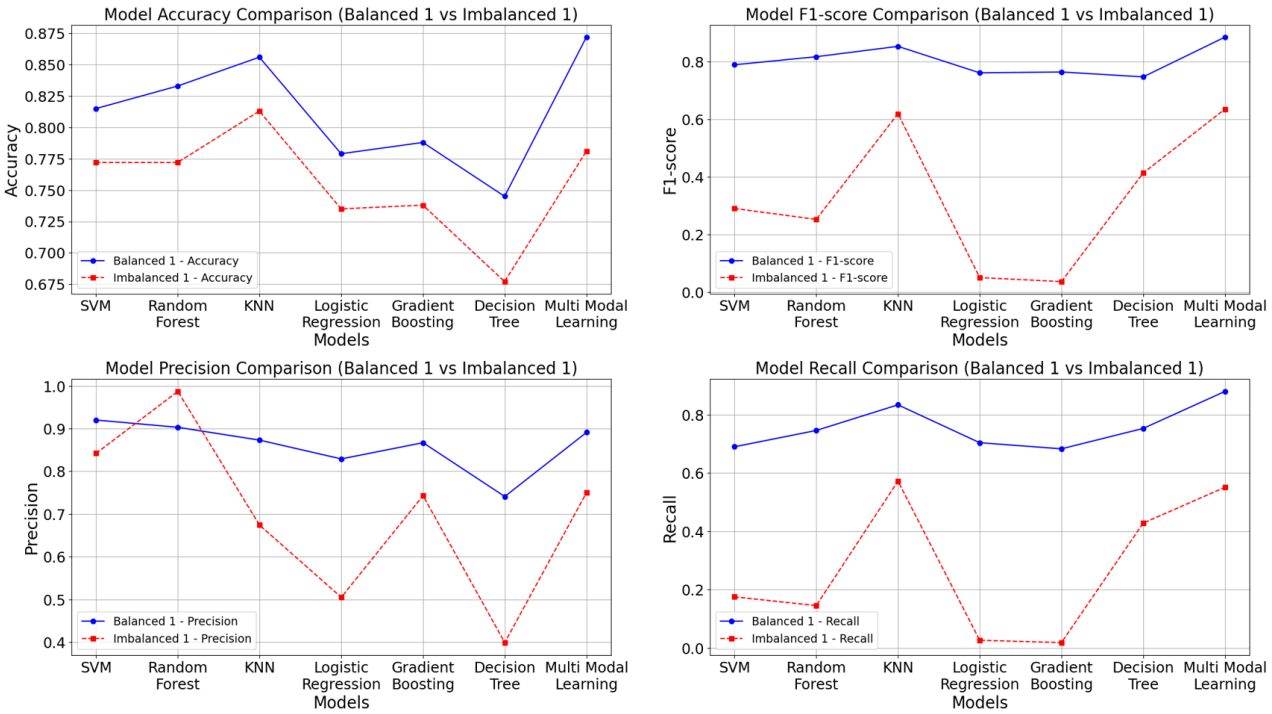


FIGURE 17: Comparison of Performance Results of Different Models on Balanced and Imbalanced Dataset.

different dataset variations are outlined in Table 12.

Further hyperparameter optimization significantly improved the performance of the multimodal learning model trained on the balanced dataset. The F1-score increased from 0.885 to 0.904, the recall from 0.880 to 0.859, and the accuracy from 0.872 to 0.910. Additionally, precision improved from 0.891 to 0.968, which indicates more reliable classification of defective samples. On the imbalanced dataset, hyperparameter tuning had a less pronounced effect: the F1-score increased from 0.635 to 0.692, while the recall remained relatively low. These results, summarized in Table 13, highlight the critical role of dataset balancing in maximizing model performance. Figure 17 shows the performance results of different models on balanced and imbalanced datasets.

In summary, balancing the dataset and optimizing the hyperparameters significantly enhanced defect detection performance, particularly for the multimodal learning model. While the baseline models benefited from data balancing, they remained less effective than the multimodal approach. These results confirm that a well-structured dataset combined with an optimized multimodal learning model ensures superior defect classification performance.

V. CONCLUSION

IHerein, we proposed multimodal learning for defect detection in optical modules, leveraging both eye-diagram images and tabular data from the optical module manufacturing process. To address dataset imbalance, we evaluated model performance on both imbalanced and balanced datasets using multiple data augmentation techniques. Furthermore, we comparatively analyzed a single-modal model, which utilized only image data, against a

multimodal learning model that integrated both image and tabular data. Based on the experimental results, the multimodal model significantly outperformed its single-modal counterpart, achieving an accuracy of 91% on the balanced dataset and 83.2% on the imbalanced dataset. These findings lead to two key conclusions: (1) using a balanced dataset enhances overall model performance, and (2) the multimodal learning approach provides superior defect detection accuracy compared with single-modal methods. This work underscores the potential of multimodal learning in improving the detection of defects in optical modules and lays the groundwork for future research on advanced deep learning techniques to streamline screen inspection and quality control.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2023-00242528).

REFERENCES

[1] F. Zhang, L. Xiong, R. Min, Y. Guo, Z. Wang, and K. Xiao, "A miniaturized optical communication module: Design, development, and testing," in *Proc. IEEE 12th Int. Conf. Inf., Commun. Netw. (ICICN)*, Guilin, China, 2024, pp. 132–139, doi: 10.1109/ICICN62625.2024.10761827.

[2] H. D. Kim, I. Kim, H.-S. Lee, J.-B. Yoon, H.-K. Shin, and S.-W. Ryu, "Low-cost CWDM transceiver module by spatial power combination and division for multimode fiber links," *IEEE Photon. Technol. Lett.*, vol. 29, no. 1, pp. 39–42, Jan. 2017, doi: 10.1109/LPT.2016.2623659.

[3] T. Sakamoto, S. Nobuo, S. Koike, K. Hadama, and K. Naoya, "4 channel/spl times/10 Gbit/s optical module for CWDM links," in *Proc. 54th Electron. Compon. Technol. Conf. (ECTC)*, Las Vegas, NV, USA, 2004, vol. 1, pp. 1024–1028, doi: 10.1109/ECTC.2004.1319465.

[4] Y.-Y. Chen, K.-W. Jwo, and R.-S. Chang, "Research of an intelligent auto-controlling system for LCD screen flicker," *J. Display Technol.*, vol. 12, no. 6, pp. 557–561, Jun. 2016, doi: 10.1109/JDT.2015.2506783.

[5] R. Patel, S. Pandey, C. I. Patel, and R. Paul, "Effective flicker detection technique using artificial neural network for video," in *Proc. ICRISET 2017 - Int. Conf. Research and Innovations in Sci., Eng. and Technol.*, Kalpa Publications in Engineering, vol. 1, pp. 442–450, 2017, doi: 10.1109/ACCESS.2019.2925554.

[6] D. K. Tizikara, J. Serugunda, and A. Katumba, "Machine learning-aided optical performance monitoring techniques: A review," *Front. Commun. Netw.*, vol. 2, Jan. 2022, doi: 10.3389/frcmn.2021.756513.

[7] J.-Y. Park, H.-S. Lee, S.-S. Lee, and Y.-S. Son, "Passively aligned transmit optical subassembly module based on a WDM incorporating VCSELs," *IEEE Photon. Technol. Lett.*, vol. 22, no. 24, pp. 1790–1792, Dec. 15, 2010.

[8] G. Sialm *et al.*, "Comparison of simulation and measurement of dynamic fiber-coupling effects for high-speed multimode VCSELs," *J. Lightw. Technol.*, vol. 23, no. 7, pp. 2318–2330, Jul. 2005, doi: 10.1109/JLT.2005.850054.

[9] J. Hancock, "Jitter—Understanding it, measuring it, eliminating it: Part 1: Jitter fundamentals," *High Frequency Electron.*, Summit Tech. Media, Apr. 2004.

[10] A. S. Das, A. K. Pattanaik, and A. S. Patra, "Simultaneous long-haul gigabit transmission in bidirectional WDM-PON," Dept. of Physics, Sidho-Kanho-Birsha Univ., Purulia, West Bengal, India, and VSSUT, Burla, Sambalpur, Orissa.

[11] R. Stephens, "Analyzing jitter at high data rates," *IEEE Commun. Mag.*, vol. 42, no. 2, pp. S6–S10, Feb. 2004, doi: 10.1109/MCOM.2003.1267095.

[12] N. Juhari, P. S. Menon, A. A. Ehsan, and S. Shaari, "4-Channel double S-shaped AWG demultiplexer on SOI for CWDM," in *Proc. 17th Int. Conf. Adv. Commun. Technol. (ICACT)*, PyeongChang, South Korea, 2015, pp. 436–440, doi: 10.1109/ICACT.2015.7224833.

[13] N. Juhari, P. S. Menon, A. A. Ehsan, and S. Sahbudin, "Design and application of 8-channel SOI-based AWG demultiplexer for CWDM-system," *AIP Conf. Proc.*, Apr. 2015, doi: 10.1063/1.4915235.

[14] P. P. Hema and A. Sangeetha, "Analysis of four channel CWDM transceiver modules based on extinction ratio and with the use of EDFA," *Int. J. Eng. Technol. (IJET)*, vol. 5, no. 3, pp. 2895–2902, Jun.–Jul. 2013.

[15] A. Priyadarshi, P. V. Ramana, C. J. Leo, S. G. Mhaisalkar, and V. Kripesh, "Link simulation of four channel CWDM transceiver modules," in *Proc. 6th Electron. Packag. Technol. Conf. (EPTC)*, Singapore, 2004, pp. 767–771, doi: 10.1109/EPTC.2004.1396711.

[16] S. Robinson and R. Nakkeeran, "Analysis of CWDM network with photonic crystal ring resonator based add-drop filter," in *Proc. Int. Conf. Emerg. Trends Sci., Eng. Technol. (INCOSET)*, Tiruchirappalli, India, 2012, pp. 308–311, doi: 10.1109/INCOSET.2012.6513923.

[17] T. Huang, Z. Fan, J. Su, and Q. Qiu, "Real-time eye diagram monitoring for optical signals based on optical sampling," *Appl. Sci.*, vol. 13, no. 3, pp. 1–10, Jan. 2023, doi:10.3390/app13031363.

[18] Quantum Data, "882E HDMI protocol analyzer generator," [Online]. Available: *https* : $//www.quantumdata.com/$

[19] Rohde & Schwarz, "A/V quality testing of set-top boxes and TVs," Rohde & Schwarz, 2024. [Online]. Available: *https* : $//www.rohde-schwarz.com/cz/applications/a-v-quality-testing-of-set-top-boxes-and-tvs-application-card_5 6279-54541.html$.

[20] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019, doi: 10.1109/TPAMI.2018.2798607.

[21] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 96–108, Nov. 2017, doi: 10.1109/MSP.2017.2738401.

[22] Y. Wang, "Survey on deep multi-modal data analytics: Collaboration, rivalry, and fusion," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 17, no. 1s, pp. 1–25, 2021, doi: 10.1145/3408317.

[23] B. Nojavanasghari, D. Gopinath, J. Koushik, T. Baltrušaitis, and L.-P. Morency, "Deep multimodal fusion for persuasiveness prediction," in *Proc. 18th ACM Int. Conf. Multimodal Interact.*, 2016, pp. 284–288, doi: 10.1145/2993148.2993176.

[24] S. Chen and Q. Jin, "Multi-modal conditional attention fusion for dimensional emotion prediction," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 571–575, doi: 10.1145/2964284.2967286.

[25] S. Poria, E. Cambria, and A. Gelbukh, "Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis," in *Proc. Conf. Empir. Methods Nat. Lang. Process.*, 2015, pp. 2539–2544, doi: 10.18653/v1/D15-1303.

[26] A. Zadeh, M. Chen, S. Poria, E. Cambria, and L.-P. Morency, "Tensor fusion network for multimodal sentiment analysis," 2017, arXiv:1707.07250.

[27] Z. Liu, Y. Shen, V. B. Lakshminarasimhan, P. P. Liang, A. Zadeh, and L.-P. Morency, "Efficient low-rank multimodal fusion with modality-specific factors," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguist. (ACL)*, Melbourne, VIC, Australia, Jul. 2018, pp. 2247–2256, doi: 10.18653/v1/P18-1209.

[28] M. Hou, J. Tang, J. Zhang, W. Kong, and Q. Zhao, "Deep multimodal multilinear fusion with high-order polynomial pooling," in *Proc. Adv.*

Neural Inf. Process. Syst., vol. 32, Vancouver, BC, Canada, Dec. 2019, pp. 12113–12122.

[29] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021, doi: 10.1109/TGRS.2020.3016820.

[30] T. Wörtwein and S. Scherer, "What really matters—An information gain analysis of questions and reactions in automated PTSD screenings," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, 2017, pp. 15–20.

[31] J. Tian, W. Cheung, N. Glaser, Y.-C. Liu, and Z. Kira, "UNO: Uncertainty-aware noisy-or multimodal fusion for unanticipated input degradation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2020, pp. 5716–5723, doi: 10.1109/ICRA40945.2020.9197266.

[32] Z. Han, C. Zhang, H. Fu, and J. T. Zhou, "Trusted multi-view classification," in *Proc. Int. Conf. Learn. Representations (ICLR)*, Virtual Event, May 2021.

[33] M. Turk, "Multimodal interaction: A review," *Pattern Recognit. Lett.*, vol. 36, pp. 189–195, Jan. 2014, doi: 10.1016/j.patrec.2013.07.003.

[34] R. Panda, C.-F. R. Chen, Q. Fan, X. Sun, K. Saenko, A. Oliva, and R. Feris, "AdaMML: Adaptive multi-modal learning for efficient video recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 7576–7585.

[35] A. Tonge and C. Caragea, "Dynamic deep multi-modal fusion for image privacy prediction," in *Proc. 2019 World Wide Web Conf. (WWW)*, San Francisco, CA, USA, May 2019, pp. 1829–1840, doi: 10.1145/3308558.3313691.

[36] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[37] L. Shen, Z. Lin, and Q. Huang, "Relay backpropagation for effective learning of deep convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 467–482, doi: 10.1007/978-3-319-46478-7_29.

[38] N. Sarafianos, X. Xu, and I. A. Kakadiaris, "Deep imbalanced attribute classification using visual attention aggregation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, Sep. 2018, pp. 708–725, doi: 10.1007/978-3-030-01252-6_42.

[39] J. Byrd and Z. C. Lipton, "What is the effect of importance weighting in deep learning?" in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, Long Beach, CA, USA, Jun. 2019, pp. 872–881.

[40] R. C. Holte, L. Acker, and B. W. Porter, "Concept learning and the problem of small disjuncts," in *Proc. 11th Int. Joint Conf. Artif. Intell. (IJCAI)*, 1989, pp. 813–818.

[41] A. More, "Survey of resampling techniques for improving classification performance in unbalanced datasets," *arXiv preprint arXiv:1608.06048*, 2016.

[42] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Netw.*, vol. 106, pp. 249–259, Oct. 2018, doi: 10.1016/j.neunet.2018.07.011.

[43] C. Elkan, "The foundations of cost-sensitive learning," in *Proc. 17th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2001, pp. 973–978.

[44] K. M. Ting, "An instance-weighting method to induce cost-sensitive trees," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 3, pp. 659–665, May–Jun. 2002, doi: 10.1109/TKDE.2002.1000348.

[45] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit, and S. Kumar, "Long-tail learning via logit adjustment," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2021.

[46] H. Cho, W. Koo, and H. Kim, "Prediction of highly imbalanced semiconductor chip-level defects in module tests using multimodal fusion and logit adjustment," *IEEE Trans. Semicond. Manuf.*, vol. 36, no. 3, pp. 425–433, Aug. 2023, doi: 10.1109/TSM.2023.3283101.

[47] L. Wang, M. Han, X. Li, N. Zhang, and H. Cheng, "Review of classification methods on unbalanced data sets," *IEEE Access*, vol. 9, pp. 64606–64634, Apr. 2021, doi: 10.1109/ACCESS.2021.3074243.

[48] M. S. Shelke, P. R. Deshmukh, and V. K. Shandilya, "A review on imbalanced data handling using undersampling and oversampling technique," *Int. J. Recent Trends Eng. Res. (IJRTER)*, vol. 3, no. 4, pp. 444–449, Apr. 2017, doi: 10.23883/IJRTER.2017.3168.0UWXM.

[49] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 07, pp. 13001–13008, 2020, doi: 10.1609/aaai.v34i07.7000.

[50] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[51] S. S. Lee, "Noisy replication in skewed binary classification," *Comput. Stat. Data Anal.*, vol. 34, no. 2, pp. 165–191, Aug. 2000, doi: 10.1016/S0167-9473(99)00095-X.

[52] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. 7th Int. Conf. Document Anal. Recognit. (ICDAR)*, Edinburgh, U.K., Aug. 2003, pp. 958–963, doi: 10.1109/ICDAR.2003.1227801.

[53] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional GAN," in *Advances in Neural Information Processing Systems*, vol. 32, 2019, pp. 7333–7343.

[54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[55] J. G. Choi, D. C. Kim, M. Chung, S. Lim, and H. W. Park, "Multimodal 1D CNN for delamination prediction in CFRP drilling process with industrial robots," *Comput. Ind. Eng.*, vol. 190, Art. no. 110074, Apr. 2024, doi: 10.1016/j.cie.2024.110074.

[56] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[57] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[58] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.

[59] D. R. Cox, "The regression analysis of binary sequences," *J. R. Stat. Soc. B (Methodol.)*, vol. 20, no. 2, pp. 215–242, 1958, doi: 10.1111/j.2517-6161.1958.tb00292.x.

[60] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.

[61] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York, NY, USA: Wiley, 1973.

[62] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Monterey, CA, USA: Wadsworth International Group, 1984.

[63] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of online learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.

[64] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, San Francisco, CA, USA, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[65] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 3146–3154.

[66] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Advances in Neural Information Processing Systems*, vol. 31, 2018, pp. 6639–6649.

[67] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics Intell. Lab. Syst.*, vol. 39, no. 1, pp. 43–62, Nov. 1997, doi: 10.1016/S0169-7439(97)00061-0.

[68] Z. Han, F. Yang, J. Huang, C. Zhang, and J. Yao, "Multimodal dynamics: Dynamical fusion for trustworthy multimodal classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 20675–20685, doi: 10.1109/CVPR52688.2022.02005.

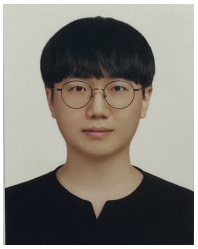
[69] C. Corbière, N. Thome, A. Bar-Hen, M. Cord, and P. Pérez, "Addressing failure prediction by learning model confidence," in *Advances in Neural Information Processing Systems*, vol. 32, 2019, pp. 2902–2913.

[70] W. Jiang, T. Li, S. Zhang, W. Chen, and J. Yang, "PCB defects target detection combining multi-scale and attention mechanism," *Engineering Applications of Artificial Intelligence*, vol. 123, Art. no. 106359, 2023, doi: 10.1016/j.engappai.2023.106359.

[71] Y. Kong and D. Ni, "A semi-supervised and incremental modeling framework for wafer map classification," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 1, pp. 62–71, Feb. 2020, doi: 10.1109/TSM.2020.2964581.

[72] Q. Xu, N. Yu, and F. Essaf, "Improved wafer map inspection using attention mechanism and cosine normalization," *Machines*, vol. 10, no. 2, p. 146, Feb. 2022, doi: 10.3390/machines10020146.

[73] A. Kumbhar, A. Chougule, P. Lokhande, S. Navaghane, A. Burud, and S. Nimbalkar, "DeepInspect: An AI-powered defect detection for manufacturing industries," *arXiv preprint arXiv:2311.03725*, Nov. 2023.



KYU-JEONG CHOI rreceived the B.S. degree in Physics from Chonnam National University in 2017. He is currently pursuing the M. S. degree in Data Science at Chonnam National University. Since 2017, he has been with Opticis Co., Ltd., Korea, where he works in the field of optical module technology and data analysis.

JIA YANG received the B.S. and M.S. degrees in Agricultural Economics and Food Distribution from Dankook University, Cheonan, Korea, in 2020 and 2021, respectively. She is currently pursuing the Ph.D. degree at the Graduate School of Data Science, Chonnam National University, Gwangju, Korea. With a background in agricultural economics, she is focusing on applying data science techniques to agricultural forecasting and market analysis

BOTAMBU COLLINS received his B.Sc. and M.Sc. degrees in Political Science from the University of Buea, Cameroon, respectively. He possesses a second master's in computer engineering from Yeungnam University, South Korea. He has contributed to over eight publications in various reputable journals/conferences. He is pursuing his Ph.D. at the Graduate School of Data Science at Chonnam National University, South Korea. His research interests include social computing, complex systems including social networks and information mining, data analytics with graph networks, graph neural networks for disease forecasting, temporal dynamic models, and collective/swarm intelligence.

SUNG-GEUN KIM received the B.S. degree in Physics from Chonnam National University in 2019 and the M.S. degree in Physics from the same university in 2021. Since 2021, he has been working at Opticis Co., Ltd. His research interests include VCSEL device processing and analysis.

DO-JIN LIM received the B.S. degree in Electronics Engineering from The University of Suwon in 2013. Since August 2016, he has been working at Opticis Co.,Ltd. His research interests include optical communication systems and signal processing.

JIN-TAEK SEONG received the B.S. degree from the University of Seoul, Seoul, Korea, in 2006 and the M.S. and Ph.D. degree from the Gwangju Institute of Science and Technology (GIST), Gwangju, Korea, in 2008 and 2014, respectively. He was an associate professor at the Department of Convergence Software, Mokpo National University from March 2018 to February 2023. Since then, he is currently working as an associate professor at the Graduate School of Data Science, Chonnam National University. His main research interests include artificial intelligence, information theory, and data analysis in multi-domain areas.