A Hybrid Model for Identifying Defect Wafer Map Patterns Based on an Optimized Deep Convolution Neural Network (OPDCNet) with a Classifier

This thesis is submitted to the Department of Electronics and Telecommunication Engineering of Chittagong University of Engineering & Technology As a partial fulfillment of the requirement for the degree of Master of Science in Electronics and Telecommunication Engineering.

Submitted By

Sharith Dhar

Student ID: 21METE001P

Supervised By

Dr. Md. Saiful Islam

Associate Professor

Department of Electronics and Telecommunication Engineering

Chittagong University of Engineering and Technology

MASTER OF SCIENCE IN ELECTRONICS AND TELECOMMUNICATION ENGINEERING

CHITTAGONG UNIVERSITY OF ENGINEERING AND TECHNOLOGY

APRIL 2025

Department of Electronics and Telecommunication Engineering
Chittagong University of Engineering and Technology

The Thesis Title: "A Hybrid Model for Indentifying Defect Wafer Map Patterns Based on an Optimized Deep Convolution Neural Network (OPDCNET) with Classifier" submitted by Sharith Dhar, Roll No. 21METE001P Session 2021-2022, has been accepted as satisfactory in partial fulfillment of the requirements for the degree of Master of Science in Electronics and Telecommunication Engineering on 10/04/2025.

BOARD OF EXAMINERS

1. ---------- Supervisor Chairman
Dr. Md. Saiful Islam
Associate Professor
Department of Electronics and Telecommunication Engineering
CUET.

2. ---------- (Ex-Officio) Member
Dr. Md. Azad Hossain
Professor & Head
Department of Electronics and Telecommunication Engineering
CUET.

3. ---------- Member
Dr. Md. Jahedul Islam
Professor
Department of Electronics and Telecommunication Engineering
CUET.

4. ---------- Member
Dr. Nursadul Mamun
Assistant Professor
Department of Electronics and Telecommunication Engineering
CUET.

5. ---------- External-Member
Dr. Mohammad Rubaiyat Tanvir Hossain
Professor
Department of Electrical and Electronic Engineering
CUET.

Declaration

I hereby declare that the work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the Thesis contains no material previously published or written by another person except where due reference is cited. Furthermore, the Thesis complies with PLAGIARISM and ACADEMIC INTEGRITY regulation of CUET.

Sharith Dhar
21METE001P
Department of Electronics and Telecommunication Engineering
Chittagong University of Engineering & Technology (CUET)

Dedication

Dedicated to

My beloved parents and Dr. Md. Saiful Islam sir
For their continuous support

List of publications

1. **Sharith Dhar** and **Md. Saiful Islam**, "A Hybrid DCNN ECOC-SVM System for Detecting Semiconductor Wafer Defects Using Histogram Equalization," *4th International Conference on Innovations in Science, Engineering and Technology (ICISET),* Chittagong, Bangladesh, 26th - 27th October 2024, pp. 1-6, DOI:10.1109/ICISET62123.2024.10941498 [Scopus Index].

2. **Sharith Dhar** and **Md. Saiful Islam**, "Convolution Neural Network (CNN) based Semiconductor Manufacturing Defect Detection System by using Real Wafer Maps," 5th *International Conference on Physics for Sustainable Development & Technology (ICPSDT),* Chittagong, Bangladesh, 7th - 8th September 2023.

3. **Sharith Dhar** and **Md. Saiful Islam**, "Fuzzy Logic-based Soft Starter for Controlling Starting Parameters of Induction Motor," *International Conference on Electrical, Computer and Communication Engineering (ECCE),* Chittagong, Bangladesh, 23rd - 25th February 2023, pp.1-6. DOI:10.1109/ECCE57851.2023.10101647 [Scopus Index].

Approval by the Supervisor

This is to certify that **Sharith Dhar** has carried out this research work under my supervision, and that she has fulfilled the relevant Academic Ordinance of the Chittagong University of Engineering and Technology, so that she is qualified to submit the following Thesis in the application for the degree of MASTER of SCIENCE in Electronics and Telecommunication Engineering (ETE). Furthermore, the Thesis complies with the PLAGIARISM and ACADEMIC INTEGRITY regulations of CUET.

Dr. Md. Saiful Islam

Associate Professor

Department of Electronics and Telecommunication Engineering

Chittagong University of Engineering & Technology

Acknowledgement

First and foremost, I'd like to thank my GOD for making things easier than I could ever have imagined. I would like to keep in mind a few very essential people without whom I would be unable to accomplish this thesis. I would like to express my gratitude to my distinguished supervisor, Dr. Md. Saiful Islam, Associate Professor, Department of ETE, CUET, for his contributions, inspiration, and motivation. Throughout my journey, he has provided regular instruction, assistance, and motivation in overcoming various technical challenges. I'd like to thank Prof. Dr. Md. Azad Hossain, Head, ETE, CUET, and my colleagues for their ongoing encouragement and drive to accomplish my work. I'm grateful to my family for inspiring me to pursue higher education in the first place. Their unwavering support and encouragement were the first things that prepared the path for this journey. Last but not least, I would like to acknowledge everyone at the Department of Electronics and Telecommunication Engineering (ETE), CUET, for their technical and administrative assistance with my thesis work.

.

Abstract

Identifying defective semiconductor wafers is a crucial and complicated aspect of the manufacturing process. Developed systems confront difficulties in capturing intricate defect patterns and the long-range relationship of defects. In addressing these issues, this research proposes a hybrid system that comprises a modified Mobilenet structure and the error-correcting output code (ECOC) technique-based support vector machine (SVM) classifier. In the proposed method, for feature extraction purposes, modified MobileNet architecture is designed 1) by using the Swish function instead of the ReLU function in the depth-wise separable convolution block of MobileNet structure for detecting complex defective patterns and 2) by incorporating the multi-head attention mechanism in MobileNet architecture for capturing long-range dependencies on defective wafers. The ECOC-SVM approach is used to classify wafer defects. Also, the histogram equalization technique is employed to improve the visibility of small defects. The real-world semiconductor wafer dataset WM-811K is used in this study. The proposed system with histogram equalization achieved a superior testing accuracy of 98.55% and better average values of AUC (99.74%), recall (93.34%), precision (95.64%) and F1-score (94.42%). On the other hand, the original version of the MobileNet model achieved testing accuracy of 95.31%. The proposed model correctly identified complex patterns in scratch, edge-ring, and donut types of defect wafers compared with other models by using the swish function in the model. The comparative analyses of the proposed system with other developed systems and state-of-the-art systems are given in this research.

বিমূর্ত

ত্রুটিপূর্ণ সেমিকন্ডাক্টর ওয়েফার সনাক্তকরণ উৎপাদন প্রক্রিয়ার একটি গুরুত্বপূর্ণ এবং জটিল দিক। উন্নত সিস্টেমগুলি জটিল ত্রুটির ধরণ এবং ত্রুটিগুলির দীর্ঘ-পরিসরের সম্পর্ক ক্যাপচার করতে অসুবিধার সম্মুখীন হয়।এই সমস্যাগুলি সমাধান করার জন্য, এই গবেষণায় একটি হাইব্রিড সিস্টেমের প্রস্তাব করে যার মধ্যে একটি পরিবর্তিত মোবাইলনেট কাঠামো এবং ত্রুটি-সংশোধনকারী আউটপুট কোড (ECOC) কৌশল-ভিত্তিক সহায়তা ভেক্টর মেশিন (SVM) শ্রেণিবদ্ধকারী রয়েছে। প্রস্তাবিত পদ্ধতিতে, বৈশিষ্ট্য নিষ্কাশনের উদ্দেশ্যে,পরিবর্তিত মোবাইলনেট আর্কিটেকচার ডিজাইন করা হয়েছে ১) জটিল ত্রুটিপূর্ণপ্যাটার্ন সনাক্ত করার জন্য মোবাইলনেট কাঠামোর গভীরতা-ভিত্তিক বিভাজক কনভোলিউশন ব্লকে ReLU ফাংশনের পরিবর্তে Swish ফাংশন ব্যবহার করে এবং২)ত্রুটিপূর্ণ ওয়েফারের উপর দীর্ঘ-পরিসরের নির্ভরতা ক্যাপচার করার জন্য মোবাইলনেট আর্কিটেকচারে মাল্টি-হেড মনোযোগ প্রক্রিয়া অন্তর্ভুক্ত করে।ওয়েফার ত্রুটিগুলিকে শ্রেণীবদ্ধ করার জন্য ECOC-SVM পদ্ধতি ব্যবহার করা হয়। এছাড়াও, ছোট ত্রুটিগুলির দৃশ্যমানতা উন্নত করার জন্য হিস্টোগ্রাম সমীকরণ কৌশল ব্যবহার করা হয়।এই গবেষণায় বাস্তব-বিশ্বের সেমিকন্ডাক্টর ওয়েফার ডেটাসেট WM-811K ব্যবহার করা হয়েছে। হিস্টোগ্রাম ইকুয়ালাইজেশন সহ প্রস্তাবিত সিস্টেমটি ৯৮.৫৫% এর উচ্চতর পরীক্ষার নির্ভুলতা এবং AUC (৯৯.৭৪%), recall (৯৩.৩৪%), precision (৯৫.৬৪%) এবং F1-score (৯৪.৪২%) এর উন্নত গড় মান অর্জন করেছে। অন্যদিকে, মোবাইলনেট মডেলের মূল সংস্করণটি ৯৫.৩১% পরীক্ষার নির্ভুলতা অর্জন করেছে। প্রস্তাবিত মডেলটি মডেলটিতে Swish ফাংশন ব্যবহার করে অন্যান্য মডেলের তুলনায় স্ক্র্যাচ, এজ-রিং এবং ডোনাট ধরণের ডিফেক্ট ওয়েফারে জটিল প্যাটার্নগুলি সঠিকভাবে চিহ্নিত করেছে। অন্যান্য উন্নত সিস্টেম এবং অত্যাধুনিক সিস্টেমের সাথে প্রস্তাবিত সিস্টেমের তুলনামূলক বিশ্লেষণ এই গবেষণায় দেওয়া হয়েছে।

Table of Contents

List of Figures

List of Tables

Nomenclature

DL	Deep Learning
ML	Machine Learning
DSC	Depth-wise Separable Convolution
DWC	Depth-wise Convolution
PWC	Pointwise Convolution
k-NN	K-Nearest Neighbors
ReLU	Rectified Linear Unit
ECOC	Error Correcting Output Code
FC	Fully Connected
GAP	Global Average Pooling
MP	Max-Pooling
FL	Flatten Layer

Chapter 1: INTRODUCTION

This chapter explains the overview of the current state of the machine learning models, deep learning models and hybrid models in defect wafer classification. Also, this chapter outlines the background in section 1.1 and the problem statements in section 1.2. The aims and objectives of this research are explained in section 1.3. After that, section 1.4 explains the scope of the study. Finally, section 1.5 describes the thesis outlines.

1.1 BACKGROUND

The semiconductor industry plays a vital role in driving advancements in consumer electronics and computing systems, which challenge the boundaries of human creativity and innovation, powering the present and future technologies. Classifying defects in semiconductor wafer manufacturing is crucial for early identification of potential flaws, facilitating prompt corrective measures and reducing expensive waste. The process used to rely significantly on manual inspection or basic automated systems. Nonetheless, the emergence of artificial intelligence (AI), machine learning (ML), and deep learning techniques has significantly enhanced the capacity to identify and categorize wafer defects. Implementing these technologies can transform the defect detection process by improving speed, accuracy, and scalability, all while minimizing dependence on human expertise. A wide range of research works has been conducted to identify defect patterns in wafers.

1.1.1 Conventional Methods

Before the emergence of contemporary machine learning and deep learning methodologies, the semiconductor manufacturing sector predominantly depended on conventional, computer-assisted, and rule-based approaches for the identification of wafer defects. The conventional approaches encompassed optical and electron beam inspection, statistical process control, and manual

visual inspection, all enhanced by computer-aided tools for data analysis and image comparison. Optical inspection emerged as one of the initial and most extensively utilized methods for detecting defects in wafers. The systems employed high-resolution cameras in both bright-field and dark-field lighting to emphasize differences in reflectance and surface topography. Computer-aided tools facilitated comparisons between die-to-die and die-to-database, allowing for the evaluation of images from test dies against reference dies or golden layout CAD data. Defects were detected through the analysis of variations in pixel intensity, shape, and geometry. Ebayyeh et al. provide [1] a detailed resource for engineers and researchers engaged in inspection automation within the electronics and semiconductor sectors. Also, provides a standard for AOI system development, emphasizing recognized methodologies and potential areas for innovation. Electron beam inspection tools were utilized in scenarios requiring higher resolution, especially for nanoscale defects that exceeded the capabilities of optical systems. The wafer was scanned by e-beam systems utilizing a focused electron beam, which detected secondary electrons emitted from the surface. Computer-aided tools in e-beam inspection offer systematic automation for alignment, precise sizing of defects, and comprehensive classification of those defects. Electron beam (e-beam) inspection tools play a crucial role in identifying nanoscale defects in semiconductor wafers, especially in scenarios where optical systems do not provide adequate resolution. The systems utilize a focused electron beam to scan wafers, detecting emitted secondary electrons to pinpoint both surface and subsurface anomalies. Computer-aided tools improve e-beam inspection through the automation of alignment processes, precise defect sizing, and systematic classification capabilities. Cass et al. [2] establish the use of the SEMSpec e-beam inspection system for detecting defects during the manufacturing process, emphasizing its features in automated alignment and the classification of defects. Statistical Process Control **(**SPC) served as a critical tool for monitoring process stability

and identifying deviations in wafer fabrication. The process utilized statistical techniques including control charts, process capability indices, and Pareto analysis, all enhanced by yield management software. The computer-aided tools gathered real-time sensor data throughout various stages of the fabrication process to examine trends, identify anomalies, and enhance yield. Statistical Process Control (SPC) serves as a fundamental method for assessing process stability and identifying deviations in the fabrication of semiconductor wafers. Utilizing statistical techniques like control charts, process capability indices, and Pareto analysis, SPC allows manufacturers to pinpoint and tackle variations that may affect product quality. The methods are frequently enhanced by yield management software that gathers real-time sensor data throughout various stages of the fabrication process to examine trends, identify anomalies, and improve yield. The implementation of SPC facilitates the distinction between normal (systematic) variations and special variations by utilizing control limits and charts at each node of the semiconductor manufacturing process. This approach enhances the quality and reliability of the manufacturing cycle. Yield management systems (YMS) enhance SPC by offering tools that facilitate the monitoring, analysis, and optimization of production yields. These systems assist manufacturers in pinpointing and resolving process inefficiencies, monitoring performance, and facilitating decisions based on data to improve production results. Integrating SPC with yield management systems enables semiconductor manufacturers to detect process deviations early, uphold high-quality standards, and enhance overall production efficiency. Trained technicians conducted manual inspections by utilizing microscopes to perform a visual examination of the wafers. This approach was primarily employed in scenarios involving low-volume production or as a verification phase subsequent to automated inspection. This approach, while demonstrating high reliability, was characterized by its labor-intensive nature, lack of scalability, and susceptibility to human error. Historically, manual inspection has been crucial in

detecting defects on semiconductor wafers, especially in low-volume production environments or as a verification process after automated inspections. The procedure entails skilled technicians employing microscopes to conduct a visual inspection of wafers for any defects present. Manual inspection provides a level of reliability attributed to human judgment; however, it is characterized by labor intensity, limited scalability, and vulnerability to human error. Nonetheless, manual inspection methods encounter a variety of challenges. Their inefficiency and unreliability frequently stem from inconsistencies in inspection procedures and the limitations inherent in human performance. Elements like fatigue, personal judgment, and differences in skill levels can contribute to varying outcomes. The increasing complexity and miniaturization of semiconductor devices render manual inspection increasingly impractical and inefficient. Moreover, with the increasing complexity and miniaturization of semiconductor devices, the shortcomings of manual inspection have become increasingly evident. Traditional microscopy techniques and the human eye frequently fall short in identifying nanoscale defects, prompting a transition to automated inspection systems that provide enhanced precision and scalability. In conclusion, although manual inspection has served as a fundamental approach in wafer defect detection, its inefficiencies, scalability challenges, and accuracy issues have prompted the semiconductor industry to progressively embrace automated inspection technologies to satisfy the requirements of contemporary manufacturing processes. Computer-Aided Tools (CATs) have significantly enhanced traditional defect detection systems in semiconductor manufacturing by automating and refining various inspection processes. In Optical Inspection, CATs enable the acquisition of images, enhance contrast, and apply morphological filtering, thereby improving the identification of surface anomalies. In electron beam (e-beam) systems, their management of scanning paths and data interpretation facilitates the accurate identification of subsurface defects. In statistical process control (SPC) and yield analysis, CATs analyze

extensive in-line sensor data, presenting statistical control charts and defect density maps to assess process stability and enhance yield optimization. In a die-to-database comparison, the tools facilitate the alignment of scanned images with CAD models and execute automated pixel-wise subtraction to identify discrepancies.

1.1.2 Machine Learning (ML) based Methods

Traditional machine learning (ML) techniques find patterns, predict, and understand structured data. Traditional machine learning requires hand-feature engineering and lower computational expenses. Linear regression, logistic regression, SVMs, decision trees, KNN, and k-means clustering are important methods. SVM is the most powerful and popular conventional machine learning approach used for classification purposes. This technique finds the best hyperplane to separate data points from different classes. The purpose is to optimize the margin, defined as the distance separating the hyperplane from the nearest data points of each class. Nearest points are called support vectors because they define the hyperplane [3]. Ma et al. [4] analyze several aspects of SVM applications, including its implementation for addressing imbalanced data set issues. Cervantes et al. [5] provide a summary of the challenges of SVM implementation in engineering fields and the real-world problem-solving capability of SVM. The k-Nearest Neighbors (kNN) is a conventional machine learning algorithm frequently employed for classification and regression tasks. kNN is an instance-based learning algorithm that does not assume any underlying distribution. It classifies data points by examining the classes of their closest neighbors. The parameter k denotes the quantity of nearest data points (neighbors) considered during the classification decision-making process [6]. S. Zhang [7] investigates the challenges of using KNN algorithms with unbalanced datasets and suggests classification rules for solving imbalanced data issues. Sehih et al. [8] developed a machine learning-based system that identifies white

pixel defects in wafers. Mohiuddin et al. [9] suggested a wavelet entropy-based optimization approach for bearing fault classification, employing four machine learning algorithms, including SVM and KNN, for classification purposes. Hoque et al. [10] developed a crop yield prediction system that used three machine-learning techniques.

1.1.3 Deep Learning (DL) based Methods

Deep learning, especially Convolutional Neural Networks (CNNs), has transformed defect detection by automating feature extraction and learning directly from unprocessed data, including images or sensor inputs. Convolutional Neural Networks (CNNs) are proficient at learning spatial hierarchies, rendering them very effective for image-based defect detection in semiconductor production. The deep CNN based model for an autonomous wafer map defect identification system developed by Zheng et al. [11] performs the best compared to other models. In this model, the random sampling method is applied. A small-scale defect identification system has been suggested by Byun et al. [12] using the rotated defects (RoD) transform technique and a modified version of median filtering. Misra et al. [13] used soft voting to improve performance after acquiring the deep ensemble characteristics for defective wafer identification. The attention mechanism and cosine normalization-based wafer map analysis method developed by Xu et al. [14] effectively addresses the problem of dataset imbalance. E. Sin and C. D. Yoo [15] present a wafer defect classification method that requires no high-performance hardware. Chen et al. [16] designed a dual-source DCNN architecture to detect defect patterns of wafers. A contactless inspection system constructed by Jeon et al. [17] used thermal images and CNN to detect printed circuit board assembly (PCBA) defects. The multi-scale residual dilated convolution attention mechanism network super-resolution reconstruction algorithm used by X. Sun et al. [18] for defect wafer detection. An automatic visual defect identification

system developed by Limam et al. [19] uses different types of CNN architectures. M. Shahroz et al. [20] proposed a hierarchical attention module-based CNN model to detect hotspots in semiconductor wafer manufacturing. Deng et al. [21] developed a light-weight neural network-based system for the identification of mixed-type wafer defects. Hou et al. [22] developed a mixed defect type detection and classification system that used improved version of DCNN architecture which defined as multi-path DCNN. Ma et al. [23] developed a method that identifies tiny defects by using spatial attention block and residual architecture. López et al. [24] proposed a semiconductor wafer defect detection method that merged computer vision technique with lightweight Squeeze Net CNN. A mixed defect wafer classification method proposed by J. Shim and S. Kang [25] uses the single defect wafers training dataset. The mix-up, random rotation, and noise filtering are used in this suggested technique to create synthetic wafer maps from the training dataset. The parallel CNN-based brain tumor detection & classification system is developed by Rahman et al. [26]. Also, Rahman et al. [27] proposed a dilated parallel deep convolutional neural network model for brain tumor classification, which extracts both coarse and fine characteristics from MRI images. Hossain et al. [28] developed the depth-wise separable convolution-based modified and reduced Mobile Net model that recognized plant leaf diseases.

1.1.4 Hybrid Methods

Hybrid methods integrate traditional machine learning and deep learning techniques, seeking to capitalize on the advantages inherent in both approaches. Hybrid systems frequently combine traditional machine learning algorithms with deep learning models to enhance performance. Chen et al. [29] developed a hybrid system for the defective wafer map patterns recognition that used the improved version of the convolution block attention module with DCNN architecture for feature extraction and error-correcting output code-based SVM

for classification purposes. Schlosser et al. [30] proposed a hybrid semiconductor wafer fault inspection system that integrates classical computer vision techniques for fault localization and deep neural networks (DNNs) for classification. This system focuses on detecting very tiny defects in wafers. Biswas et al. [31] proposed a hybrid model for the classification of brain tumors. The CNN model is utilized as the feature extractor and the support vector machine (SVM) classifier is used for classification purposes. For satellite image classification tumpa et al. [32] proposed an SVM-based lightweight parallel CNN model that used the EuroSAT dataset and the RSI-CB256 dataset.

The semiconductor defective wafer detection-related research works that are mentioned above emphasize general patterns of defect wafers and use several processing techniques for improving defective wafers classification accuracy. However, developed methods encounter difficulties in detecting tiny defects in wafers. Also, existing methodologies rely on fundamental feature extraction structures that cannot address complex and irregular patterns present in defective wafers. Another notable limitation of developed research methods is the inability to identify spatial dependencies among defects that are located far apart on the wafer map. The lack of long-range relationships in the models results in difficulties in capturing the context and correlations among various regions on defective wafers, which leads to erroneous classifications.

This research is motivated by the vital need to overcome these challenges to the development of a more effective and efficient hybrid system for detecting and classifying wafer defects. The reason behind this work is consistent with the industry's goal of attaining high precision and dependability in defect analysis while decreasing computing overhead.

1.2 PROBLEM STATEMENT

In the semiconductor manufacturing sector, imperfections in wafer maps frequently manifest as subtle differences in intensity. Developed methods

emphasize general patterns and overlook intricate specifics, resulting in challenges in identifying small defects. The absence of high-resolution feature extraction and insufficient focus on minor variations in wafer maps plays a significant role in this problem. Also, existing methodologies rely on fundamental feature extraction structures that cannot address complex and irregular patterns present in defective wafers. Another notable limitation of developed research methods is the inability to identify spatial dependencies among defects located far apart on the wafer map. The lack of long-range relationships in the models results in difficulties in capturing the context and correlations among various regions on defective wafers, which leads to erroneous classifications. Several state-of-the-art models such as MobileNet V1, MobileNet V2, ResNet-18, and ResNet-50 are used for detecting defect wafers, but these models have a large number of parameters and high computational complexity.

1.3 MOTIVATION

Traditional methods of wafer defect detection struggle to detect small defects. Also, human errors are involved in the manual inspection process. In the modern world, engineers need to know the types of defects as well as the density of defects with less time consumption in the semiconductor fabrication industry. The conventional detection techniques require a significant amount of time, which degrades the overall process's performance. The advanced wafer detection methods related to deep learning and machine learning algorithms encounter challenges in identifying intricate patterns in defective wafers. Also, these methods struggle in detecting long-distance relationships of defects in wafers. The tiny defects in wafers are not identified correctly by these advanced methods. This research is motivated by the vital need to overcome these challenges to the development of an effective hybrid model for identifying different types of wafer defects.

1.4 AIMS AND OBJECTIVES

The main objective of this research is to develop a hybrid system that can efficiently detect different types of defects in semiconductor wafers. The hybrid system comprised a modified MobileNet with the multi-head attention mechanism and error-correcting output code (ECOC) -based SVM classifier.

The objectives of this research are specified below:

i. To modify the depth-wise separable convolution (DSC) block of the original version of MobileNet by using the swish function instead of using the ReLU activation function for capturing features of complex patterns and employing multi-head attention mechanism in modified Mobile Net architecture for capturing long-range dependencies in defects wafer maps.

ii. To design the hybrid system using the ECOC-based SVM classifier with modified Mobile Net architecture. This classifier solves the class imbalance problem besides the classification task.

iii. To apply the histogram equalization technique on the wafer map dataset to enhance the visibility of tiny defects in wafers which facilitates the proposed system's ability to identify small defects.

iv. To use the real-world data set WM-811K for experimental validation of the proposed hybrid system.

v. To establish a concrete trade-off among accuracy, parameters, and computational latency for identifying defective wafers for several state-of-the-art models alongside the developed hybrid model.

1.5 SCOPE OF THE STUDY

Current methodologies encounter challenges in detecting micro-level defects in wafers, which negatively affect the accuracy of the model's

classification. A further limitation is that existing methods unable to extract features from intricate defective patterns in wafers. Furthermore, these established models encounter challenges in accurately representing the spatial relationships among defects which distributed across the wafer map. This research focuses on defective wafer detection, providing a novel approach to the challenges faced by current models. The modified version of MobileNet and the inclusion of multi-head attention of the hybrid system of this research address the limitations of existing systems. The scope of this study extends to an assessment of multiple classifiers and different optimizers to determine the most effective one for improving the overall performance of wafer defect detection. This research aims to offer novel insights applicable to an effective solution for semiconductor manufacturing. The potential influence beyond the resolution of issues in defective wafer identification facilitates enhanced advancements in related fields.

1.6 THESIS OUTLINE

The subjects of this thesis have been split into five chapters. This is a concise summary of the chapters:

- Chapter 1, the introductory chapter, encompasses the research motivation, problem description, objectives, and organization of the thesis. The chapter outlines the motivation for the research, the study's objectives, and the organization of the thesis paper.
- Chapter 2 presents the theoretical framework aligned with the literature review relevant to the research. This will offer a concise summary of the theories, formulas, and literature necessary for the audience to cultivate interest in and understand the topic.
- Chapter 3 meticulously establishes the methodological framework employed in the investigation. A detailed overview of the methodological framework is provided to facilitate a clear

understanding of the study's implementation. The chapter contains details on design, including design criteria and figures.

- Chapter 4 encompasses simulation results, performance evaluation, and a comparison analysis. This chapter contains data and output figures.
- Chapter 5 concludes with a summary of the research methodology, results, and findings. Challenges and limitations are also addressed. Finally, this thesis concludes with future recommendations.

Chapter 2: LITERATURE REVIEW

This chapter gives the reader a thorough understanding by offering a broad theoretical framework supported by a significant collection of literature. Section 2.1 discusses the details of semiconductor wafers and several types of defect classes. Section 2.2 then investigates the operational concepts of convolutional neural network, including depth-wise separable convolutional. Section 2.3 describes the pre-trained models including different versions of MobileNet, and ResNet. Finally, Section 2.4 describes various types of classifiers, including the SoftMax, Support Vector Machine (SVM), and K-Nearest Neighbour (K-NN).

2.1 INTRODUCTION TO SEMICONDUCTOR WAFERS

Semiconductor wafers are essential components of today's electronics, acting as substrates for integrated circuits (ICs) and microelectronic devices. These wafers consist of thin slices of semiconductor material, predominantly silicon, and are produced through a series of meticulously controlled processes to attain electrical, physical, and mechanical characteristics [33]. They are crucial in various applications, spanning computing, telecommunications, healthcare, and automotive systems [34].

There are several steps involved in the fabrication of semiconductor wafers:

a. **Crystal Growth:** The initial phase involves the formation of a single-crystal ingot through techniques like the Czochralski (CZ) or the Float-Zone (FZ) process. Silicon of high purity undergoes melting and solidification in meticulously regulated environments to yield monocrystalline formations with few defects.

b. **Wafer Slicing:** The cylindrical ingot is sectioned into thin discs, referred to as wafers, with diamond wire saws. The thickness and

diameter of the wafers fluctuate based on the application and the dimensions of the ingot.

c. **Lapping and Polishing:** The rough wafers undergo lapping and chemical-mechanical polishing (CMP) to attain a smooth, flat surface devoid of saw marks or irregularities. The refined surface is essential for the following lithographic procedures [35].

d. **Doping and Annealing:** Introducing impurities, like phosphorus or boron, modifies the electrical characteristics of silicon. Subsequently, annealing processes are implemented to rectify any damage incurred during doping and to guarantee a consistent distribution of dopants [36].

e. **Surface Treatment:** Wafers are subjected to cleaning and oxidation processes to ensure they are adequately prepared for device fabrication. This phase guarantees that the surface is devoid of impurities and possesses the required electrical properties.

These procedures are necessary to give the wafers the necessary electrical and physical properties [37]. Defects can occur at any level, compromising wafer quality and device performance even with strict production controls.

2.1.1 Types of Wafer Defects

There are several types of defects that remain in wafers that are described below:

a. **Center Defect:** Center faults are found in the middle of the wafer and are frequently brought on by uneven crystal formation, uneven polishing, or temperature gradients. These flaws show themselves as discolorations, voids, or abnormalities on the surface. Localized electrical inconsistencies result from these

abnormalities, which disturb the regularity of thin-film deposition [37].

Fig. 2.1: Center defect class wafer

b. **Random Defects:** Random defects, which dispersed unevenly throughout the wafer, are frequently caused by environmental influences, equipment failures, or particle contamination. Microcracks, embedded particles, and pits are a few examples. In ICs, these defects may result in signal interference, short circuits, or open circuits . Random defects are often found and categorized using sophisticated inspection methods including scanning electron microscopy (SEM) and laser-based devices.

Fig. 2.2: Random defect class wafer

c. **Edge Local Defect:** This type of defect, limited to the wafer's edge, is usually caused by mishandling, incorrect edge trimming, or mechanical stress. These defects, which include delamination, chips, and fractures, might weaken the wafer's structure and raise the possibility that it will break during processing.

Fig. 2.3: Edge Local defect class wafer

d. **Donut Defects:** Donut defects, frequently brought on by non-uniform deposition rates or an uneven temperature distribution during thermal processing, manifest as ring-like patterns on the

wafer. The performance of devices made on the wafer may be impacted by these flaws, which result in differences in layer thickness and electrical characteristics.

Fig. 2.4: Donut defect class wafer

e. **Edge Ring Defect:** Edge ring defects are usually the result of imperfections in the etching or chemical vapor deposition (CVD) processes, generating concentric patterns close to the wafer edge. These flaws lower the total wafer yield by compromising the integrity of IC components in edge areas.

Fig. 2.5: Edge-ring defect class wafer

f. **Scratch Defect:** Scratches are linear indentations or grooves resulting from mechanical abrasion during handling, cleaning, or equipment maintenance. They can impair thin coatings, obstruct electrical channels, and result in device malfunctions

Fig. 2.6: Scratch defect class wafer

g. **Near Full Defect**: Near-full defects encompass a substantial area of the wafer surface, typically arising from catastrophic failures, including large equipment breakdowns or severe contamination incidents. These defects significantly diminish wafer yield and are often regarded as irreversible. Preventive maintenance of

equipment and cleanroom environmental controls is crucial to prevent such incidents [37].

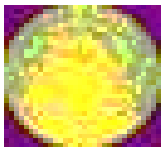


Fig. 2.7: Near Full defect class wafer

h. **Local Defect:** Local defects are little, discrete irregularities such as pits, voids, or micro-cracks. These may originate from material defects or localized mechanical stress. Despite their seemingly trivial nature, local defects can significantly affect high-density integrated circuits where accuracy is essential.

Fig. 2.8: Local defect class wafer

2.1.2 Importance of Semiconductor Wafer

A wafer map, which graphically depicts the distribution of flaws, yield fluctuations, and process anomalies throughout a wafer, is a crucial tool in the semiconductor manufacturing industry. It assists engineers in examining defect patterns that point to underlying problems in manufacturing processes including lithography, etching, and deposition, such as center flaws, edge rings, scratches, and clusters [38]. Recognizing these patterns allows manufacturers to categorize defects into systematic (those that recur due to process issues) and random (those that are sporadic and unpredictable) classifications, thereby enhancing process control and facilitating root cause analysis [39]. Wafer maps play a crucial role in semiconductor manufacturing by facilitating yield optimization. They allow engineers to pinpoint areas with low yield and track persistent defects, which in turn improves manufacturing efficiency through methods such as statistical process control (SPC) and predictive maintenance. Advancements in machine learning (ML) and artificial intelligence (AI) have led

to the increased utilization of wafer maps for automating defect classification and yield prediction. This integration enhances accuracy while significantly reducing the time required for manual inspection [40].

By integrating wafer maps with Industry 4.0 technologies, such as big data analytics and real-time monitoring, semiconductor manufacturers can enhance quality control, reduce production losses, and ensure high-performance semiconductor devices.

2.2 CONVOLUTION NEURAL NETWORK (CNN)

Convolutional Neural Networks (CNNs) represent a category of deep learning models meticulously crafted for the analysis of data structured in a grid-like format, including images. Their success in tasks such as image classification, object detection, and semantic segmentation is noteworthy, attributed to their capacity to automatically and adaptively learn spatial hierarchies of features from raw input data. The fundamental architecture of CNN is depicted in Fig 2.9. The standard architecture of a CNN includes several essential layers:

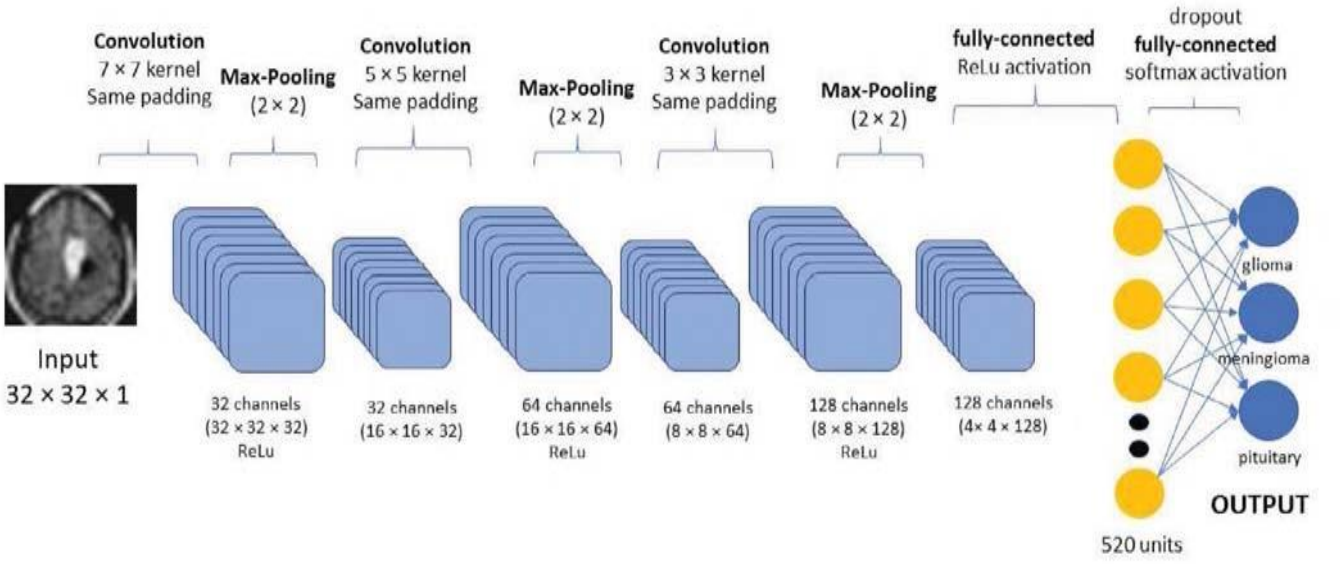


Fig. 2.9: Basic architecture of CNN [41]

a) **Convolution Layer:** The convolutional layer serves as the fundamental building block of CNN. The process involves applying convolution operations to the input, utilizing a collection of learnable filters (or kernels) to identify spatial patterns such as edges, textures, or components of objects. The mathematical representation of the operation used in the convolution layer is as follows [42]:

$$Y(i, j) = \sum_{m=1}^{M} \sum_{n=1}^{N} X(i+m, j+n).W(m,n) + b \qquad (2.1)$$

The output value of the feature map at the location (i,j). This is referred to as the 'convolved feature'. The double summation systematically traverses the dimensions of the filter, denoted as (M, N) or kernel. The value at the input is located at positions (i+m, j+m). This denotes the specific area of the input that is presently under analysis. The value is located at the coordinates (m,n) within the filter. Every filter comprises adjustable weights that are refined throughout the training process.

b) **Max-Pooling Layer:** A max-pooling layer is a prevalent function in CNN, facilitating the downsampling of feature maps' spatial dimensions. The process involves identifying the highest value within a designated window (kernel) and systematically shifting this window across the feature map at a predetermined stride. The main objectives of max pooling are to decrease spatial dimensions, boost computational efficiency, and improve feature invariance. The output feature map Y resulting from max pooling is determined for a feature map X with dimensions H x W and a pooling window of size M x N as follows [43]:

$$Y(i, j) = \max \{X(i+m, j+n) \mid 0 \leq m < M, 0 \leq n < N\} \qquad (2.2)$$

The pooled value at the position (i,j) is represented as Y (i,j). X(i+m, j+n) denotes the values within the current pooling

window. The stride defines the distance the pooling window moves over the input feature map.

c) **Global Average Pooling (GAP):** Global Average Pooling (GAP) serves as a pooling operation frequently utilized in Convolutional Neural Networks (CNNs) to diminish the spatial dimensions of a feature map, all the while preserving its channel-wise information. This approach offers a straightforward and efficient method for linking the feature extraction layers to the output layer in a convolutional neural network. Let F ϵ $R^{H\times W\times C}$ be the input feature map, where H represents the height, W denotes the width, and C indicates the number of channels. GAP(F) generates a vector y ϵ R^{C}, where each element corresponds to a channel, for the c-th channel, the GAP output is [44]:

$$y_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{i,j,c} \quad (2.3)$$

Where $F_{i,j,c}$ represents the value of the feature map at the locations (i,j) within channel c.

d) **Fully Connected Layer:** The fully connected layer (FCL) in a CNN is generally located near the conclusion of the architecture. In contrast to the convolutional and pooling layers that emphasize spatial feature extraction, the fully connected layer synthesizes all the features derived from the input to generate predictions. This layer establishes connections between each neuron in the preceding layer and every neuron in the current layer, creating a dense network.

e) **Activation Functions:** In a Convolutional Neural Network (CNN), an activation function is a non-linear mathematical function that is applied to a neuron's (or layer's) output to add non-linearity to the network. It is essential to the network's ability to represent intricate relationships in the data. Activation

functions of various types, including the Swish activation function and Rectifier Linear Units (ReLUs), are employed.

Fig.2.10 presents a comparative analysis of the ReLU and Swish activation functions. The advantages of the swish function can be attributed to its lower bound and unbounded upper limit, as well as its non-monotonic nature.

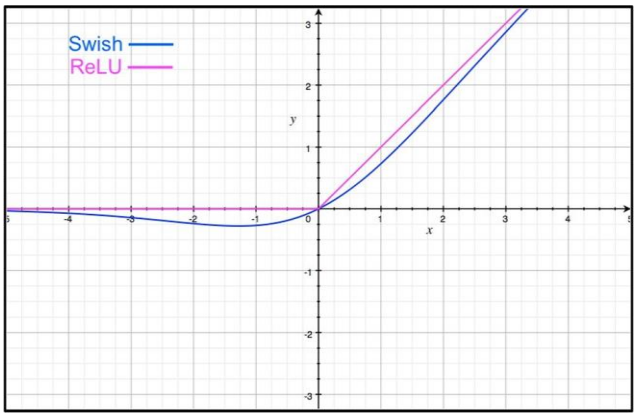


Fig. 2.10: Graphical representations of ReLU and Swish [45]

These characteristics enable it to surpass ReLU in deep networks and prevent the occurrence of inactive neurons within the neural network.

2.2.1 Depth-wise Separable Convolution:

Depthwise separable convolution represents a computationally efficient alternative to standard convolution, commonly employed in lightweight neural network architectures such as MobileNet. The convolution operation is divided into two distinct stages: depthwise convolution and pointwise convolution [46] In contrast to traditional convolution, which applies filters to all input channels, depth-wise convolution applies a single convolutional filter to each separate input channel. This step lowers computing complexity because cross-channel interactions are not processed directly. Subsequently, the pointwise convolution, which is a 1x1 convolution, integrates the outputs from the

depthwise convolution across channels, efficiently merging channel information and producing the intended output features [47]. Fig. 2.11 illustrates traditional convolution, which makes a single output feature map by applying a filter (or kernel) concurrently to each input channel. Every filter captures correlations between channels and spaces. Because this approach has so many variables and procedures, it is computationally costly.

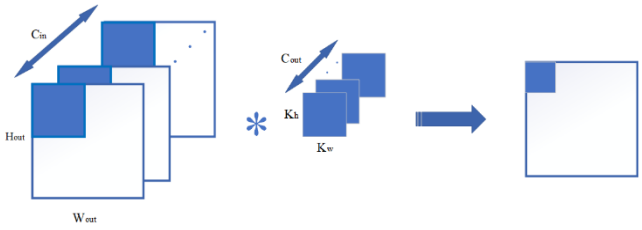


Fig. 2.11: Traditional convolution operation

The computational cost of traditional convolution is given below:

$$\text{Computation}_{\text{TD}} = (H_{out} \times W_{out} \times C_{in} \times K_h \times K_w \times C_{out}) \quad (2.4)$$

Where, H_{out} is the height of the output feature map W_{out} is the weight of the output feature map, C_{in} is the number of input channels, C_{out} is the number of output channels (or filters), K_h is the height of the filter kernel, K_w is the width of the filter kernel. The depth-wise convolution operation is depicted in Fig.2.12.

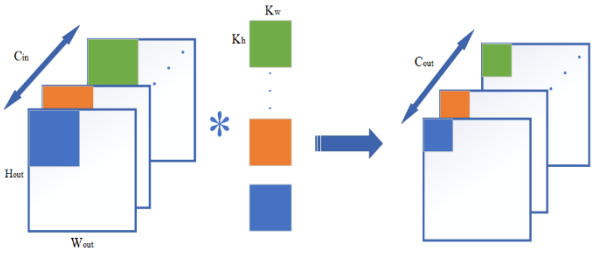


Fig. 2.12: Depth-wise convolution operation

The computational cost of depth-wise convolution is given below:

$$\text{Computation}_{\text{DW}} = (H_{out} \times W_{out} \times C_{in} \times K_h \times K_w) \quad (2.5)$$

A point-wise convolution follows the depth-wise convolution, depicted in Fig.2.13. In point-wise convolution, 1x1 convolution is applied to combine the information across channels, allowing interaction between channels.

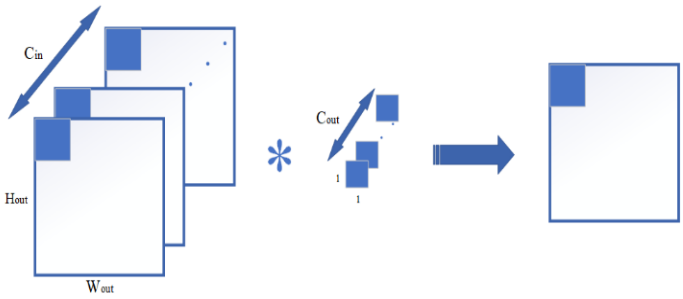


Fig. 2.13: Pointwise convolution operation

The computational cost of point-wise convolution is given below:

$$\text{Computation}_{PW} = (H_{out} \times W_{out} \times C_{in} \times 1 \times 1 \times C_{out}) \quad (2.6)$$

The total computational cost of depth-wise separable convolution

$$\text{Total} = (H_{out} \times W_{out} \times C_{in} \times K_h \times K_w) + (H_{out} \times W_{out} \times C_{in} \times 1 \times 1 \times C_{out}) \quad (2.7)$$

This depth-wise separable convolution significantly reduces the computational cost, and the number of parameters compared to traditional convolution.

Conventional deep learning architectures, like Convolutional Neural Networks (CNNs), excel in extracting local features; however, they face challenges in effectively capturing long-range dependencies within the input data. In the realm of wafer defect detection, various defect patterns—such as donut, scratch and edge-ring shaped defects—necessitate a comprehension of the spatial relationships among distant areas of the wafer map. CNNs function by employing small convolutional kernels, such as 3×3 or 5×5, thereby constraining their receptive field to localized areas. While deeper CNNs increase the receptive field, the relationships between distant features tend to be indirect and frequently lack strength. Additionally, the implementation of pooling layers, such as max pooling, diminishes spatial resolution and may result in the forfeiture of essential structural information required for the identification of intricate or nuanced patterns. Fully connected layers, despite their dense connections, have limited spatial awareness and are unable to explicitly represent distance-based interactions.

2.3 IMAGE PRE-PROCESSING TECHNIQUES

In machine learning and deep learning applications, image pre-processing is an essential phase as it improves image quality and increases its suitability for analysis. Pre-processing methods assist in eliminating noise, fixing distortions, and enhancing feature extraction, thereby enhancing model performance. There are different image pre-processing techniques used in wafer defect detection systems. In this thesis, we used histogram equalization techniques in the image pre-processing step, which is briefly described in the next chapter. Other widely used techniques are discussed below:

2.3.1 ROTATED DEFECTS (ROD) TRANSFORM

Rotated Defects (RoD) Transform is a sophisticated image preprocessing method aimed at improving the visibility of defects, particularly in scenarios where these defects display rotational variations. This approach establishes a consistent orientation for defects by converting the image into an invariant representation of rotation (RoD Transform). The process of RoD transform is depicted in Fig. 2.14 [12].

- **Rotation Normalization:** The wafer image undergoes rotation to ensure consistent alignment of defects, thereby minimizing variations that arise from differing orientations.
- **Feature Enhancement:** Following the implementation of rotation correction, the process of feature extraction exhibits enhanced robustness, leading to an increase in the accuracy of defect classification.
- **Reduction of Geometric Distortions:** RoD Transform addresses distortions arising from varying defect positions, thereby facilitating the generalization of defect patterns in machine learning models.

- **Integration with CNN Models:** RoD Transform can be incorporated into deep learning pipelines, guaranteeing that convolutional networks are presented with standardized defect patterns, thereby enhancing detection performance.

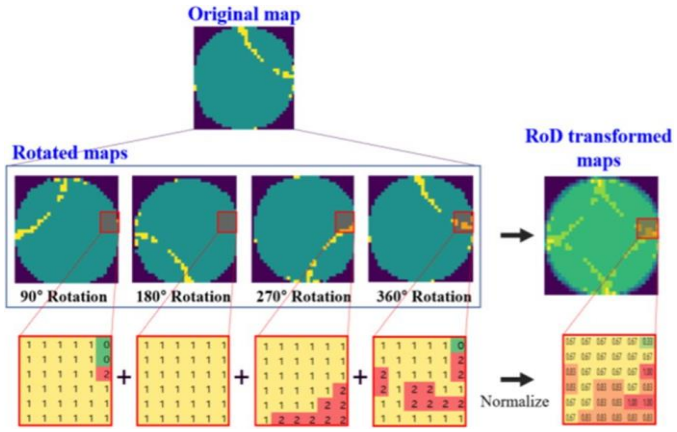


Fig. 2.14: The process of RoD transformation [12]

2.3.2 MEDIAN FILTERING

Median filtering represents a non-linear digital filtering approach frequently utilized in image processing to eliminate noise while maintaining critical image structures. In contrast to mean filtering, which computes the average of pixel values and can result in edge blurring, median filtering substitutes each pixel's intensity with the median value derived from the neighboring pixels within a specified kernel. This approach demonstrates significant efficacy in eliminating impulse noise, including salt-and-pepper noise, which has the potential to obscure defect patterns in wafer images. Median filtering is used in wafer defect detection system [16].

In semiconductor wafer inspection, the attainment of superior image quality is essential for the precise identification and classification of defects. Median filtering plays a crucial role in improving image pre-processing through the following methods:

- **Reducing Impulse Noise:** Wafer inspection images exhibit vulnerability to salt-and-pepper noise, which can arise from factors such as dust, inconsistencies in fabrication, and artifacts during transmission. Median filtering systematically removes these distortions, leading to enhanced clarity in defect visualization.
- **Preserving Edge Details:** A variety of wafer defects, including scratches, edge ring defects, and local defects, demonstrate subtle structural variations. In contrast to conventional smoothing filters, median filtering preserves sharp transitions at defect edges, facilitating the detection of small and intricate defect features.
- **Enhancing Defect Contrast:** Some wafer defects can display minimal contrast when viewed against the wafer background. Median filtering enhances defect visibility by reducing noise and improving local intensity variations, which facilitates more effective segmentation and feature extraction.

2.3.3 C-MEANS FILTERING

C-Means Filtering is an image processing technique that utilizes clustering principles, originating from the C-Means Clustering Algorithm. C-Means Filtering distinguishes itself from conventional filtering techniques like median or Gaussian filtering by adjusting to the local intensity variations within an image. It achieves this by clustering pixels according to similar measures. This technique systematically minimizes noise while preserving the structural details of defects in wafer images. This filter is employed in wafer failure recognition system [14].

During wafer inspection, raw images can be affected by sensor noise, inconsistent illumination, and variations in the background. C-Means Filtering enhances the accuracy of defect detection through the following ways:

- **Noise Reduction:** C-Means Filtering reduces noise by grouping like pixels, consequently maintaining essential image features.
- **Edge and Detail Preservation:** C-Means Filtering preserves crisp transitions in defect structures, unlike conventional smoothing filters that could blur defect edges, therefore helping to enable reliable segmentation.
- **Adaptive Image Enhancement:** Adapting C-Means Filtering to local variations results in enhanced defect contrast relative to the wafer background, thereby improving classification accuracy.

2.4 PRE-TRAINED MODELS

Pre-trained models are fundamental in contemporary machine learning, especially within deep learning applications. They provide considerable benefits by utilizing insights gained from extensive datasets and transferring that knowledge to similar tasks. The training of these models usually involves extensive datasets, including ImageNet or COCO, which allows for the extraction of generalizable features such as edges, textures, and object representations in images [48]. Models such as ResNet, and MobileNet, which have been trained on ImageNet, serve as effective feature extractors for various tasks, including image classification, object detection, and semantic segmentation [49]. The primary advantage of employing pre-trained models is that transfer learning allows for the adaptation of knowledge acquired from a source domain (such as generic object recognition) to be refined for a specific target domain (like defect detection in semiconductor wafers). This method effectively decreases the duration of the training and lessens the requirement for extensively annotated datasets within the target domain [50]. Another notable benefit of pre-trained models is their capacity to reduce overfitting, particularly in situations involving limited datasets. Pre-trained models inherently capture general patterns, resulting in a reduced need for task-specific

data points to attain satisfactory performance [51]. From the mathematical viewpoint, pre-trained models demonstrate superior performance attributed to their initialization process. The network initiates training with weights that have been optimized for general features, rather than commencing from arbitrary weights. The initialization facilitates a more rapid convergence of the optimization algorithm, frequently leading to improved generalization on previously unseen data [52]. Moreover, utilizing large-scale pre-training minimizes the necessity for costly training from the ground up, which proves particularly advantageous in environments with limited resources [53].

2.4.1 MobileNet V1

Howard et al. [46] developed the MobileNet V1 model. MobileNet is a lightweight CNN architecture. This architecture was introduced to meet the increasing need to implement deep learning models on devices with limited resources. The principal breakthrough of MobileNet V1 is the use of depthwise separable convolutions (DSC), which significantly decrease computing cost and parameter count while preserving competitive accuracy. The Structure of MobileNet V1 is illustrated in Fig. 2.15.

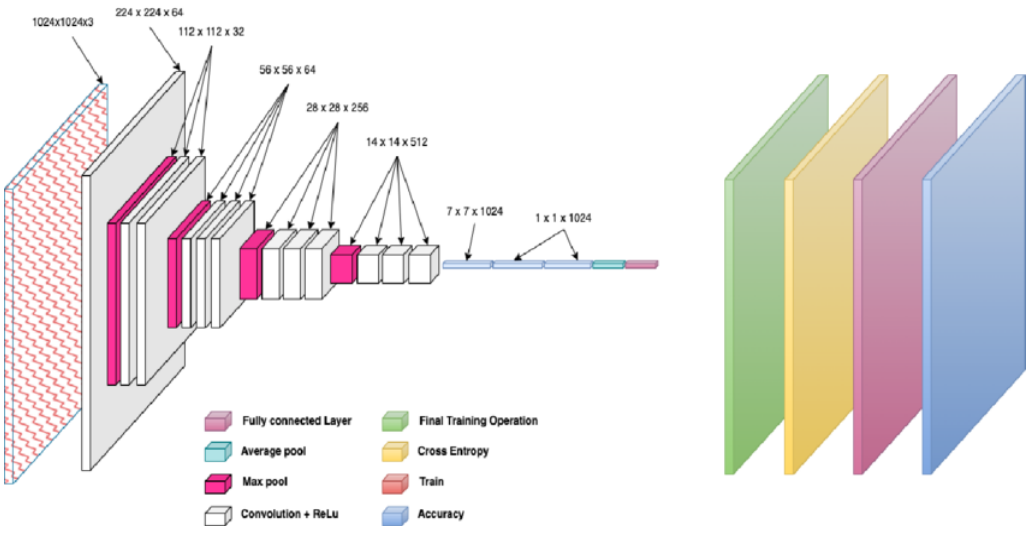


Fig. 2.15: Architecture of MobileNet V1 [54]

2.4.2 MobileNet V2

MobileNet V2 represents an advancement compared to MobileNet V1, as introduced by Mark Sandler et al. [55] This version presented two significant architectural advancements: inverted residual blocks and Linear Bottlenecks. Fig.2.16 represents the architecture of MobileNet V2. MobileNet V2 employs inverted residual blocks, initiating the process by expanding feature maps to higher dimensions through a 1×1 convolution. A depth-wise convolution for processing follows this, and the feature maps are compressed back to lower dimensions via another 1×1 convolution. Skip connections are incorporated between the input and output when the stride is 1, facilitating feature reutilization and effective gradient propagation. In contrast to conventional architecture, the final layer of each bottleneck block does not use ReLU activation. This aids in maintaining detailed information that is typically overlooked, improving the network's effectiveness for tasks that demand accuracy in results.

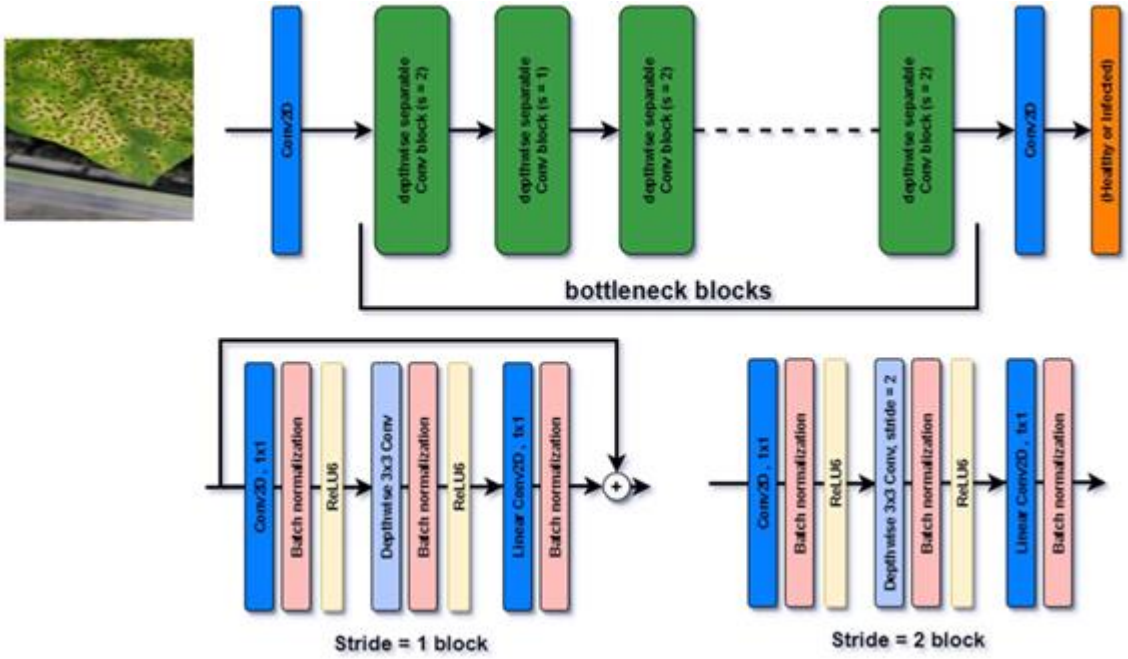


Fig. 2.16: Architecture of MobileNet V2 [56]

Bottleneck blocks with a stride of 1 preserve spatial dimensions, whereas blocks with a stride of 2 reduce the size of the feature maps, enabling the network to capture hierarchical representations.

2.4.3 ResNet-18

ResNet-18 is a well-established convolutional neural network architecture developed by Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun [57]. The ResNet family was created to tackle the challenges associated with vanishing gradients and the decline in performance observed in extremely deep networks. Fig. 2.17 represents the basic architecture of ResNet-18. The ResNet-18 is composed of 18 layers, which include convolutional layers, batch normalization, ReLU activations, and shortcut connections that establish a direct link between the input and output of a layer. The implementation of these shortcuts allows the network to focus on learning residual mappings rather than directly correlating inputs with outputs, thereby enhancing the training process for deeper models. This design has had a profound impact on the fields of computer vision and deep learning research, evidenced by its attainment of leading results on benchmark datasets such as ImageNet.

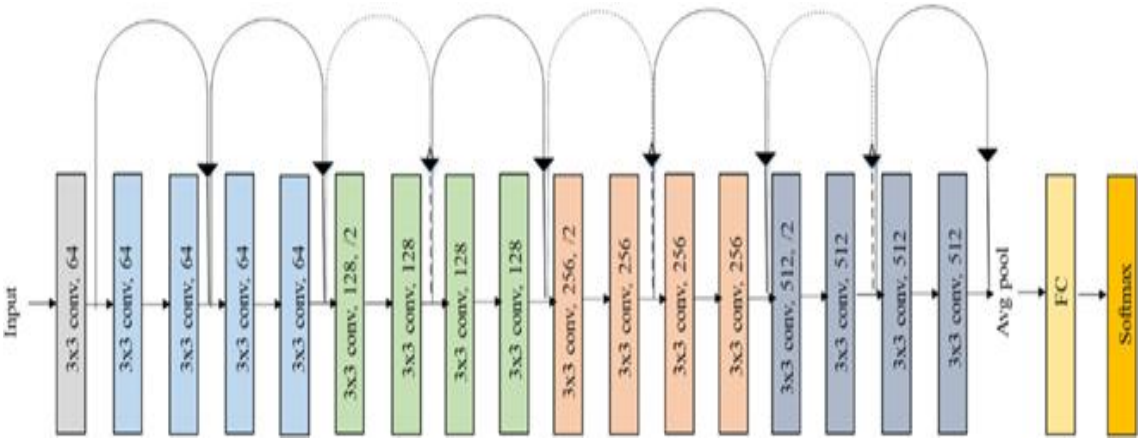


Fig. 2.17: Architecture of ResNet-18 [58]

2.4.4 ResNet-50

ResNet-50, an advanced version of the ResNet architecture, was developed by Kaiming He et al. [57]. ResNet-50 is composed of 50 layers and employs bottleneck residual blocks, distinguishing it from the simpler residual blocks utilized in ResNet-18. The structure of each bottleneck block is composed of three convolutional layers (1×1, 3×3, 1×1), facilitating a reduction in computational cost while preserving the ability to extract deep features effectively. The implementation of identity and projection shortcuts enhances gradient flow within the network, thereby improving optimization and accuracy on extensive datasets such as ImageNet. Fig. 2.18 represents the basic architecture of ResNet-50. The architecture initiates with a 7×7 convolutional layer comprising 64 filters and a stride of 2, subsequently incorporating batch normalization and an activation function (ReLU). This layer facilitates the extraction of fundamental features such as edges and textures from the input image. A 3×3 max pooling layer is then applied to decrease the spatial dimensions, effectively maintaining critical features.

Fig.2.18: Architecture of ResNet-50 [59]

The architecture comprises two distinct types of residual blocks: the Identity Block and the convolutional block. A direct shortcut connection is used in the Identity Block. In scenarios where the input and output dimensions are consistent, a 1×1 convolution is used in the shortcut path of the Convolutional Block to ensure dimensional alignment. When there is a discrepancy between the input and output dimensions. Residual connections mitigate the issue of vanishing gradients and facilitate improved gradient flow, thereby simplifying the training process for deep networks. ResNet-50 consists of several stages, each characterized by progressively larger filter sizes. The stages systematically augment the quantity of feature maps, thereby improving feature representation.

2.5 CLASSIFIERS

A classifier serves as a crucial element in machine learning, systematically categorizing input data by leveraging identified patterns and relationships. Classifiers are essential in data science for automated decision-making, allowing systems to effectively analyze, interpret, and organize complex datasets. Models are generally trained on labelled datasets, allowing them to identify correlations between input features and their respective class labels via supervised learning methods. In image recognition tasks, classifiers undergo training using datasets that consist of images marked with labels. Throughout the training process, the algorithm identifies and assimilates unique features, including edges, textures, and shapes, that define various object categories. After training, the model applies its acquired knowledge to categorize new images by determining the most likely classification [3] Classifiers utilize sophisticated mathematical and statistical methods, such as probability theory, feature extraction, and optimization algorithms, to improve their predictive accuracy. The application of these methodologies spans multiple domains, including computer vision, natural language processing, medical diagnostics,

and semiconductor defect classification. The capacity of classifiers to automate pattern recognition and generate precise predictions has played a crucial role in the progress of intelligent systems within industrial and scientific fields.

In this section, we discuss the softmax classifier, k-nearest neighbors (k-NN) classifier and support vector machine (SVM) classifier.

2.5.1 Softmax Classifier

The Softmax classifier serves as a prevalent algorithm for classification tasks, especially in scenarios involving multiple classes. This method expands upon logistic regression, allowing for the classification of multiple categories instead of being limited to binary outcomes. The Softmax function, recognized as the normalized exponential function, converts raw logits (unnormalized output scores) into a probability distribution across various class labels, guaranteeing that the total of probabilities sums to one [60]. Mathematically, using a set of input characteristics x and a weight matrix W, the raw class scores (logits) for each class j are calculated as:

$$z_j = W_j \,.\, \mathrm{x} + b_j \tag{2.8}$$

Where, W_j and b_j denotes the weight and bias parameters for class j. The Softmax function subsequently transforms these scores into probabilities through the formula.

$$\mathrm{P}\,(\mathrm{y} = \mathrm{j}|\mathrm{x}) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{2.9}$$

K denotes the aggregate count of classes. This formulation guarantees that every output probability is confined within the interval (0,1) and that the aggregate of all class probabilities totals one, thereby facilitating a robust probabilistic interpretation [61].

The Softmax classifier is the final activation function in neural network architectures, particularly in deep learning applications like image

classification, speech recognition, and natural language processing. In semiconductor defect classification, a convolutional neural network (CNN) systematically extracts hierarchical features, and the Softmax layer methodically assigns probabilities to various defect categories, including scratches, edge ring defects, and local defects. Softmax exhibits certain limitations, including a heightened sensitivity to outliers and a tendency towards overconfidence in its predictions. To tackle this issue, methods such as temperature scaling or different loss functions, including cross-entropy loss, are employed to enhance generalization and robustness.The Softmax classifier, while facing certain challenges, continues to be a crucial element in contemporary classification models, offering a probabilistic framework that facilitates decision-making in high-dimensional data contexts.

2.5.2 K-Nearest Neighbors (KNN) Classifier

The K-Nearest Neighbors (kNN) algorithm represents a straightforward yet effective non-parametric, instance-based learning approach applicable to classification and regression tasks. The operation is grounded in the concept that analogous data points are located near each other within a specified feature space. In contrast to parametric models, k-NN operates without assuming any underlying distribution. It retains the complete training dataset and classifies new data points by examining their k-nearest neighbors, utilizing a selected distance metric, which is commonly the Euclidean distance [6]. Mathematically, the Euclidean distance between two points x_1 and x_2 in an n-dimensional space is given by:

$$\mathrm{d}\,(x_1, x_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2} \tag{2.10}$$

where, x_{1i} and x_{2i} represent the feature values of the two points in dimension i. The algorithm categorizes a new data point by identifying the most prevalent class among its k-nearest neighbors through a majority voting

process in classification tasks. In regression, the kNN method determines the target value by calculating the average of the k-nearest values.

Although kNN is straightforward, it demonstrates significant effectiveness across a range of applications, such as image recognition, medical diagnosis, and semiconductor defect classification. In the context of wafer defect detection, kNN operates by classifying semiconductor defects through the comparison of feature vectors derived from defect patterns, subsequently identifying the nearest matches within a labelled dataset. Nonetheless, kNN presents certain constraints, including substantial computational demands when dealing with extensive datasets and a pronounced sensitivity to the selection of k and the distance metric, both of which can greatly influence classification outcomes [62]. Techniques such as KD-Trees and Ball Trees are employed to enhance efficiency in conducting rapid nearest-neighbor searches [63].

2.5.3 Support Vector Machine (SVM) Classifier

The Support Vector Machine (SVM) is a robust supervised learning algorithm applicable to classification and regression tasks. This method demonstrates significant efficacy in handling high-dimensional datasets, especially in scenarios where a distinct margin of separation can be identified between various classes. SVM functions by identifying the optimal hyperplane that maximizes the margin between various class boundaries, which enhances generalization to previously unseen data [3]. In mathematical terms, considering a labelled training dataset $\{(x_i, y_i)\}_{i=1}^{n}$ where x_i represents the feature vector and y_i denotes the class label ($y_i \in \{-1,1\}$ for binary classification), the optimal hyperplane is defined as [3]:

$$w \cdot x + b = 0 \tag{2.11}$$

Where w is the weight vector, x represents the input data, and b indicates the bias factor. The margin, defined as the distance between the closest support

vectors of each class, is maximized by addressing the subsequent optimization problem [3]:

$$\min 0.5 \|w\|^2, \text{ subject to } y_i \text{ (w. } x_i + \text{b)} \geq 1 \quad (2.12)$$

The constraints guarantee that every data point is accurately classified, maintaining a margin of no less than 1 [64].

The extension of SVM to accommodate non-linearly separable data is achieved through the introduction of kernel functions. These include Radial Basis Function (RBF), polynomial, and sigmoid kernels, which facilitate the mapping of input features into a higher-dimensional space. In this transformed space, a linear hyperplane can be effectively applied [65]. SVMs demonstrate significant utility in various applications, including image classification, bioinformatics, and the detection of defects in semiconductor wafers.

2.6 SUMMARY

In this chapter, the significance of semiconductor wafers is discussed. Several types of wafer defects are illustrated with figures. Elaborately discussed convolution neural network (CNN) with depth-wise separable convolution. Different pre-trained models which are Mobile Net V1, MobileNet V2, ResNet-18 and ResNet-50 are explained in detail. Widely used image pre-processing techniques: C-mean filtering, median filtering and RoD transform are briefly explained in this chapter. Also, the operation of well-known classifiers (soft-max classifier, KNN classifier and SVM classifier) are described at the end of this chapter.

Chapter 3: MATERIALS AND METHODOLOGY

This chapter discusses detail methodology and materials used in the research. Section 3.1 provides a general overview of the method, which will be briefly discussed in the parts that follow. Section 3.2 describes the datasets used in the study. Section 3.3 explains data pre-processing techniques in which the histogram equalization process is discussed briefly. A brief explanation of the proposed modified MobileNet architecture is given in Section 3.4. Section 3.5 introduces the proposed hybrid system including the ECOC-SVM classifier.

3.1 METHODOLOGY

Numerous techniques have been developed to identify wafer defects in the semiconductor manufacturing sector. These approaches face difficulties in identifying minor faults and intricate patterns in semiconductor wafer maps. Furthermore, these methodologies struggle to capture long-range correlations in defective wafers. To solve these limitations this research work represents a hybrid system.

The proposed method is depicted in Fig.3.1. The process begins with the raw wafer map images. In the pre-processing step, the histogram equalization technique is applied to these images to address tiny defect detection issues. This process improves the contrast in the defect wafer images. This step improves the visual clarity of both defect and non-defect wafer images, thereby increasing their appropriateness for feature extraction. The histogram-equalized wafer images are separated into two datasets: 80% training set and 20% testing set. The training and testing sets are processed through a multi-head attention-based modified MobileNet feature extraction structure. The swish function is used within modified MobileNet architecture to identify intricate defect patterns in wafers and the multi-head attention mechanism integrates with modified MobileNet so that the proposed system can capture long-range

relationships in defective wafers. This combination extracts features from defective and non-defective wafers. For training and testing purposes, the extracted features are stored separately.

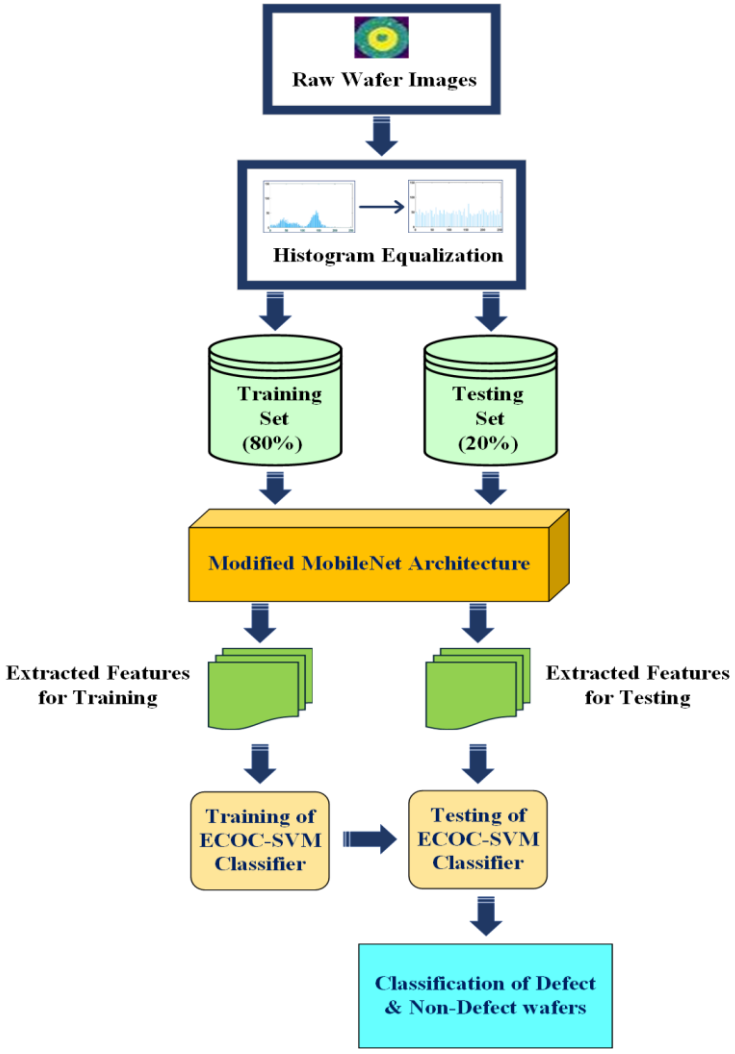


Fig. 3.1: Methodology of the proposed hybrid system

The ECOC-SVM approach is employed in this hybrid system for classification purposes. This approach uses several binary SVM classifiers to address multi-class classification problems. The features obtained from the training set are used to train the ECOC-SVM classifier. The features obtained from the testing set are input into the trained ECOC-SVM classifier. The classifier model generates predictions that classify wafer maps into defect and non-defect categories. Finally, the analysis of the proposed method's performance is conducted to assess its effectiveness in accurately identifying

wafer defects, utilizing performance metrics which include accuracy, precision, recall, and AUC values.

The SVM with non-linear kernels, like the radial basis function (RBF), are recognized for their effectiveness in classification tasks that require handling complex decision boundaries. However, their application to high-dimensional data, such as wafer maps, demonstrates notable limitations. SVMs depend significantly on manually designed feature extraction, which might not sufficiently represent the complex spatial patterns and variations found in semiconductor defect images. Conversely, deep learning models, especially CNNs, possess the capability to autonomously acquire hierarchical feature representations directly from unprocessed wafer map data. This removes the necessity for manual feature engineering and allows the model to identify both local and global defect characteristics effectively. Additionally, although SVMs demonstrate strong performance on smaller datasets, their effectiveness frequently declines or incurs significant computational costs as both the dataset size and input dimensionality expand. Deep learning models, particularly those augmented with attention mechanisms, demonstrate superior capability in managing extensive and intricate datasets, such as the 55,143 wafer samples utilized in this study. Their design facilitates comprehensive learning and enhanced scalability, which is essential for effective real-time defect detection in industrial settings. Consequently, deep learning presents a superior and more effective approach for the classification of semiconductor wafer defects compared with conventional SVM with non-linear kernels.

3.2 DATASET

A dataset for real-time wafer, WM811K [66], is used in this study. Many research studies use this data collection [11-12] [14-16][22][29]. The pie chart of Fig.3.2 represents the overall distribution of defect classes. 55,143 real wafer map images are used in this study, among them 6.27% are center defect class, 0.74%

are donut defect class, 4.38% are edge-local defect class, 15.51% are edge-ring defect class, 2.93% are local defect class, 1.10% random defect class, 0.27% are near-full defect class, 2.16% are scratch defect class, and 66.60% are non- defect classes.

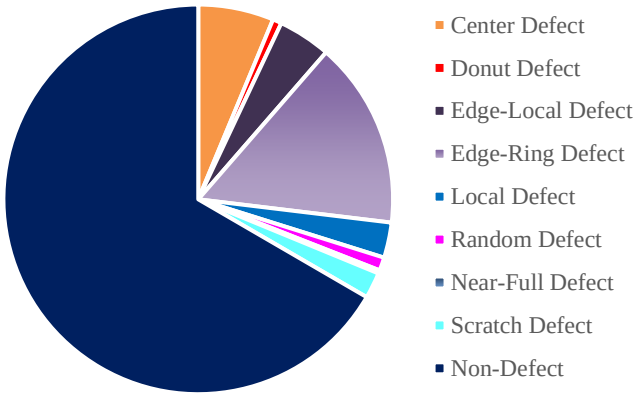


Fig. 3.2: Distribution of Defect Classes in Dataset

Although there are imbalanced classes of defects and non-defect wafer images, the ECOC-based SVM classifier solves this class imbalance problem. Fig.3.3 shows eight types of defect classes: center, donut, local, edge-local, edge ring, near full, random, scratch, and one non-defect class defined as none. All of them are non-histogram equalized wafer map images.

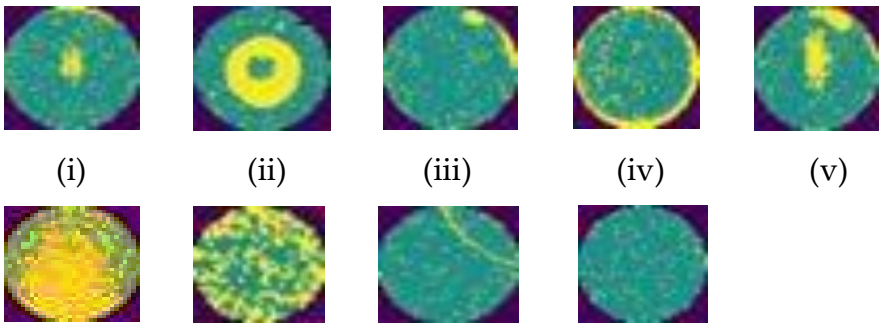


Fig. 3.3: Defect and non-defect wafer patterns before HE: (i) Center (ii) Donut (iii) Edge local (iv) Edge ring (v) Local (vi) Near full (vii) Random (viii) Scratch (ix) None

3.3 DATA PRE-PROCESSING

Non-equalized defect wafer images exhibit inconsistent contrast because pixel intensity values are confined to a limited range, which makes subtle features and tiny defects of wafer images difficult to identify. For this reason, defective wafer detection systems face difficulties in identifying tiny defects in wafers. To solve this problem, the proposed research method uses the histogram equalization technique in the image processing step applied to real defective wafer images. This technique distributes pixel intensity values to improve contrast in the wafer image.

3.3.1 HISTOGRAM EQUALIZATION (HE) TECHNIQUE

Pixel intensities in a defect wafer map usually indicate defect density; histogram equalization highlights the distinctions between defect and non-defect regions. The pixel intensity value of the defect wafer map image is I(x,y) where x,y are the spatial coordinates on the wafer map. The number of pixels in the image with intensity i is represented by the histogram $h(i)$. The total number of pixels in the wafer image is [67]:

$$T_p = \mathrm{M} \times \mathrm{N} \tag{3.1}$$

where M and N are the dimensions of the wafer map. The intensity levels are redistributed using the cumulative distribution function (CDF). For a specific intensity level, the CDF is defined as [67]:

$$CDF(\mathrm{i}) = \sum_{\mathrm{j}=0}^{\mathrm{i}} h(\mathrm{j}) \tag{3.2}$$

where h(j) is the number of pixels with intensity j. The normalized CDF is given by [67]:

$$CDF(\mathrm{i})_{norm} = \frac{CDF(\mathrm{i})}{T_p} \tag{3.3}$$

The total number of pixels is represented by T_p. The pixel values of the original image are then mapped to the new ones by transforming the pixel

intensity values using the normalized CDF. With an intensity of I(x,y), the transformation function for that pixel is [67]:

$$I(\mathrm{x},\mathrm{y})_{new} = I_{min} + (\ CD\mathrm{F(i)}_{norm}(\mathrm{I(x,y)})\ \mathrm{x}\ (\mathrm{L}-1)) \tag{3.4}$$

Here, I_{min} is the minimum pixel intensity of the image, L is the maximum possible intensity value. $I(\mathrm{x},\mathrm{y})_{new}$is the new intensity value after histogram equalization. The histogram equalized defective and non-defective wafer map images are illustrated in Fig 3.4.

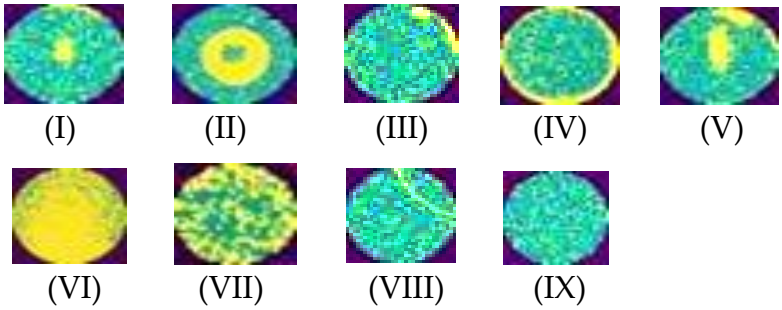
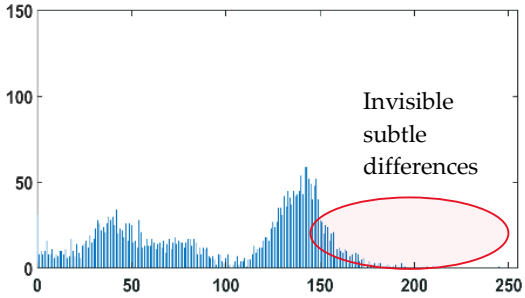
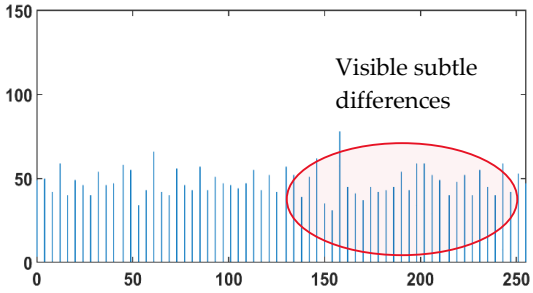


Fig.3.4: Defect and non-defect wafer patterns after HE: (I) Center (II) Donut (III) Edge local (IV) Edge ring (V) Local (VI) Near full (VII) Random (VIII) Scratch (IX) None

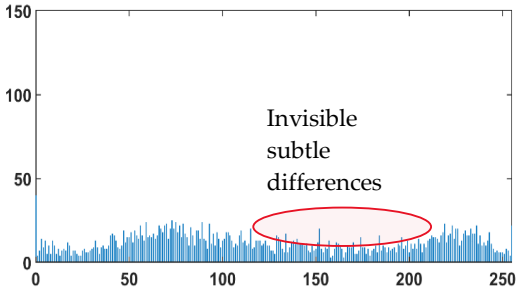
Fig. 3.5 shows the histogram of un-equalized and equalized wafer images in which the x-axis denotes the range of intensity values & the y-axis denotes the number of pixels with a specific intensity value of the images. Subtle differences are marked in histograms for both un-equalized and equalized wafer images. The histogram of non-equalized defect images indicates that most pixel intensities are concentrated within a specific range, with a fewer number of pixels at both lower and higher intensity levels. Because of this concentration of intensity levels, the image's brightness varies less, which adds to the low contrast. For this reason, the identification of tiny defects in wafer maps is challenging by using un-equalized defect images. Conversely, equalized image histograms exhibit a more consistent distribution throughout the whole intensity range. Previously underrepresented intensity levels are now enhanced due to this pixel intensity redistribution. As a result, subtle differences in image intensities are magnified and small defects in wafer maps are visible.



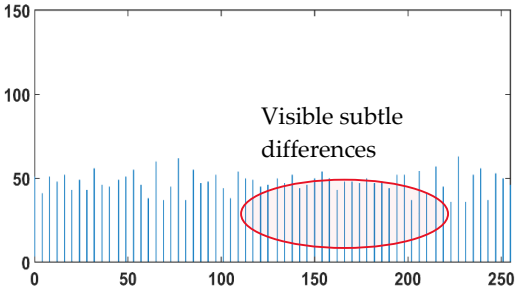
(a): Unequalized center defect



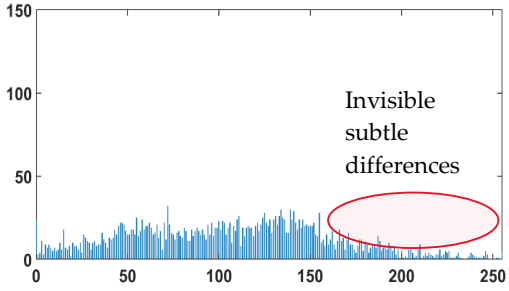
(b) Equalized center defect



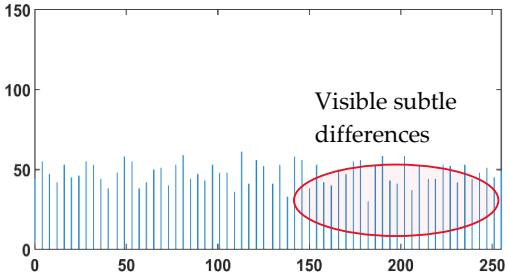
(c) Uequalized donut defect



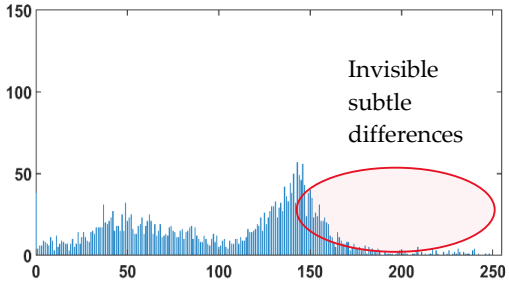
(d) Equalized donut defect



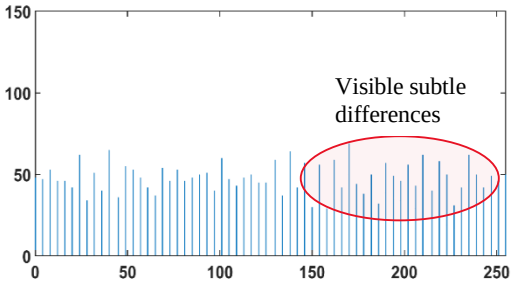
(e): Unequalized edge-local defect



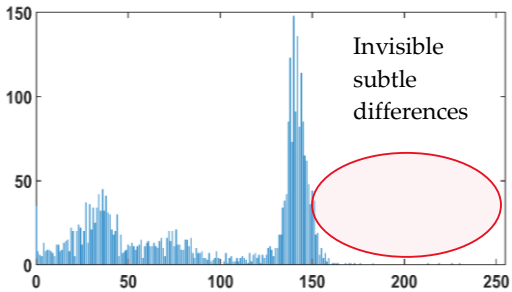
(f): Equalized edge-local defect



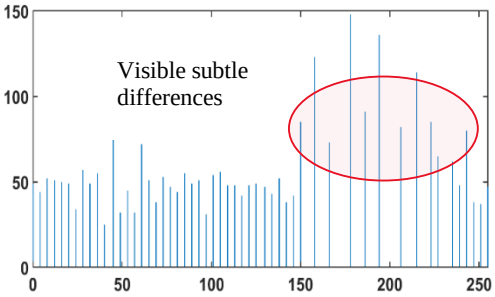
(g): Unequalized edge-ring defect



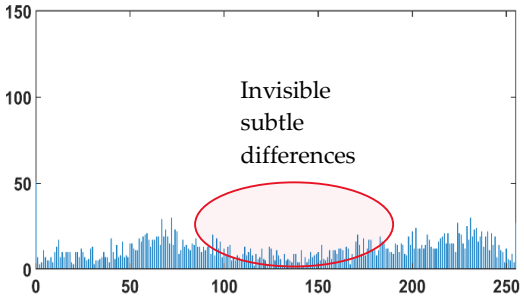
(h): Equalized edge-ring defect



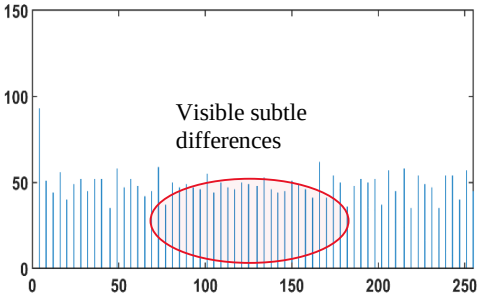
(i): Unequalized local defect



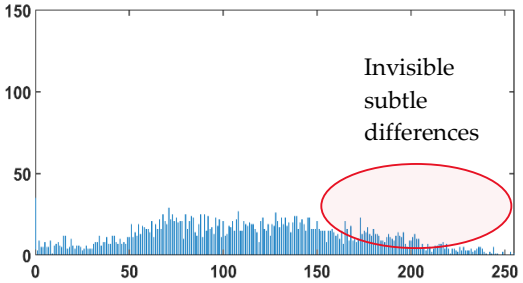
(j): Equalized local defect



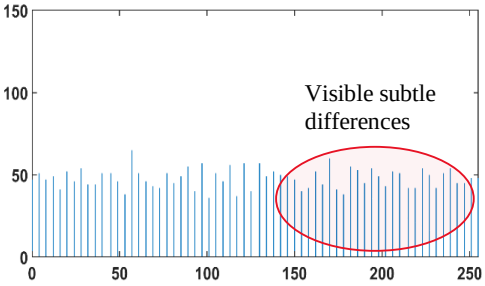
(k): Unequalized near-full defect



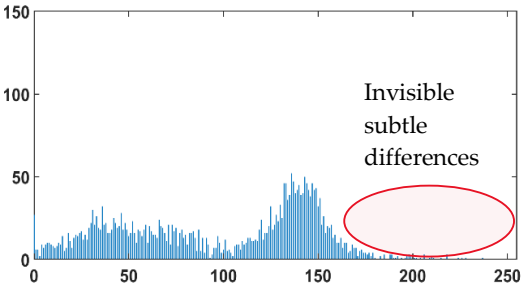
(l): Equalized near-full defect



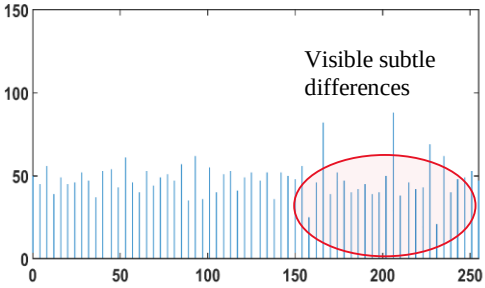
(m): Unequalized random defect



(n): Equalized random defect



(0): Unequalized scratch defect



(p): Equalized scratch defect

Fig.3.5: The histogram of Un-Equalized and equalized wafers

Fig. 3.5 (a), Fig. 3.5 (e), Fig. 3.5 (g), Fig. 3.5 (i) Fig. 3.5 (m) and Fig. 3.5 (o) represent the histogram of non-equalized images of center defect, edge local defect, edge ring defect, local defect, random defect and scratch defect respectively. The histograms indicate that the majority of pixel intensities are clustered in the mid-range, suggesting that a significant segment of the image exhibits similar intensities. This results in numerous intricate differences within the defect region being diminished or turned invisible. Following the equalization of these defect images, the pixel intensities are redistributed more uniformly across the entire range shown in Fig. 3.5 (b), Fig. 3.5 (f), Fig. 3.5 (h), Fig. 3.5 (j), Fig. 3.5 (n) and Fig. 3.5 (p). The histogram of equalized defect

images demonstrates spikes that signify regions with a high concentration of defects, while an irregular distribution of intensity values points to the presence of defect areas. The resulting histograms exhibit numerous minor spikes and fluctuations. The histogram of equalized images exhibits minor spikes in the ranges of 0 to 50, 100 to 150, and so forth. The spikes align with areas in the image that exhibit slight differences in intensity. The small intensity variations emphasize additional details in the defect regions. The difference between defect areas and their surroundings becomes more apparent following the histogram equalization process. The subtle differences in pixel intensities remain undetectable in the histogram of non-equalized defect images.

The histogram of original images of donut defect and near-full defect are depicted in Fig.3.5 (c) and Fig. 3.5 (k) respectively. Despite the majority of pixel intensities not being concentrated at the center of the histogram, the subtle variations in pixel intensities remain invisible. After applying the histogram equalization process on these un-equalized donut defect and near full defect images, the small differences in pixel intensities detected effectively are shown in Fig. 3.5 (d) and Fig. 3.5 (l).

Donut, edge local, edge ring and near full defects patterns exhibit complex patterns. Fig. 3.5 (d), Fig. 3.5 (f), Fig. 3.5 (h), and Fig. 3.5 (l) depict the histogram of equalized donut, edge local, edge ring and near full defect images, respectively. The histogram of the equalized images shows pixel intensities distributed across a wider range. Also, these histograms depict that spikes fluctuate between different heights, and pixel intensities vary from one region to another. These fluctuations reflect small variations in intensities of the donut, edge-local, edge ring and near-full defect images.

3.4 PROPOSED MODIFIED MOBILENET ARCHITECTURE

The basic MobileNet comprises the depth-wise separable convolution (DSC) block that uses the ReLU activation function. MobileNet is useful for

lower parameter counts, and lower computation costs are required. However, in detecting wafer defects, the MobileNet architecture confronts challenges to recognize complex defective wafers precisely. This thesis presents a modified version of MobileNet that modifies the DSC block by implementing the swish function in place of the ReLU function to detect complex defective patterns. Additionally, the integration of the multi-head attention mechanism with MobileNet architecture facilitates the capturing of long-range dependencies in wafer defects.

3.4.1 Basic structure of MobileNet

Fig.3.6 shows the basic architecture of MobileNet [46]. This structure comprises depth-wise separable convolution (DSC) blocks, global average pooling, a fully connected layer, and a classification layer using a softmax classifier. The DSC blocks are repeated numerous times (from DSC 1 to DSC 13), each possibly incorporating an increasing number of filters and periodically utilizing a stride of 2 to diminish the spatial dimensions further. Each DSC block consists of a 3x3 depth-wise convolution followed by batch normalization and ReLU, as well as a 1x1 pointwise convolution accompanied by batch normalization and ReLU. However, the MobileNet confronts problems in detecting complex defect patterns. The MobileNet architecture used the $ReLU$ activation function in the DSC block. the ReLU activation function is expressed as:

$$\text{ReLU}\,(x) = \max\,(0,x) \tag{3.5}$$

For positive values of x, ReLU outputs x (linear activation). For negative values of x, ReLU outputs 0 (non-linear, effectively deactivating the neuron).

The derivative of ReLU activation function is:

$$\text{for } x < 0, \quad \frac{d}{dx}\,\text{ReLU}\,(x) = 0 \tag{3.6}$$

$$\text{for } x > 0, \quad \frac{d}{dx}\,\text{ReLU}\,(x) = 1 \tag{3.7}$$

ReLU passes the gradient unchanged for positive inputs on the other hand for x< 0 the gradient of ReLU is zero, which means that during backpropagation, the weights connected to these neurons do not get updated.

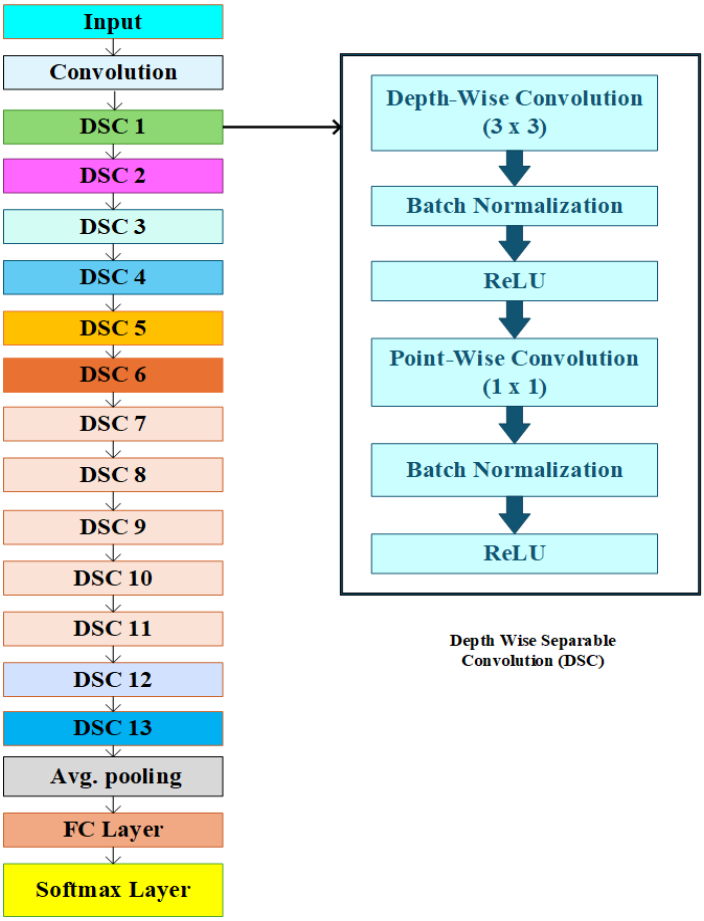


Fig. 3.6: Structure of original MobileNet [44]

This can lead to the "dying ReLU" problem, where a significant portion of neurons might stop learning, especially when defects are complex and involve subtle negative activations

3.4.2 Modified MobileNet Architecture

The basic Mobile Net comprises the depth-wise separable convolution (DSC) block that uses the ReLU activation function. In detecting wafer defects the MobileNet architecture fails to identify complex defective wafer preciously. We propose a modified version of MobileNet that changes the DSC block by using the swish function instead of the ReLU activation function to detect

intricate defective patterns. Fig.3.7 illustrates the modified structure of MobileNet.

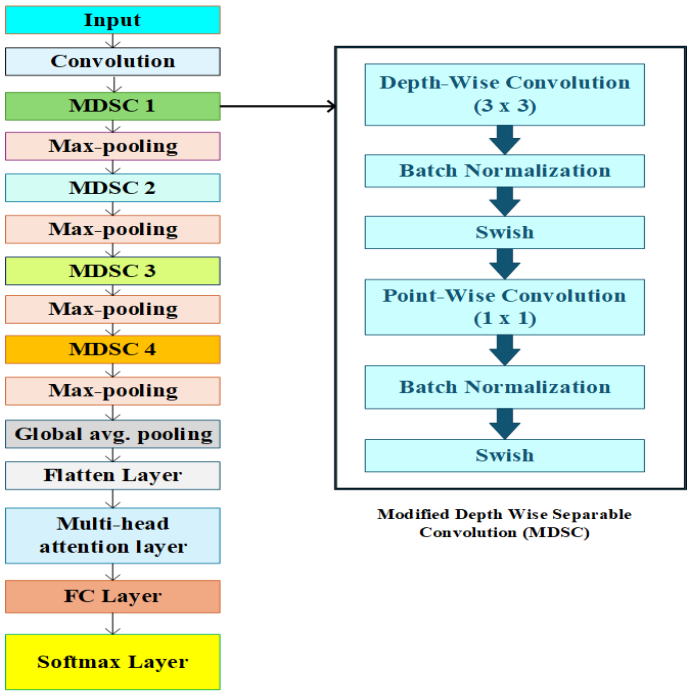


Fig. 3.7 : Architecture of Modified MobileNet

The proposed Modified MobileNet architecture encompasses four MDSC blocks. Except for the swish activation function, other elements of the MDSC block are identical to those of the DSC block of the original version of MobileNet. The max-pooling layers are strategically positioned following each MDSC, for minimizing spatial dimensions. Also, in modified MobileNet architecture global average pooling is used to reduce the overall parameter count and computational complexity of the developed hybrid model.

On the other hand, the swish function is defined as:

$$\text{Swish}\ (x) = x.\ \sigma(x) = \frac{x}{1+e^{-x}} \tag{3.8}$$

Swish is a smooth, non-linear function that outputs values in the range (−∞, ∞). For large positive x, swish is approximately linear, similar to ReLU. For large negative x, swish outputs a small negative value close to zero, but not exactly zero. For values of x around zero, swish behaves non-linearly, with a smooth

transition from negative to positive values This can lead to better convergence during training and allow the proposed modified version of MobileNet architecture to capture more complex and subtle defect patterns. The derivative of the swish function is:

$$\frac{d}{dx}\text{Swish}(x) = \sigma(x) + x.\ \sigma(x).\ (1-\sigma(x)) \quad (3.9)$$

Where σ(x) is the sigmoid function. This smooth derivative varies with input x, allowing for more nuanced weight updates during backpropagation. The term x. σ(x). (1-σ(x)) introduces non-linearities, which help the proposed model to capture complex patterns. These analyses indicate that for capturing complex defect wafers the swish activation function performs better than the ReLU activation function because of its smoother transitions and capacity to permit small negative outputs for improved gradient propagation

However, this modified MobileNet architecture is not capable of capturing long-range dependencies in defective wafers. This modified MobileNet exhibits limitations in its ability to concentrate on distinct defect patterns within the wafers, resulting in challenges at the time of dealing with prioritizing critical areas of the image over less relevant or redundant regions of the wafers. Consequently, this architecture fails to capture the global context and the relationships among different parts of the defective wafer image. that reduces the overall effectiveness of identifying wafer defects.

3.4.3 Multi-Head Attention Mechanism

The Modified MobileNet architecture encounters challenges in correctly capturing long-range dependencies. This method incorporates a multi-head attention mechanism into the modified MobileNet architecture to solve this problem. The implementation of multi-head attention makes the modified MobileNet's capable of identifying long-range dependencies by enabling the modified MobileNet to concentrate on various areas of the wafer image

simultaneously. This mechanism allows the modified architecture to focus on relevant defect regions of the wafers and measures the interactions between distant regions on the defective wafer. A small defect on one side of the wafer potentially affects the features observed on the opposite side. This attention mechanism effectively identifies this relationship, which is challenging to determine through convolution operations alone. In multi-head attention mechanisms, the input feature maps of wafers are segmented into several smaller regions, referred to as "heads." Each attention head functions autonomously, enabling the proposed MobileNet to simultaneously identify various relationships and patterns present in defect wafer feature maps.

Each attention head comprises three distinct sets of parameter matrices: Query (Q), Key (K), and Value (V). Query, Key, and Value Vectors are given values [68].

$$Q = XW^Q \tag{3.10}$$
$$K = XW^K \tag{3.11}$$
$$V = XW^V \tag{3.12}$$

In this context, W^Q, W^K and W^V represent the weight matrices that are learned, while X denotes the input. The Q denotes the feature representation of a specific region of the wafer map that the model emphasizes (current area), K indicates the feature representations of all regions (including the current and surrounding areas), and V encompasses the actual information to be aggregated according to the attention scores. The dimensionality of these matrices dictates the size of the vectors utilized for computing attention scores. Fig.3.8 depicts the process of the multi-head attention mechanism. This illustration depicts the structure of a multi-head attention mechanism utilized in transformer models, featuring 8 heads. The input consists of three distinct components Query (Q), Key (K), and Value (V). All these components undergo a linear transformation. The dimensions of the query and key are specified as d_k while the dimension of the value is defined as d_v [68].

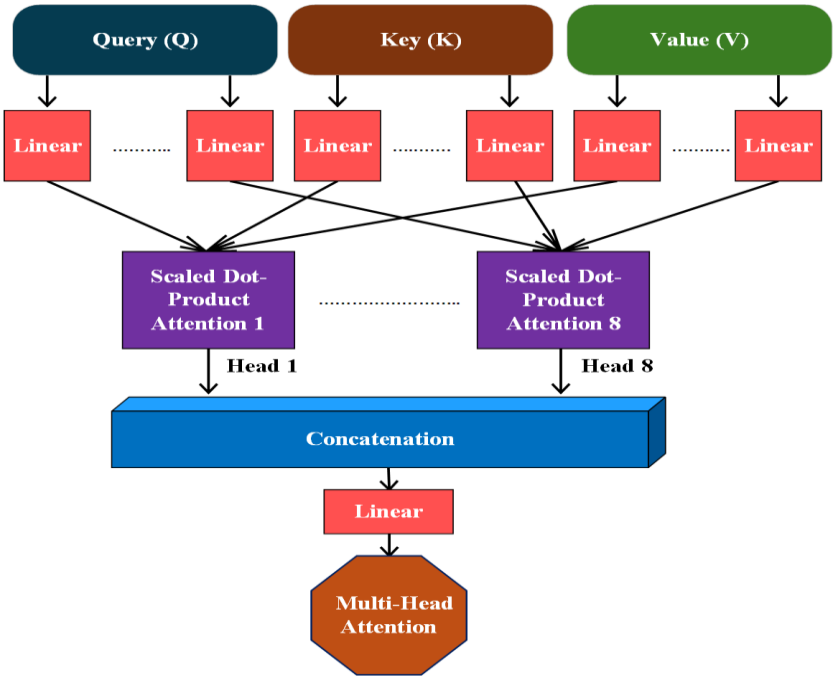


Fig.3.8: The multi-head attention process

Each component (Q, K, and V) is replicated 8 times, one for each attention Each component, namely Q, K, and V, undergoes replication 8 times, corresponding to each attention head. After the linear transformations, the model systematically calculates the scaled dot-product attention for each head separately. This operation entails comparing the queries and the keys, utilizing the outcome to assign weights to the values, thereby enabling the model to concentrate on various areas of the defect wafers. The presence of 8 distinct attention heads arises from the advantage of employing multiple attention functions rather than relying on a singular attention mechanism with d_{model} dimensional keys, values, and queries. This approach involves linearly projecting the queries, keys, and values n times using various learned linear projections to dimensions d_k, d_k, and d_v, respectively [68]. Each head independently calculates its attention weights, allowing the model to gather contextual information from diverse representations of defect wafers across multiple positions. The outputs from all 8 attention heads are concatenated, effectively integrating the information derived from each head. Subsequently, an additional linear layer processes the combined outputs to generate the

ultimate multi-head attention output. The multi-head attention mechanism calculates the attention score for each region in the wafer map corresponding to every defect pattern which given below [68].

$$Attention\ (Q, K, V) = softmax\ (\frac{QK^T}{\sqrt{d_k}})V \tag{3.13}$$

The softmax operation serves to normalize the attention weights in this context. This scaled dot product attention emphasizes the relevant parts of the wafer map while diminishing the impact of irrelevant regions. Once the attention is calculated for each head, the outputs from all heads are combined through concatenation [68].

$$Output = Concat\ (head_1\ , \dots, head_8)W^O \tag{3.14}$$

Where,

$$head_i\ = Attention\ (QW_i^Q\ , KW_i^K\ , VW_i^V), i = 1 \dots 8 \tag{3.15}$$

and

$$W_i^Q \epsilon R^{d_{model}\ x\ d_k} \tag{3.16}$$
$$W_i^K \epsilon R^{d_{model}\ x\ d_k} \tag{3.17}$$
$$W_i^V \epsilon R^{d_{model}\ x\ d_v} \tag{3.18}$$
$$W^O \epsilon R^{d_{model}\ x\ nd_v} \tag{3.19}$$

Where W^O is the final transformation weight matrix. In this proposed model, the d_{model} is 512 and the multi-head attention mechanism has 8 heads. Now, d_k and d_v are computed as:

$$d_k = d_v = \frac{d_{model}}{n} = \frac{512}{8} = 64$$

The overall computational cost is similar to single-head attention with full dimensionality because each head has a smaller dimension [68].

To implement the multi-head attention mechanism, the wafer map of the center defect is segmented into n regions, where the number of regions corresponds to the number of heads, specifically n = 8. The attention mechanism generates a new representation for each region by taking into account all other regions through the following process:

Firstly, for each input x_i, the multi-head attention mechanism computes three vectors

$$q_i = x_i W^Q \tag{3.20}$$

$$k_i = x_i W^k \tag{3.21}$$

$$v_i = x_i W^v \tag{3.22}$$

Afterwards, the attention mechanism evaluates the attention scores for each region i concerning every other region j, irrespective of their spatial proximity within the wafer of the central defect. The calculation of the attention score for region i to region j is performed as follows:

$$a_{ij} = \frac{exp(\frac{q_i k_j}{\sqrt{d_k}})}{\sum_{j=1}^{n} exp(\frac{q_i k_j}{\sqrt{d_k}})} \tag{3.23}$$

New representation for region i

$$Output_i = \sum_{j=1}^{n} a_{ij} v_j \tag{3.24}$$

Here, a_{ij} Assess the impact of region j on region i within the wafer map demonstrating the central defect pattern. By determining Output_i,the multi-head attention mechanism enables the modified MobileNet model to effectively capture long-range dependencies in center wafer defect patterns. This mechanism assesses the impact of surrounding non-defective regions on the defective central region of the wafer map, without regard of their spatial distance. The attention mechanism effectively identifies long-range dependencies in donut, edge local, edge ring, and near-full defects, which exhibit intricate patterns. The donut defect exhibits a circular configuration, characterized by an absence or clear area at its center. Multi-head attention effectively identifies dependencies between the outer and inner boundaries of the ring, demonstrating the relationships between edge pixels and the remaining parts of the wafer, spite of their spatial distance. The defect pattern is nearly comprehensive, covering almost the entire wafer, with only minor

regions of non-defective areas present. The multi-head attention effectively identifies correlations between small defect-free areas and the larger surrounding defect regions, thereby ensuring that the relationships across the entire wafer are fully considered. The localized defects are primarily concentrated along the edges of the wafer. Multi-head attention enables the allocation of distinct heads to concentrate on various segments of the edge, facilitating the model's ability to combine information from distant regions of the edge that are not near one another. The edge ring, similar to the donut defect, forms a circular pattern, but restricted to the perimeter of the wafer. The multi-head attention mechanism systematically analyzes the relationships between points along the ring by employing multiple heads that concentrate on distinct angular positions. This approach enables the proposed MobileNet structure to effectively capture interactions among various segments of the ring, irrespective of their spatial separation. Furthermore, this attention mechanism captures long-range dependencies related to local, scratch, and random defective wafers. In each instance, numerous attention heads enable the model to recognize a variety of positional relationships across different scales, making it suitable for identifying both local features and extensive patterns typical of intricate defect maps.

3.4.3 Proposed Hybrid System

The proposed hybrid architecture is illustrated in Fig.3.9. In this architecture, the multi-head attention mechanism integrates with modified MobileNet that extracts features from defects in wafers efficiently. The multi-head attention in the proposed mobile net allows the model to investigate interactions between distant wafer regions. The proposed method replaces the modified MobileNet architecture classification layer with the ECOC-SVM classifier to build a hybrid system.

Four modified depth-wise separable (MDSC) blocks are used in this architecture. Every MDSC block consists of a depth-wise convolution followed by a pointwise convolution. The depth-wise convolution utilizes an individual filter for each input channel, thereby minimizing computational requirements. Point-wise convolution employs 1x1 convolutions to integrate channel-wise features, thereby enhancing the model's depth. All the depth-wise convolution and point-wise convolution contain batch normalization and swish activation functions. The Max-pooling (2x2) is implemented following each MDSC block to halve the spatial dimensions, thereby enhancing the receptive field.

At first, the input images with dimensions 32x32 pixels of 3 color channels are fed to the initial convolution layer with 32 filters of size 3x3. This layer identifies fundamental features of the image and produces feature maps measuring 32x32x32.

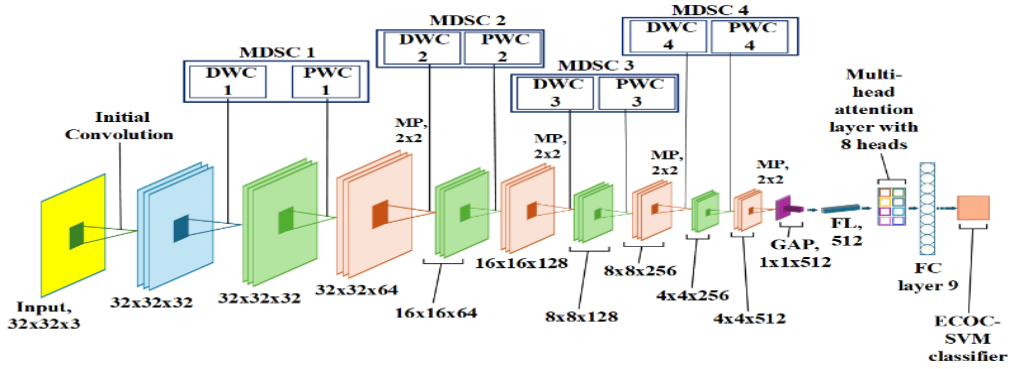


Fig. 3.9: Details Architecture of proposed Hybrid System

Then, in the MDSC1 block, spatial filtering and channel combination are separated by a depth-wise convolution with 32 filters of size 3x3, further lowering the computational cost. Outputs of depth-wise convolution 1 are fed to pointwise convolution 1. This 1x1 convolution combines the outputs from the depth-wise convolution by mixing information across channels using 64 filters and producing feature maps with dimensions 32x32x64. This feature maps dimension reduced to 16x16x64 through max pooling operation. After that, the depth-wise convolution 2 of the MDSC2 block, uses 64 filters to refine each

channel's spatial feature and the pointwise convolution 2 of the same block, mixes the 64-channel output to produce 128 feature maps of 16x16 dimension by using 128 filters. Again, the max pooling operation is executed following the MDSC 2 block, resulting in a downsampling of the feature maps by a factor of 2, transforming the feature maps into dimensions of 8x8x128.

Identically, MDSC 3 generates 256 feature maps measuring 4×4, utilizing 128 filters in depth-wise convolution and 256 filters in pointwise convolution. Also, MDSC4 blocks produce a total of 512 feature maps with dimensions of 2x2 after the max pooling operation. The MDSC4 consists of depth-wise convolution utilizing 256 filters, along with pointwise convolution employing 512 filters. This architecture employs global average pooling, which condenses the spatial dimension of 2x2x512 into a single vector of size 1x1x512 by averaging the spatial features across each channel. This guarantees a concise representation that maintains information on a channel-by-channel basis.

The flatten layer flattens the output from the global pooling layer into a one-dimensional vector with 512 features. It significantly lowers the total number of parameters. The output of flatten layer is then passed into the multi-head attention layer. This layer applies attention mechanisms using 8 attention heads and an output dimension of 64. This layer is used for capturing long-range dependencies across the defect wafers. Two fully connected (FC) layers are used in this proposed architecture. The multi-head attention layer's output is linked to the fully connected layer, which has 9 neurons and demonstrates the total number of defect and non-defect wafer map classes. From this fully connected layer, the features of the histogram equalized defective and non-defect wafers of both the training set and testing set are extracted. Features from the training set are used to train the ECOC-SVM classifier, then it uses the features from the testing set to classify wafer maps as either defect or non-defect.

3.5 ECOC TECHNIQUE-BASED SVM CLASSIFIER

SVM act as a supervised machine learning algorithm specifically designed for binary classification tasks. The method involves directly identifying a hyperplane to separate two classes. The primary goal of SVM is to optimize the distance between each class's closest data points (support vectors) and the hyperplane. However, the multiclass classification problems cannot be handled using SVM. To solve this issue, the proposed method is using the ECOC-based SVM classifier. The implementation of the ECOC-SVM classifier in the hybrid model is shown in Fig.3.9. This classifier classifies eight defect classes and one non-defect class and also solves the class imbalance problem. The SVM is primarily employed as a binary classifier. The ECOC method integrates binary SVM classifiers to address multiclass issues [69]. A multi-class classification problem is transformed into several binary classification problems by ECOC. This decomposition addresses class imbalance by directing each binary classifier's attention towards specific class pairs, thereby redistributing focus and mitigating the effects of imbalance within the framework of each binary problem. The ECOC–SVM technique comprises the ECOC encoding and the ECOC decoding stages. A coding strategy is necessary to train SVM binary learners during the encoding stage. In this study, one vs one coding design [70] is employed for the categorization of defect classes. Instead of learning the entire multi-class issue, each binary classifier learns to determine whether a wafer image belongs to a subset of defect classes during training. These codes error-correcting features add redundancy, improving the classifier's ability to overcome class imbalance problems. The final classification of defect and non-defect classes is conducted during the decoding stage by combining the outputs of all binary SVM classifiers.

The total number of binary SVM classifiers for a one-vs-one structure with nine classes is calculated as follows:

Total no. of binary classifiers = $\frac{n(n-1)}{2} = \frac{9(9-1)}{2} = 36$

where n (n = 9) is the number of classes. To differentiate between the two classes, a total of 36 binary SVM classifiers are trained. Nine rows, one for each class, and thirty-six columns, one for each binary classifier, make up the ECOC code matrix. Fig.3.10 shows the block diagram of the ECOC code matrix. With +1 denoting that the class is the positive class for that binary classifier, -1 denoting that the class is the negative class for that binary classifier, and 0 denoting that the class is excluded from this binary classification task, each entry in the code matrix is set to +1, -1, or 0.

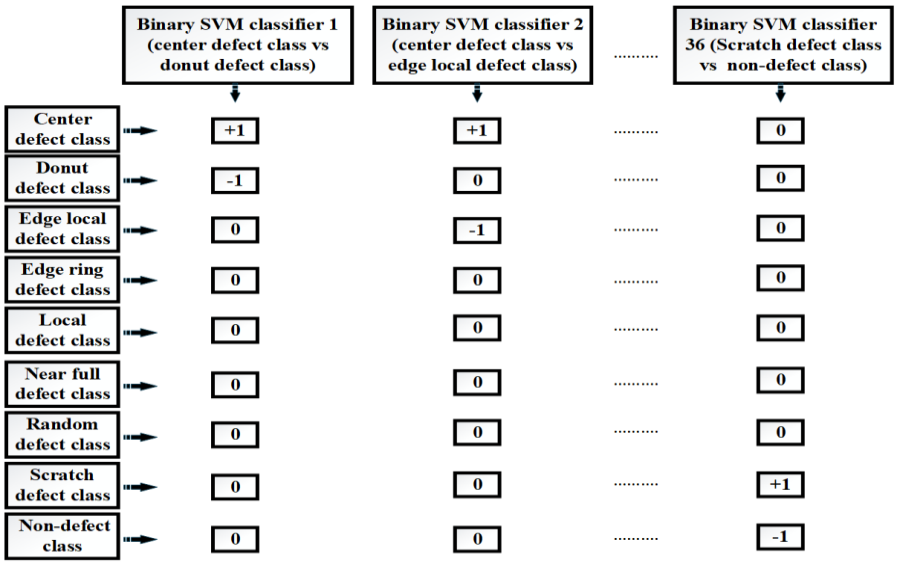


Fig. 3.10: The block diagram of the ECOC coding matrix

A binary SVM classifier is trained for every column in the code matrix to distinguish between the two classes with non-zero values. The training process for each SVM is as follows: For a given classifier, given a feature vector x and binary labels y ϵ {+1, −1}, the SVM maximizes the margin between the two classes to choose the best hyperplane separating them. Each SVM classifier solves the following optimization problem to determine the best separation hyperplane between the two classes [3].

$$\min_{w,b,\xi} 0.5 \|w\|^2 + C \sum_{i=0}^{N} \xi_i \qquad (3.25)$$

subject to:

$$y_i\,(\mathrm{w}.\,x_i\ +\mathrm{b}) \geq 1-\xi_i\ ,\ \xi_i \geq 0 \qquad (3.26)$$

Where w is the weight vector, b is the bias, $\xi_i$are slack variables, C is the regularization parameter, and y_i are the labels derived for each binary classifier. The SVM assigns a label of +1 or −1 to each test sample, determined by its position relative to the hyperplane.

To classify a new instance x, each of the 36 binary SVM classifiers produces a binary output of either +1 or -1, which is determined by the position of x relative to the hyperplane. This output is compiled as a 36-bit code for the specific instance. The output code undergoes a comparison with each codeword in the coding matrix, and the class corresponding to the codeword that exhibits the closest distance to the output code is identified as the predicted class.

To determine the nearest distance, the decoding stage uses the hamming distance D. For each class C_k (where k ranges from 1 to 9), Calculate the Hamming distance (D) between the output code and the binary code of the class. The predicted class is determined by selecting the class that exhibits the lowest Hamming distance to the output code [68].

$$C_{pred}\ =\ arg\min_{k} hamming\ distance\ (b_k\,, output\ code) \qquad (3.27)$$

Where, b_k represents binary code for class C_k .

3.5 SUMMARY

This chapter mainly focused on the methodology of the proposed model. Related to the methods, the effects of histogram equalization techniques on raw wafers are discussed in this chapter. The overall view of the distribution of wafer defect classes is illustrated by using the pie chart in this chapter. The proposed OPDCNet (modified MobileNet) is explained with the impact of the swish activation function. The multi-head attention mechanism is illustrated elaborately in this chapter. Finally, the ECOC- SVM classifier with the encoding and decoding strategy of ECOC technique are explained with block diagram.

Chapter 4: RESULT AND ANALYSIS

The performance of the proposed system is investigated in this chapter. Section 4.1 explains the experimental configuration. Section 4.2 describes the performance of the proposed system with several optimizers. Section 4.3 provides a comparative study of the original MobileNet and modified MobileNet. The effects of histogram equalization on proposed model analyses in the Section 4.4. Section 4.5 discusses the confusion chart matrix of the proposed model. The performance analysis of the proposed model through using various classifiers are briefly discussed in Section 4.6. Section 4.7 introduces a comparison of the proposed model and pre-trained models in terms of testing accuracy, total learning parameters and FLOPs. Finally, Section 4.8 compares the outcomes of the proposed system with other developed systems of the same research field.

4.1 EXPERIMENTAL CONFIGURATION

The experiment is performed using a laptop equipped with the 11th Generation Intel Core i5-1135G7 Processor. The hardware resource involves a single CPU. The MATLAB R2023a version is selected for this experiment. The proposed model and other pretrained models are implemented in this platform. The mini-batch size is set to 512, and the initial learning rate is established at 0.001. Five performance indicators—accuracy, precision, recall, AUC, and F1 score—are designed to evaluate the model's effectiveness.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{4.1}$$

$$Precision = \frac{TP}{TP+FP} \tag{4.2}$$

$$Recall = \frac{TP}{TP+FN} \tag{4.3}$$

$$F1 - Score = \frac{2*(Recall*Precision)}{Recall+Precision} \tag{4.4}$$

4.2 PERFORMANCE ANALYSIS WITH DIFFERENT OPTIMIZERS

Three optimizers are employed for functioning the proposed model: ' SGDM', 'ADAM' and 'RMSProp' optimizers. The Stochastic Gradient Descent with Momentum (SGDM) optimizer minimizes computational redundancy [71]. The momentum parameter of this optimizer facilitates the acceleration of the optimization process. The Root Mean Square Propagation (RMSProp) optimizer employs the concept of a moving average of the squared gradients across consecutive mini batches for each weight [72]. The Adaptive Moment Estimation (ADAM) optimizer integrates the properties of AdaGrad and RMSProp and requires lower memory [73]. The line graph in Fig.4.1 illustrates the comparison of the performance of the proposed hybrid system based on three different types of optimizers. The proposed system achieved the highest testing accuracy (98.55%) with the SGDM optimizer. The line graph indicates that using SGDM, the testing accuracy of the suggested hybrid technique is close in level to the training accuracy, suggesting that the system effectively resolves overfitting issues. Batch normalization is employed in the modified MobileNet architecture, assisting the proposed hybrid framework in mitigating overfitting problems.

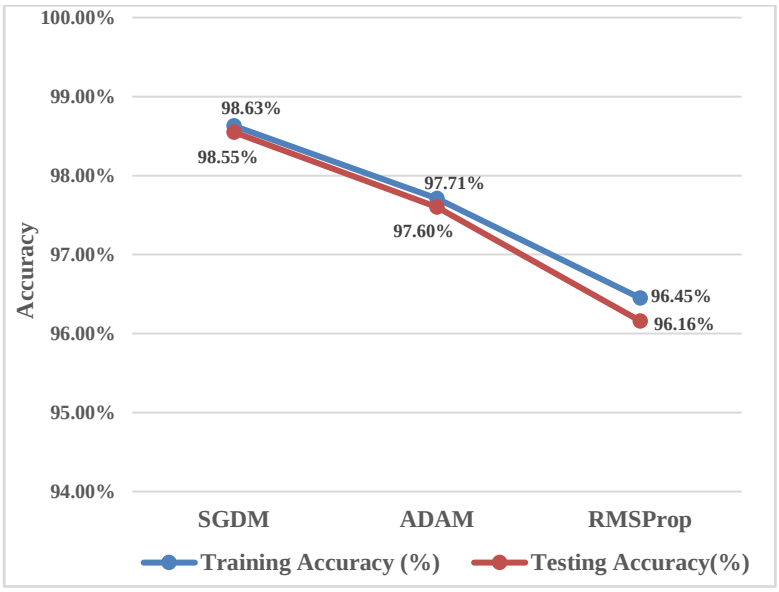


Fig. 4.1: Analysis of the proposed model with several optimizers

4.3 COMPARATIVE STUDY

In this thesis work, the performance between modified MobileNet and basic Mobile Net architecture is compared by using five performance indicators: accuracy, precision, recall, AUC, and F1 score, and provides a comparative analysis with the original Mobile Net. The area under the receiver operating characteristic (ROC) curve is known as the AUC.

The model's capacity to recognize intricate patterns in semiconductor wafer maps is demonstrated by the AUC values. The ablation study of the proposed model is included in Table 4.1. The original MobileNet model is the first model, and the suggested hybrid model with a modified MobileNet architecture is the second. In Table 4.1, the best results are shown in bold. The F1-score, precision, recall, and AUC values for eight defects and one non-defect wafer are shown in this table. This table's data shows that the suggested hybrid system attains an excellent average precision value of 0.9564. This suggests that the hybrid approach outperforms the original MobileNet in precisely locating defective areas in the semiconductor wafers compared to the original Mobile Net model, the suggested hybrid framework has a better average recall value of 0.9442, indicating that the suggested system is capable of detecting a small percentage of defects in defective wafers. The suggested hybrid model produces a higher average F1 score of 0.9334, as indicated in Table 4.1, guaranteeing precise defect class detection and lowering the probability of missed detections. Additionally, Table 4.1 shows that the suggested hybrid model outperforms the original Mobile Net version with an average AUC value of 0.9974. This indicates the suggested method is better at recognizing intricate patterns and capturing distant relationships in wafer maps. The suggested hybrid system achieves the maximum precision value (1) for near-full defects in Table 4.1, demonstrating that the model appropriately avoids false positives for defects with significant spatial distribution.

Table 4.1: Performance Analysis of Proposed Model

Models	Measure	Center	Donut	Edge - Local	Edge - Ring	Local	Near Full	None	Random	Scratch	Avg.
Original MobileNet + Histogram Equalization + Softmax	Precision	0.9310	0.9362	0.8910	0.9889	0.7007	0.9857	0.9801	0.9583	0.4321	0.8671
	F1-Score	0.9527	0.6822	0.8403	0.9895	0.6783	0.9857	0.9839	0.8440	0.4663	0.8248
	Recall	0.9754	0.5366	0.7950	0.9901	0.6574	0.9857	0.9877	0.7541	0.5063	0.7987
	AUC	0.9970	0.8758	0.9938	0.9998	0.9712	0.9857	0.9958	0.9497	0.9457	0.9683
Modified MobileNet with Swish + Histogram Equalization +ECOC-SVM	Precision	0.9587	0.9076	0.9432	0.8801	0.8828	0.9237	0.9653	0.9128	0.6300	0.8893
	F1-Score	0.9554	0.9152	0.9297	0.9042	0.8487	0.9146	0.9620	0.9226	0.6051	0.8842
	Recall	0.9522	0.9231	0.9167	0.9297	0.8358	0.9058	0.9589	0.9328	0.5821	0.8819
	AUC	0.9437	0.9558	0.9351	0.9647	0.9587	0.9583	0.9482	0.9439	0.9033	0.9457
Modified MobileNet with Swish + Histogram equalization +Multi-head Attention + Softmax	Precision	0.9656	0.8324	0.8747	0.9854	0.7754	0.8600	0.9791	0.8239	0.5970	0.8548
	F1-Score	0.9698	0.7846	0.8409	0.9869	0.7133	0.8091	0.9803	0.8864	0.4290	0.8222
	Recall	0.9740	0.7415	0.8095	0.9883	0.6605	0.7333	0.9917	0.9590	0.3347	0.7991
	AUC	0.9623	0.8445	0.9567	0.9342	0.9432	0.7945	0.9864	0.9798	0.8841	0.9206
Modified MobileNet with ReLU + Histogram equalization +Multi-head Attention + ECOC-SVM	Precision	0.9766	0.9044	0.8997	0.9148	0.8994	0.9780	0.9712	0.8819	0.6922	0.9020
	F1-Score	0.9699	0.9062	0.8974	0.9109	0.8374	0.9252	0.9741	0.8949	0.6386	0.8838
	Recall	0.9605	0.9082	0.8952	0.9072	0.7834	0.8779	0.9772	0.9083	0.5928	0.8689
	AUC	0.9439	0.9081	0.9497	0.9363	0.9262	0.9085	0.9694	0.9587	0.8611	0.9291
Modified MobileNet with Swish + Histogram equalization +Multi-head Attention + ECOC-SVM (Proposed hybrid model)	Precision	0.9857	0.9512	0.9665	0.9971	0.9159	**1.0000**	0.9917	0.9440	0.8551	**0.9564**
	F1-Score	0.9907	0.9512	0.9604	**0.9965**	0.8942	0.9474	0.9945	0.9555	0.8079	**0.9442**
	Recall	0.9957	0.9512	0.9545	**0.9959**	0.8735	0.9900	0.9073	0.9672	0.7657	**0.9334**
	AUC	0.9999	0.9999	0.9984	0.9998	0.9948	**1.0000**	0.9967	0.9998	0.9876	**0.9974**

The proposed framework appears to accurately capture the distributed dependencies of the ring, as indicated by the precision value (0.9971) for edge ring defect patterns. The model successfully captures almost all true positives, indicating that the suggested model accurately covers the long-range dependencies on the defective wafer. This is demonstrated by the highest recall value (0.9959) obtained for the edge-ring defect class. High precision, recall, and AUC across these classes illustrate the suggested model's great capacity to

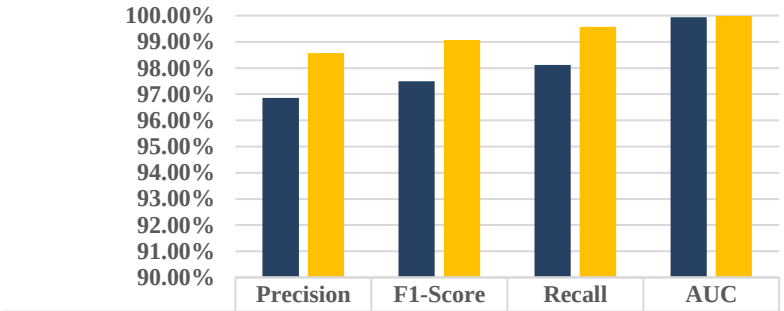
capture long-range dependencies employing multi-head attention for defects with scattered spatial patterns, as the findings above show.

The highest AUC value (1) is obtained while detecting the near-full defect class in wafer maps using the suggested model. The model's capacity to reliably differentiate between these complex defects and normal wafers is demonstrated by the near-perfect AUC values for the donut defect class (0.9999) and edge-ring defect class (0.9998). These AUC values for defect classes show that, throughout complicated and irregular spatial arrangements, the swish activation function aids the suggested model in detecting complex defect patterns. This table indicates that the proposed hybrid model with the multi-head attention mechanism outperforms the hybrid model without using this mechanism in terms of capturing long-range dependencies in donut, scratch and edge-ring defects classes. The outcomes demonstrated that the hybrid model with ECOC-SVM classifier solves the class imbalance problem by focusing on the small-sized dataset of defect classes, such as donut and near-full defect classes, unlike the softmax classifier, which focuses on majority samples defect classes, such as the non-defect class. The hybrid model with Relu activation function fails to detect intricate patterns in donut, scratch and edge ring defect classes. On the other hand, the inclusion of swish activation function enables the hybrid model to identify these complex patterns, which is proven by the results of proposed hybrid model in Table 4.1.

4.4 EFFECTS OF HISTOGRAM EQUALIZATION ON PROPOSED MODEL

The histogram equalization (HE) plays an important role in recognizing tiny defects in defective wafers, which are discussed in this section. Among of eight defect wafer types of center, edge-ring and scratch wafer defects exhibit small defects mostly. The bar chart of Fig.4.2 illustrates that the histogram equalization markedly improves the model's performance in identifying the center defect class. The results indicate an enhancement in precision from 96.86%

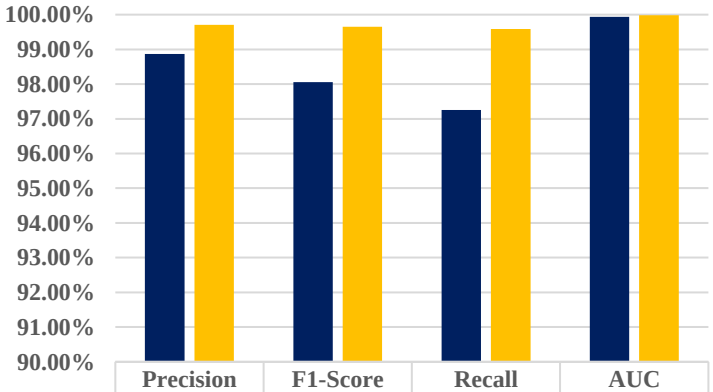
to 98.57%, an increase in F1-Score from 97.49% to 99.07%, and an improvement in recall from 98.12% to 99.57%. Additionally, there is a slight rise in AUC from



	Precision	F1-Score	Recall	AUC
Without Histogram Equalization	96.86%	97.49%	98.12%	99.94%
With Histogram Equalization	98.57%	99.07%	99.57%	99.99%

Fig. 4.2: Impact of HE on center defect class

99.94% to 99.99%. This bar chart in Fig. 4.3 demonstrates the effect of histogram equalization on the model's ability to identify the edge ring defect class in wafer maps.



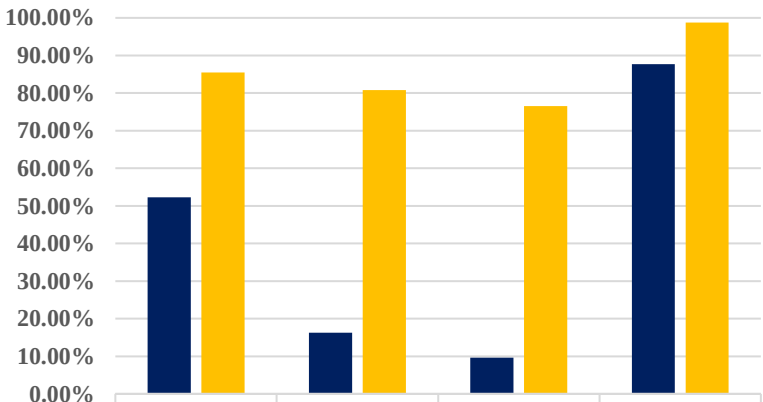
	Precision	F1-Score	Recall	AUC
Without Histogram Equalization	98.87%	98.06%	97.25%	99.94%
With Histogram Equalization	99.71%	99.65%	99.59%	99.98%

Fig. 4.3: Impact of HE on edge-ring defect class

The Precision value improved from 98.87% (without histogram equalization) to 99.71% (with histogram equalization), demonstrating enhanced accuracy in predicting true positivesThe F1-Score increased from 98.06% to 99.65%, indicating a better balance between precision and recall. The Recall value rose from 97.25% to 99.59%, indicating an improved rate of accurately

identified edge ring defects. The AUC value showed a marginal improvement from 99.94% to 99.98%.

Fig. 4.4 depicts that the proposed model can effectively identify difficult, thin defect lines of scratch defect class using the histogram equalization technique. The precision rises from 52.27% without histogram equalization to 85.51% with it, demonstrating a greater percentage of accurately identified scratch defect class. The F1-score increases from 16.25% to 80.79%, indicating an enhancement in overall performance.



	Precision	F1-Score	Recall	AUC
Without Histogram Equalization	52.27%	16.25%	9.62%	87.68%
With Histogram Equalization	85.51%	80.79%	76.57%	98.76%

Fig. 4.4: Impact of HE on scratch defect class

The recall value markedly rises from 9.62% to 76.57%, indicating that the model can identify a greater proportion of actual scratch defect types. The AUC value demonstrates a rise from 87.68% to 98.76%, reflecting an improved ability to differentiate between scratch and non-scratch defect wafers. These analyses prove that the suggested model detects tiny defects in defective wafers by employing histogram equalization correctly. In this thesis work, the overall impact of the histogram equalization technique on the performance of the proposed model is also analysed. The bar chart in Fig.4.5 analyses the effects of incorporating the histogram equalization technique on the overall performance of the proposed hybrid model. The findings demonstrate that the integration of

histogram equalization markedly improves model performance. The proposed model utilizing histogram equalization demonstrates an average precision of 95.64%, an average F1-Score of 94.42%, an average recall of 93.93%, and an exceptionally high average AUC of 99.74%.

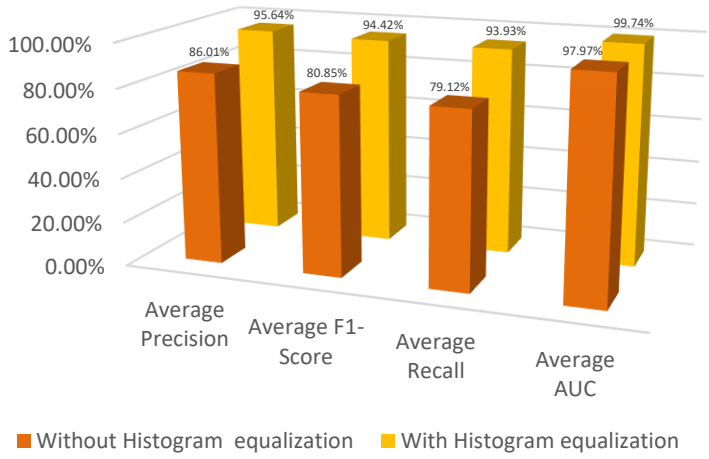


Fig. 4.5: Overall performance evaluation of the proposed model based on the HE

The proposed model without histogram equalization exhibits substantially lower performance, attaining an average precision of 86.01%, an average F1-Score of 80.85%, an average recall of 79.12%, and an average AUC of 97.97%. The findings indicate the significant impact of histogram equalization on enhancing the classification accuracy of the proposed model.

4.5 THE CONFUSION MATRIX CHART OF THE PROPOSED MODEL

The effectiveness of the proposed system is demonstrated by using the confusion chart matrix in this thesis work. Fig.4.6 illustrates the confusion matrix chart of the proposed hybrid system. This confusion chart presents the count of accurately classified defect wafer map classes along with details regarding misclassification by the proposed hybrid system. The confusion chart matrix demonstrates diagonal entries that indicate the count of accurately classified images for each defect class, whereas the off-diagonal entries reflect the misclassifications among defect classes. In the confusion chart, 689 in the starting row and the first column signifies that 689 images of the defect wafer map class

center are accurately categorized as the center defect class. This chart shows that the value 1 in the first row and the 8th column represents 1 image from the defect class center misclassified as the defect class random (false positives)

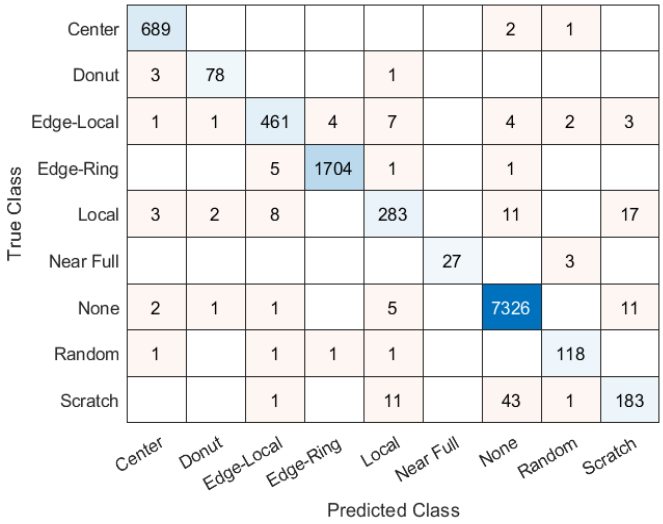


Fig. 4.6: Confusion matrix chart of the proposed hybrid model

Conversely, the 9th row and the 3rd column indicate that 1 image belonging to the edge-local class has been misclassified as part of the defect class center (false negatives). A total of 1704 samples have been accurately classified as edge-ring. The proposed model demonstrates strong performance on edge ring defects, as evidenced by the limited number of images classified as edge-local and local defect classes. In the same way, the counts of accurately identified defect classes are as follows: donut - 78, edge-local - 461, local - 283, near full - 27, none - 7326, random - 118, and scratch - 183. The confusion matrix chart illustrates a significantly higher rate of accurately classified defect classes compared to the misclassification rate. This suggests that the proposed hybrid system effectively identifies various types of wafer defects, including those with intricate patterns and tiny defects, with a notable level of precision.

4.6 PERFORMANCE EVALUATION USING DIFFERENT CLASSIFIERS

We use different classifiers for the classification of defective wafers. Table 4.2 shows the effectiveness of the proposed hybrid models with classifiers. The

hybrid model with an ECOC-SVM classifier attains the best overall precision at 95.64%, surpassing the Softmax classifier model at 88.04% and the KNN classifier model at 76.32%.

Table 4.2: Performance analysis with different classifiers

Model with Classifiers	Overall Precision	Overall Recall	Overall F1-Score	Testing Accuracy
Hybrid model with Softmax	88.04%	82.14%	84.45%	95.62%
Hybrid model with KNN	76.32%	74.52%	75.33%	93.23%
Hybrid model with ECOC-SVM	95.64%	93.34%	94.42%	98.55%

This signifies that the ECOC-SVM classifier model improves at reducing false positives, making it more dependable in cases where accurate classification is critical. The ECOC-SVM classifier demonstrates superior overall recall at 93.34%, with the softmax classifier following at 82.14%, and the KNN classifier at 74.52%. The higher recall in the ECOC-SVM classifier-based model indicates its enhanced capability to identify all positive instances, a vital aspect in defect detection in cases where missing defects can lead to severe consequences. The ECOC-SVM classifier model demonstrates superior performance, with an F1 score of 94.42%, outpacing the Softmax classier model, with an F1 score of 84.45%, and the KNN classifier model, with an F1 score of 75.33%. The superior accuracy of the ECOC-SVM model highlights its reliability and effectiveness in accurately classifying defect wafer maps.

4.7 COMPARATIVE ANALYSIS OF THE PROPOSED MODEL WITH PRE-TRAINED MODELS

Comparisons among the proposed model with various pre-trained models are analysed in this thesis work. The testing accuracy of many deep learning

models, including two Mobile Net variations, two ResNet variants, and a proposed hybrid model, is displayed in the bar chart in Fig.4.7. The suggested model outperforms MobileNet V1 (95.31%), MobileNet V2 (94.01%), ResNet-18 (96.34%), and ResNet-50 (95.04%) with the highest testing accuracy of 98.55%. The usefulness of the hybrid method in capturing intricate defect patterns and long-range dependencies is demonstrated by the fact that the suggested hybrid model performs better than all other pre-trained models included in this bar chart.

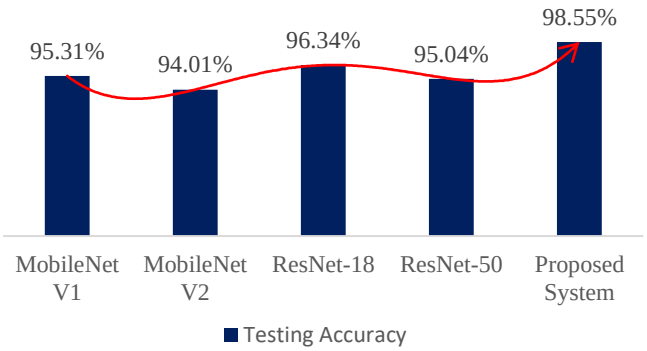


Fig. 4.7: Comparative analysis based on testing accuracy

Fig. 4.8 illustrates a bar chart that compares the total parameters (in millions) of several models. The suggested hybrid model has 1 million parameters, considerably less than the other models. The ResNet-18 and ResNet-50 models comprise 23.5 million and 11.1 million parameters, respectively.

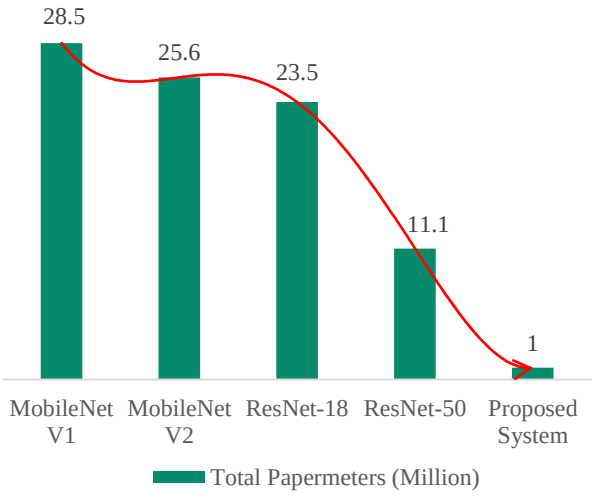


Fig. 4.8: Comparative analysis based on total parameters

MobileNet V1 possesses the greatest amount of parameters (28.5 million) compared to the other models illustrated in this bar chart. The MobileNet V2 model has the second-largest parameter count in this bar chart.

Fig.4.9 illustrates a line graph that compares the computational complexity, quantified in FLOPs (Floating Point Operations per Second), of the Proposed Hybrid Model with several pre-trained models.

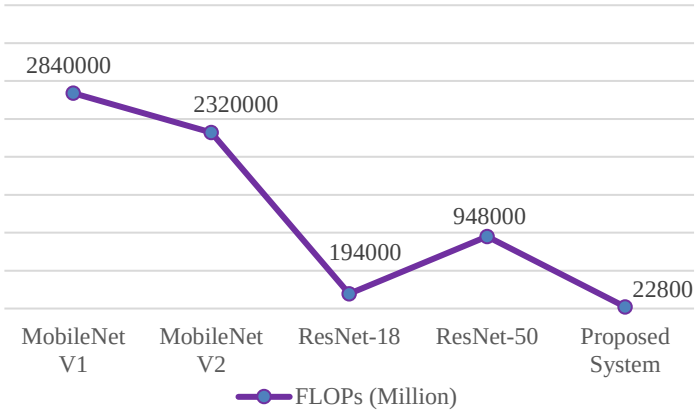


Fig. 4.9: Comparative analysis based on FLOPs

FLOPs measure the operations necessary for a single forward pass, indicating the model's computational cost. The Proposed Hybrid Model has a minimal computational complexity at 22,800 million FLOPs. ResNet-50 exhibits a considerably higher computing requirement of 948,000 FLOPs (million). ResNet-18 necessitates 194,000 million FLOPs, significantly above the proposed model albeit lower to ResNet-50. MobileNet V2 constitutes one of the most computationally demanding models, requiring 2,320,000 FLOPs (million), which indicates the expense of its operations despite optimized parameters. MobileNet V1: The model with the greatest computing need, totalling 2,840,000 FLOPs (million). This result demonstrates that the proposed hybrid model provides superior outcoms with less computational complexity and outperforms other pre-trained models in terms of measuring Flops.

4.8 COMPARATIVE ANALYSIS OF THE PROPOSED HYBRID SYSTEM WITH SEVERAL DEVELOPED SYSTEMS

We use several research works for comparative studies with the proposed system. Table 4.3 compares several defect wafer map categorization methods. All the methods presented in this Table utilize the real wafer map dataset WM811K [66].

Table 4.3: Comparison of the proposed hybrid approach with other methods based on accuracy.

Methods	Accuracy (%)
[11]	93.75 %
[14]	95.46 %
[16]	98.34%
[29]	96.96 %
Proposed method	98.55%

The results in this table demonstrate that the proposed hybrid approach achieves superior accuracy (98.55%) compared to the other methods listed. The accuracy of the lightweight network-based defective wafer classification system [16] is nearest to the proposed system's accuracy, 98.34%. An accuracy of 96.96% is attained by the system [29] that used a modified attention mechanism. Table 4.4 compares the performance of the proposed model with other developed models based on average values of precision, recall and F1-score. The proposed hybrid model outperforms the listed models in terms of all performance matrics.

Table 4.4: Comparison of the proposed hybrid approach with other methods based on average values of performance metrics.

Methods	Avg. Precision (%)	Avg. Recall (%)	Avg. F1-score (%)
[11]	93.81	93.79	93.76
[16]	93.30	89.10	90.71
Proposed method	**95.64**	**93.34**	**94.42**

Table 4.5 represents the comparison of performance metrics for defect classes between the proposed hybrid model and the developed method in [16]. This table demonstrates that the proposed model performs better than the other mentioned model [16], in the context of detecting complex patterns and capturing long-range dependencies in the scratch defect class, donut defect class, and the edge ring defect class.

Table 4.5: Comparison of the proposed hybrid approach with other system-based on classifications of different defect classes.

Methods	Measure	Center (%)	Donut (%)	Edge-Local (%)	Edge-Ring (%)	Local (%)	Near-Full (%)	None (%)	Random (%)	Scratch (%)
[16]	Precision	98.85	94.02	92.22	99.57	89.80	93.57	96.40	98.99	76.31
	Recall	98.96	93.88	89.88	99.32	81.91	92.33	95.23	100	50.40
	F1-score	98.90	93.91	90.99	99.44	85.53	92.13	95.80	99.49	60.17
Proposed method	Precision	98.57	**95.12**	96.65	**99.71**	91.59	100	99.17	94.40	**85.51**
	Recall	99.57	**95.12**	95.45	**99.59**	87.35	99.00	90.73	96.72	**76.57**
	F1-score	99.07	**95.12**	96.04	**99.65**	89.42	94.74	99.45	95.55	**80.79**

Table 4.6 shows the comparison between the proposed methods with other wafer defect detection method [29] based on the total parameters count in million. The proposed hybrid model requires only 1 million parameters to correctly detect eight defect classes and one non-defect class. The overall performance studies indicate that the proposed model detects the different types of defect patterns efficiently with lower learning parameters.

Table 4.6: Comparison of the proposed hybrid approach with another method based on total parameters.

Methods	Total parameters (Million)
[29]	28.03
Proposed method	1

The effectiveness of the proposed hybrid model is evaluated for handling imbalanced data by conducting a comparison with MobileNet, which was

trained on a balanced dataset. The proposed hybrid model surpassed MobileNet in recall and F1-score metrics. The increased recall of the hybrid model suggests that it effectively detects a larger percentage of minority defects, which are typically more challenging to identify in imbalanced datasets. The enhanced F1-score indicates that the hybrid model effectively balances the identification of defects while reducing the occurrence of false positives. This suggests that, even without dataset balancing, the proposed hybrid model with ECOC-SVM classifier exhibits superior effectiveness in identifying donut, near-full, scratch and random defects, which are a small percentage of the total wafer image dataset.

Table 4.7: Comparison of the proposed hybrid model and the base MobileNet model based on dataset balancing.

Models	Measure	Center (%)	Donut (%)	Edge-Local (%)	Edge-Ring (%)	Local (%)	Near-Full (%)	None (%)	Random (%)	Scratch (%)
Original MobileNet Model with the balanced dataset	Precision	94.15	94.12	90.13	97.88	76.72	97.37	97.11	93.73	70.22
	Recall	96.77	83.37	92.48	96.25	74.02	93.67	96.19	90.40	60.63
	F1-score	95.44	88.42	91.23	97.06	75.34	95.48	96.65	92.03	65.07
Proposed model with the imbalanced dataset	Precision	98.57	95.12	96.65	99.71	91.59	100	99.17	94.40	85.51
	Recall	99.57	**95.12**	95.45	99.59	87.35	**99.00**	90.73	**96.72**	**76.57**
	F1-score	99.07	**95.12**	96.04	99.65	89.42	**94.74**	99.45	**95.55**	**80.79**

4.8 SUMMARY

The outcomes of the proposed model and comparative studies are given in this chapter. The proposed hybrid model analyses using different optimizers. Based on recall, precision, F1-score and AUC, the proposed model is compared with the original MobileNet version. The effects of applying the histogram equalization technique in identifying wafer defects are depicted by bar charts. The correct classification rate of the proposed model is represented by the confusion chart matrix. The performance of several pre-trained models

(MobileNet V1, MobileNet V2, ResNet-18 and ResNet-50) are compared with the proposed model in terms of accuracy, learning parameters and Flops. At the end of the chapter, the comparative analysis of the proposed model with other related developed models is given.

Chapter 5: CONCLUSIONS

This chapter analyse the findings and offers ideas for future study possibilities. A summary of the thesis and an interpretation of the findings leading to a broad conclusion are given in Section 5.1. The limitations and challenges involved in developing a detection system for semiconductor wafer defects are highlighted in Section 5.2. Finally, suggestions for further study in this area are given in Section 5.3.

5.1 CONCLUSION

Due to the numerous variances and sophisticated geometries of semiconductor defective wafer maps, noise and unpredictability in the production process make it difficult to identify small defects and complex patterns. Furthermore, as defects frequently interact intricately across several spatial regions, it is critical to capture long-range dependencies inside defective wafers for accurate defect identification. This research develops a modified MobileNet-based hybrid system to classify wafer defects using an ECOC-SVM classifier, that effectively addresses the identified issues. The modified MobileNet architecture uses the swish activation function for detecting complex defect patterns. Also, the modified MobileNet incorporates a multi-head attention mechanism, enabling the model to effectively comprehend spatial relationships across considerable distances in defective wafers. The higher average AUC value (0.9974) attained by the proposed MobileNet compared to the original MobileNet and the model attains an impeccable AUC of 1 for identifying the near-full defect class that is 1.43% higher than the AUC value achieved by the original MobileNet. These investigations indicate that the proposed Modified version of MobileNet effectively detects intricate defective wafers.

The case of detection of the edge ring defect class shows a notable improvement using the modified MobileNet instead of using the original MobileNet with the F1 score rising from 0.9895 to 0.9965 and the recall value increasing from 0.9901 to 0.9959. Furthermore, the proposed model achieves the highest average F1 score and recall value, demonstrating that integrating the multi-head attention mechanism within the modified MobileNet architecture significantly enhances the ability to capture long-range relationships in defective wafers.

In proposed method, the histogram equalization technique is employed on defective wafer images to ensure precise identification of small defects. The model that includes the histogram equalization process indicated an improvement of 9.63% in average precision, 13.57% in F1 score, 14.81% in recall, and 1.77% in AUC compared to the model without the histogram equalization process. These outcomes indicate that the inclusion of the histogram equalization technique improves the ability of the proposed system to identify tiny defects.

The ECOC-SVM classifier is employed in the proposed hybrid system for classification tasks. The ECOC strategy is used in the SVM classifier model which enhances the binary classification capabilities of SVMs to deal with multiple classes efficiently. The hybrid model using ECOC-SVM classifiers surpassed the softmax and KNN classifiers-based models in every evaluation metric, attaining the highest precision, recall, F1-score, and testing accuracy. This indicates the ECOC-SVM classifier is the most robust and reliable for different defective wafer classification challenges.

The proposed hybrid model performs better than the state-of-the-art models, including MobileNet V1, MobileNet V2, ResNet-18, and ResNet-50, particularly regarding accuracy, learning parameters, and computational complexity. The suggested hybrid model requires a very small percentage of

the FLOPs in comparison to the MobileNet V1 (over 120 times less) and the MobileNet V2 (over 100 times less). The hybrid model attains a significant decrease in learning parameter count, with reductions of roughly 96.5% and 96.1% less parameters compared to MobileNet V1 and MobileNet V2, respectively. The Proposed Hybrid Model is well-optimized for computational efficiency while reducing resource consumption. Notwithstanding the significant reduction in learning parameters, it retains competitive performance (accuracy) relative to pre-trained models, demonstrating its design efficiency and suitability for environments with limited resources. Various optimizers are used to function in the proposed hybrid model, among which SGDM optimizer provides the best results. The proposed hybrid system with SGDM optimizer attains an accuracy of 98.55%, surpassing the accuracy of the technique in [11] by 4.80%, the accuracy of the method in [14] by 3.09%, the defect wafer recognition system's accuracy in [16] by 0.21%, and the wafer map defect pattern detection system's accuracy in [29] by 1.59%. The performance study and comparative analysis above demonstrated that this hybrid system efficiently handled defective wafer map identification challenges and recognized wafer defects more accurately. In addition, the proposed model outperforms the pre-trained models discussed in this study regarding classification accuracy, parameter size and computational cost.

5.2 LIMITATIONS AND CHALLENGES

Although the combination of the ML approach and DL approach provides significant improvements in identifying defective wafers, there are some limitations and challenges to applying the hybrid method to defect wafer recognition and classification which are described below:

- **Limited Dataset Diversity**: The evaluation of the proposed hybrid system was conducted using a specific dataset of semiconductor defective wafer maps. This dataset may not comprehensively capture the

diversity and complexities of defect patterns found in wafer maps across various production environments. The model's ability to generalize across multiple datasets from different semiconductor fabrication processes or industries is still questioned. A comprehensive assessment across various datasets could strengthen the system's robustness and adaptability.

- **Time Consumption:** The training process is notably time-consuming. Although the proposed system shows considerable performance enhancements, the training of the modified MobileNet, which incorporates multi-head attention, histogram equalization, and ECOC-SVM classifiers, requires substantial computational resources. The training process, particularly when dealing with large-scale datasets, demands significant computational resources and extended processing durations. This could impede the implementation of real-time or large-scale industrial applications, highlighting the need for additional optimization to accelerate training processes.
- **Limited Preprocessing Technique:** The study utilizes histogram equalization as a fixed preprocessing technique for enhancing small defects; however, it does not investigate alternative methods or adaptive thresholding, which could potentially yield further performance enhancements.
- **Hyperparameter Sensitivity:** The performance of the model could be influenced by the selection of hyperparameters, such as attention heads, SVM parameters, and optimizer settings. Conducting a more thorough hyperparameter optimization may lead to additional improvements in performance.

5.3 FUTURE STUDY

To improve the effectiveness, generalizability, and efficiency of the system, future investigations should concentrate on the following aspects:

- **Enhancement of Dataset Variety:** Future investigations should focus on gathering and synthesizing diverse datasets from a range of semiconductor fabrication processes, encompassing various wafer sizes, materials, and types of defects. An expanded and more varied dataset would enhance the system's resilience and ability to generalize across various production environments.

- **Optimizing Training Time:** Future investigations should delve into methods for enhancing training duration, including model pruning, quantization, knowledge distillation, and effective parallel processing techniques. Furthermore, exploring hardware acceleration alternatives, such as specialized GPUs, TPUs, or edge computing solutions, may enhance the capability for real-time deployment in industrial applications.

- **Analysis of Alternative and Adaptive Preprocessing Techniques:** Future research may explore adaptive preprocessing methods, including Contrast Limited Adaptive Histogram Equalization (CLAHE), wavelet-based denoising, or preprocessing techniques grounded in deep learning. These methods could enhance adaptability to diverse defect patterns and increase classification precision.

- **Improved Hyperparameter Optimization Approaches:** The effectiveness of the hybrid model is significantly influenced by the selection of hyperparameters, such as the number of attention heads, SVM kernel parameters, learning rate, and optimizer configurations. Future research

ought to integrate sophisticated hyperparameter optimization methods, including Bayesian optimization, genetic algorithms, or grid and random search, to refine the model for peak performance. Automated hyperparameter tuning frameworks can be utilized to methodically investigate various configurations.

- **Integration of Explainability and Interpretability Mechanisms:** Grasping and demonstrating the decision-making processes of deep learning models in defect classification is vital for their acceptance in industrial applications. Future work may incorporate explainability techniques like SHAP (Shapley Additive Explanations), Grad-CAM, or attention heatmaps to elucidate the model's processes for detecting and classifying defects. This would improve confidence and integration of the hybrid system in manufacturing environments.

References

[1] A. A. R. M. A. Ebayyeh and A. Mousavi, "A review and analysis of automatic optical inspection and quality monitoring methods in electronics industry," IEEE Access, vol. 8, pp. 183192–183271, 2020, doi: 10.1109/ACCESS.2020.3029362.

[2] T. R. Cass, D. Hendricks, J. Jau, H. J. Dohse, A. D. Brodie and W. D. Meisburger, "Application of the SEMSpec electron-beam inspection system to in-process defect detection on semiconductor wafers," Microelectronic Engineering, vol. 30, no. 1–4, pp. 567–570, 1996.

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

[4] Y. Ma and G. Guo, Support vector machines applications. Cham Switzerland: Springer, 2014

[5] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," Neurocomputing, vol. 408, no. 1, pp.189–215,Sep.2020,doi:https://doi.org/10.1016/j.neucom.2019.10.118.

[6] T. Cover and P. Hart, "Nearest neighbor pattern classification," in IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, January 1967, doi: 10.1109/TIT.1967.1053964.

[7] S. Zhang, "Challenges in KNN Classification," in IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 10, pp. 4663-4675, 1 Oct. 2022, doi: 10.1109/TKDE.2021.3049250

[8] D.-H. Shih, C.-Y. Yang, T.-W. Wu, and M.-H. Shih, "Investigating a Machine Learning Approach to Predicting White Pixel Defects in Wafers—A Case Study of Wafer Fabrication Plant F," Sensors, vol. 24, no.10,pp. 3144–3144, May 2024, doi: https://doi.org/10.3390/s24103144.

[9] M. Mohiuddin, M. S. Islam, and J. Uddin, "Feature Optimization for Machine Learning Based Bearing Fault Classification," Indonesian Journal of Electrical Engineering and Informatics (IJEEI), vol. 12, no. 3, Sep. 2024, doi: https://doi.org/10.52549/ijeei.v12i3.5671.

[10] M. J. Hoque et al., "Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning," in IEEE Access,vol.12,pp.47768-47786,2024,doi: 10.1109/ACCESS.2024.3383309.

[11] H. Zheng, S. W. A. Sherazi, S. H. Son, and J. Y. Lee, "A Deep Convolutional Neural Network-Based Multi-Class Image Classification for Automatic Wafer Map Failure Recognition in Semiconductor Manufacturing," Applied Sciences, vol. 11, no. 20, p. 9769, Oct. 2021

[12] Y. Byun and J.-G. Baek, "Pattern Classification for Small-Sized Defects Using Multi-Head CNN in Semiconductor Manufacturing," International Journal of Precision Engineering and Manufacturing, Aug. 2021.

[13] S. Misra, D.-G. Kim, J. Kim, W. Shin, and C. Kim, “A voting-based ensemble feature network for semiconductor wafer defect classification,” Scientific Reports, vol. 12, no. 1, Sep. 2022.

[14] 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.

[15] E. Sin and C. D. Yoo, “Efficient Convolutional Neural Networks for Semiconductor Wafer Bin Map Classification,” Sensors, vol. 23, no. 4, p. 1926, Feb. 2023.

[16] S. Chen, Y. Zhang, X. Hou, Y. Shang, and P. Yang, “Wafer map failure pattern recognition based on deep convolutional neural network,” Expert Systems with Applications, vol. 209, p. 118254, Dec. 2022, doi: https://doi.org/10.1016/j.eswa.2022.118254.

[17] M. Jeon, S. Yoo, and S.-W. Kim, “A Contactless PCBA Defect Detection Method: Convolutional Neural Networks With Thermographic Images,” IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 12, no. 3, pp. 489–501, Mar. 2022, doi: https://doi.org/10.1109/tcpmt.2022.3147319.

[18] X. Sun et al., "A Multiscale Attention Mechanism Super-Resolution Confocal Microscopy for Wafer Defect Detection," in IEEE Transactions on Automation Science and Engineering, vol. 22, pp. 1016-1027, 2025, doi: 10.1109/TASE.2024.3358693

[19] K. Limam, S. Cheema, S. Mouhoubi and F. D. Freijedo, "Deep Learning-Based Visual Recognition for Inline Defects in Production of Semiconductors," in IEEE Journal of Emerging and Selected Topics in Industrial Electronics, vol. 5, no. 1, pp. 203-211, Jan. 2024, doi: 10.1109/JESTIE.2023.3326092.

[20] M. Shahroz et al., "Hierarchical Attention Module-Based Hotspot Detection in Wafer Fabrication Using Convolutional Neural Network Model," in IEEE Access, vol. 12, pp. 92840-92855, 2024, doi: 10.1109/ACCESS.2024.3422616.

[21] G. Deng and H. Wang, “Efficient Mixed-Type Wafer Defect Pattern Recognition Based on Light-Weight Neural Network,” Micromachines, vol. 15, no. 7, p. 836, Jun. 2024, doi: https://doi.org/10.3390/mi15070836.

[22] X. Hou, M. Yi, S. Chen, M. Liu and Z. Zhu, "Recognition and Classification of Mixed Defect Pattern Wafer Map Based on Multi-Path DCNN," in IEEE Transactions on Semiconductor Manufacturing, vol. 37, no. 3, pp. 316-328, Aug. 2024, doi: 10.1109/TSM.2024.3418520.

[23] C. Ma, L. Jie, Y. Yao, T. Xu and L. Hu, "Spatial Attention Enhanced Wafer Defect Classification Algorithm for Tiny Defects," 2023 IEEE 15th International Conference on Advanced Infocomm Technology (ICAIT), Hefei, China, 2023, pp. 380-384

[24] F. López, J. L. Gómez-Sirvent, R. Morales, R. Sánchez-Reolid, and A. Fernández-Caballero, “Defect detection and classification on semiconductor wafers using two-stage geometric transformation-based data augmentation and SqueezeNet lightweight convolutional neural network,” Computers & Industrial Engineering, vol. 183, pp. 109549–109549, Sep. 2023

[25] J. Shim and S. Kang, “Learning from single-defect wafer maps to classify mixed-defect wafer maps,” Expert Systems with Applications, vol. 233, pp. 120923–120923, Jul. 2023

[26] T. Rahman and M. S. Islam, “MRI brain tumor detection and classification using parallel deep convolutional neural networks,” Measurement: Sensors, p. 100694, Feb. 2023

[27] T. Rahman, M. S. Islam, and J. Uddin, “MRI-Based Brain Tumor Classification Using a Dilated Parallel Deep Convolutional Neural Network,” Digital, vol. 4, no. 3, pp. 529–554, Sep. 2024, doi: https://doi.org/10.3390/digital4030027.

[28] S. M. M. Hossain, K. Deb, P. K. Dhar, and T. Koshiba, “Plant Leaf Disease Recognition Using Depth-Wise Separable Convolution-Based Models,” Symmetry, vol. 13, no. 3, p. 511, Mar. 2021

[29] S. Chen, M. Liu, X. Hou, Z. Zhu, Z. Huang, and T. Wang, “Wafer map defect pattern detection method based on improved attention mechanism,” Expert Systems with Applications, vol. 230, p. 120544, Nov. 2023.

[30] T. Schlosser, M. Friedrich, F. Beuth, and D. Kowerko, “Improving automated visual fault inspection for semiconductor manufacturing using a hybrid multistage system of deep neural networks,” Journal of Intelligent Manufacturing, vol. 33, no. 4, pp. 1099–1123, Jan. 2022, doi: https://doi.org/10.1007/s10845-021-01906-9.

[31] A. Biswas and Md. Saiful Islam, “A Hybrid Deep CNN-SVM Approach for Brain Tumor Classification,” Journal of Information Systems Engineering and Business Intelligence, vol. 9, no. 1, pp. 1–15, Apr. 2023

[32] P. P. Tumpa and M. S. Islam, "Lightweight Parallel Convolutional Neural Network With SVM Classifier for Satellite Imagery Classification," in IEEE Transactions on Artificial Intelligence, vol. 5, no. 11, pp. 5676-5688, Nov. 2024, doi: 10.1109/TAI.2024.3423813.

[33] S. M. Sze, Y. Li, and K. K. Ng, “Physics of Semiconductor Devices”, 4th ed. Hoboken, NJ, USA: John Wiley & Sons, 2021.

[34] S. Wolf, Silicon Processing for the VLSI Era, in LATTICE, pp. 559-581, 1995.

[35] L. M. Cook, "Chemical processes in glass polishing," Journal of Non-Crystalline Solids, vol. 120, no. 1–3, pp. 152–171, 1990, doi: 10.1016/0022-3093(90)90296-F.

[36] J. D. Plummer, Silicon VLSI Technology: Fundamentals, Practice, and Modeling, New Delhi, India: Pearson Education India, 2009.

[37] M. J. Madou, Fundamentals of Microfabrication and Nanotechnology, Three-Volume Set. CRC Press, 2018.

[38] R. Wang and S. Wang, "Similarity searching for fault diagnosis of defect patterns in wafer bin maps," Computers & Industrial Engineering, vol. 185, p. 109679, Nov. 2023, doi: 10.1016/j.cie.2023.109679.

[39] C.-W. Liu and C.-F. Chien, "An intelligent system for wafer bin map defect diagnosis: An empirical study for semiconductor manufacturing," in Engineering Applications of Artificial Intelligence, vol. 26, no. 5–6, pp. 1479–1486, May 2013, doi: 10.1016/j.engappai.2012.11.009.

[40] M. Jariya, P. Kumar, R. Devi, and B. Singh, "Silicon wafer defect pattern detection using machine learning," in Materials Today: Proceedings, May 2023, doi: 10.1016/j.matpr.2023.04.233.

[41] Z. A. Sejuti and M. S. Islam, "An Efficient Method to Classify Brain Tumor using CNN and SVM," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Jan. 2021, doi: https://doi.org/10.1109/icrest51555.2021.9331060

[42] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.

[43] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization, and detection using convolutional networks," *arXiv* preprint, arXiv:1312.6229, Dec. 2013.

[44] M. Lin, Q. Chen, and S. Yan, "Network in network," arXiv preprint arXiv:1312.4400, Dec. 2013.

[45] J. Sharma, "Experiments with SWISH activation function on MNIST dataset," Medium,October,20,2017.https://medium.com/%40jaiyamsharma/experiments-with-swish-activation-function-on-mnist-dataset-fc89a8c79ff7 (accessed Feb. 26, 2025)

[46] Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv 2017, arXiv:1704.04861.

[47] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 1800-1807, 2017 doi: 10.1109/CVPR.2017.195

[48] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, 2009, doi: 10.1109/CVPR.2009.5206848

[49] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, 2016 doi: 10.1109/CVPR.2016.90

[50] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, Oct. 2010, doi: 10.1109/TKDE.2009.191

[51] Yosinski, J., Clune, J., Bengio, Y. and Lipson, H., "How transferable are features in deep neural networks?" Advances in neural information processing systems, 27. 2014

[52] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, Massachusetts: The MIT Press, 2016. Available: https://www.deeplearningbook.org/

[53] M.Tan and Q.V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019

[54] Saini, Kirti, and Reeta Devi. "A systematic scoping review of the analysis of COVID-19 disease using chest X-ray images with deep learning models." J. Auton. Intell 7.10.32629 ,2023.

[55] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474

[56] Shahoveisi, F., Taheri Gorji, H., Shahabi, S. et al. Application of image processing and transfer learning for the detection of rust disease. Sci Rep 13, 5133 ,2023. https://doi.org/10.1038/s41598-023-31942-9

[57] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90

[58] F. Ramzan et al., "A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks," Journal of Medical Systems, vol. 44, no. 2, Dec. 2019, doi: https://doi.org/10.1007/s10916-019-1475-2

[59] Q. Ji, J. Huang, W. He, and Y. Sun, "Optimized Deep Convolutional Neural Networks for Identification of Macular Diseases from Optical Coherence Tomography Images," Algorithms, vol. 12, no. 3, p. 51, Feb. 2019, doi: https://doi.org/10.3390/a12030051

[60] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006

[61] K. P. Murphy, Machine learning: a probabilistic perspective. Cambridge (Ma): Mit Press, 2012

[62] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," The American Statistician, vol. 46, no. 3, p. 175, Aug. 1992, doi: https://doi.org/10.2307/2685209

[63] Muja, M., & Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. International Conference on Computer Vision Theory and Applications (VISAPP)

[64] Bernhard Scholkopf and A. J. Smola, Learning with kernels : support vector machines, regularization, optimization, and beyond. Cambridge, Mass.: Mit Press, 2002

[65] Nello Cristianini and J. Shawe-Taylor, Support vector machines and other kernel-based learning methods. Cambridge (Inglaterra): Cambridge, 2004. Available: https://dl.acm.org/citation.cfm?id=345662

[66] Qingyi, W. WM-811K Wafer Map. Kaggle. Available online: https://www.kaggle.com/datasets/qingyi/wm811k-wafer-map (accessed on 26th March 2025).

[67] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.

[68] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob,Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. Advances in neural information processing systems, 30, 2017.

[69] T. G. Dietterich and G. Bakiri, “Solving multiclass learning problems via error-correcting output codes,” Journal of Artificial Intelligence Research, vol. 2, pp. 263–286, Jan. 1995.

[70] T. Hastie and R. Tibshirani, “Classification by pairwise coupling,” The Annals of Statistics, vol. 26, no. 2, pp. 451–471, Apr. 1998.

[71] S. Ahlawat, A. Choudhary, A. Nayyar, S. Singh, and B. Yoon, “Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN),” Sensors, vol. 20, no. 12, p. 3344, Jun. 2020, doi: https://doi.org/10.3390/s20123344.

[72] T. Kurbiel, S. Khaleghian, “Training of deep neural networks based on distance measures using RMSProp”, Mathematics Computer Science ArXiv, 2017. doi: https://doi.org/10.48550/arXiv.1708.01911.

[73] D. P. Kingma, J. L. Ba, “ADAM: a method for stochastic optimization”, 3rd International Conference for Learning Representations San Diego, 2015, pp. 1-15. https://doi.org/10.48550/arXiv.1412.6.

thesis

ORIGINALITY REPORT

26%	20%	21%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1%
2	Submitted to University of Warwick Student Paper	1%
3	Submitted to Chittagong University of Engineering and Technology Student Paper	1%
4	arxiv.org Internet Source	1%
5	beei.org Internet Source	<1%
6	Submitted to University of Nottingham Student Paper	<1%
7	e-journal.unair.ac.id Internet Source	<1%
8	Submitted to Queen Mary and Westfield College Student Paper	<1%
9	www.cesr.ncsu.edu Internet Source	<1%
10	www.nature.com Internet Source	<1%
11	Submitted to Higher Education Commission Pakistan Student Paper	<1%
12	lib.buet.ac.bd:8080 Internet Source	<1%