

Reducing measurement error in magnetic particle inspection through the optimization of process parameters and artificially intelligent solutions

by

Sharon Lau

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Industrial Engineering

Program of Study Committee:
Frank E. Peters, Major Professor
Matthew Frank
Gül E. Kremer
David Eisenmann
Peter Collins

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2022

Copyright © Sharon Lau, 2022. All rights reserved.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
LIST OF TABLES	ix
ACKNOWLEDGMENTS	x
ABSTRACT	xi
CHAPTER 1. GENERAL INTRODUCTION	1
Introduction	1
Dissertation Organization	6
References	7
CHAPTER 2. THE DEVELOPMENT OF AN IMAGE ANALYSIS PROTOCOL TO DEFINE NOISE IN WET MAGNETIC PARTICLE INSPECTION	8
Abstract	8
Introduction	9
Standardizing Test Parameters	11
Sample preparation	12
MPI parameters	14
Image Procedure	16
Calculation of Metrics	19
<i>G</i> Calculation	19
Noise Area Percentage Calculation	19
Preliminary Evaluation of Method	20
Experiment 1 – Quantifying the collection of particles on the surface	20
Experiment 2 - Quantifying the difficulty in identifying discontinuities	22
Conclusion	25
References	26
Appendix: Statement of Authorship	27
CHAPTER 3. MEASUREMENT ERROR DUE TO HUMAN FACTORS IN WET MAGNETIC PARTICLE INSPECTION	28
Abstract	28
Introduction	29
Methods	34
R&R Calculations	34
Testing Procedure	37
Data Collection and Analysis	39
Conclusions	43
References	44
Appendix: Statement of Authorship	44

CHAPTER 4. INVESTIGATING THE INFLUENCE OF PROCESS PARAMETERS ON THE EFFECTIVENESS OF WET MAGNETIC PARTICLE INSPECTION.....	46
Abstract.....	46
Introduction	47
Impact of geometry on magnetic field strength.....	50
Impact of surface roughness, orientation, current type and magnetic field strength on the effectiveness of wet MPI.....	54
Creating Manufactured Defects.....	57
Experimental Procedure and Design	58
Software Tool.....	60
Results and Discussion.....	62
Impact of geometry on the direction of the magnetic field lines.....	65
Conclusions	70
References	71
Appendix A: Statement of Authorship	72
Appendix B: Supporting Documentation	73
Flat geometry with irregular contact points	73
Flat geometry with flush contact points	74
Curved geometry with irregular contact points.....	75
Curved geometry with regular contact points	76
CHAPTER 5. INSPECTOR AID FOR WET MAGNETIC PARTICLE INSPECTION	78
Abstract.....	78
Introduction	79
Methods	82
Grid-style labeling.....	83
Shannon’s entropy.....	84
Experiments and Results	87
Dataset.....	88
Evaluation metric	88
Training details.....	89
Design of experiments.....	90
Proposed Implementation	94
Software.....	94
Hardware	96
Conclusions	97
References	98
Appendix: Statement of Authorship.....	100
CHAPTER 6. GENERAL CONCLUSION.....	102
Summary.....	102
Recommendation and Future Research	104

LIST OF FIGURES

	Page
Figure 1-1. Simplified illustration of the sand-casting process [3].....	2
Figure 1-2. Dry MPI inspection [9].	4
Figure 2-1. The two types of castings which were used in Experiment 1	13
Figure 2-2. The sectioned casting used in Experiment 2	13
Figure 2-3. An example of classified regions with A1, A2, and A3 roughness for Experiment 1	13
Figure 2-4. An example of the manufactured discontinuity with a depth of 0.254 mm (0.01 in) and a diameter of 1.78 mm (0.07 in).....	14
Figure 2-5. Part setup using angle meter	15
Figure 2-6. Camera with modified surrounding UV Lights	16
Figure 2-7. Step by step procedure to obtain G. A) represents the marked region, B) represents the cropped image and C) represents the software tool created to analyze the cropped image. The parts represented in this image are the same parts defined in section 2.1	17
Figure 2-8. The region containing discontinuity, A, and the representative surface patch near the discontinuity, B.....	18
Figure 2-9. The smallest rectangle containing the known discontinuity.	18
Figure 2-10. The G results for A1, A2, and A3 surface classifications	22
Figure 2-11. The surface classifications on the samples.....	23
Figure 2-12. The noise area percentage results for A1, A2, A3, and A4 surface classifications	25
Figure 3-1. Wet MPI process	29
Figure 3-2. Types of variation in a system	31
Figure 3-3. Results were obtained from the same casting inspected twice by each of the two operators at Foundry 2 [1].	33

Figure 3-4. Search zone coefficient example with default values of 2.5 for the master cluster [1].	33
Figure 3-5. The blue “x” represents Operator 1 Part 1 Trial 1 and the red circles represent Operator 1 Part 1 Trial 2.....	35
Figure 3-6. The brown triangles represent the union of Operator 1 Part 1 all Trials and the green squares represent the union of Operator 2 Part 1 all Trials.	35
Figure 3-7. A) The brown triangles represent the union of Operator 1 Part 1 all Trials, green squares represent the union of Operator 2 Part 1 all Trials, and the black stars represent the union of Operator 3 Part 1 all Trials. B), C), and D) represent the resulting pairwise comparison based on A).....	37
Figure 3-8. Magnets are used to mark the beginning and end of the crack.	37
Figure 3-9. A flowchart of the process.	38
Figure 3-10. A) Operator 1 Part 1 Trial 1 Front view; B) Operator 1 Part 1 Trial 2 Front view	39
Figure 3-11. A) Operator 1 Part 1 Region A Trial 1; B) Operator 1 Part 1 Region A Trial 2.....	41
Figure 3-12. A) Operator 1 Part 1 Region A Trial 1 defect is visible and identified;	42
Figure 3-13. Indirect magnetization was performed at Foundry 3	42
Figure 3-14. Placement variability when using indirect magnetization	43
Figure 4-1. This chapter investigates the MPI testing step of the wet MPI process.....	49
Figure 4-2. (a) Labeled concave geometry part at orientation 1 and (b) Results for concave geometry	51
Figure 4-3. (a) Labeled concave geometry part at orientation 2 and (b) Results for concave geometry	52
Figure 4-4. (a) Labeled flat geometry part and (b) Results for flat geometry	53
Figure 4-5. (a) Labeled convex geometry part and (b) Results for convex geometry	53
Figure 4-6. Rectangular plate cut into 8 square samples.	54
Figure 4-7. A) 0-degree orientation between the magnetic field line direction and the crack; B) The part is rotated 15 degrees counterclockwise from the part in A), however the contact point between the copper pads and the part is shifted	

hence it is hard to determine the orientation between the magnetic field lines and the crack when rotating a square part.	55
Figure 4-8. A puck-shaped sample machined from casting plate.	56
Figure 4-9. A) 0-degree orientation between the magnetic field line direction and the crack; B) The part is rotated 15 degrees counterclockwise from the part in A, since the contact points between the copper pads and the part is the same as in A, the magnetic field direction remains the same as in A.	56
Figure 4-10. The dimensions of the drilled hole which was utilized as sub-surface defects.	57
Figure 4-11. The step by step procedure to create puck shaped samples containing surface defects.	58
Figure 4-12. The puck-shaped samples containing sub-surface and surface defects.	58
Figure 4-13. A) The smallest rectangle containing discontinuity and the representative surface patch near the discontinuity; B) Five linear points along the defect and the representative surface patch near the discontinuity.	60
Figure 4-14. A standard NAP software tool session is shown, with the most important elements of the user interface labeled.	61
Figure 4-15. Left: The results for surface-breaking defect for orientation for A1 roughness; Right: The results for surface-breaking defect for orientation for A3 roughness. ...	63
Figure 4-16. Based on the results in Figure 4-15, a surface defect might be missed at 15 degree orientation, however, if tested at right angles from the initial setup, it will be detected.	63
Figure 4-17. Left: The results for sub-surface defect for orientation for A1 roughness; Right: The results for sub-surface defect for orientation for A3 roughness.	64
Figure 4-18. 1) Flat geometry with flush contact points, 2) Curved geometry with flush contact points 3) Flat geometry with irregular contact points, 4) Curved surface with irregular contact points.	65
Figure 4-19. Transverse probe rotated along 4 axes.	66
Figure 4-20. Illustration of how the magnetic field line is determined.	66
Figure 4-21. The mapping of the magnetic field direction on the part based on magnetic field strength.	67

Figure 4-22. Left: Rectangular block setup on bench; Right: Orientation of magnetic field direction on the tested points of the rectangular block.....	68
Figure 4-23. Left: Horizontal setup with the orientation of magnetic field direction on the tested points of rectangular block with curved surfaces; Right: Vertical setup with the orientation of magnetic field direction on the tested points of rectangular block with curved surfaces	68
Figure 4-24. The orientation of magnetic field direction on the tested points of flat surface with irregular contact points.....	69
Figure 4-25. Irregular contact point illustrated in a computer aided design (CAD) drawing to show current flow	69
Figure 4-26. Left: Horizontal setup with the orientation of magnetic field direction on the tested points of curved surface with irregular contact points; Right: Vertical setup with the orientation of magnetic field direction on the tested points of curved surface with irregular contact points	69
Figure 4-27. Summary of findings with the green arrows indicating that the actual magnetic field line matches the expected whereas the red arrows indicating that the actual magnetic field line does not match the expected	70
Figure 5-1. Left: One bounding box per crack; Right: Grid style labeling (25 by 25 pixels)	84
Figure 5-2. Two graphs represent the grayscale intensities of two different images. The graph on the right is more compact which means it will have a lower Shannon's entropy compared to the image on the left.	85
Figure 5-3. MPI image color-coded to represent various percentile ranges for Shannon's entropy for a grid size of 25 pixels.....	86
Figure 5-4. Process flow for dynamic selection of grid size using Shannon's entropy	87
Figure 5-5. These images show the high variability in part geometry, illumination, crack characteristics, and image quality.....	88
Figure 5-6. Left: One bounding box per crack; Middle: Smaller bounding boxes; Right: Grid style labeling (25 by 25 pixels)	90
Figure 5-7. Type of augmenters applied.....	91
Figure 5-8. Data augmentation on the training set.....	91
Figure 5-9. Data augmentation on the test set.....	92

Figure 5-10. Bar graph of grid-style labeling experiments and data augmentation	93
Figure 5-11. Deep learning model's performance versus human inspection	93
Figure 5-12. A standard Inspector Aid software tool session is shown, with each element of the user interface labeled	95
Figure 5-13. The new session is shown, where the user can select the folder of images that they want to label using the grid-style method.....	95
Figure 5-14. Once the folder is selected, each image is loaded to the picture box to be labeled.....	96
Figure 5-15. The plot of the labeled images in sets of four	96
Figure 5-16. Live prediction hardware setup	97

LIST OF TABLES

	Page
Table 3-1. Experimental layout and R&R results for all foundries.	40
Table 4-1. The explanatory variables and the levels tested in this experiment	59

ACKNOWLEDGMENTS

I would like to start by thanking my major professor, Dr. Frank E. Peters, for taking me in and mentoring me during this journey. Next, I would like to thank my committee members, Dr. Gül E. Kremer, Dr. Matthew Frank, Dr. David J. Eisenmann, and Dr. Peter Collins for their guidance throughout my time at Iowa State University. I would like to recognize department staff: Aaron Jordan, Matthew Janowicz, Devna Popejoy-Sheriff, Krista Briley, and Nick Fetty for their support during my time here. Additionally, I would like to acknowledge my research group: Daniel Schimpf, Eric Weflen, Landon Getting, Michael Ginther, and the four undergraduate research students: Aron Mitchell, Grace Peterson, Katie Wyatt, and Colton Richardson. Thank you for all your help and encouragement in this process.

I want to take the time to thank the Steel Founders Society of America and the Defense Logistics Agency for their support and funding. Their feedback helped shape this research and was extremely beneficial in providing a clear direction during the study. A big thank you to the people at Tarleton State University, especially to my mentor, Dr. Billy Gray, who advised me to pursue studies beyond the bachelor's degree level and has been there for me since the beginning of my college studies. I am so thankful for the Fulfer family, former coach, teammates, and friends that I have met while I was studying in Stephenville, Texas. I want to take the time to thank former managers and mentors during my time at Fibergrate Composite Structures, Nike, Bunn, and Apple for giving me the opportunity to learn and grow. Everything I have learned through working in the three different industries has helped me immensely throughout my graduate studies. Lastly, this endeavor would not have been possible without my family, thank you for everything.

ABSTRACT

This dissertation will cover the evaluation of measurement error in wet magnetic particle inspection (MPI) for metal castings. Wet MPI is one of the most commonly used nondestructive test (NDT) method in the steel casting industry for detecting surface defects. However, this method is highly subjective because it relies on a human operator to locate the defect. Chapter 2 presents an objective method to quantify the effectiveness of wet MPI when detecting defects. This method is called the noise area percentage (NAP) which uses the green intensity of the crack and the surrounding area to determine relative contrast. Additionally, the NAP method was validated through the evaluation of surface roughness on the effectiveness of wet MPI. It was found that the rougher the surface, the harder it was to locate the defect.

In Chapter 3, the percent match repeatability and reproducibility (R&R) of wet MPI was investigated across four foundries and was found to be 27% and 29%, respectively. In one foundry, the defects that were identified in the first trial were found to be missing in the second trial. This led to the investigations in Chapter 4, where different process parameters were studied to identify the cause behind “disappearing” defects between trials. In Chapter 4, the orientation of the defect and the magnetic field lines, the current type, the magnetic field strength, and the surface roughness were investigated on surface breaking and subsurface defects. A modified version of the NAP metric was used to evaluate the impact of each factor on the effectiveness of wet MPI. The orientation of the defect with respect to the magnetic field lines and the surface roughness was found to play a role in the ability to find defects. For surface-breaking defects, it was found that all defects can be found by magnetizing the part in two directions. For subsurface defects, it was found that only at a 90-degree orientation to the magnetic field lines could the defect be identified.

Lastly, the human operator introduces variability due to different levels of qualifications, vision acuity, and fatigue. Chapter 2's findings show that the current visual inspection method in wet MPI equates to around a 70% escape rate. Hence, in Chapter 5, an artificially intelligent solution was developed to aid the human operator when identifying defects. The goal here is to reduce the risk of a defect going undetected which could lead to catastrophic failure in the field. This study leverages the findings from past research by using a deep learning model that has been successful in similar applications. Faster regional convolutional neural network (R-CNN) was the architecture used in this study. A novel labeling and bounding box post-processing approach was used to improve the accuracy of the model. The deep learning model achieved an escape rate of 13% which is a significant improvement from the current human visual inspection at an escape rate of 70%.

Overall, this dissertation will present findings to reduce measurement error in wet MPI through process optimization and artificially intelligent solutions.

CHAPTER 1. GENERAL INTRODUCTION

Introduction

Metal casting is a manufacturing process of forming parts from molten metal using a mold to create a near-net shape resultant piece. There are several advantages to the metal casting process. Since the molten metal will be formed based on the shape of the mold, complex external and internal geometries can be created through the casting process. Additionally, a part can be made from a variety of materials and in different sizes through this process. However, there are also several limitations to the metal casting process. The metal casting process is a near-net-shape process which means that generally, the part needs to undergo further processing to be ready for final use. It is also a labor-intensive process that requires skilled human operators to produce the casting. The amount of post-processing work will likely be dependent on the type of metal casting method used.

Metal casting methods can be broken down into 1) reusable mold and 2) expendable mold. Examples of reusable molds methods are permanent molds and die casting, while examples of expendable molds methods are sand casting, shell molding, and investment casting. Sand casting is the most commonly used metal casting method because sand is inexpensive, abundant, malleable, and can handle high temperatures [1]. The general steps of sand casting involve creating a pattern, creating the mold from the pattern, pouring the molten metal into the mold, and once the metal is solidified, the casting can be removed from the mold. Lastly, the casting will undergo finishing processes to improve the dimensional accuracy and aesthetic value of the part [2]. Figure 1-1 shows the sand-casting process in a simplified way.

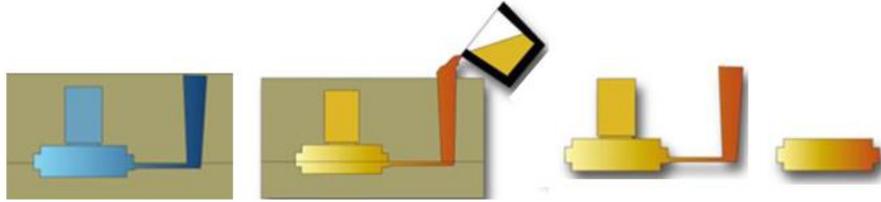


Figure 1-1. Simplified illustration of the sand-casting process [3].

Prior to delivering the final part, an inspection should be done to ensure it meets the requirements specified by the customer. [4]. The inspection of a casting can be broken down into five different groups: 1) casting finish, 2) dimensional accuracy, 3) mechanical testing, 4) chemical composition, and 5) casting soundness. The casting soundness or quality inspection is important to identify any casting defects that could result in the failure of the part.

Nondestructive test (NDT) methods are used to locate defects present in the castings to provide insight into casting quality. Common NDT methods used in the steel casting industry include visual inspection, magnetic particle inspection (MPI), liquid penetrant testing, ultrasonic testing, and radiographic testing. Compared to other NDT methods, MPI is inexpensive, fast, and easy to use, which is why it is widely used in various industries. In the steel casting industry, it is the most commonly used method to locate surface cracks. However, there is a lack of research into how reliable this method is on steel castings and how to further improve its reliability [5]. This is important because defects that escape MPI testing might result in catastrophic failures in the field.

MPI is capable of detecting surface and sub-surface defects in ferromagnetic parts through the use of magnetization. This NDT method works by first magnetizing the part by applying a current, and if a discontinuity is present, it will disrupt the magnetic field lines and result in magnetic flux leakage. Magnetic particles are applied to the part and will be attracted to the area of flux leakage resulting in a collection of magnetic particles on the surface of the part.

Human operators can identify the defect through visual inspection. Limitations of this NDT method include that it can only be used on ferromagnetic materials and sub-surface defects have to be very close to the surface to be detected [6].

MPI can be performed using either wet or dry methods, and the steel casting industry typically uses both methods to identify discontinuities. There are several advantages of wet MPI over dry MPI, including 1) good for fine surface cracks, 2) able to test large parts quickly, 3) can be automated, 4) good particle mobility on smooth surfaces, and 5) various geometries can be tested [7]. In the steel casting industry, wet MPI is typically used first since the entire part can be quickly tested. Once defects are found, they are marked and sent to get reworked to fix all the defects. Once the defects are fixed, dry MPI can be used to run localized testing in the areas that were reworked to ensure the defects are no longer there.

Dry MPI utilizes an electromagnetic yoke to induce a magnetic field in the part. The dry particles would then move to the area where flux leakage is present. Dry particles are available in many colors, shapes, and sizes. The various colors include red, black, gray, and yellow. The optimal color to select is the one that will provide the most contrast between the part and the particles. The size of the particles is another important consideration. Larger particles will be less sensitive to fine cracks but will be more effective on rougher surfaces since the particles are less likely to be caught in the surface valleys. Particle sizes typically range from 50 μm to 150 μm and are used with a mixture of sizes because different particle sizes help locate various defect sizes. Additionally, larger particles are included to lessen the dusty nature of the particles. Dry MPI is typically done on smaller parts or after wet MPI is performed to test localized areas where the defects are found. Dry MPI is also more effective on rougher surfaces due to larger particle size,

which is less likely to get caught in the surface textures [8]. Overall, dry MPI is still used for small parts or localized testing, but the wet method is the primary inspection method for MPI.

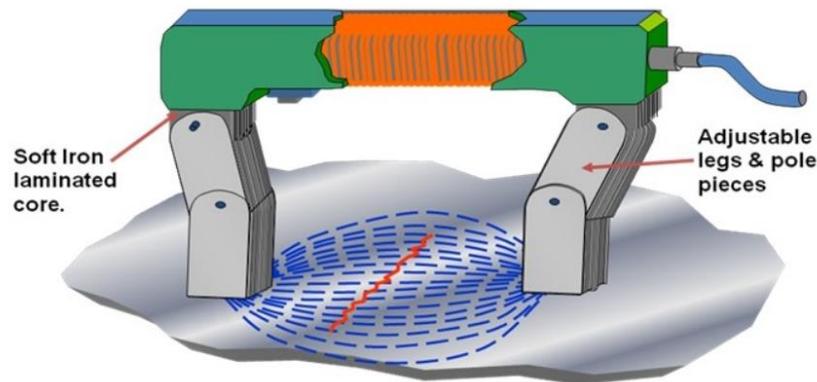


Figure 1-2. Dry MPI inspection [9].

While wet MPI is not as effective on rougher surfaces, it is more widely used in the steel casting industry because it is more sensitive to fine cracks and is able to test large casting efficiently. Wet MPI utilizes a stationary horizontal bench that can magnetize the part using two different methods, direct and indirect. Direct magnetization or head shot is performed by clamping the part between two copper pads and subsequently applying current to generate a magnetic field. Indirect magnetization or coil shot is performed by placing the part on the coil, and current is applied to generate a magnetic field in the coil. The part is indirectly magnetized since it is in close proximity to the coil. While the part is being magnetized, fluid containing magnetic particles is then applied to the part. Fluorescent particles are used in wet MPI because they can provide good contrast. Particle size is another important consideration. Particle size for wet MPI is smaller than dry MPI because particles in wet MPI have more mobility because of the use of fluid making it less likely for the particles to be caught in the surface valleys compared to dry particles. The average particle size in wet MPI is around 10 μ m and is used with a mixture of sizes because different particle sizes help locate various defect sizes [10]. Wet MPI is

typically the test that dictates which defects get reworked in the steel casting industry, which makes this a crucial process to investigate.

Many aspects of the quality of the steel casting can be evaluated using wet MPI. Undetected defects could cause catastrophic failure of the casting in the field. To mitigate this risk, the goal of this research is to reduce the escape rate of defects caused by process variability. To do this, the effectiveness of wet MPI would need to be quantifiable. Chapter 2 covers the work that was done to develop a metric that quantifies the level of difficulty in identifying a defect. Additionally, this method was validated by evaluating the effect of surface roughness on the effectiveness of wet MPI.

Variability in wet MPI can come from several factors such as human error, procedure, equipment, and part. Human error can impact the process in several ways. Human operators working with MPI inspection typically require certification through the American Society of Nondestructive Testing (ASNT), where three levels are available. Operators with Level 1 or Level 2 certifications can run MPI inspections with basic knowledge of the process and principles. The responsibilities of a Level 3 operator include training Level 1 and Level 2 operators and writing the procedure for a new part. Human error in wet MPI can stem from Level 1 and 2 operators missing defects during visual inspection of the part or improper procedure setup by a Level 3 operator. The variability caused by human error can be captured through an attribute gauge R&R study which will be covered in Chapter 3.

Procedure variation is another main factor contributing to measurement error in wet MPI. Procedure variations such as bench preparation, orientation, magnetic field strength, current type, and magnetization type can reduce the effectiveness of wet MPI. A clear understanding of the optimal parameters to set up and test a part is needed to increase the chances of finding

discontinuities. The characteristics of the part being tested are known to have an impact on the measurement error. Hence, it is important to understand the impact of the part characteristics such as surface roughness, geometry, and alloy on the effectiveness of wet MPI to increase the chances of locating all the defects on the part. Chapter 4 will cover the investigation of measurement error caused by the procedure and part characteristics in wet MPI. The equipment and consumables could also contribute to variability in MPI. The bench, particles, carrier fluid, and lighting need to be checked for functionality and compatibility with the application. In this investigation, the equipment and consumables adhered to the American Society for Testing and Materials (ASTM) 1444/3024.

Wet magnetic particle inspection (MPI) relies on a human operator to do a visual inspection after the wet MPI process to identify defects. The human operator introduces variability due to different levels of qualifications, vision acuity, and fatigue. Smaller or faint defects are harder to locate compared to larger defects. In the steel casting industry, the defects that are identified will be ground and welded to meet the quality specifications of the customer. Defects that go undetected during MPI could result in returns from the buyer or failure in the field. In Chapter 3, there were several instances in the percent match R&R study where defects were visible in testing but not identified by the human operator, equating to approximately 70% escape rate of defects. Hence, Chapter 5 aims to reduce the escape rate of defects by providing a system to prompt the operator to the location of the defects.

Dissertation Organization

The following chapters will cover four topics revolving around reducing measurement error in wet MPI. Chapter 2 will cover the development and validation of a metric to quantify the effectiveness of wet MPI. Additionally, the effect of surface roughness on wet MPI was evaluated using the proposed metric. Chapter 3 will cover the evaluation of measurement error

due to the human operator through percent match gauge R&R testing. Chapter 4 will cover the investigation of measurement error due to process and part characteristics in wet MPI. Chapter 5 will cover a proposed solution to help the human operator locate defects. Lastly, Chapter 6 will summarize the findings, provide insight on how the findings can impact the steel foundries to improve the effectiveness of wet MPI and discuss the potential for future work.

References

- [1] Reliance Foundry, “Metal Casting Methods.” [Online]. Available: <https://www.reliance-foundry.com/blog/metal-casting-methods>.
- [2] J. Cook, “Introduction to Metalcasting,” 2019. [Online]. Available: <https://blog.eaglegroupmanufacturers.com/introduction-to-metalcasting>.
- [3] D. Olsen, “What Is Sand Casting And How Does It Work?” [Online]. Available: <https://www.metaltex.com/blog/what-is-sand-casting/>.
- [4] ClubTechnical, “Casting Definition, Advantages, Limitations, Defects, Types,” 2019. [Online]. Available: <https://clubtechnical.com/casting>.
- [5] A. Zolfaghari, A. Zolfaghari, and F. Kolahan, “Reliability and sensitivity of magnetic particle nondestructive testing in detecting the surface cracks of welded components,” *Nondestruct. Test. Eval.*, vol. 33, no. 3, pp. 290–300, 2018.
- [6] “Casting Inspection Methods 101,” 2019. [Online]. Available: <https://www.castingsource.com/articles/2019/04/04/casting-inspection-methods-101>.
- [7] “Wet Suspension Inspection,” 2021. [Online]. Available: <https://www.nde-ed.org/NDETechniques/MagParticle/TestingPractices/WetSuspension.xhtml>.
- [8] “Dry Particle Inspection,” 2021. [Online]. Available: <https://www.nde-ed.org/NDETechniques/MagParticle/TestingPractices/DryParticle.xhtml>.
- [9] Mathers Gene, “Nondestructive Evaluation (NDE) Part 1 Liquid Penetrant and Magnetic Particle Inspection.” [Online]. Available: <https://www.twi-global.com/technical-knowledge/job-knowledge/non-destructive-examination-nde-part-1-liquid-penetrant-and-magnetic-particle-inspection-122>.
- [10] “Magnetic Particles,” 2021. [Online]. Available: <https://www.nde-ed.org/NDETechniques/MagParticle/Equipment/Particles.xhtml>.

CHAPTER 2. THE DEVELOPMENT OF AN IMAGE ANALYSIS PROTOCOL TO DEFINE NOISE IN WET MAGNETIC PARTICLE INSPECTION

Sharon May Yen Lau*, David Eisenmann**, Frank Earl Peters*

*Department of Industrial and Manufacturing Systems Engineering, Iowa State University

**Department of Agricultural and Biosystems Engineering, Iowa State University

Modified from a manuscript published in *International Journal of Metal Casting*

Based on my master's degree work

Abstract

This study presents a novel method to quantify the effectiveness of wet magnetic particle inspection (MPI) when detecting possible defects. Wet MPI is an established method utilizing magnetic fields to locate possible areas of defects in ferromagnetic parts. The capability of this method has been evaluated in the past, most notably using the probability of detection (POD) graphs. However, MPI requires a large amount of data and is subjective because it is based on human operators' evaluations. The method proposed in this paper is an objective method to determine the effectiveness of the MPI test based on how well discontinuities can be delineated in the image. This approach utilizes the intensity of the particle illumination in the defect area and compares it to its surroundings. This analysis generates a value to represent how well a discontinuity can be identified objectively. This method was then used to validate the effect of surface roughness on the effectiveness of wet MPI using two experiments. The first experiment was conducted to test for the collection of particles on varying surface roughness levels, and the second experiment was used to evaluate the effect of surface roughness when detecting a subsurface discontinuity. Results indicate that there is a significant increase in particle collection

as roughness increases, and as the surface roughness increases, the harder it is to locate discontinuities. This method provides a quantitative measure that could be used to aid parameter selection.

Introduction

Many aspects of the quality of steel castings can be evaluated through the use of nondestructive testing (NDT) techniques. Magnetic particle inspection (MPI) is one of these NDT methods commonly used for ferromagnetic materials because of its simplicity, speed of testing, and ability to identify the presence at or near the casting surface for indications such as cracks, oxides, and other variations within the material [1]. MPI is a combination of two methods: 1) magnetic flux leakage (MFL) and 2) visual inspection. MFL will occur if the discontinuity is on the surface or close enough to the surface. The magnetic particles will be attracted to the MFL area on the surface above the discontinuity [2]. A human inspector will then visually inspect the part to look for any collection of particles. A false positive in this paper refers to the identification of a defect that is not present in the part, while true discontinuity refers to defects that are present in the part. Many factors could lead to the human inspector not being able to discern a discontinuity, such as particles collecting in the surface texture that hinder the inspector from locating true discontinuities. Additionally, particle collection in the valley of the surface may result in bright areas that create noise that makes it difficult for the human operator to see discontinuities. This study focuses on wet MPI because the particle sizes used in wet MPI are smaller than dry MPI [3]; smaller particles will tend to adhere to the surface textures more readily than larger particles.

The magnetic particles are suspended in the carrier (either oil or water) to allow for mobility of the particles when applied to the part. A horizontal stationary magnetic particle

machine is used to test various sized parts with two different types of magnetization methods. Direct magnetization, which creates a circular magnetic field around the part, is used in this study because it has a more consistent magnetic field strength since the current is directly induced into the part [4]. The advantages of wet MPI versus dry MPI include 1) a more uniform layer of particles, 2) higher particle mobility, 3) higher sensitivity to fine surface discontinuities, and 4) the use of fluorescent particles [5]. However, wet MPI is potentially less effective compared to dry MPI when testing rougher surfaces because the smaller particles tend to adhere to the valleys of the surface. This phenomenon is known to reduce the visual contrast between the discontinuity and its surrounding area, which may deter the inspector from finding true discontinuities on the part. This has led to the need for this study. Per ASTM E3024, “The surface of the part to be examined shall be essentially smooth, clean, dry, and free of oil, scale, machining marks, or other contaminants or conditions that might interfere with the efficiency of the examination,” suggesting that the surface texture of a part may affect the reliability of the test. Reliability in this paper refers to the effectiveness of the test in finding true discontinuities. There has not been any quantifiable research published looking at surface roughness as a factor.

Currently, probability of detection (POD) is a method used to determine the capability of an NDT process. POD is a mathematical model that predicts the likelihood that the inspector will detect a discontinuity of a certain size. In most cases, a 90% POD value is required, which is a mathematical estimate of the defect size that a highly skilled inspector can detect 9 out of 10 times. There is also a 95% confidence bound associated with the POD value, which provides information about the variability of the POD experiment. The POD model values allow for inspection strategies to be optimized and the quantification of inspection systems and abilities of inspectors [6]. However, this process is time-consuming due to the amount of testing required.

Hence, POD trials are typically not conducted for individual NDT procedures; rather, it is done to assess the limitation of the technique [7].

With the new method proposed in this paper, values are objective and determined by the illumination of the particles on the part. This method can provide an objective value for the level of difficulty in identifying a discontinuity which will be helpful for quick identification of the best parameters for a specific part procedure. In this paper, the methods are explained, and experiments were conducted to validate each method. There are two metrics introduced in this study: 1) average G value (\bar{G}) and 2) the noise area percentage. \bar{G} is a metric used to measure the average green intensity of an area in the image. The noise area percentage quantifies the level of difficulty in finding a discontinuity. Noise area percentage can be used to determine wet MPI capability for specific surface textures. In this study, the noise area percentage was determined by the percentage of pixels surrounding the discontinuity, which have higher green intensity compared to the average green intensity of the discontinuity itself.

Standardizing Test Parameters

To reliably determine the impact of specific variables on the MPI process, a detailed test procedure was developed. For example, when using the \bar{G} metric, if images of an MPI test were taken with a camera at different proximities, then the intensity of the image would be affected by the difference in distance from the light source rather than factors that are being investigated (e.g., depth, diameter, surface roughness, MPI parameters). This section outlines the experimental setup for repeatable results and the calculation of the metrics employed in the proposed method.

Sample preparation

The samples used for Experiments 1 and 2 are as shown in Figure 2-1 and Figure 2-2, respectively. For Experiment 1, the low alloy steel castings were produced using normal commercial methods including chemically bonded sand molds. For Experiment 2, the samples were cast using commercial chemically bonded sand process and were made of WCB cast alloy (cast equivalent to 1020). The surface roughness of the samples for Experiment 2 was developed by scratching the sand mold before it was fully cured. The American Society for Testing and Materials (ASTM) A802 standard was used to classify the surface roughness. For the \bar{G} , the area of interest was marked in the shape of a rectangle with dimensions of 20 mm (0.8 in) by 50 mm (2.0 in) using a permanent marker. The ASTM A802 surface roughness classification is then written above the marked area of the as-cast surface as shown in Figure 2-3.

For Experiment 2, discontinuities need to be present in the part with a known location to validate the noise area percentage metric. Holes were drilled, parallel to the surface being assessed, as manufactured discontinuities for this study. The dimensions of the manufactured defect were at a depth of 0.254 mm (0.01 in) and a diameter of 1.78 mm (0.07 in). The diameter of the manufactured discontinuity followed the Aerospace Standard (AS) 5282 tool steel ring specimen which is a system performance measuring device containing drilled holes at a diameter of 1.78 mm (0.07 in) and varying depths. The depth of the manufactured discontinuity was drilled as close to the surface as possible without breaking the surface hence a distance of 0.254 mm (0.01 in) from the surface to be inspected was used, Figure 2-4.



Figure 2-1. The two types of castings which were used in Experiment 1

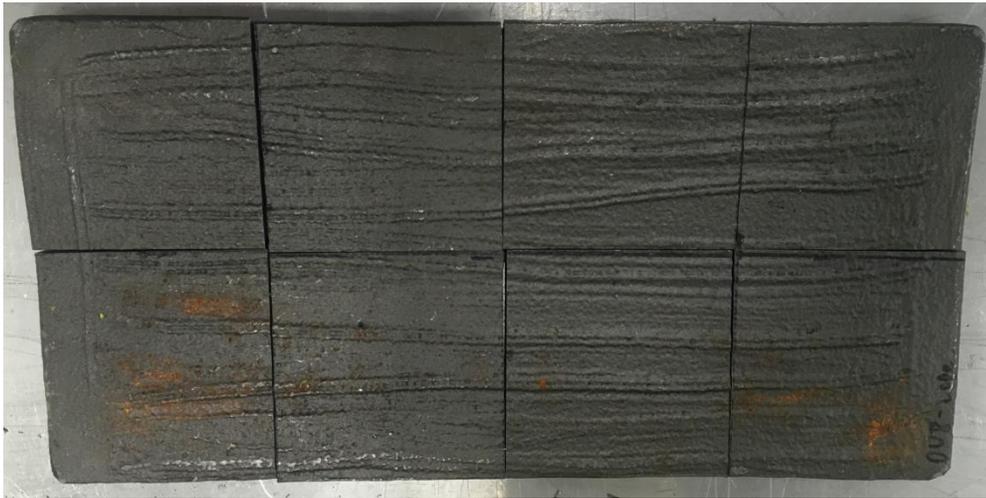


Figure 2-2. The sectioned casting used in Experiment 2



Figure 2-3. An example of classified regions with A1, A2, and A3 roughness for Experiment 1

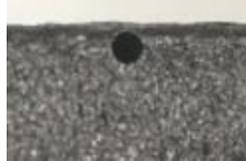


Figure 2-4. An example of the manufactured discontinuity with a depth of 0.254 mm (0.01 in) and a diameter of 1.78 mm (0.07 in)

MPI parameters

After the horizontal stationary magnetic particle machine (MD3-2060, Magnaflux®) was switched on, the sides of the reservoir were scrubbed, and then was left to run at least 30 minutes to ensure the particles (CAS# 1309-37-1, Magnaflux®) were evenly circulating through the system [8]. A 100 ml graduated centrifuge tube (14-A, Magnaflux, Illinois) was used to monitor the particle concentration, particle condition, and suspension contamination. The mean particle size was 6 μm with an estimated level of indication detection of 8 to 9 on an AS 5282 tool steel ring specimen as defined in the ASTM 1444 and a recommended settling volume of 0.1 to 0.4 mL in a 100 ml graduated centrifuge tube. Since particle collection on the surface texture is being investigated in this study, the particle concentration was kept at a level between 0.3% to 0.4%, and the percentage of suspension contamination was kept below 50% per the ASTM E3024 standard. These levels were adjusted by adding particles or oil when the particle concentration or the suspension contamination moved out of the desired range.

An angle meter (700, Johnson, Wisconsin) was used to ensure the part was at a minimum of 45 degrees when mounted on the bench as specified by the ASTM E3024 standard, as shown in Figure 2-5. It is important that the part was handled carefully because excessive handling of the part may cause residual particles from the gloved hand to be imparted to the part. If the same part was retested, a stainless-steel brush (54022SP, Osborn, Indiana) was used to remove particles

from previous testing or handling. After the part was mounted and scrubbed, a Gaussmeter (5180, FW Bell, Oregon) was used to ensure the magnetic field strength was within 30 (± 1) gauss as per ASTM 3024. For the experiments containing no discontinuities, alternating current (AC) magnetization was used because of its ability to detect surface-breaking discontinuities. For the experiments containing manufactured sub-surface defects, direct current (DC) magnetization was used because of its ability to detect sub-surface discontinuities [9].

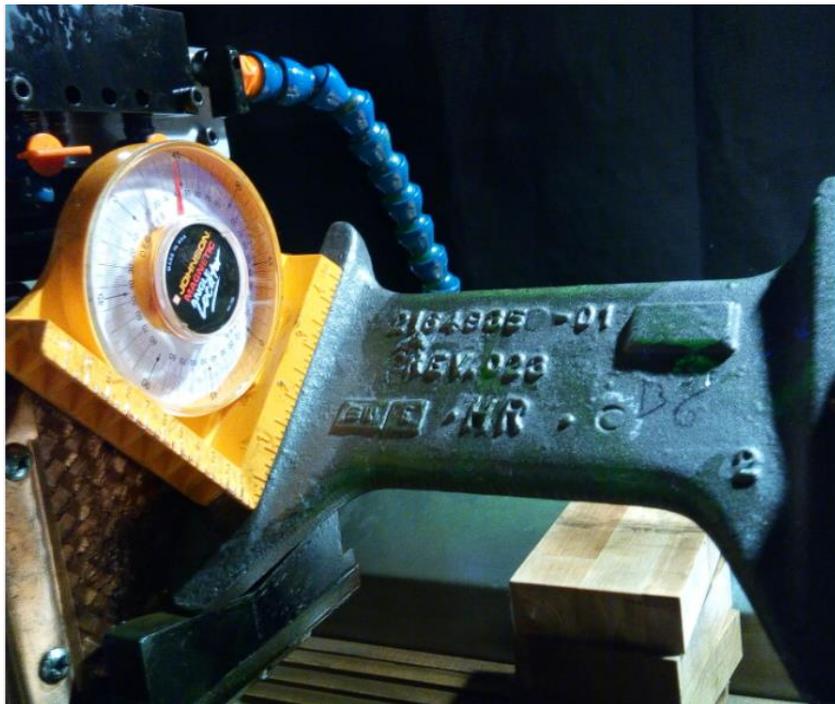


Figure 2-5. Part setup using angle meter

The current was induced into the part three times with fluid running over the surface of the part and copper pads. Images were then captured under ultraviolet light (UV) with a camera (PX-45, Crack Check, China) at a distance of 508 mm (20 in) from the surface of the part to the end of the camera for the experiments with no discontinuities, and at a distance of 305 mm (12 in) for the experiments with discontinuities. The distance was changed for the experiment with discontinuities because of the resolution of the discontinuity was not sufficient at the farther

distance. The two experiments were not compared to one another rather two distinct experiments to test each metric. The three UV lights integrated with the camera created a visible bright spot which made it difficult to detect discontinuities due to its high intensity. Hence, diffusing filter (see Figure 2-6) was used to disperse the bright spot and create even illumination of the surface. The camera was set to ISO 200 to standardize the camera's sensitivity to light.



Figure 2-6. Camera with modified surrounding UV Lights

Image Procedure

After all the samples were tested, image processing was utilized. The image with the best quality which was determined subjectively by the author was selected for each setting since multiple images were taken. The selected images were then cropped, leaving only the region within the 20 mm (0.8 in) by 50 mm (2.0 in) area for the \bar{G} to be computed. Then, the image was loaded to the software created by the author. This process is illustrated in Figure 2-7.

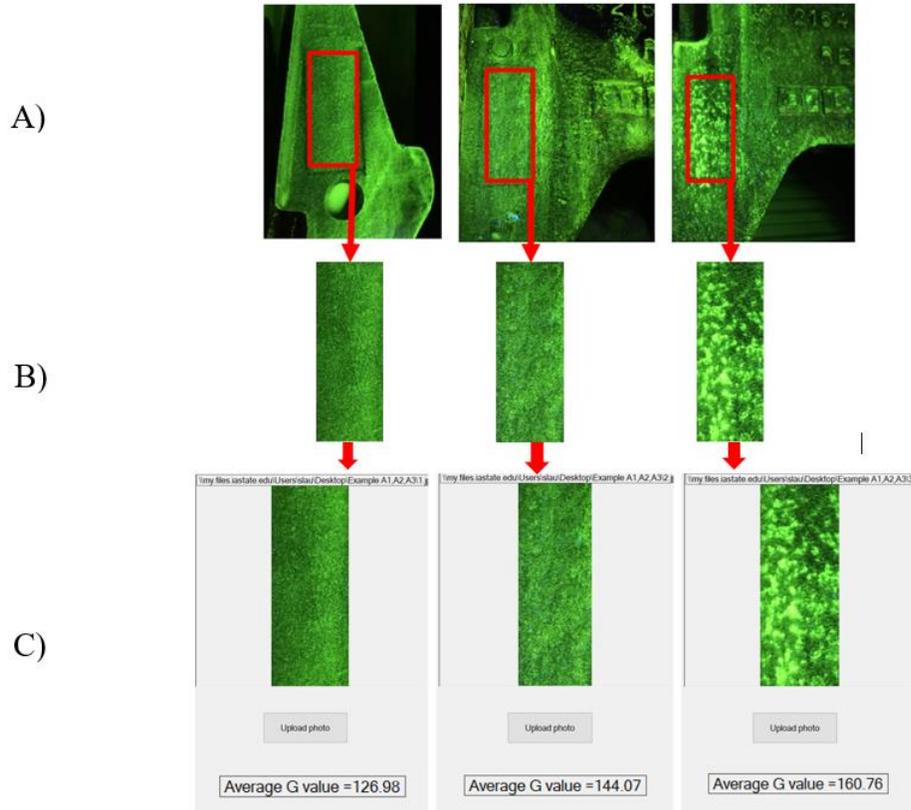


Figure 2-7. Step by step procedure to obtain \bar{G} . A) represents the marked region, B) represents the cropped image and C) represents the software tool created to analyze the cropped image. The parts represented in this image are the same parts defined in section 2.1

The noise area percentage metric was utilized for experiments containing known discontinuities. Two regions of the image would need to be cropped: the smallest rectangle encompassing the discontinuity (see Figure 2-8 A and Figure 2-9), and a representative surface patch near the discontinuity (see Figure 2-8 B). The area near the discontinuity should have the same surface roughness classification as the area of the discontinuity and should be selected as close to the discontinuity as possible. If the discontinuity is visible, then the smallest rectangle encompassing the discontinuity is manually drawn by selecting two points in the image to denote the minimum and maximum of the rectangle. If the discontinuity is not visible, then the software will automatically draw the smallest rectangle encompassing the discontinuity based on where the discontinuity is located in the part. The two cropped images will then be processed by the

software created by the author to obtain the noise area percentage. The software is designed to first analyze the smallest rectangle encompassing the discontinuity to obtain the \bar{G} associated with the image. Next, the software compares the \bar{G} of the smallest rectangle encompassing the discontinuity to the representative surface patch near the discontinuity. The noise area percentage is computed based on comparing the threshold set by the average green intensity of the discontinuity to its surrounding area.

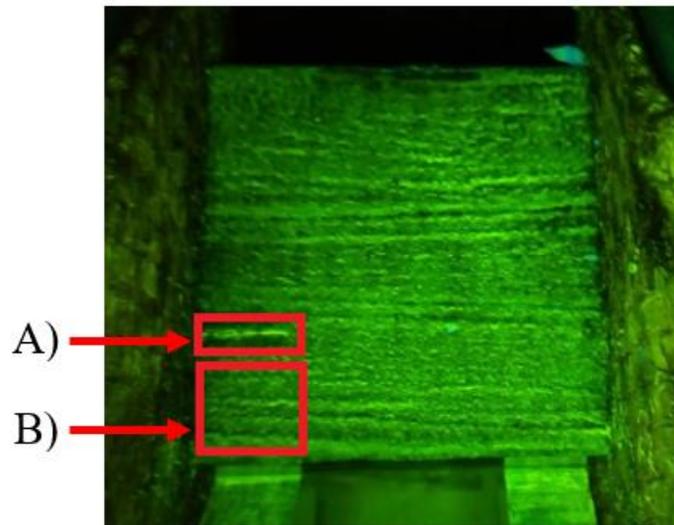


Figure 2-8. The region containing discontinuity, A, and the representative surface patch near the discontinuity, B.

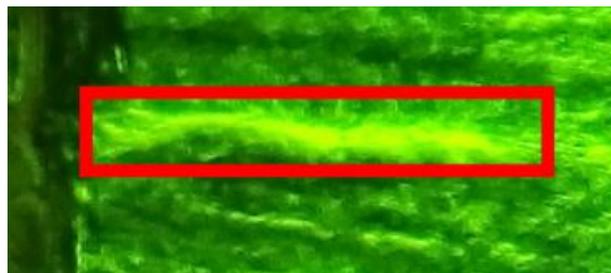


Figure 2-9. The smallest rectangle containing the known discontinuity.

Calculation of Metrics

\bar{G} Calculation

The average G value or \bar{G} is a metric developed to quantify the collection of particles on the surface of a part. The higher the \bar{G} , the higher accumulation of fluorescent particles on the surface of the part which may lead to false positives or create noise which may prevent the human inspector from finding the true discontinuities. The \bar{G} is calculated by summing the green values of all the pixels in the image and dividing it with the total number of pixels in the image.

The formula is:

$$\bar{G} = \frac{\sum_{i=1}^n Gi}{n}$$

where Gi represents the green value of the i^{th} pixel from the RGB scale, and n represents the total number of pixels in the image. The Gi value can range from 0 to 255.

Noise Area Percentage Calculation

The noise area percentage is a metric developed to quantify the noise surrounding discontinuities which could result in false positives. The higher the noise area percentage, the more difficult it would be for a human operator to identify the true discontinuities. To measure the noise area percentage, the \bar{G} of the discontinuity was first calculated and this value is set as the threshold, t . Each pixel in the representative surface patch near the discontinuity is then compared to the threshold value. The noise area percentage is the percentage of pixels in the representative surface patch near the discontinuity that is higher than the threshold.

$$Pi = \begin{cases} 0, & Gi \leq t \\ 1, & Gi > t \end{cases}$$

$$\text{Noise Area Percentage, \%} = \frac{\sum_{i=1}^n Pi}{n} * 100$$

where G_i represents the green value of the i^{th} pixel from the RGB scale, t represents the threshold, and n represents the total number of pixels in the image.

Preliminary Evaluation of Method

Experiment 1 – Quantifying the collection of particles on the surface

There were 10 samples of each part type shown in Figure 2-1 used for Experiment 1. The surface roughness on the part was classified by two different subjects using the ASTM A802 standard; the surfaces were determined to be A1, A2, and A3. If there were disagreements in classification between the two operators, the surface would need to be reevaluated until an agreement is reached. Experiment 1 was conducted to quantify the collection of particles in the surface textures through the intensity of the green color in the image after MPI testing. The higher the \bar{G} , the more particles got caught in the valleys of the surface which could result in higher rates of false positives or increased difficulty in identifying true discontinuities. The values from the image analysis software were processed and analyzed using R (Boston, MA, USA). The model's assumptions of normality, constant variance, and independence were checked using diagnostic plots. The analysis using the scale location diagnostics plot identified that there was increasing variance in the data as the response variable, \bar{G} increased, hence a logarithmic scale was used to transform the data to meet the model's assumption of constant variance. The results are as shown in Figure 2-10 where the logarithm of the \bar{G} was plotted for surface roughness levels A1, A2, and A3. A one-way ANOVA was conducted to compare the effect of the different surface classifications on the \bar{G} . There was a significant difference found between the \bar{G} for all three surface roughness levels when evaluated at a significance level of 0.05; $F = 106.9$; $p = 1.48 \times 10^{-13}$. The estimated marginal means of the \bar{G} for A1, A2, and A3

are 26 (95% CI [22, 31]), 88 (95% CI [74, 105]), and 148 (95% CI [124, 176]), respectively. The estimated marginal means is the mean of the \bar{G} that is adjusted based on the model. However, since the surface roughness classification is a categorical variable, the estimated marginal mean is the same as the mean from the data. Additionally, a Tukey multiple comparisons test was performed between the surface classification levels. There was a significant difference in the multiplicative effect on the mean from A1 to A2 and from A1 to A3 when analyzed at a significance level of 0.05. The multiplicative effect on the median from A1 to A2 is 3 (95% CI [2, 5]) for the \bar{G} . On the other hand, the multiplicative effect on the median from A1 to A3 is 6 (95% CI [4, 8]) for the \bar{G} . The multiplicative effect is used to explain the comparison between the levels A1 to A2 and A1 to A3 because the data for the response variable was in logarithm form. An example of how the multiplicative effect is interpreted is considering the median of the \bar{G} for A1 is 25, the multiplicative effect of A1 to A2 is 3, hence the median of the \bar{G} for A2 is $25 \times 3 = 75$. Likewise, the multiplicative effect of A1 to A3 is 6, hence the median of the \bar{G} for A3 is $25 \times 6 = 150$.

The results in Experiment 1 support the findings in past research that the wet MPI method is not as effective when performed on parts with rougher surfaces [10]. In this experiment, the results show that rougher surfaces led to a higher collection of particles in the valleys of the surface which leads to higher UV intensity emitted from the part surface to the camera. Although the \bar{G} can quantify the collection of particles on the surface texture, it is not indicative of the difficulty in finding a discontinuity.

This is because when a discontinuity is present, MFL occurs which may attract the particles from the area surrounding the discontinuity. Hence, to appropriately quantify the difficulty of visually identifying a discontinuity, the noise area percentage is used. In Experiment 2, parts with manufactured discontinuities were used to evaluate the noise area percentage metric.

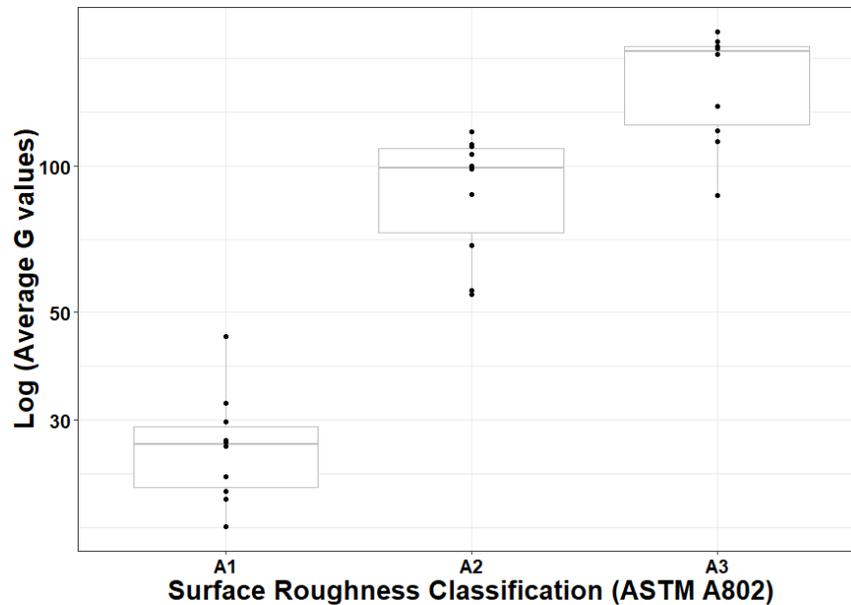


Figure 2-10. The \bar{G} results for A1, A2, and A3 surface classifications

Experiment 2 - Quantifying the difficulty in identifying discontinuities

In Experiment 2, holes were drilled into the part to create discontinuities to be identified through the inspection process. A rectangular cast steel plate with dimensions of 356 by 175 mm (14.00 by 6.875 in) was cut into eight pieces as shown in Figure 2-2. The surface roughness was generated by adding purposefully scratching the mold. The surface roughness was then classified, and surface roughness levels A1, A2, A3, and A4 were found on the samples per ASTM A802. The surface classifications for each region are shown in Figure 2-11. Only three of the eight parts contained all four surface classifications, A1 to A4. Hence, these three pieces

were utilized in this experiment to test the influence of surface roughness on noise area percentage when detecting a sub-surface discontinuity. The manufactured discontinuity was drilled at a diameter of 1.78 mm (0.07 in) and a depth from the surface of 0.254 mm (0.01 in). Figure 2-4 shows an example of a manufactured discontinuity.

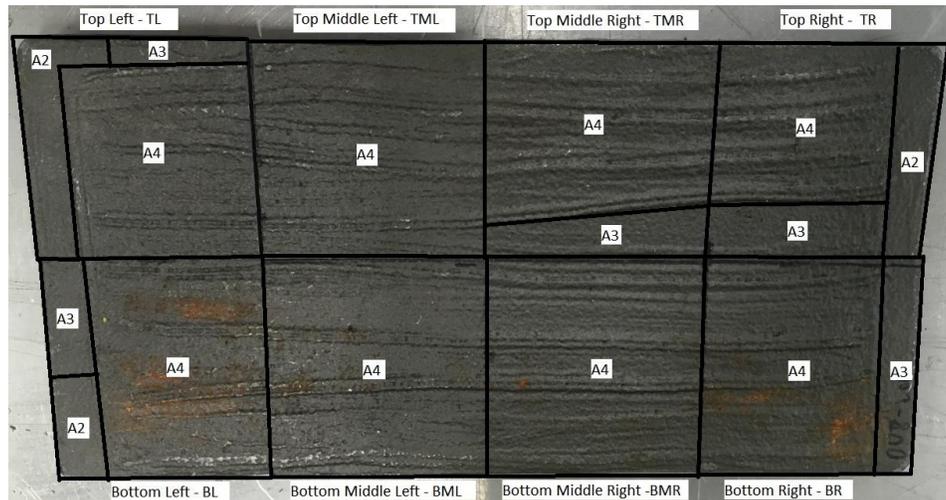


Figure 2-11. The surface classifications on the samples

The results from Experiment 1 show that as the surface roughness increases particle collection on the surface increases. Experiment 2 investigates the effect of surface roughness on the reliability of wet MPI when detecting discontinuities. The drilled holes are expected to create MFL; if there is an adequate amount of MFL, particles will be pulled from the area surrounding the discontinuity which will alter the noise area percentage [11]. Hence, Experiment 2 was planned to investigate the influence of surface roughness on the noise area percentage when a discontinuity is present in the part.

The results are as shown in Figure 2-12. A one-way ANOVA was utilized to compare the effect of the different surface classification levels on the noise area percentage. The model's assumptions of normality, constant variance, and independence were checked using diagnostic plots. The analysis using the scale location diagnostics plot identified that there was increasing

variance in the data as the response variable, noise area percentage increased, hence a logarithmic scale was used to transform the data to meet the model's assumption of constant variance. No significant differences were found between the noise area percentage for all four surface roughness levels when evaluated at a significance level of 0.05; $F= 3.261$; $p = 0.0805$. The estimated marginal means for the noise area percentage are 16% (95% CI [7%, 36%]), 50% (95% CI [23%, 100%]), 29% (95% CI [13%, 64%]), and 65% (95% CI [30%, 100%]) for A1 to A4, respectively. The estimated marginal means is the mean of the noise area percentage that is adjusted based on the model. However, since the surface roughness classification is a categorical variable, the estimated marginal mean is the same as the mean from the data. Additionally, a Tukey multiple comparisons test was performed between the surface classification levels. The multiple comparisons test was utilized regardless of the result of the F-test to provide further insights to aid the development of future research. The multiplicative effect on the median of noise area percentage from A1 to A2, A1 to A3, and A1 to A4 is 3.07 (95% CI [0.66, 14.40]), 1.81 (95% CI [0.39, 8.47]), and 4.03 (95% CI [0.86, 18.86]), respectively. The multiplicative effect is used to explain the comparison between the levels A1 to A2, A1 to A3, and A1 to A4 because the data for the response variable was in logarithm form. The results of this experiment show a general increase in the noise area percentage as surface roughness increases.

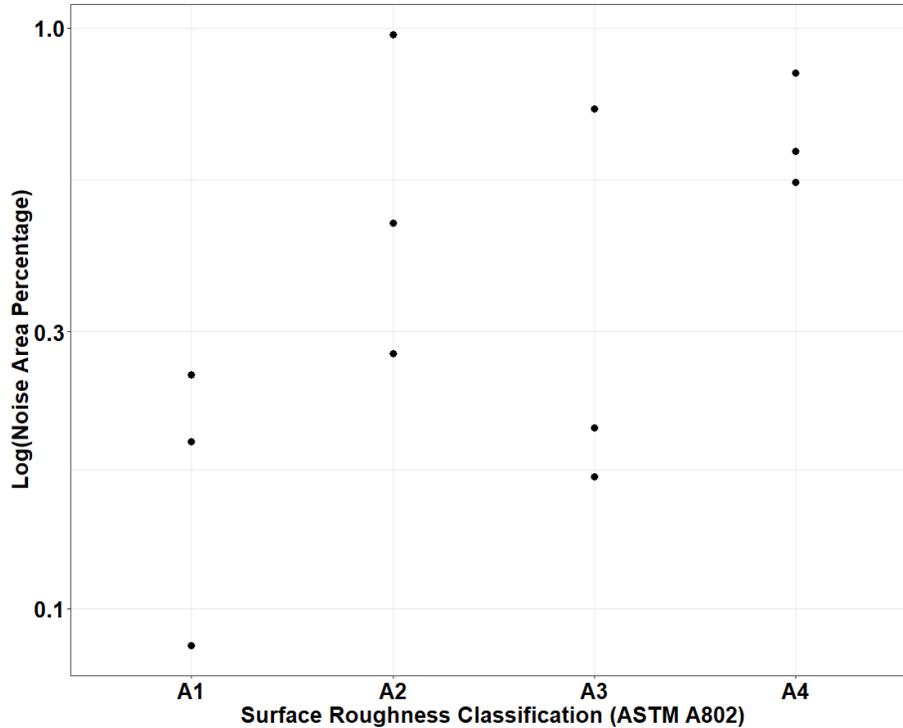


Figure 2-12. The noise area percentage results for A1, A2, A3, and A4 surface classifications

The findings in Experiment 2 support the research from past studies indicating that rougher surfaces reduce the reliability of wet MPI [10]. As shown in Figure 2-12, the noise area percentage for the A2 sample was higher than the A3 sample; this inconsistency may likely be due to the subjectivity of determining the surface roughness. The two human operators agree that the A2 and A3 comparator plates were tough to differentiate. Furthermore, research by Schimpf on digital surface characterization confirms that the differences between A2 and A3 are difficult to discern [12]. Overall, there is a general increase in noise area percentage as surface roughness increases. This means that as surface roughness increases, the difficulty in identifying the discontinuity increases.

Conclusion

The proposed method can quantify the level of difficulty to identify a true discontinuity. This can help with the development of optimal parameters for a specific part procedure. For example,

in setting up a new part to be tested in wet MPI, a procedure can optimize the parameter settings based on which combination of settings results in the lowest noise area percentage. Experiment 1 utilized the \bar{G} metric to show that surface roughness has significant influence (p-value = 1.48×10^{-13}) on the particle collection on the surface (see Figure 2-10). This shows that a rougher surface will have more particles adhering to the surface texture which may lead to higher rates of false-positive or reduce the likelihood of detecting true discontinuity. Experiment 2 utilized the noise area percentage that showed a general increase (p-value = 0.0805) in the level of difficulty in detecting discontinuities as surface roughness increase (see Figure 2-12). Experiments 1 and 2 show how the metrics developed in this paper could be used to quantify factors that influence the reliability of wet MPI. Several limitations in this paper should be considered in future research. First, the surface roughness classification method used in this research was subjective. This created uncertainty in the results due to the variability in the surface roughness classifications. Second, this study only investigated the effect of surface roughness on the effectiveness of wet MPI. Future work could utilize the noise area percentage metric to quantify the effect of other factors in wet MPI. Despite the limitations, this paper established a new method for quantifying the level of difficulty in identifying discontinuities using image analysis.

References

- [1] J. Y. Lee, S. J. Lee, D. C. Jiles, M. Garton, R. Lopez, and L. Brasche, "Sensitivity Analysis of Simulations for Magnetic Particle Inspection Using the Finite-Element Method," *IEEE Trans. Magn.*, 2003.
- [2] M. Ibrahim *et al.*, "Magnetic Flux Leakage (MFL) Capsule material selection and its robustness analysis in oil and gas pipeline," *Sensors and Transducers*, 2013.
- [3] D. Lovejoy and D. Lovejoy, "Magnetic particles, their characteristics and application," in *Magnetic Particle Inspection*, 1993.

- [4] M. A. A.-A. Abd-Allah, “Determination of Length , Thickness and Elements of Some Samples defects By Using magnetic particles and Ultrasound testing.,” no. May, pp. 1–48, 2016.
- [5] “6 Advantages of Wet Suspension Magnetic Particle Testing,” *Non-Destructive Testing (NDT) Blog - TSH*. [Online]. Available: <http://tspndt.com/non-destructive-testing-industrial-supplies-blog/6-advantages-of-wet-suspension-magnetic-particle-testing>.
- [6] M. Wright, “How to implement a PoD into a highly effective inspection strategy,” *NDT Canada 2016 6th Int. CANDU In-Service Insp. Work. Nov 15-17, Burlington, (NDT Canada 2016)*, 2016.
- [7] C. A. Harding and G. R. Hugo, “Guidelines for Interpretation of Published Data on Probability of Detection for Nondestructive Testing,” pp. 1–20, 2011.
- [8] D. Lovejoy and D. Lovejoy, “The history and basis of the magnetic particle testing method,” in *Magnetic Particle Inspection*, 1993.
- [9] D. Lovejoy and D. Lovejoy, “Choosing a magnetic particle inspection method,” in *Magnetic Particle Inspection*, 1993.
- [10] D. J. Eisenmann, D. D. Enyart, D. Kosaka, and C. Lo, “Fundamental Engineering Studies of Magnetic Particle Inspection and Impact on Standards and Industrial Practice,” 2014.
- [11] E. Li, Y. Kang, J. Tang, and J. Wu, “A New Micro Magnetic Bridge Probe in Magnetic Flux Leakage for Detecting Micro-cracks,” *J. Nondestruct. Eval.*, 2018.
- [12] D. Schimpf and F. Peters, “Variogram Roughness Method for Casting Surface Characterization,” *Int. J. Met.*, 2020.

Appendix: Statement of Authorship

Conceptualization - Sharon Lau; methodology - Sharon Lau; software - Sharon Lau;
 validation - Sharon Lau; formal analysis - Sharon Lau; investigation - Sharon Lau; data curation
 - Sharon Lau; writing (original draft preparation) - Sharon Lau; writing (review and editing) -
 Sharon Lau, David Eisenmann, and Frank Peters.

CHAPTER 3. MEASUREMENT ERROR DUE TO HUMAN FACTORS IN WET MAGNETIC PARTICLE INSPECTION

Sharon May Yen Lau*, David Eisenmann**, Frank Earl Peters*

*Department of Industrial and Manufacturing Systems Engineering, Iowa State University

**Department of Agricultural and Biosystems Engineering, Iowa State University

Modified from a manuscript be submitted to *International Journal of Metal Casting*

Abstract

In wet magnetic particle inspection (MPI), the human operator introduces variability due to different levels of qualifications, vision acuity, and fatigue. This can result in high escape or overkill rates of defects leading to failure of the part in the field or unnecessary rework being done on a part. Additionally, a substandard process could cause a defect to not be visible. Because of this, it is important to understand the capability of wet MPI to produce reliable steel castings. Hence, the goal of this research is to determine the repeatability and reproducibility of the wet MPI process in the steel casting industry to create a baseline for the current state. The method used to determine the percent match repeatability and reproducibility (R&R) is based on past research on visual surface inspection of steel castings. This work builds on that research with several modifications to allow for the repeatability and reproducibility to be measured in wet MPI. For example, the work by Daricilar et. al. used circular-shaped stickers to mark the defects that the operator identified via visual inspection [1]. In this paper, magnets were used instead of stickers since stickers will not be able to adhere to the part due to the wet nature of the testing. The suspension of magnetic particles in the carrier fluid will cause the collection of particles in the area of the magnets which can influence subsequent inspections. Hence, shot-

blasting was done between inspections to ensure this does not impact the results. The R&R results were reported as the ‘percent of matches’ instead of the percent variation. The repeatability and reproducibility were validated at four different steel foundries and the averages were 27% and 29% match, respectively. The results from this study indicate that further work is needed in optimizing the process parameters in wet MPI and creating an inspector aid to help guide human operators to find defects.

Introduction

In the metal casting industry, the evaluation of an acceptable part is often subjective and defined by the buyer. Nondestructive testing (NDT) methods are used as a way of gauging some aspects of the quality of the castings. Visual inspection, radiograph, magnetic particle inspection, dye penetrant, and ultrasonic testing are commonly used NDT methods in the metal casting industry. In this chapter, a method to determine the measurement error due to human factors in wet magnetic particle inspection (MPI) was developed. Advantages of wet MPI include its cost, speed of testing, and ability to be used on a variety of different sized parts. A ferromagnetic material is required for testing using wet MPI because the part needs to be magnetized and remained magnetized. Once the part is magnetized, magnetic flux leakage forms around the area of the defect. A suspension of fluorescent magnetic particles is then applied to the part and the magnetic particles are attracted to the area of magnetic flux leakage. A human inspector then shines an ultraviolet (UV) light on the part to visually locate the defect by identifying the collection of particles. A simple breakdown of the wet MPI process is shown in Figure 3-1.

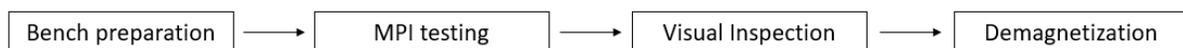


Figure 3-1. Wet MPI process

The measurement error of a process can be determined through gauge R&R testing which has been widely used across various industries. It is used to determine the variability due to the measurement system. More specifically, it evaluates the precision of the measurement system where less variation equates to better precision (see Figure 3-2). There exist three sources of variation: the actual variation within the part, the measurement system, and the operator. The variation due to the measurement device is called repeatability and the variation due to the operator is called reproducibility. The variability due to the operator can stem from the different types of certifications such as the American Society for Nondestructive Testing (ASNT) or in-house certification provided by the specific foundry. Both types have 3 levels of certification available to the operators. A level 3 MPI operator is responsible for creating documents that outline standard procedures that operators should follow for specific parts and for the training of the level 1 and 2 operators. The level 1 and 2 operators are responsible for the day-to-day testing of the parts which can introduce measurement errors due to the different levels of expertise. Additionally, since wet MPI is performed in a dark room, the visual ability to discern shape and color is an important factor. Hence, human visual acuity is another factor contributing to measurement error in wet MPI. Fatigue is also known to affect work performance so a human operator running MPI at the beginning of the shift might perform differently compared to the same operator at the end of the shift [2].

A majority of the past work in gauge R&R was developed for continuous data such as length, depth, etc. where the total gauge R&R is the sum of the repeatability and reproducibility. A total gauge R&R of more than 30% is typically deemed unacceptable for most applications.

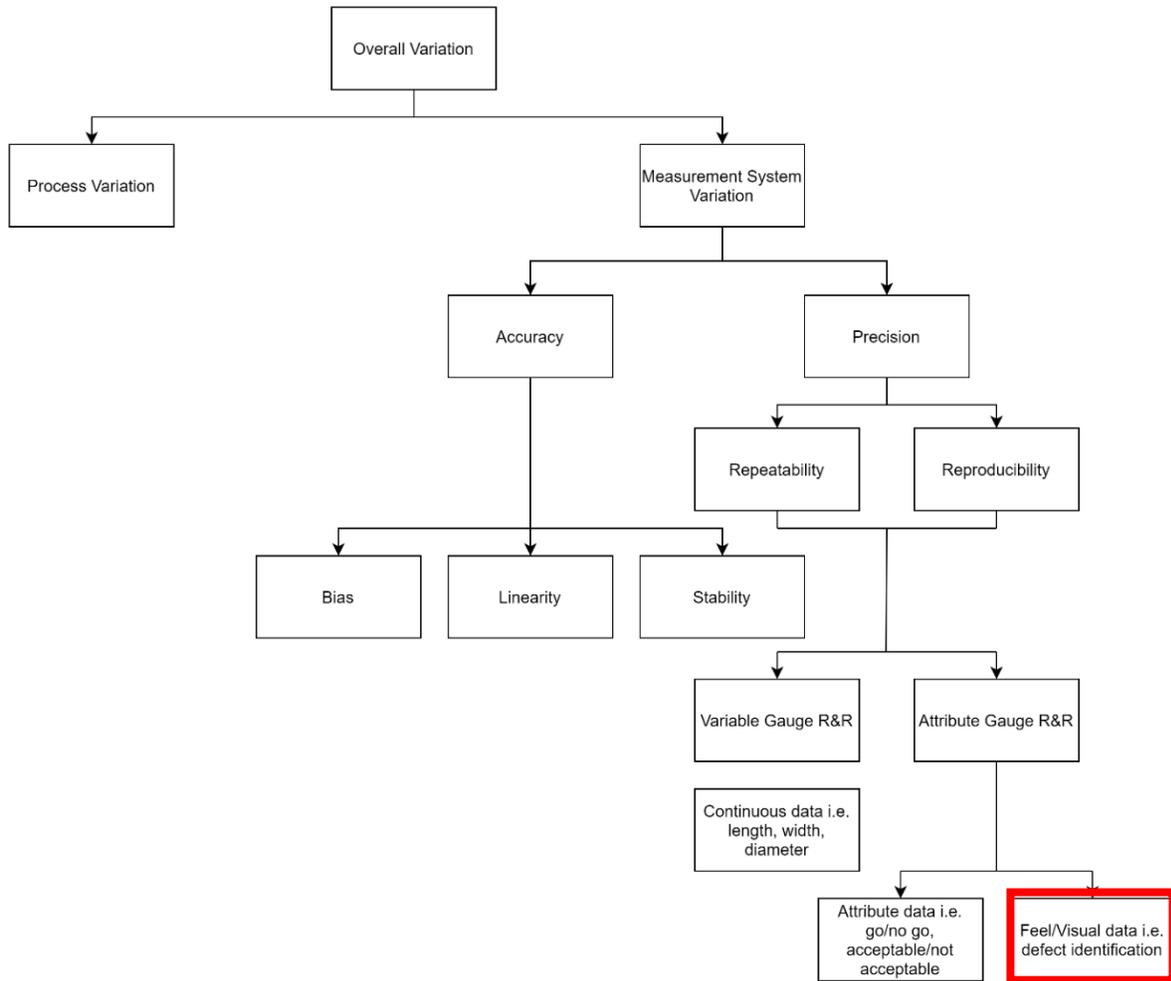


Figure 3-2. Types of variation in a system

Unlike continuous data, visual inspection data are often categorical or count data. For visual inspection, attribute gauge R&R is typically used to evaluate the precision of the measurement system. An example of attribute gauge R&R is when a part is tested under a pass or fail criteria. A human operator must compare the combination of attributes of the part to a standard to determine whether the part passes or fails. If operator A inspects 10 parts twice and has the same accept or reject outcome for 8 out of the 10 parts inspected, the percent agreement for repeatability is 80%. In attribute gauge R&R, a higher value equates to a higher precision. Similarly, for reproducibility, if operator A and operator B inspects the same 10 parts and have

the same accept or reject outcome for 7 out of the 10 parts inspected, the percent agreement for reproducibility is 70%. This example uses the simple percent agreement as the metric for measuring the agreement between the operators.

Cohen's kappa was a metric that was developed by Jacob Cohen to create a more robust measurement of agreement between two operators by including the probability that the agreement occurred by chance [3]. A variation of this metric coined Fleiss' kappa was developed to measure the agreement between more than two operators [4]. However, Cohen's and Fleiss' kappa could not be applied to this study because the identification of defects is not a simple yes or no but also includes the location. Since there are many possibilities of where the defect could be found on a part, adding probability due to chance is not suitable. Hence, this paper used the simple percent agreement method as the metric for measuring the agreement within and between the operators.

However, there is a need for more research to evaluate the precision of visual inspections because visual inspection of metal castings is not a typical go or no-go operation. Multiple defects are typically identified for each part and the part undergoes a combination of grinding and welding operations to remove the identified defects. Hence, the metric of acceptable or unacceptable is not used, rather the precision needs to be determined based on the defects identified. Daricilar et. al. created a method for visual surface inspection of metal casting by using circular stickers to determine clusters based on a search zone coefficient. The search zone size was larger than the circular markers to accommodate minor errors in the placement of the marker and differences in the quality and angle between images. By calculating the percentage of sticker overlap, the percent match repeatability and reproducibility can be obtained (see Figure

3-3 and Figure 3-4). Since percent match is the metric used, the higher values represent better repeatability and reproducibility [1].

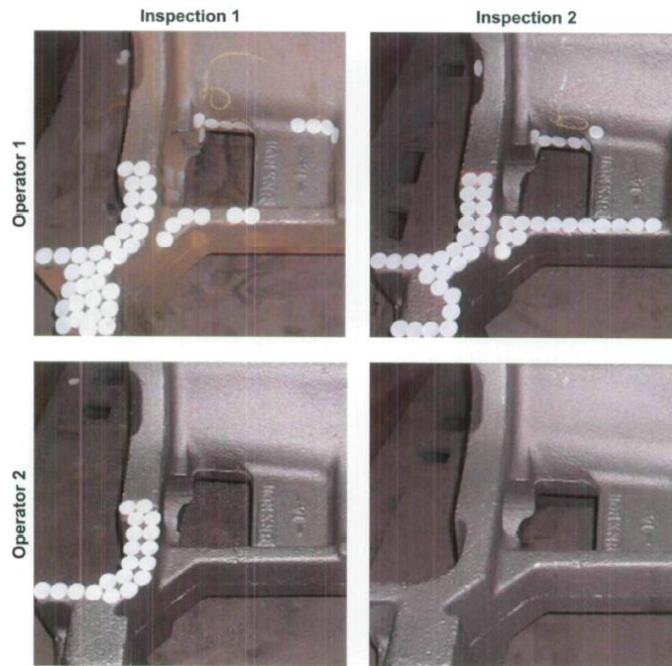


Figure 3-3. Results were obtained from the same casting inspected twice by each of the two operators at Foundry 2 [1].

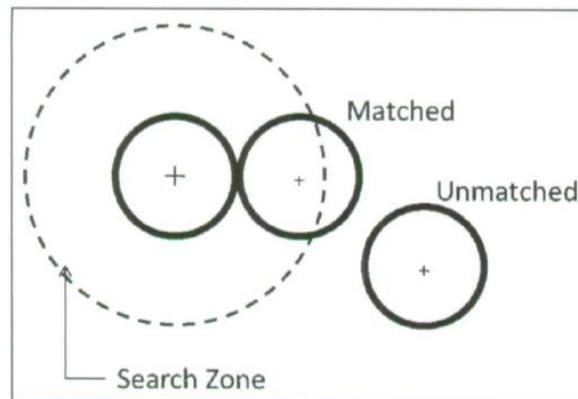


Figure 3-4. Search zone coefficient example with default values of 2.5 for the master cluster [1].

The use of circular stickers as markers to determine the percent matching R&R was not feasible in wet MPI. The use of carrier fluid containing magnetic particles in wet MPI would mean that the stickers would not adhere to the part. Additionally, the process of using stickers to

mark the defects identified by the operators is slow and accurate results would rely heavily on the placement of each sticker as well as the consistency in image quality and angle. Due to these reasons, the method used for visual inspection will not work for wet MPI.

For wet MPI, the proposed solution uses magnets to mark the identified defects which address the limitation of using stickers since it is a wet environment. Another limitation of the sticker method is the speed of marking since a sticker is placed over each part of the defect. In the proposed approach, only two magnets are used to mark the beginning and end of each defect which significantly speeds up the marking time. Additionally, since the magnets are only placed at the beginning and end of the defect, the location and the defect can be captured in the image whereas in the sticker method only location information is available because the defect is covered. In the proposed solution, the identified defect is mapped onto the part using software which allows for efficient calculation of the R&R values. Overall, the proposed method provides an effective solution for the precision of wet MPI to be evaluated which will allow for the variability in the system to be determined and improved upon.

Methods

R&R Calculations

Repeatability is the measurement of variability under the same conditions with multiple trials. In this study, two trials for the same part, operator, and setup were compared. The repeatability was calculated by identifying the number of defects in the same part that matched between the two trials, divided by the total number of unique defects found by the operator in both trials. Figure 3-5 shows an example where the operator found three defects during the first inspection and four defects the second time. However, only two of the defects were identified in both inspections. Since there was a total of five unique defects found, the percent match repeatability is 40%.

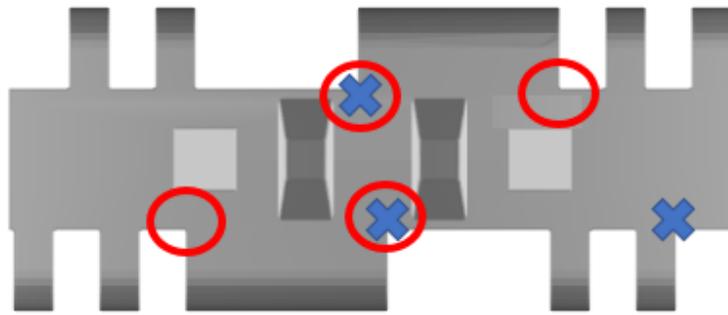


Figure 3-5. The blue “x” represents Operator 1 Part 1 Trial 1 and the red circles represent Operator 1 Part 1 Trial 2.

Reproducibility is the measurement of variability under the same conditions with different operators. Since each operator ran two trials, Trial 1 and Trial 2 were combined for each operator and part. The reproducibility was determined by the number of defects on the same part that matched between inspectors, divided by the total number of unique defects present across both inspectors. In Figure 3-6, four defects match between operator 1 and operator 2. Lastly, the number of defects that were matched between operator 1 and operator 2 would be divided by the total number of unique defects, which is six, resulting in a percent match for reproducibility of approximately 67%.

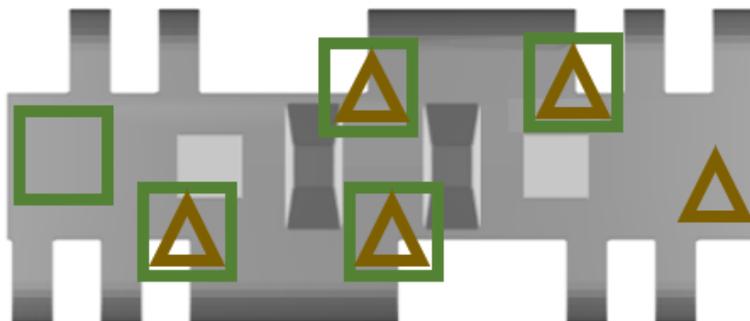


Figure 3-6. The brown triangles represent the union of Operator 1 Part 1 all Trials and the green squares represent the union of Operator 2 Part 1 all Trials.

For more than two operators, a pairwise agreement was used and an average of the reproducibility values for each pair was taken to represent the final reproducibility value in this study. This approach is different from the typical attribute gauge R&R approach where an overall agreement is used. All operators must agree for a defect to be considered a match using the overall agreement method. Using the overall agreement method for the example in Figure 3-7A would result in a 1 out of 6 match which is a percent match reproducibility of 16.67%. This paper proposes using a pairwise agreement where Figure 3-7B is comparing the agreement between operator 1 and operator 2, Figure 3-7C is comparing the agreement between operator 2 and operator 3, and Figure 3-7D is comparing the agreement between operator 1 and operator 3. This would result in three pairwise reproducibility results of 66.67%, 16.67%, and 40%, respectively. The pairwise results are useful in considering which operators in the foundry have higher agreement percentages. This data could be used to evaluate other factors that may cause higher agreement percentages between certain operators. For example, if it were found that the operators with more than 10 years' experience in wet MPI had higher percent agreement reproducibility, this might indicate more training could be provided to the less experienced operators to increase reproducibility. If a single value is needed to represent the reproducibility of the foundry, the average of all the pairwise reproducibility result could be used. For the example used in Figure 3-7, the average pairwise agreement reproducibility is 41.11%.

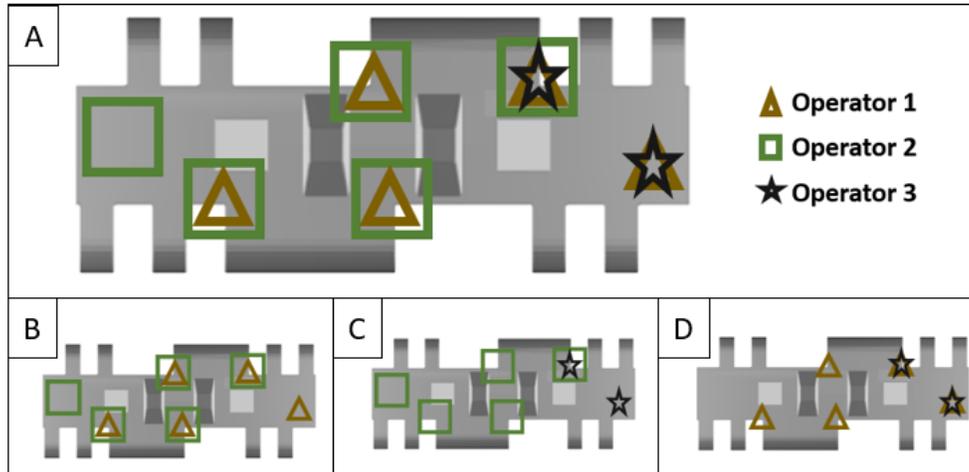


Figure 3-7. A) The brown triangles represent the union of Operator 1 Part 1 all Trials, green squares represent the union of Operator 2 Part 1 all Trials, and the black stars represent the union of Operator 3 Part 1 all Trials. B), C), and D) represent the resulting pairwise comparison based on A)

Testing Procedure

The measurement error with the MPI process includes the differences in how an operator would process the part (e.g. setup, current, orientation) and the visual interpretation of the resultant part. For all the experiments, the operator communicated the presence of a defect by placing magnets on the casting (see Figure 3-8). The location of the magnet was then captured via a camera. Prior to the placement of the magnets, the casting surface was imaged under UV light to capture what the operator saw. After the inspection, the number of defects found for that test was noted. A complete flowchart of the process is shown in Figure 3-9.



Figure 3-8. Magnets are used to mark the beginning and end of the crack.

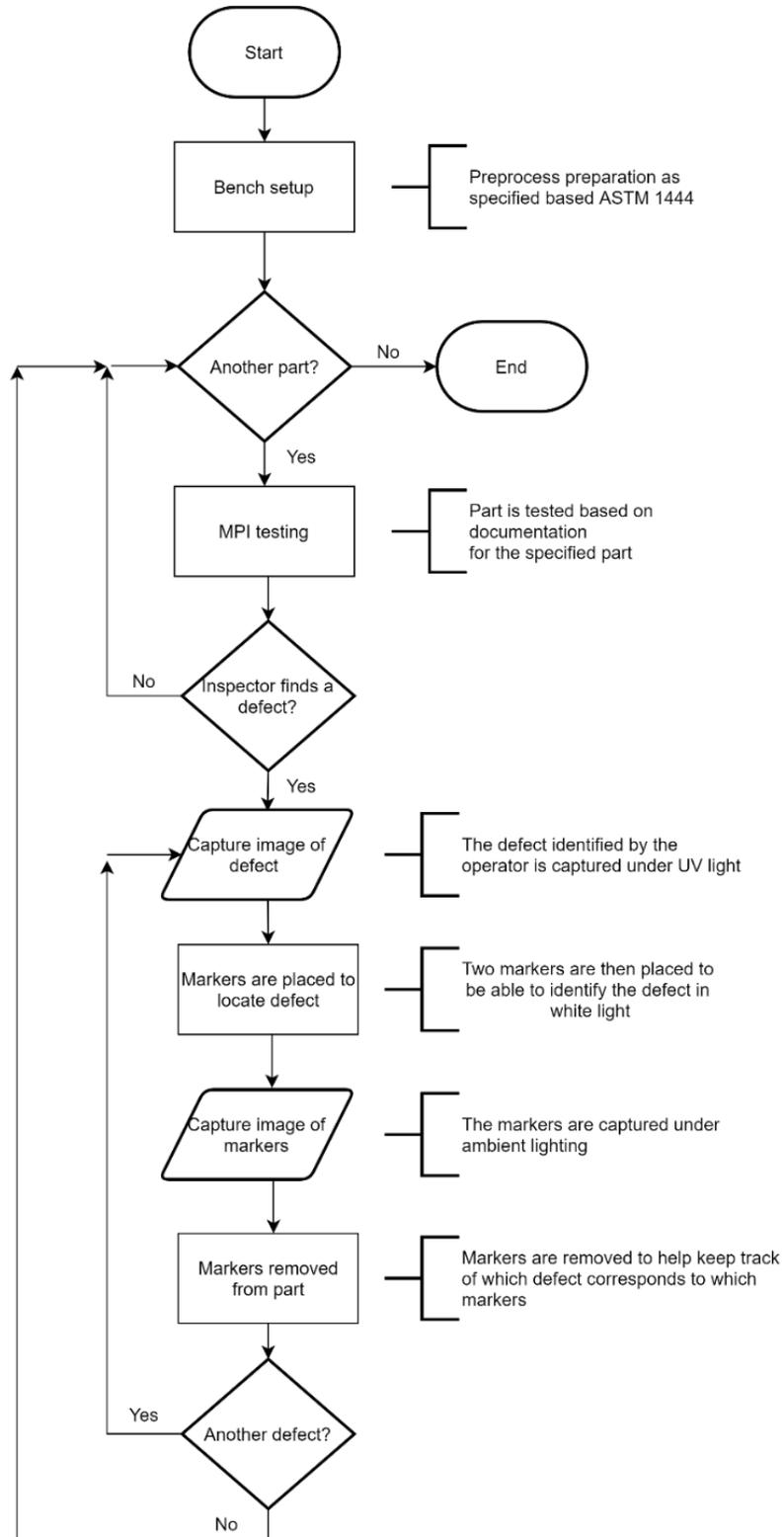


Figure 3-9. A flowchart of the process.

After the MPI testing was completed, the defects that were identified were mapped onto a three-dimensional (3D) computer-aided design (CAD) model. The CAD model was created based on basic measurements to represent the relevant geometry of the part. Casting mapper software which was developed at Iowa State University was used to mark the defects. Using this software, a CAD model can be imported and two-dimensional (2D) views of the part will be generated which can be marked for defects. An example of the mapping of the defects on the part is shown in Figure 3-10. The mapping of defects from the casting mapper software was then exported as a PDF to be analyzed for repeatability and reproducibility measurements.

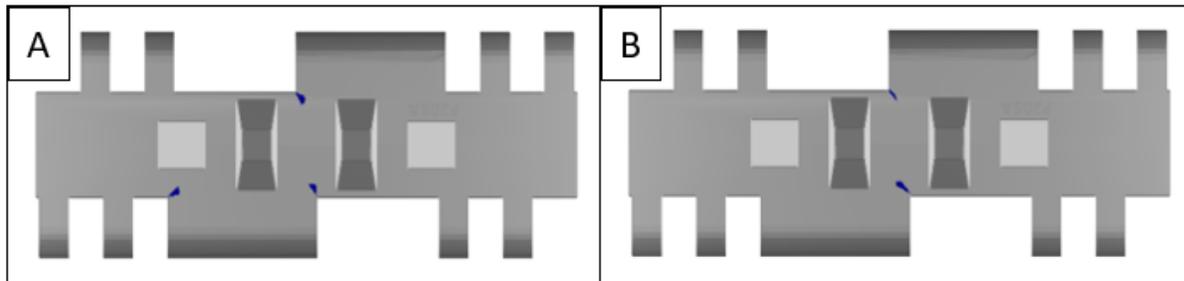


Figure 3-10. A) Operator 1 Part 1 Trial 1 Front view; B) Operator 1 Part 1 Trial 2 Front view

Data Collection and Analysis

Each defect in a part is captured independently, hence each image only contains one identified defect. Because of this, it is hard to visualize all the defects identified on one part and to match the defects across two trials. Hence, casting mapper software was used to allow for a quick way to visualize all the defects that were identified on a specific part. The process for analyzing the images gathered from the gauge R&R study is as follows:

1. A CAD drawing of the part is created.
2. The CAD drawing is imported into casting mapper software.
3. The location of defects is mapped from the images to the casting mapper software.
4. PDF reports containing the mapping of all defects for a specific part, operator, and trial are saved.

5. Once all the PDF outputs are created, repeatability and reproducibility can be evaluated.
6. By analyzing the PDF report, the defects in the same region between the two trials will be further validated through the images of the defect under UV light to compare the characteristics of the defect for similarities.
7. The repeatability and reproducibility values are then calculated based on the methods outlined in the R&R Calculation section.

This method was then validated across four steel foundries and the results are shown below.

Table 3-1. Experimental layout and R&R results for all foundries.

Foundry	Parts	Total Number of Operator(s)	Trials	% Match Repeatability	% Match Reproducibility
*1	6	3	2	73%	48%
2	6	2	2	33%	25%
3	4	1	2	15%	N/A
4	6	2	2	33%	33%
<i>Note: *Parts were not blasted in between sets hence the data were not included as a part of the averages of R&R</i>					

Foundry 1's result was not taken into account because parts in between trials were not shot-blasted, and the magnets that were used to identify the location of the defects were found to leave marks on the part, which affected the results of the study. In Foundry 3, there was evidence of the defects not appearing on the photographs, so the repeatability was not due to problems with the visual perception of the operator (see Figure 3-11). Prior to examining the results from Foundry 4, the reasoning behind this phenomenon was unclear, with speculations of possible changes in parameters of testing between trials (e.g., orientation, magnetic field strength, etc.) or shot blasting causing the defects to be covered up. In Foundry 4, it was found that the defects do not get covered up by shot blasting, as shown in Figure 3-12. All the defects identified by one

operator but not by the other were still visible in the images captured in both trials. In other words, the defects were all present and visible even after multiple iterations of shot-blasting. The defects in Foundry 4 were surface-breaking discontinuities. Since the defects in Foundry 3 show up faintly and were not visible upon close inspection under ambient light, they were likely subsurface defects. The ‘disappearing’ defects between trials that occurred for Foundry 3 could be explained by slight orientation changes since the part was tested using indirect magnetization (on the coil rather than clamped) which allowed for more variability in the placement of the part (see Figure 3-13 and Figure 3-14).

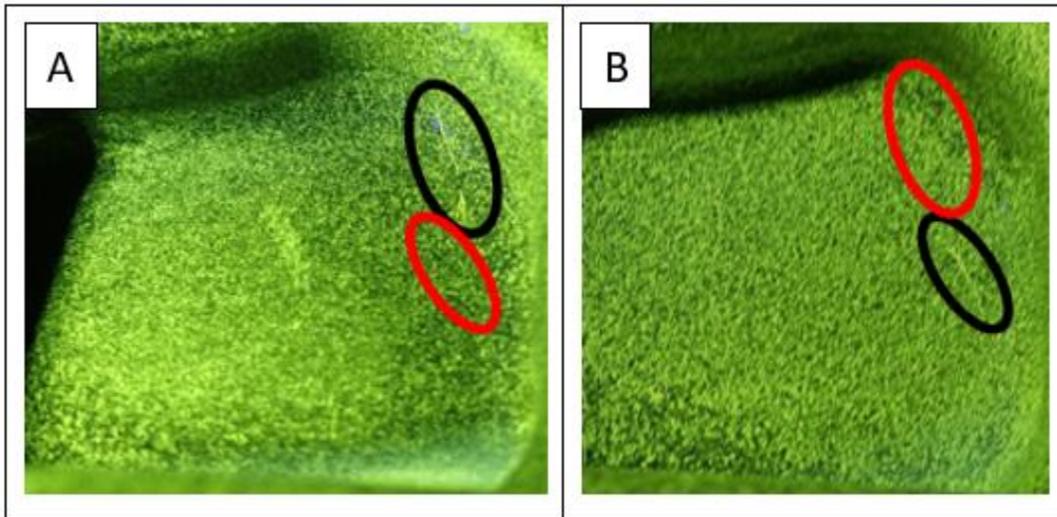


Figure 3-11. A) Operator 1 Part 1 Region A Trial 1; B) Operator 1 Part 1 Region A Trial 2

Note: The black oval is the defect the operator identified, the red oval is the same spot they located a defect in another trial however defect is not visible and not identified in the current trial.

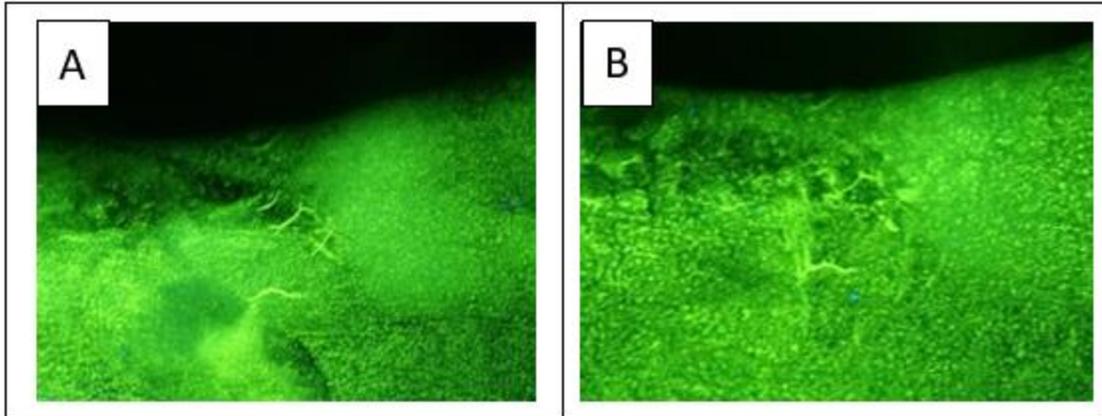


Figure 3-12. A) Operator 1 Part 1 Region A Trial 1 defect is visible and identified;
B) Operator 2 Part 1 Region A Trial 1 is visible and not identified



Figure 3-13. Indirect magnetization was performed at Foundry 3

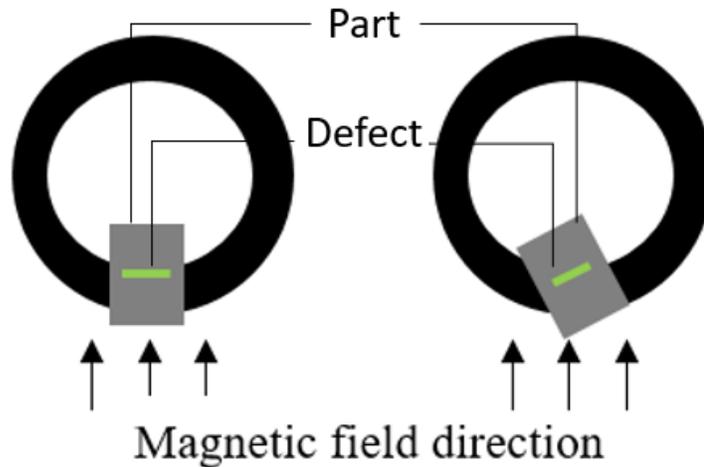


Figure 3-14. Placement variability when using indirect magnetization

Conclusions

Understanding the amount of variation present in a measurement system is important to drive process improvements. While there are many studies on how to evaluate the measurement error in a system, the majority of the existing work is on continuous data and not visual identification tasks. Therefore, there has been no readily available method to evaluate measurement error in wet MPI. Past research developed a method to measure process variation for visual inspections in metal castings using a percentage match method for clusters defined by small circular stickers. This study proposed a revised method to adapt this approach to wet MPI by 1) changing the marking system to magnets instead of stickers, 2) using a software tool to speed up the calculation of R&R, and 3) proposing a solution that enables the consideration of more than 2 operators for reproducibility. Overall, the findings across four steel foundries resulted in an average percent match repeatability and reproducibility of 27% and 29%, respectively.

Additionally, the type of defect can be determined by examining the part under ambient light. The parts inspected in Foundry 2 and Foundry 4 contained surface cracks whereas parts in

Foundry 3 likely contained sub-surface defects. It was discovered that sub-surface defects that were visible in one trial but not in the second trial could be attributed to minor changes in orientation (see Figure 3-14). To mitigate variation in the orientation of the part, direct magnetization where the part can be clamped into a fixed location and orientation is preferred when possible. Additionally, it was found that surface cracks are not as sensitive to minor changes in orientation. All surface cracks were visible in both trial 1 and trial 2, which indicates that the defects that were not identified were due to human error in the visual inspection step. Chapter 4 will investigate the impact of different process parameters on the effectiveness of wet MPI since it was discovered that orientation may have played a role in disappearing sub-surface defects between trials. Chapter 5 will discuss the development of an inspector aid to mitigate defects missed due to human error. Additionally, the inspector aid solution's performance will be benchmarked against current visual inspection performance in wet MPI.

References

- [1] G. Daricilar and F. Peters, "Methodology for assessing measurement error for casting surface inspection," *Int. J. Met.*, 2011.
- [2] K. O'Keefe, S. Hodder, and A. Lloyd, "A comparison of methods used for inducing mental fatigue in performance research: individualised, dual-task and short duration cognitive tests are most effective," *Ergonomics*, vol. 63, no. 1, pp. 1–12, 2020.
- [3] J. Cohen, "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit," *Psychol. Bull.*, 1968.
- [4] J. L. Fleiss and J. Cohen, "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability," *Educ. Psychol. Meas.*, 1973.

Appendix: Statement of Authorship

Conceptualization - Sharon Lau; methodology - Sharon Lau; software - Sharon Lau; validation - Sharon Lau; formal analysis - Sharon Lau; investigation - Sharon Lau; data curation - Sharon Lau, Landon Getting, Carley Haus, and Ryan Utterback; writing (original draft

preparation) - Sharon Lau; writing (review and editing) - Sharon Lau, David Eisenmann, and Frank Peters.

CHAPTER 4. INVESTIGATING THE INFLUENCE OF PROCESS PARAMETERS ON THE EFFECTIVENESS OF WET MAGNETIC PARTICLE INSPECTION

Sharon May Yen Lau*, Aron Mitchell**, David Eisenmann***, Frank Earl Peters*

*Department of Industrial and Manufacturing Systems Engineering, Iowa State University

**Department of Mechanical Engineering, Iowa State University

***Department of Agricultural and Biosystems Engineering, Iowa State University

Modified from a manuscript to be submitted to *International Journal of Metal Casting*

Abstract

The characteristics of the part and the process parameters play a role in the effectiveness of wet magnetic particle inspection (MPI). One factor that affects the ability to locate a defect is having sufficient magnetic field strength to create adequate flux leakage to attract magnetic particles to the area of the defect which creates higher contrast between the defect and its surroundings. This study investigated the impact of various geometries on magnetic field strength. Concave geometry was found to have lower magnetic field strength compared to flat and convex geometries. Additionally, this investigation introduced a new metric to evaluate the visibility of defects in wet magnetic particle inspection (MPI). Wet MPI is a commonly used nondestructive test (NDT) method to identify flaws in ferromagnetic parts. Probability of detection (POD) graphs have been used historically to determine the capability of this process. However, creating POD graphs requires a substantial amount of data and is a subjective method. A new method, noise area percentage (NAP), has recently been proposed (see Chapter 2) to replace POD graphs to provide a quick snapshot of the effectiveness of a wet MPI test using an image-based approach to evaluate the level of difficulty in identifying a defect [1]. The NAP method effectively determined the level of difficulty in finding a defect; however, the method

did not consider how a human would identify defects. Hence, this study implemented some modifications to the NAP method which will provide an objective value for how well defects can be identified and is more representative of what the human operator would see. Additionally, this chapter displays a user-friendly open-source graphical user interface (GUI) that was developed for obtaining the NAP metric, which will aid data analysis and encourage the use of the technique. This approach was then used to validate the effect of surface roughness, orientation, current type, and magnetic field strength on surface-breaking and sub-surface discontinuities. The results indicate that for surface-breaking defects, the part should be set up in at least two orthogonal orientations to detect discontinuities. In contrast, sub-surface defects would require the magnetic field lines and the discontinuity to be perpendicular to be detected, which may require more than two setups to locate the defect. Surface roughness was also found to influence the ability to detect defects. Since geometry is known to impact magnetic field direction, various geometries on the orientation of the magnetic field lines were investigated.

Introduction

Several aspects of the integrity of metal castings can be gauged through the use of nondestructive testing (NDT) methods. Magnetic particle inspection (MPI) is an NDT technique that is commonly used for several reasons; 1) speed of testing, 2) relatively low expense compared to the other NDT methods, 3) ability to identify surface and sub-surface defects, and 4) ability to be used on small or large parts [2]. MPI is used on ferromagnetic parts, which are materials that can be magnetized. With MPI, first the part is magnetized, then if there is a defect present, magnetic flux leakage (MFL) will occur. Magnetic particles are then introduced to the part; if the MFL is adequate, the particles will collect on the surface where MFL is present. The human operator can then identify the defect through the collection of particles [3].

There are two types of MPI testing, wet and dry. Dry MPI tends to perform better on rougher surfaces due to larger particle size, which reduces the possibility of particles getting caught in the valleys of the surface texture [4]. An electromagnetic yoke is used to generate the pulsating magnetic field in a small area while dry magnetic particles are applied to the part. Wet MPI is performed on a horizontal stationary magnetic particle machine where the magnetic particles are suspended in oil or water to aid the movement of the particles on the part [5]. Wet MPI is the method investigated in this study because it is the most commonly used MPI method in steel foundries.

When inspecting rougher surfaces, wet MPI is not as effective as dry MPI because the smaller particle sizes used in wet MPI tend to get caught in the surface of the part. However, for larger castings, wet MPI is more efficient than dry MPI because the entire part can be tested using the suspension containing carrier fluid and magnetic particles. Using dry MPI, only small areas can be tested at one time, which can be time-consuming for larger castings [6].

Additionally, the use of carrier fluid in wet MPI leads to higher particle mobility when compared to dry MPI. This allows for the particles to easily move to the area of flux leakage, increasing the effectiveness of illuminating the area containing the defects. Fluorescent coated particles used mainly in wet MPI also provide more contrast between the particles and the background.

Although the smaller-sized particles in wet MPI tend to create noise, it allows for the detection of finer defects [7].

Particle collection on the surface of the part does not only occur in the area of the defect. Surface roughness, geometry, or residual particles left on the part through handling could result in a collection of particles that could lead to false positives or create noise that would hinder the human operator from finding true defects. In this paper, a false positive refers to finding a

discontinuity that does not exist in the part while true defects refer to discontinuities that exist in the part. There exist multiple reasons that contribute to a human operator missing true defects such as variability due to the process, the part, the equipment, and the human operator [8]. This chapter investigates two of those main factors in MPI testing which are the process and the part using a novel metric to quantify the capability of the MPI.

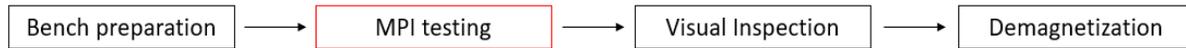


Figure 4-1. This chapter investigates the MPI testing step of the wet MPI process

In this chapter, the impact of geometry on magnetic field strength and magnetic field direction were investigated. Flat, convex, and concave geometries were tested. Additionally, the impact of surface roughness, orientation, magnetic field strength, current, and magnetization type on the effectiveness of wet MPI were investigated. Results from Chapter 3 on the gauge repeatability and reproducibility (R&R) of MPI showed low R&R and evidence of defects disappearing from one trial to the next, which indicates variability in the process. This led to the development of a metric to provide an objective value for defect visibility in MPI which allowed for the investigation of the variability due to process and part.

Probability of detection (POD) is an approach used to evaluate the capability of various NDT methods. The likelihood that a defect of a certain size will be identified by an inspector can be predicted through this method. In a majority of cases, a 90% POD value is needed, which equates to a skilled inspector being able to detect a given defect 9 out of 10 times based on a mathematical estimate. The variability of the POD experiment is captured through the 95% confidence interval accompanied with the POD value. The purpose of the POD approach is to provide a metric for inspection parameters to be optimized through the quantification of the inspection system and abilities of the human operator [9]. The downside to this methodology is

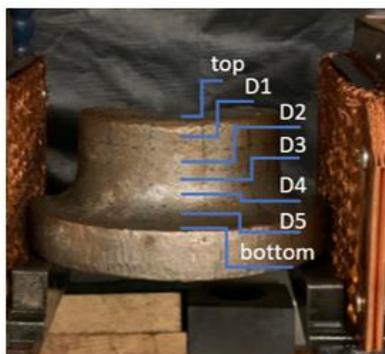
that it is time intensive to acquire sufficient data due to the number of tests required. Because of this, the POD method is not used to evaluate individual parameters for each NDT technique, rather it is used to evaluate the overall limitation of the technique [10].

In this paper, the impact of geometry on the magnetic field strength was investigated. In our previous work, an objective method to quantify the level of difficulty in identifying defects in wet MPI was developed based on comparing the illumination of the particles in the background against the illumination of particles in the location of the discontinuity using image analysis. This metric is called noise area percentage (NAP) and can be used as a way to quantify the effects of the different parameters on the effectiveness of wet MPI [1]. In this paper, the NAP metric is further refined to better represent how a human operator identifies a defect. Subsequently, the improved NAP metric is then validated through a series of experiments to gain an understanding of the variability associated with the key parameters in wet MPI. Better understanding of the capabilities of wet MPI can lead to better defect detection standards between the foundries and the buyers which will save time and improve the quality of inspection.

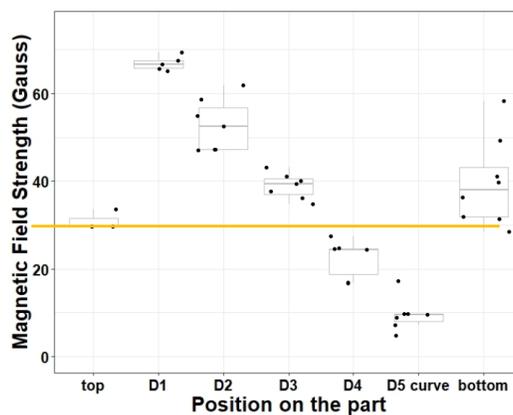
Impact of geometry on magnetic field strength

The capability of MPI is impacted by the characteristics of the tested part. In this study, the geometry of the part was investigated. Geometry is known to affect the magnetic field strength since it defines the arrangement of magnetic domains [11]. This study is important to determine which geometries contain lower magnetic field strength which may lead a defect in that region to go undetected. Part containing three different geometries, concave, flat, and convex geometries were tested at two orientations. The magnetic field strength was measured using a gauss meter where a baseline of 30 gauss (per the ASTM 1444 standard) was set by tuning the current levels on a flat region in each part.

For concave geometry, the results for orientation 1 as shown in Figure 4-2 indicate that the magnetic field strength decreases as the radius of the concave region increases. The plot in Figure 4-2 (b) shows the magnetic field strength decreasing from position D1 to D5. Position D4 and D5 had magnetic field strength values of less than 30 gauss. For orientation 2, the results in Figure 4-3 show a similar trend with the magnetic field strength decreasing from position D1 to D5. This means the region with concave geometry may be at a lower than the recommended magnetic field strength compared to the flat regions where the baseline of 30 gauss was set. This may lead to defects in those regions having lower magnetic flux leakage leading to more difficulties in identifying the discontinuities due to lack of contrast. Additionally, defects tend to occur in concave regions post heat treatment if excessive warping that occurs. In creating the procedure for the amount of current to use on a part, concave regions should be evaluated to ensure these regions have sufficient magnetic field strength to meet the ASTM guidelines.

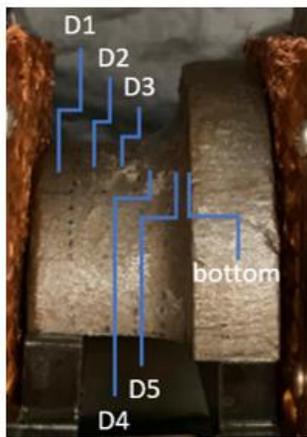


(a)

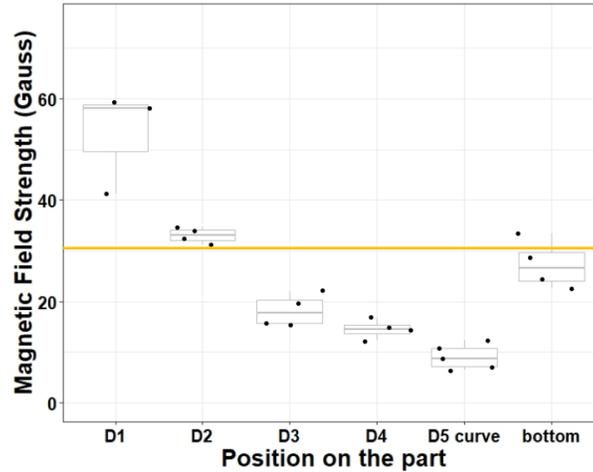


(b)

Figure 4-2. (a) Labeled concave geometry part at orientation 1 and (b) Results for concave geometry



(a)



(b)

Figure 4-3. (a) Labeled concave geometry part at orientation 2 and (b) Results for concave geometry

For flat and convex geometry, the results are shown in Figure 4-4 and Figure 4-5. The sample containing flat regions also contained sharp corners, however, flat regions or corners did not show any significant trends. This results for flat regions were as expected since the flat regions were used to determine the baseline magnetic field strength on each part. On the other hand, the part containing a convex geometries did not show any meaningful trends. Different radiuses for convex geometry should be evaluated to validate that convex geometries have no significant impact on the magnetic field strength.

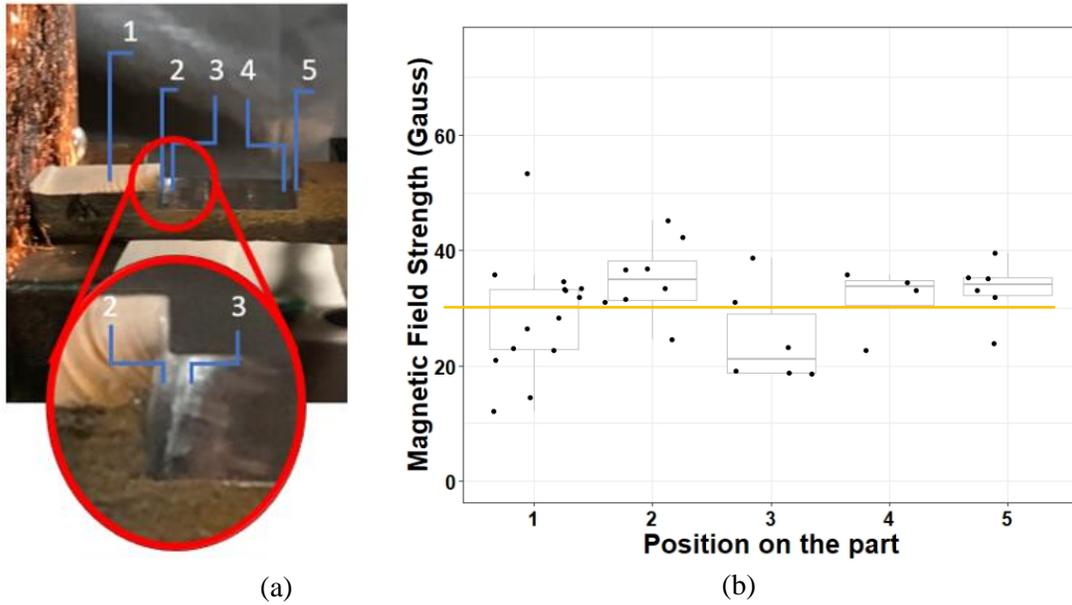


Figure 4-4. (a) Labeled flat geometry part and (b) Results for flat geometry

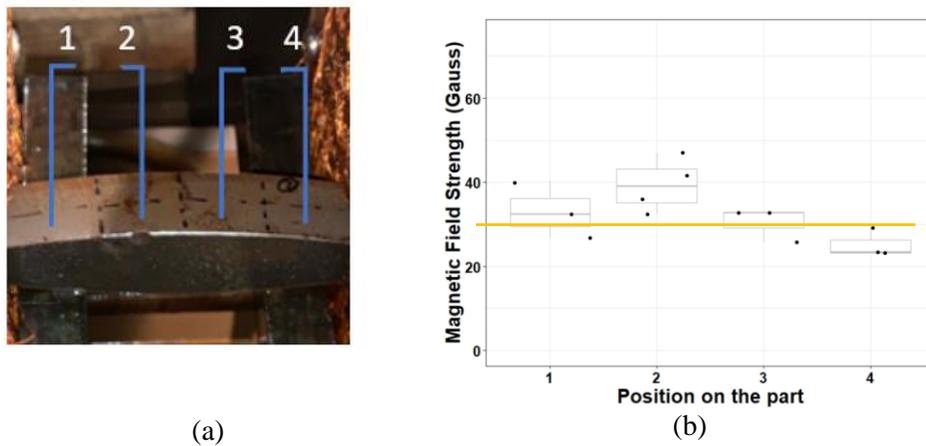


Figure 4-5. (a) Labeled convex geometry part and (b) Results for convex geometry

Overall, the results in this section show that for concave geometry, there was a decrease in magnetic field strength as the radius increases. This is important because defects usually occur in concave regions and may be missed because of the lower magnetic field strength in those regions. Hence, prior to testing the part, the current that produces sufficient magnetic field strength across all geometries of the part would need to be determined to increase the probability

of locating defects in concave regions. This can be done using Quantitative Quality Indicators (QQIs) across the relevant areas on the part or using a gauss meter.

Impact of surface roughness, orientation, current type and magnetic field strength on the effectiveness of wet MPI

The results from Chapter 3, suggests that process parameter play a role in the repeatability and reproducibility of wet MPI. Hence, this study is conducted to further understand which parameter influences the effectiveness of wet MPI. Rectangular plates with the dimensions of 355.6 by 174.6 mm by 12.70 mm (14.00 in by 6.875 in by 0.5 in) were acquired from an industrial foundry which were cast using commercial chemically bonded sand process and were made of WCB cast alloy (cast equivalent to 1020). The rougher than normal surface texture of the plates was created by intentionally scraping the sand mold before it was fully cured. The surface roughness of the samples was classified using The American Society for Testing and Materials (ASTM) A802 standard. The rectangular plate was cut into 8 square samples as shown in Figure 4-6.

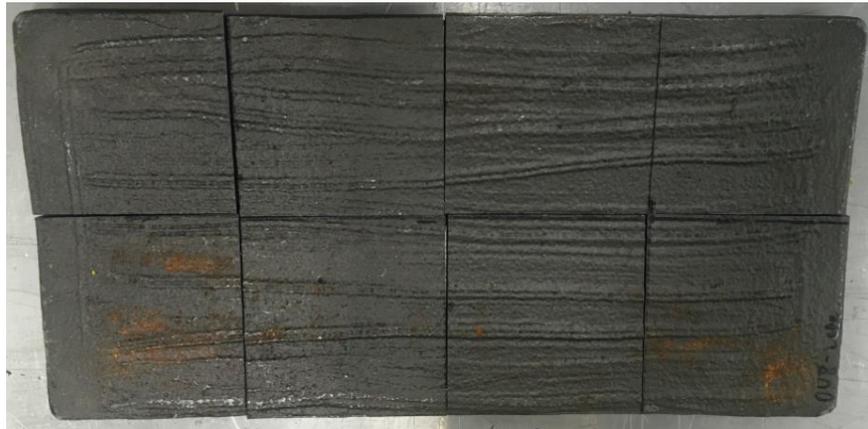


Figure 4-6. Rectangular plate cut into 8 square samples.

However, the square samples could not actually be used to evaluate the surface roughness, orientation, current type, and magnetic field strength. This is because the orientation

between the magnetic field lines and the crack would be difficult to be determined due to the changing contact points between the copper pads and the part as the part is rotated due to the shape of the part. This phenomenon can be illustrated in Figure 4-7A, when the square part is placed flat, the orientation between the magnetic field line direction and the crack is 0-degrees. However, when the part is rotated 15 degrees counterclockwise as shown in Figure 4-7B, the contact point between the copper pads and the part changes causing a shift in current direction, magnetic field direction and crack orientation. This makes it difficult to determine the orientation between the magnetic field lines and the crack when rotating a square part.

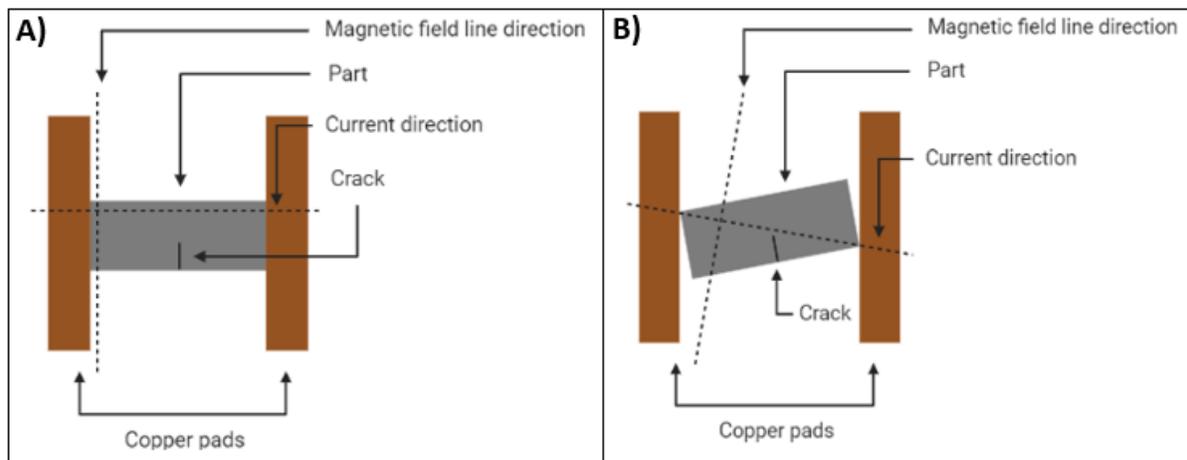


Figure 4-7. A) 0-degree orientation between the magnetic field line direction and the crack; B) The part is rotated 15 degrees counterclockwise from the part in A), however the contact point between the copper pads and the part is shifted hence it is hard to determine the orientation between the magnetic field lines and the crack when rotating a square part.

To ensure the orientation between the magnetic field lines and the defect can be controlled in the experiments, the square samples were machined into round puck-shaped parts as shown in Figure 4-8. This will allow for the contact points between the copper pads and the part to remain in the same location when rotating the part which will keep the current direction and magnetic field direction constant. The only changing parameter as the part is rotated is the defect orientation which simplifies the process of testing different orientations between the

magnetic field direction and defect significantly. This can be visualized in Figure 4-9A, when the puck-shaped part is placed with the crack in line with the magnetic field direction, the orientation between the magnetic field line direction and the crack is 0-degrees and when the part is rotated 15 degrees counterclockwise as shown in Figure 4-9B, the orientation between the magnetic field lines and the crack is 15 degrees. Hence, surface and subsurface defects were manufactured on a 2-inch diameter puck-shaped samples to ensure the orientation parameter can be controlled during experiments.



Figure 4-8. A puck-shaped sample machined from casting plate.

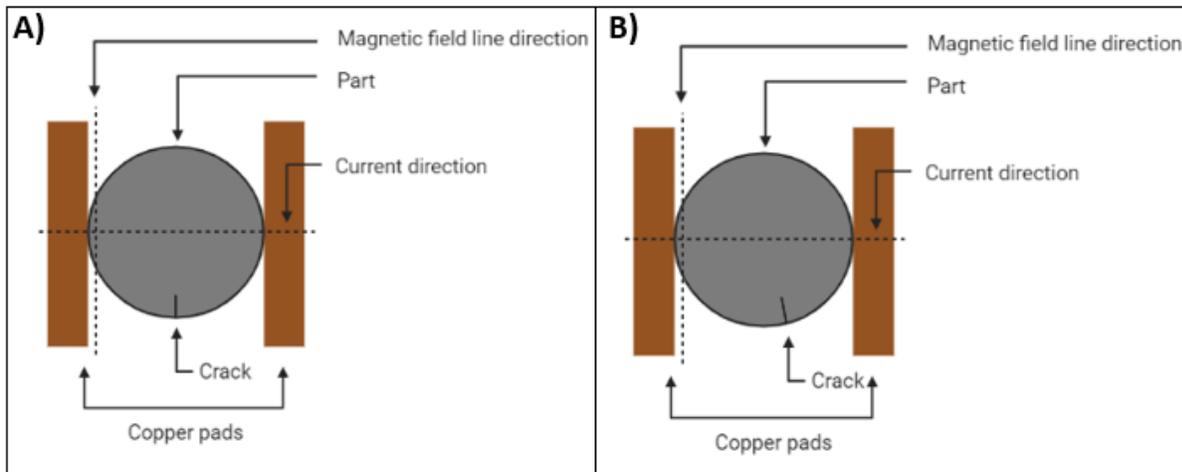


Figure 4-9. A) 0-degree orientation between the magnetic field line direction and the crack; B) The part is rotated 15 degrees counterclockwise from the part in A, since the contact points between the copper pads and the part is the same as in A, the magnetic field direction remains the same as in A.

Creating Manufactured Defects

Sub-surface indications were manufactured in the form of drilled holes with a diameter of 1.78mm (0.07 in) and a depth of 1.78mm (0.07 in) from the inspected surface of the part as shown in Figure 4-10. Generating surface defects required a more extensive procedure compared to manufacturing sub-surface defects. Figure 4-12 outlines the step-by-step procedure used to manufacture close-lipped surface cracks. First, a 1-inch slot was cut using a band saw and pressure was applied on the opposite end of the part using a hydraulic press to initiate the crack. The desired length of the crack can be determined in this step by slowly applying pressure and stopping when the desired length of crack is reached. In step 2, the part is rotated 180 degrees and pressure was applied directly on the 1-inch slot using a hydraulic press to close the crack. Step 3 illustrates how the part should look after steps 1 and 2. In step 4, material was removed from the rectangular plate to create a 2-inch in diameter puck-shaped part containing the close-lipped crack. Step 5 shows the puck-shaped part containing the crack where one side has as-cast surface and the other side is machined. In this paper, 4 puck-shaped samples were made with 2 containing subsurface indications and 2 containing surface cracks (see Figure 4-12).

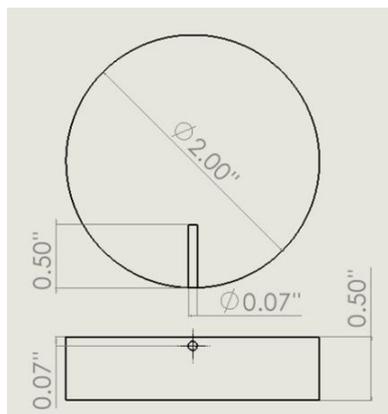


Figure 4-10. The dimensions of the drilled hole which was utilized as sub-surface defects.

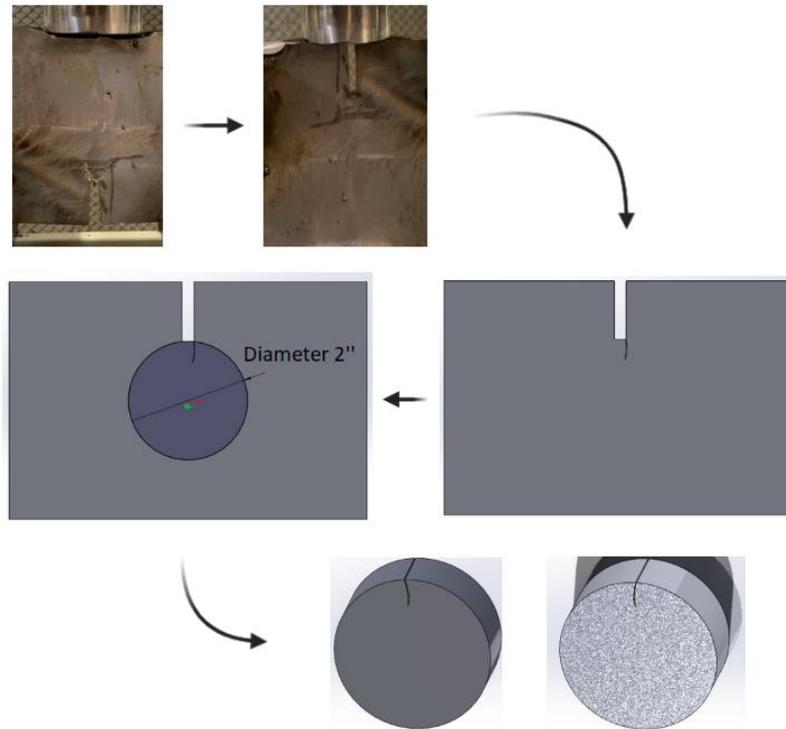


Figure 4-11. The step by step procedure to create puck shaped samples containing surface defects.



Figure 4-12. The puck-shaped samples containing sub-surface and surface defects.

Experimental Procedure and Design

A horizontal stationary magnetic particle machine (MD3-2060, Magnaflux®) was used in this study. To ensure the particles (CAS# 1309-37-1, Magnaflux®) were circulating evenly, the sides of the reservoir were scrubbed, and the machine was left running for a minimum of 30 minutes. The MPI test procedure in this study adhered to the ASTM 1444 standard. The

magnetic particles used in this study were fluorescent coated particles with a mean particle size of 6 μ m which corresponds to an estimated level of indication detection of 8 to 9 on an AS 5282 tool steel ring specimen. An angle meter (700, Johnson, Wisconsin) was used to ensure the part was at a minimum of 45 degrees when mounted on the bench as specified by the ASTM 1444 standard. The ASTM A802 was used to classify the surface roughness of the samples.

Excessive handling of the sample was avoided because residual particles from the glove can be imparted to the part which will introduce unwanted noise during image analysis. Hence, the samples were brushed with a stainless-steel brush (54022SP, Osborn, Indiana) once mounted in between the cooper pads to remove particles due to handling. A gauss meter (5180, FW Bell, Oregon) was used to measure the magnetic field strength. After each MPI tests, photographs were taken under ultraviolet light (UV) using a camera (PX-45, Crack Check, China) at a fixed distance of 305 mm (12 in) from the surface of the part to the camera. The camera was set to ISO 400 instead of automatic to standardize the camera's sensitivity to light. The explanatory variables in this study were surface roughness, orientation, current type, and magnetic field strength. Table 4-1 outlines all the explanatory variables along with the levels tested.

Table 4-1. The explanatory variables and the levels tested in this experiment

Explanatory variables	Levels
Surface roughness	A1, A2, A3
Orientation (degrees)	0, 15, 30, 45, 60, 75, 90
Current type	Alternating, Direct
Magnetic field strength (gauss)	15, 30, 45

The response variable used in this study was a refined version of the noise area percentage (NAP) metric developed by the author in a past study. The NAP metric quantifies the

level of difficulty in identifying a defect by comparing the average green intensity of the smallest rectangle encompassing the discontinuity to the representative surface patch near the discontinuity (see Figure 4-13A) [10]. A revised version of the NAP metric was developed for this study to better emulate how a human operator would identify defects. Humans typically identify discontinuities by identifying a pattern of bright areas that resemble common defects. For example, if a crack was present, the human operator would notice bright areas in the pattern of a line rather than the smallest rectangle surrounding the crack. Hence, this paper used a modified version of NAP which compares the average green intensity of five linear points along the discontinuity to the representative surface patch near the discontinuity (see Figure 4-13B).

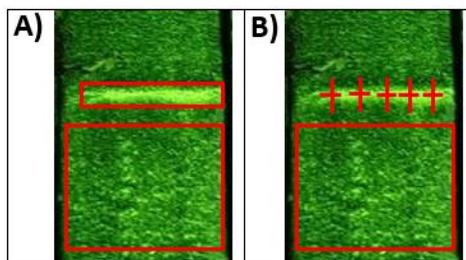


Figure 4-13. A) The smallest rectangle containing discontinuity and the representative surface patch near the discontinuity; B) Five linear points along the defect and the representative surface patch near the discontinuity.

Software Tool

The programming language used to make the software tool to obtain NAP was C#. The primary reasons behind using this programming language were familiarity with the language and ease of software implementation with Windows systems. Figure 4-14 shows a standard session of the NAP software tool with the key elements of the graphical user interface highlighted.

Element 1 is the “Instructions” button where a complete documentation outlining how to use the software will appear as a new window. Element 2 is the segment of the user interface containing four buttons to trigger events such as 1) to upload the image, 2) select the surrounding area, 3) select the area containing the defect, and 4) calculate the noise area percentage metric. Element 3

displays the coordinates and green intensity value for the selected points along the defect. This allows for the user to check if the correct areas of the image were selected, and the intensity of the point selected as a sanity check. For example, if there is a sudden drop in green intensity value might indicate that the user did not select an area containing the defect. Element 4 provides the user with the average green value of the selected points along the defect. Element 5 shows the percentage of pixels in the rectangle that have higher green intensity values than the average of the defect, which is also known as the NAP value. Element 6 allows the user to add or delete the data for each image. Lastly, the user can export the data as an Excel file by using the button in Element 7.

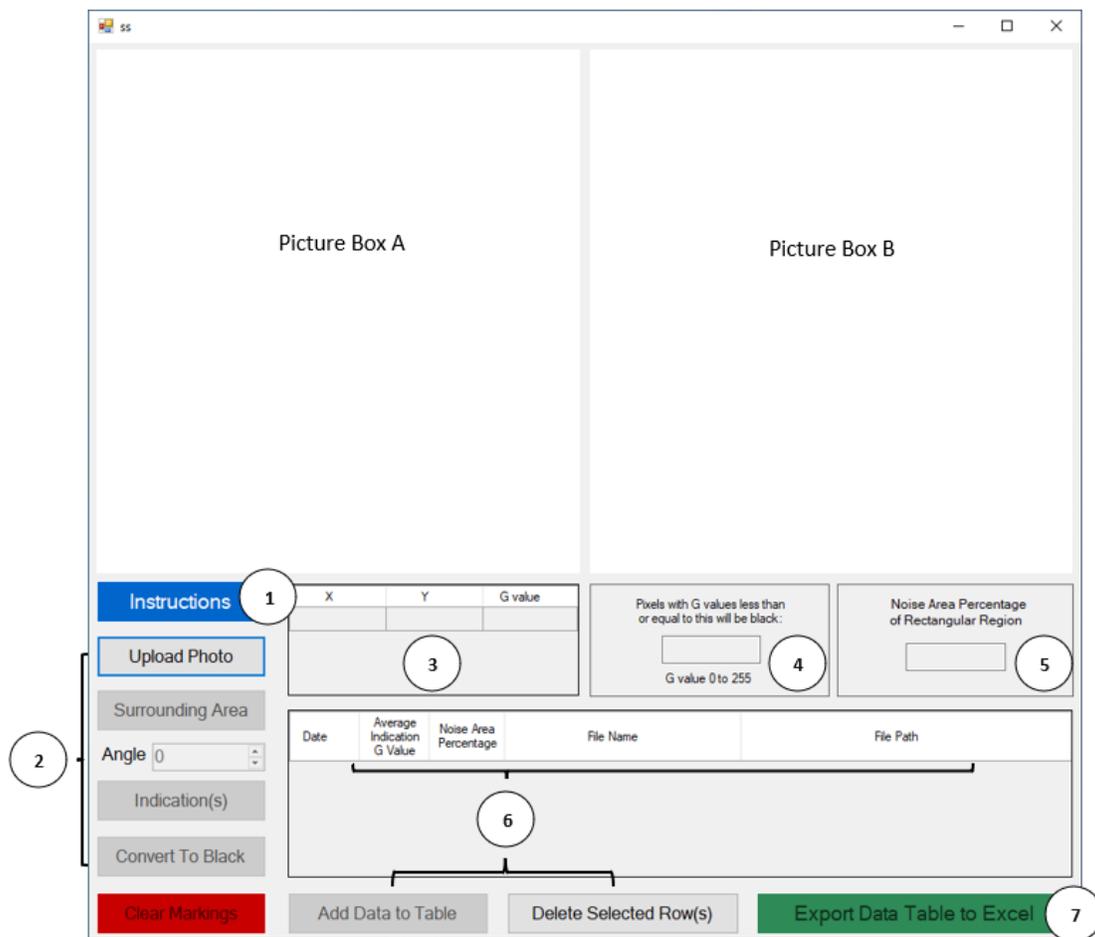


Figure 4-14. A standard NAP software tool session is shown, with the most important elements of the user interface labeled.

Results and Discussion

The ANOVA analysis found orientation, surface roughness, and the interaction between orientation and surface roughness to have a significant impact on the level of difficulty in identifying surface-breaking and sub-surface defects. Any defects with a NAP of 20% or higher were not able to be identified visually in the image. Hence, a red line is placed at this threshold. Orientation in this section refers to the orientation between the magnetic field lines and the defect. Current type and magnetic field strength were not found to be significant factors. The results for surface-breaking and subsurface defects will be discussed next.

Surface-breaking defects

Figure 4-15 shows the effect of orientation between the defect and magnetic field direction on the level of difficulty in identifying surface-breaking defects on two different surface roughness levels. For the smoother surface (A1), the result shows that the surface-breaking discontinuity cannot be identified when orientation between the defect and magnetic field lines is 0 degrees and for a portion of the instances when the orientation between the defect and magnetic field lines is at 15 degrees. For the rougher surface (A3), the result shows that the surface-breaking discontinuity cannot be identified when orientation between the defect and magnetic field lines is 0 degrees and a portion of the instances when orientation between the defect and magnetic field lines is at 15 and 30 degrees. This means that a surface-breaking defect at 30 degrees orientation to the magnetic field line on an A1 surface will be detected whereas this defect might be missed on rougher surfaces.

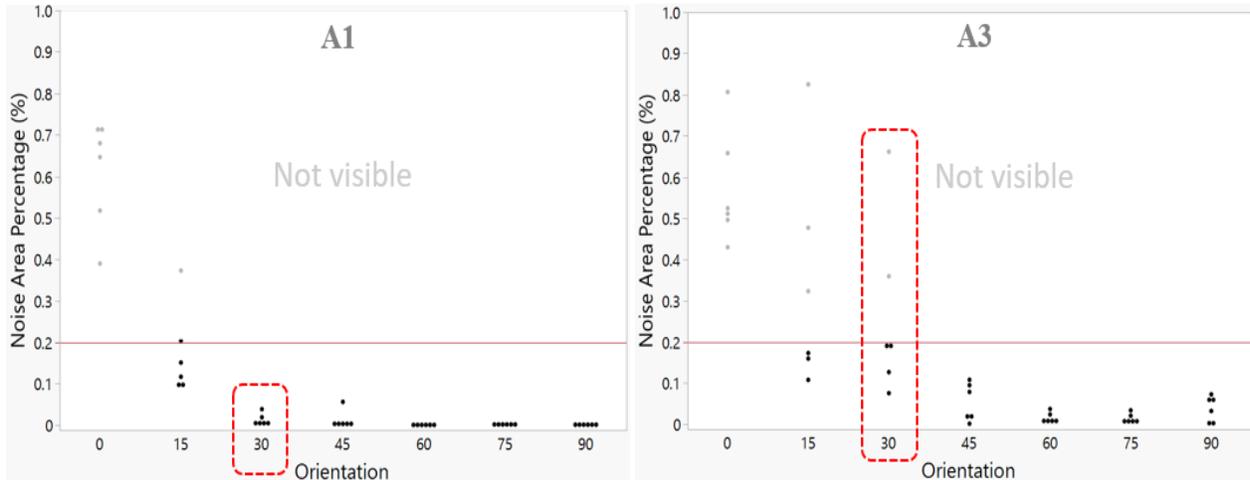


Figure 4-15. Left: The results for surface-breaking defect for orientation for A1 roughness; Right: The results for surface-breaking defect for orientation for A3 roughness.

Overall, the results for surface breaking defects corroborates the guidelines for section 6.2.5 in ASTM 1444 which states “To ensure the detection of discontinuities in any direction, each part must be magnetized in a minimum of two directions at approximately right angles to each other”. Based on the results in this study, if a part is magnetized in two directions, the surface-breaking defect should be able to be detected (see Figure 4-16). Surface roughness was also found to be significant where a higher surface roughness was found to increase level of difficulty in identifying discontinuities.

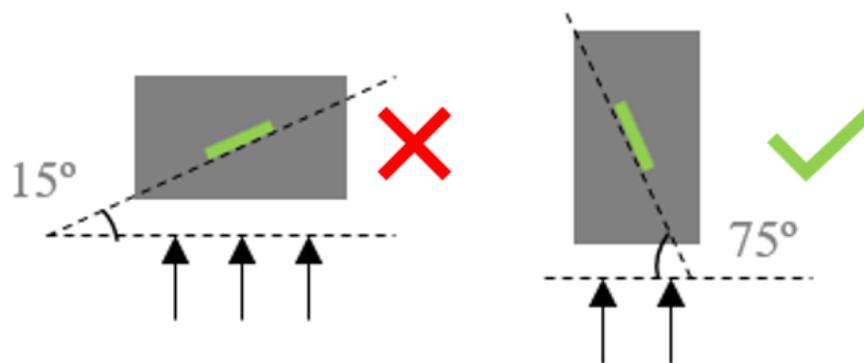


Figure 4-16. Based on the results in Figure 4-15, a surface defect might be missed at 15-degree orientation, however, if tested at right angles from the initial setup, it will be detected

Sub-surface defects

Figure 4-17 shows the effect of orientation on the level of difficulty in identifying sub-surface defects on two different surface roughness levels. For the smoother surface (A1), the result shows that the sub-surface discontinuity can be identified when the defect is at 90 degrees to the magnetic field lines and a portion of the instances when the defect is at a 75 degrees orientation to the magnetic field lines. For the rougher surface (A3), the result shows that the sub-surface discontinuity can only be identified when the defect is at 90 degrees to the magnetic field lines. This means that a sub-surface defect at 75 degrees orientation to the magnetic field line on an A1 surface have a chance to be detected whereas this defect would be missed on rougher surfaces.

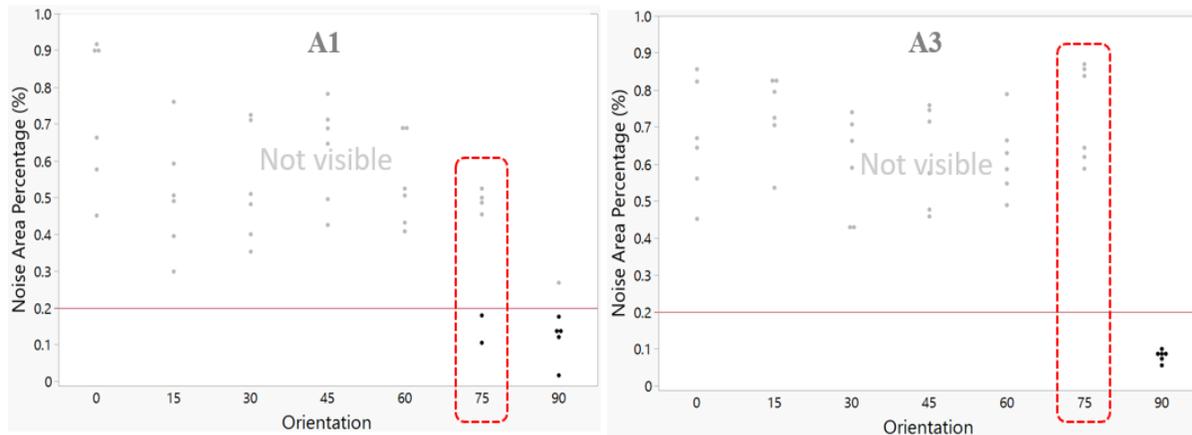


Figure 4-17. Left: The results for sub-surface defect for orientation for A1 roughness; Right: The results for sub-surface defect for orientation for A3 roughness.

The results for sub-surface defects show that the magnetic field lines and the defect have to be at 90 degrees to be detected. This makes identifying sub-surface defects in wet MPI difficult. It is important for steel foundries to be aware of this to be able to set reasonable standards. Based on the results, sub-surface defects of the size used in this experiment should not

be part of the requirement of defects to be identified by the operator since the process is not capable of doing this effectively.

Impact of geometry on the direction of the magnetic field lines

The orientation between the magnetic field lines and the defect is shown to have a significant impact on the ability to find the defect. For sub-surface defects, the results show that the defect is only visible at a 90-degree orientation between the magnetic field line and the defect. Since orientation is an important factor, there is a need to understand effect of geometry on the orientation of the magnetic field lines since castings have complex geometries. To investigate the impact of geometry on the magnetic field direction, four parts were used (see Figure 4-18).



Figure 4-18. 1) Flat geometry with flush contact points, 2) Curved geometry with flush contact points 3) Flat geometry with irregular contact points, 4) Curved surface with irregular contact points

A gauss meter with a transverse probe is used in this experiment to provide the magnetic field strength which can be used to estimate the magnetic field direction. The transverse probe which has a flat tip was placed orthogonal to the part's surface and rotated in four different orientations as shown in Figure 4-19. Theoretically, the orientation where the magnetic field direction is perpendicular to the probe will produce the highest magnetic field strength (an example shown in Figure 4-20). The magnetic field direction can then be mapped onto the tested part as shown in Figure 4-21. Several locations of the part will be tested for magnetic field direction in the experiment which will be illustrated by a blue dot. The statistically analysis to

determine the highest magnetic field strength for each point on the part can be found in the Appendix.

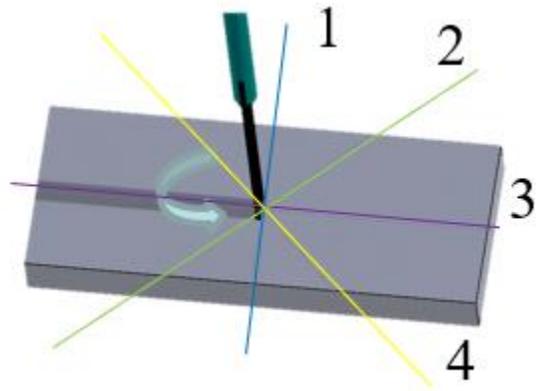


Figure 4-19. Transverse probe rotated along 4 axes

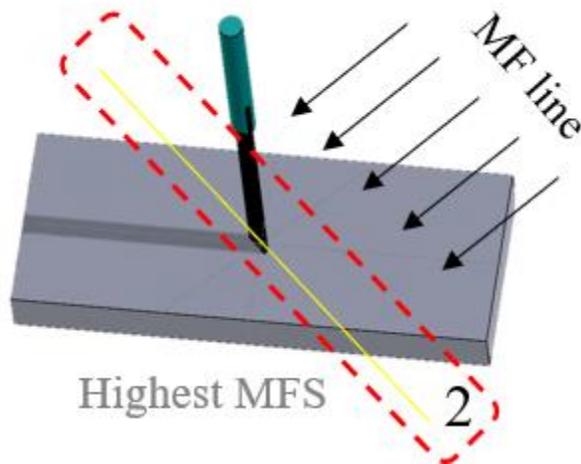


Figure 4-20. Illustration of how the magnetic field line is determined

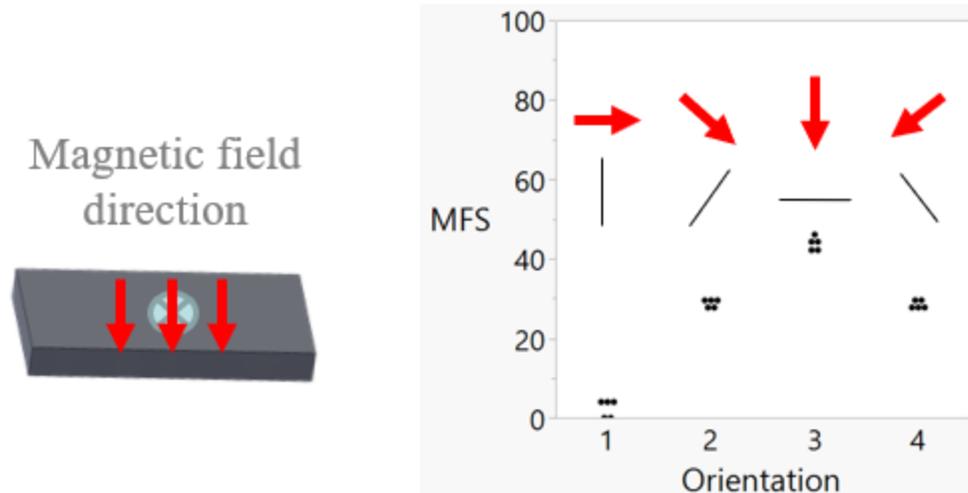


Figure 4-21. The mapping of the magnetic field direction on the part based on magnetic field strength

The first two samples that were tested were rectangular blocks with flat or curved surfaces. The rectangular sample allows for the contact points of the part on the copper pad to be flush and even across the whole part. The rectangular block with flat surfaces was setup horizontally on the bench (see Figure 4-22 Left). The results of the magnetic field lines were as expected based on how the part was setup (see Figure 4-22 Right). Next, the rectangular block with curved surfaces was tested with horizontal and vertical setup on the bench. For the horizontal setup, the results of the magnetic field lines were as expected based on how the part was setup (see Figure 4-23 Left). For the vertical setup, the result of the magnetic field lines for Point A was as expected based on how the part was setup, however, the magnetic field direction for Point B was not in line with the direction that would be expected based on the setup (see Figure 4-23 Right). This was a concern because a subsurface defect in the tight corner would be missed since the true orientation of the magnetic field line is not the same as the expected magnetic field line.



Figure 4-22. Left: Rectangular block setup on bench; Right: Orientation of magnetic field direction on the tested points of the rectangular block

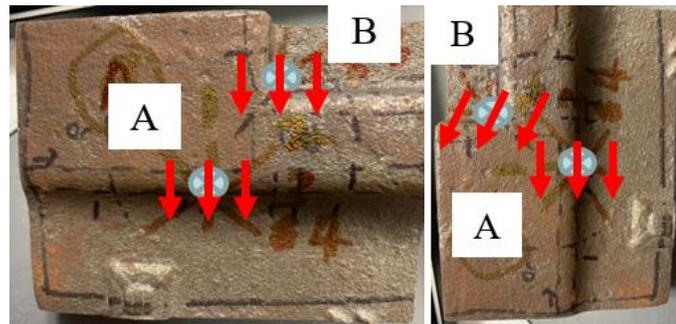


Figure 4-23. Left: Horizontal setup with the orientation of magnetic field direction on the tested points of rectangular block with curved surfaces; Right: Vertical setup with the orientation of magnetic field direction on the tested points of rectangular block with curved surfaces

Castings often have complex geometries which tends to create irregular contact points when setting up the part to be tested for wet MPI. Therefore, irregular contact points with flat and curved surfaces were tested. For the flat surface with irregular contact points, the result of the magnetic field lines for Point A was as expected based on how the part was setup, however, the magnetic field direction for Point B and Point C was not in line with the direction that would be expected based on the setup (see Figure 4-24). The magnetic field direction for Point B and Point C were found to be caused by the irregular contact between the copper pads and the part as shown in Figure 4-25. For curved surfaces with irregular contact points in horizontal and vertical setup, the result of the magnetic field lines for Point A was as expected with based on how the part was setup however the magnetic field direction for Point B and Point C was not in line with

the direction that would be expected based on the setup (see Figure 4-26). This could be caused by the curved surfaces and/or the irregular contact points. This shows that subsurface defects for casting will be difficult to locate because the magnetic field line direction varies based on the geometry of the surface and the contact points with the copper pads.

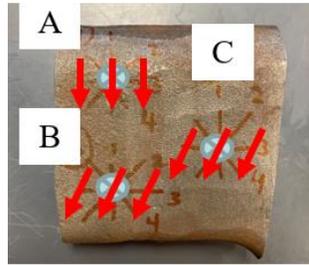


Figure 4-24. The orientation of magnetic field direction on the tested points of flat surface with irregular contact points

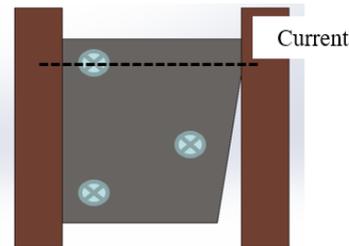


Figure 4-25. Irregular contact point illustrated in a computer aided design (CAD) drawing to show current flow

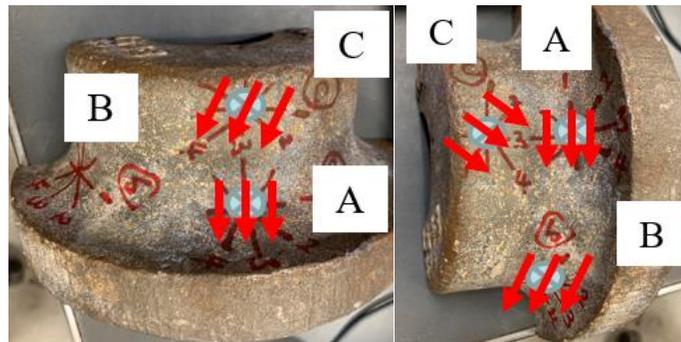


Figure 4-26. Left: Horizontal setup with the orientation of magnetic field direction on the tested points of curved surface with irregular contact points; Right: Vertical setup with the orientation of magnetic field direction on the tested points of curved surface with irregular contact points

Overall, the only setup where the magnetic field direction matches the expected magnetic field direction is with the part containing a flat geometry and flush contact with the copper pads. The other setups had areas on the part where the magnetic field direction did not match the expected magnetic field direction. The overview of the magnetic field direction mapped for each sample is as shown in Figure 4-27. For sub-surface indications where the magnetic field direction and the defect would need to be perpendicular to be detected, this experiment shows that it would be difficult to locate sub-surface defects even if the sub-surface defect's location and orientation were known prior to testing because the geometry of the part causes variability in the magnetic field direction.

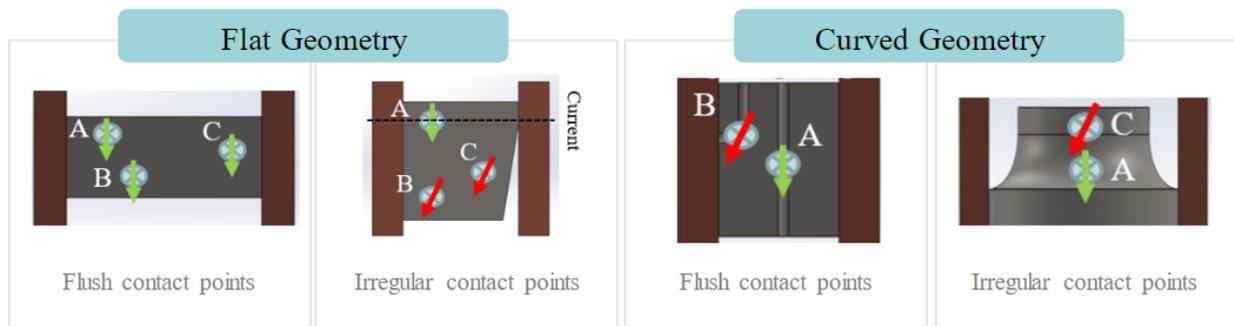


Figure 4-27. Summary of findings with the green arrows indicating that the actual magnetic field line matches the expected whereas the red arrows indicating that the actual magnetic field line does not match the expected

Conclusions

Understanding impact of process parameters and part characteristic on the effectiveness of wet MPI is important to drive process improvement. For the effect of geometry on the magnetic field strength, the results show that magnetic field strength decreases as the radius of the concave region increases. This finding is important to note because defects usually occur in concave or corner regions in casting. For the experiment on the impact of surface roughness, orientation, current type, and magnetic field strength, the results indicate that for surface-

breaking and sub-surface defects, surface roughness and orientation were found to impact the level of difficulty in identifying the discontinuity. It was found that for surface-breaking defects, a crack in any orientation can be detected by magnetizing a part in two orientations. A surface-breaking defect at 30 degrees orientation to the magnetic field line on a smoother surface (A1) can be detected whereas this defect might be missed on rougher surfaces. On the other hand, for sub-surface defects, only at a 90-degree orientation between the defect and the magnetic field lines can a majority of the discontinuities be identified. A sub-surface defect at 75 degrees orientation to the magnetic field lines on an A1 surface have a chance to be detected whereas this defect would be missed on rougher surfaces. For both surface-breaking and sub-surface defects, rougher surfaces make it harder to identify discontinuities.

A major advantage of metal casting is its' ability to produce complex geometry. However, geometry impacts the magnetic field direction. Because orientation between the defect and the magnetic field lines is shown to be a significant factor in defect identification, the effect of geometry on the magnetic field direction was investigated. The results show that sub-surface defects can most likely only be identified on parts containing flat geometry and flush contact points. Additionally, the location and orientation of the sub-surface defect need to be known prior to testing to be able to detect the indication. This is important for sub-surface defects because the orientation between the defect and the magnetic field lines need to be exactly perpendicular to be detected. Overall, this means that subsurface defects cannot be effectively identified in castings because of their complex geometry.

References

- [1] S. M. Y. Lau, D. Eisenmann, and F. E. Peters, "Development of an Image Analysis Protocol to Define Noise in Wet Magnetic Particle Inspection," *Int. J. Met.*, vol. 15, pp. 1317–1325, Jan. 2021.

- [2] K. Woolley, T. Woolley, and B. Banfield, “A fresh initiative on the use of daylight magnetic particle inspection for the inspection of underwater steel structures,” *Underw. Technol.*, 2016.
- [3] P. Staněk and Z. Škvor, “Automated Magnetic Field Evaluation for Magnetic Particle Inspection by Impulse,” *J. Nondestruct. Eval.*, 2019.
- [4] R. Goodwin, “Choose the most reliable MPI method for your business: magnetic particle inspection methods vary in terms of ease of use as well as the level of detail,” *Quality*, vol. 53, no. 10, 2014.
- [5] D. Lovejoy and D. Lovejoy, “Magnetic particles, their characteristics and application,” in *Magnetic Particle Inspection*, 1993.
- [6] D. Eisenmann and D. Enyart, “Review of progress in magnetic particle inspection,” in *AIP Conference Proceedings*, 2014, p. 1581.
- [7] D. J. Eisenmann, D. D. Enyart, D. Kosaka, and C. Lo, “Fundamental Engineering Studies of Magnetic Particle Inspection and Impact on Standards and Industrial Practice,” 2014.
- [8] B. L. Luk and A. H. S. Chan, “Human Factors and Ergonomics in Dye Penetrant and Magnetic Particles Nondestructive Inspection Methods,” *Adv. Ind. Eng. Oper. Res.*, vol. 0948, no. August, pp. 127–142, 2008.
- [9] M. Wright, “How to implement a PoD into a highly effective inspection strategy,” *NDT Canada 2016 6th Int. CANDU In-Service Insp. Work. Nov 15-17, Burlington, (NDT Canada 2016)*, 2016.
- [10] C. A. Harding and G. R. Hugo, “Guidelines for Interpretation of Published Data on Probability of Detection for Nondestructive Testing,” pp. 1–20, 2011.
- [11] J. Y. Lee, S. J. Lee, D. C. Jiles, M. Garton, R. Lopez, and L. Brasche, “Sensitivity Analysis of Simulations for Magnetic Particle Inspection Using the Finite-Element Method,” *IEEE Trans. Magn.*, 2003.

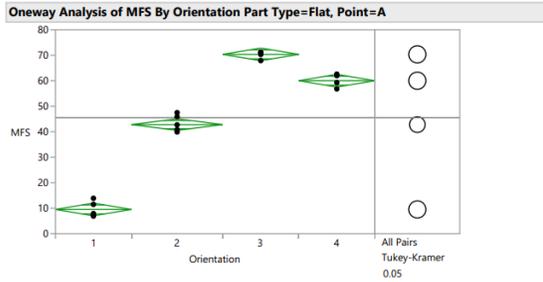
Appendix A: Statement of Authorship

Conceptualization - Sharon Lau; methodology - Sharon Lau and David Eisenmann;
 software - Sharon Lau and Aron Mitchell; validation - Sharon Lau; formal analysis - Sharon Lau;
 investigation - Sharon Lau; data curation - Sharon Lau, Grace Peterson, and Colton Richardson;

writing (original draft preparation) - Sharon Lau; writing (review and editing) - Sharon Lau, David Eisenmann, and Frank Peters.

Appendix B: Supporting Documentation

Flat geometry with irregular contact points



Oneway Anova

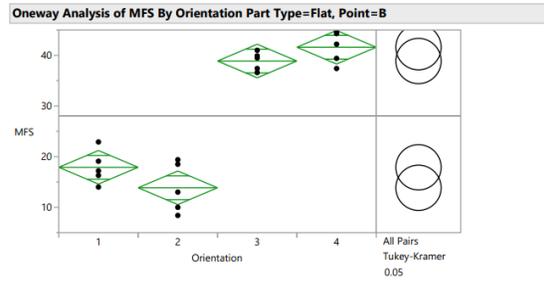
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Orientation	3	10655.799	3551.93	507.6408	<.0001*
Error	17	118.948	7.00		
C. Total	20	10774.747			

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
3	1	60.82000	1.672954	56.06453	65.57547	<.0001*
4	1	50.50000	1.672954	45.74453	55.25547	<.0001*
2	1	33.28000	1.601732	28.72698	37.83302	<.0001*
3	2	27.54000	1.601732	22.98698	32.09302	<.0001*
4	2	17.22000	1.601732	12.66698	21.77302	<.0001*
3	4	10.32000	1.672954	5.56453	15.07547	<.0001*



Oneway Anova

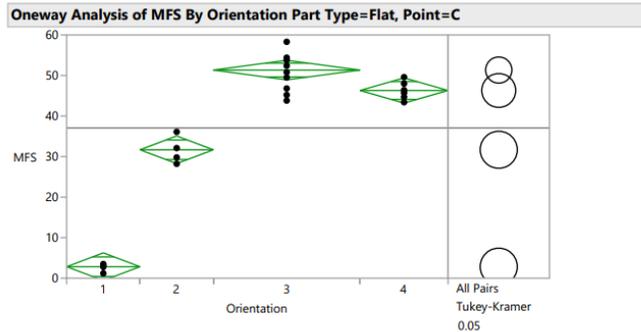
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Orientation	3	3026.3480	1008.78	82.5940	<.0001*
Error	16	195.4200	12.21		
C. Total	19	3221.7680			

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
4	2	27.74000	2.210317	21.4162	34.06376	<.0001*
3	2	25.02000	2.210317	18.6962	31.34376	<.0001*
4	1	23.70000	2.210317	17.3762	30.02376	<.0001*
3	1	20.98000	2.210317	14.6562	27.30376	<.0001*
1	2	-4.04000	2.210317	-2.2838	10.36376	0.2969
4	3	2.72000	2.210317	-3.6038	9.04376	0.6174



Oneway Anova

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Orientation	3	8551.2114	2850.40	210.3591	<.0001*
Error	22	298.1040	13.55		
C. Total	25	8849.3154			

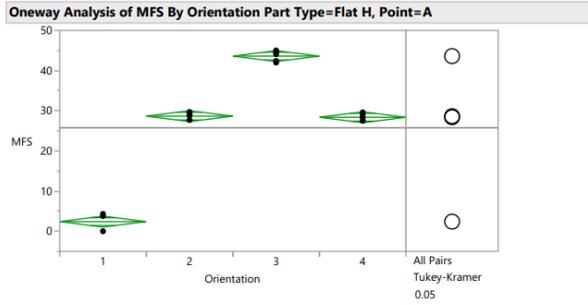
Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
3	1	48.50000	2.016198	42.9013	54.09865	<.0001*
4	1	43.46000	2.228991	37.2705	49.64954	<.0001*
2	1	28.84000	2.328105	22.3752	35.30477	<.0001*
3	2	19.66000	2.016198	14.0613	25.25865	<.0001*
4	2	14.62000	2.228991	8.4305	20.80954	<.0001*
3	4	5.04000	1.900890	-0.2385	10.31846	0.0648

Flat geometry with flush contact points



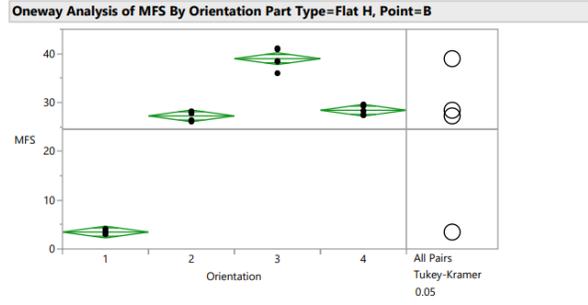
Oneway Anova

Analysis of Variance					
Source	DF	Squares	Mean Square	F Ratio	Prob > F
Orientation	3	4396.1480	1465.38	711.3508	<.0001*
Error	16	32.9600	2.06		
C. Total	19	4429.1080			

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Ordered Differences Report						
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
3	1	41.20000	0.9077445	38.6029	43.79708	<.0001*
2	1	26.26000	0.9077445	23.6629	28.85708	<.0001*
4	1	25.98000	0.9077445	23.3829	28.57708	<.0001*
3	4	15.22000	0.9077445	12.6229	17.81708	<.0001*
3	2	14.94000	0.9077445	12.3429	17.53708	<.0001*
2	4	0.28000	0.9077445	-2.3171	2.87708	0.9894



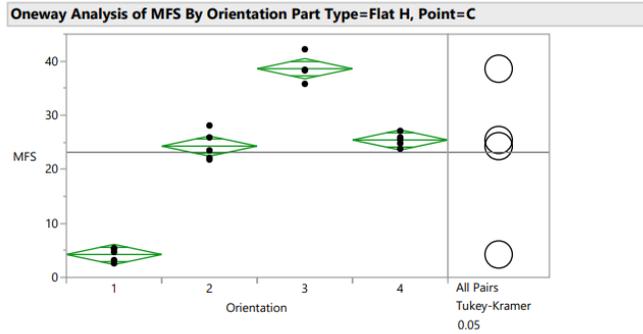
Oneway Anova

Analysis of Variance					
Source	DF	Squares	Mean Square	F Ratio	Prob > F
Orientation	3	3376.6620	1125.55	694.2507	<.0001*
Error	16	25.9400	1.62		
C. Total	19	3402.6020			

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Ordered Differences Report						
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
3	1	35.52000	0.8052950	33.2160	37.82397	<.0001*
4	1	24.96000	0.8052950	22.6560	27.26397	<.0001*
2	1	23.80000	0.8052950	21.4960	26.10397	<.0001*
3	2	11.72000	0.8052950	9.4160	14.02397	<.0001*
3	4	10.56000	0.8052950	8.2560	12.86397	<.0001*
4	2	1.16000	0.8052950	-1.1440	3.46397	0.4939



Oneway Anova

Analysis of Variance					
Source	DF	Squares	Mean Square	F Ratio	Prob > F
Orientation	3	3013.5975	1004.53	258.1185	<.0001*
Error	16	62.2680	3.89		
C. Total	19	3075.8655			

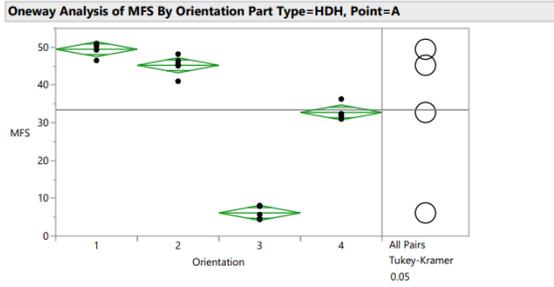
Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Ordered Differences Report						
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
3	1	34.36000	1.247678	30.7904	37.92963	<.0001*
4	1	21.18000	1.247678	17.6104	24.74963	<.0001*
2	1	20.04000	1.247678	16.4704	23.60963	<.0001*
3	2	14.32000	1.247678	10.7504	17.88963	<.0001*
3	4	13.18000	1.247678	9.6104	16.74963	<.0001*
4	2	1.14000	1.247678	-2.4296	4.70963	0.7979

Curved geometry with irregular contact points

Horizontal setup



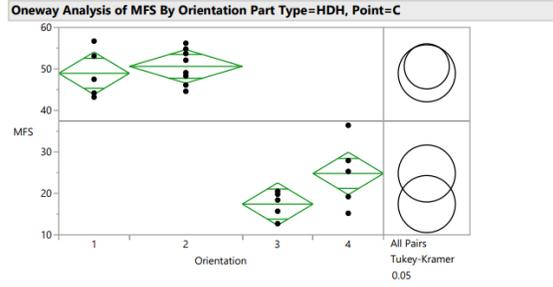
Oneway Anova

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Orientation	3	5711.4760	1903.83	433.5251	<.0001*
Error	16	70.2640	4.39		
C. Total	19	5781.7400			

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
1	3	43.34000	1.325368	39.54810	47.13190	<.0001*
2	3	39.10000	1.325368	35.30810	42.89190	<.0001*
4	3	26.60000	1.325368	22.80810	30.39190	<.0001*
1	4	16.74000	1.325368	12.94810	20.53190	<.0001*
2	4	12.50000	1.325368	8.70810	16.29190	<.0001*
1	2	4.24000	1.325368	0.44810	8.03190	0.0258*



Oneway Anova

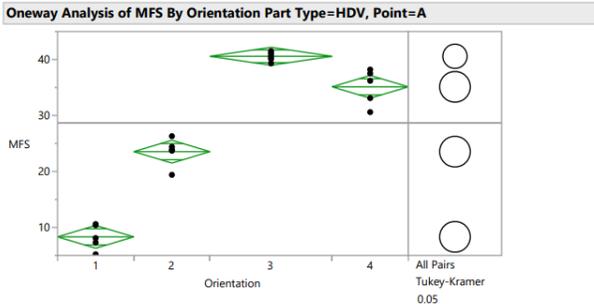
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Orientation	3	4852.2104	1617.40	54.0131	<.0001*
Error	19	568.9488	29.94		
C. Total	22	5421.1591			

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
2	3	33.19250	3.119618	24.4206	41.96439	<.0001*
1	3	31.52000	3.460906	21.7885	41.25154	<.0001*
2	4	25.81250	3.119618	17.0406	34.58439	<.0001*
1	4	24.14000	3.460906	14.4085	33.87154	<.0001*
4	3	7.38000	3.460906	-2.3515	17.11154	0.1788
2	1	1.67250	3.119618	-7.0994	10.44439	0.9491

Vertical setup



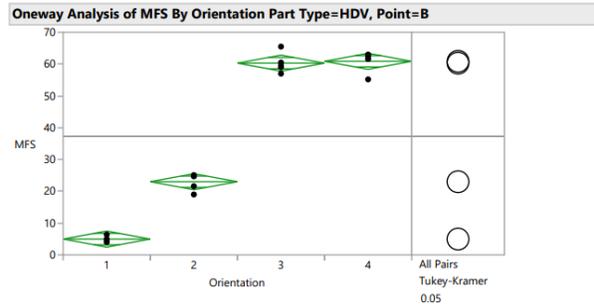
Oneway Anova

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Orientation	3	3544.0413	1181.35	248.3936	<.0001*
Error	19	90.3630	4.76		
C. Total	22	3634.4043			

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
3	1	32.25500	1.243255	28.75916	35.75084	<.0001*
4	1	26.80000	1.379268	22.92171	30.67829	<.0001*
3	2	17.03500	1.243255	13.53916	20.53084	<.0001*
2	1	15.22000	1.379268	11.34171	19.09829	<.0001*
4	2	11.58000	1.379268	7.70171	15.45829	<.0001*
3	4	5.45500	1.243255	1.95916	8.95084	0.0017*



Oneway Anova

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Orientation	3	11667.586	3889.20	548.7207	<.0001*
Error	16	113.404	7.09		
C. Total	19	11780.990			

Means Comparisons

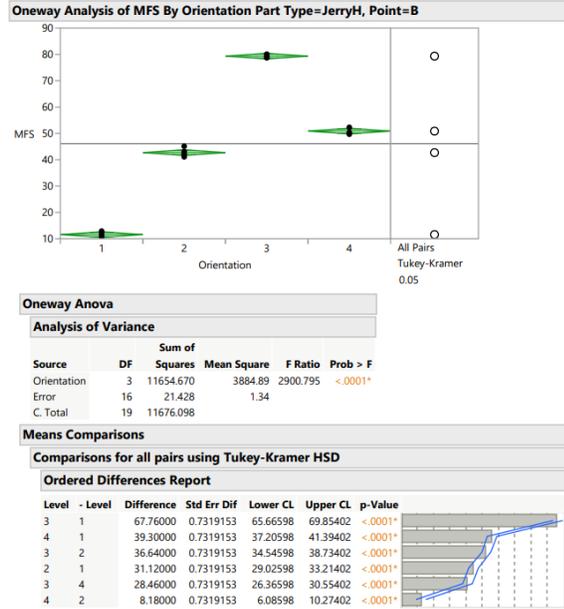
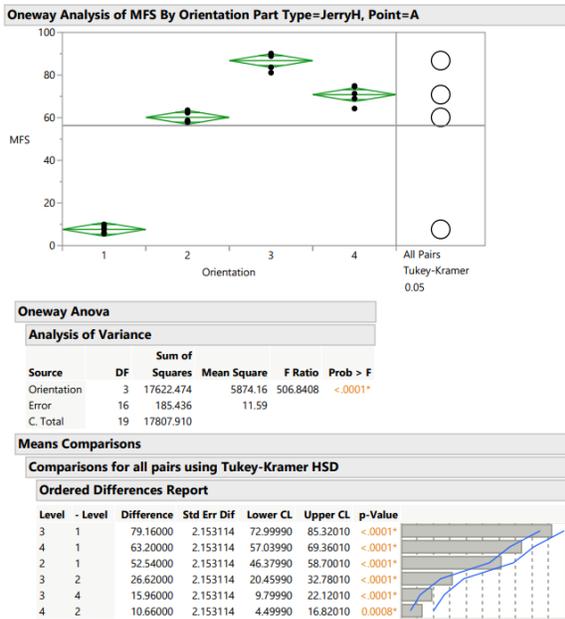
Comparisons for all pairs using Tukey-Kramer HSD

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
4	1	55.90000	1.683776	51.0827	60.71732	<.0001*
3	1	55.32000	1.683776	50.5027	60.13732	<.0001*
4	2	37.86000	1.683776	33.0427	42.67732	<.0001*
3	2	37.28000	1.683776	32.4627	42.09732	<.0001*
2	1	18.04000	1.683776	13.2227	22.85732	<.0001*
4	3	0.58000	1.683776	-4.2373	5.39732	0.9854



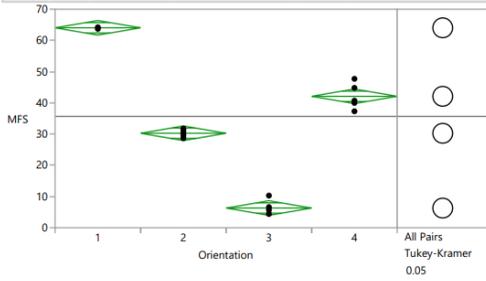
Curved geometry with regular contact points

Horizontal setup



Vertical setup

Oneway Analysis of MFS By Orientation Part Type=JerryV, Point=A



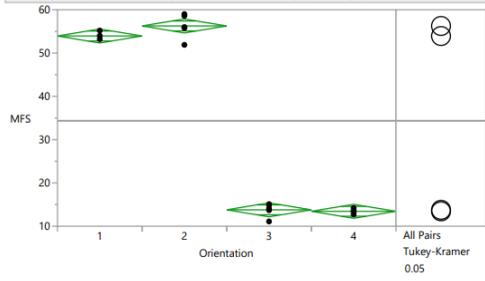
Oneway Anova					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Orientation	3	8690.3335	2896.78	474.5704	<.0001*
Error	16	97.6640	6.10		
C. Total	19	8787.9975			

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Ordered Differences Report						
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
1	3	57.74000	1.562562	53.26948	62.21052	<.0001*
4	3	35.80000	1.562562	31.32948	40.27052	<.0001*
1	2	33.78000	1.562562	29.30948	38.25052	<.0001*
2	3	23.96000	1.562562	19.48948	28.43052	<.0001*
1	4	21.94000	1.562562	17.46948	26.41052	<.0001*
4	2	11.84000	1.562562	7.36948	16.31052	<.0001*

Oneway Analysis of MFS By Orientation Part Type=JerryV, Point=B



Oneway Anova					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Orientation	3	8620.8455	2873.62	991.6712	<.0001*
Error	16	46.3640	2.90		
C. Total	19	8667.2095			

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Ordered Differences Report						
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
2	4	42.82000	1.076615	39.7398	45.90022	<.0001*
2	3	42.48000	1.076615	39.3998	45.56022	<.0001*
1	4	40.50000	1.076615	37.4198	43.58022	<.0001*
1	3	40.16000	1.076615	37.0798	43.24022	<.0001*
2	1	2.32000	1.076615	-0.7602	5.40022	0.1783
3	4	0.34000	1.076615	-2.7402	3.42022	0.9887

CHAPTER 5. INSPECTOR AID FOR WET MAGNETIC PARTICLE INSPECTION

Sharon May Yen Lau*, David Eisenmann**, Frank Earl Peters*

*Department of Industrial and Manufacturing Systems Engineering, Iowa State University

**Department of Agricultural and Biosystems Engineering, Iowa State University

Modified from a manuscript to be submitted to *The International Journal of Advanced Manufacturing Technology*

Abstract

In the steel casting industry, magnetic particle inspection (MPI) is a common nondestructive testing method used to detect cracks that are normally invisible to the naked eye. This is achieved by creating a cluster of magnetic particles at the crack openings which allows the cracks to be detected via visual inspection. However, the reliability of visual inspection has been shown in past studies to be poor with a high escape rate of around 70%. To reduce the escape rate, an automated crack detection method was proposed in this paper. The proposed method uses an entropy-driven grid labeling approach to optimize the tradeoff between time and accuracy of the annotations. The deep learning model used was Faster R-CNN with a ResNet-101 backbone for training and feature extraction. Faster R-CNN was a viable baseline model that could handle the various environmental conditions present in wet MPI images with an escape rate of 13% and an overkill rate of 38%. The deep learning model reduced the escape rate of defects by 57% when compared to the current human inspection method. Additionally, the implementation strategy for this model is not to fully replace human visual inspection in MPI, but to serve as an aid to the human operator. Using this hybrid approach, the advantages of each

agent can be leveraged. The automated vision system can provide stable predictions of defects whereas the human operator might perform better when it comes to discerning if a crack is acceptable or unacceptable. Overall, this study demonstrates the potential improvement in crack detection through the use of artificially intelligent vision systems for MPI in steel castings.

Introduction

Nondestructive testing (NDT) is necessary to ensure the reliable performance of the steel casting [1]. Wet MPI is an NDT method that is commonly used in the steel casting industry to detect discontinuities during the manufacturing process and to prevent failure in the field [2]. Prior to running any tests, bench preparation is required for getting good test results. With wet MPI, a suspension of magnetic particles in carrier fluid will be attracted to the area of the defect and will make the discontinuity light up. This suspension sits in the reservoir of the bench and is sent through a tube to spray liquid on the part during the test. The level of particles in the liquid is important to keep consistent so the results are accurate, reliable, and repeatable between different tests. Once bench preparation is completed, wet MPI can be performed. The optimal process parameters need to be used to increase the chances of locating defects. Visual inspection is then performed to locate the defects that become visible after wet MPI is performed. Lastly, the part is demagnetized and if there are any defects found, a combination of welding and grinding operations are then performed [3].

Inspection errors can stem from a variety of factors but generally come from task, environmental, individual, organizational, and social factors [4]. More specifically, the variability in visual inspection in wet MPI can be traced to the human operator's level of qualifications, vision acuity, and fatigue level [5]. Based on past research, most inspection tasks have around a 20% to 30% error rate [6]. However, visual inspection in wet MPI has a higher complexity level compared to regular visual inspection. For one, it is not a typical accept or

reject inspection task rather each defect needs to be identified and repaired [7]. Additionally, the visual inspection takes place in a dark room where the part is illuminated under ultraviolet (UV) light. The amount of visible light or UV light present can impact the inspector's ability to locate defects. Another factor contributing to inspector variability as a result of the lighting setup in wet MPI is the inspector's visual acuity and dark adaptation. Because the inspection takes place in a dark room, the eyes typically take 20 to 30 minutes to adapt to the darkness; however, this process will vary depending on one's eye health [8], [9]. Since many factors contribute to the variability in the visual inspection process in wet MPI [10], past researchers have conducted an empirical study to evaluate the measurement error in wet MPI and found that it equates to an escape rate of around 70% as mentioned in Chapter 3.

The current escape rate of 70% is alarming because defects that are not identified could result in the part failing in the field. Although the quality of inspection can typically be improved through training, inspection procedure, and equipment [11], it is time-consuming, expensive, and difficult to scale across all foundries. To overcome this problem, automated inspection techniques are a more appropriate solution. The goal however is not to fully replace the human operator but to assist the human operator with defect detection.

Computer vision systems for defect detection have become a vital need in many manufacturing applications. Object detection in images can be done through various traditional techniques such as 1) blob analysis, 2) contour processing, 3) matching, 4) color processing, 5) texture analysis, and 6) classification. Classification is used for object detections through methods such as gaussian models, neural networks, or support vector machines. In classification, objects are assigned to a specific set of classes defined by special attributes such as color or shape. These features have to be learned by training the model based on a set of known objects.

Once the model is trained, the classifier is able to compare the features of objects in a new image to the features of the trained classes and returns the class with the highest score. This method is commonly used in situations where objects have resemblances but there are multiple variations such as lighting, shape, orientation, or color. For an object with a defined shape, shape-based matching is typically used, however, with objects where shapes are similar, but the similarities are not distinct, classification would be a better fit. Classification could be applied to several applications such as 1) image segmentation, 2) object recognition, 3) quality control, 4) novelty detection or 5) optical character recognition. Although traditional classification approaches have been widely used to solve various applications, the use of deep learning is known to have better accuracy and flexibility [12].

The advantages of deep learning are its adaptability and robustness to variations in the images for similar objects. Convolutional Neural Networks (CNNs) represent a class of deep learning neural networks and have been the gold standard for image classification since 2012. CNNs are the commonly used artificial neural network when predicting on image data because it performs well on feature extraction and preserves the spatial information of the image. There are several challenges with this approach, for example, if the image was taken at different depths, angles, or with different aspect ratios, this would lead to the need for a larger number of images to train a model to account for all the different spatial locations and distortions of the object in the images. This would result in the need for high computational power. Hence, other variants of CNNs have been developed to reduce the computational power needed to successfully train these models.

Recent work in object detection in manufacturing applications show the potential for deep learning approaches to be adopted more widely. However, research to combine artificially

intelligent solutions and MPI in steel castings is relatively untouched. Tout et. al. developed a model called ClassNet which uses CNN with a modified ResNet-34 architecture catered to maximize performance based on the characteristics of the MPI dataset. Precision and recall were used to evaluate the performance of the model. Precision is the number of true positives divided by the total number of predictions while recall is the number of true positives divided by the total number of ground truths. ClassNet was able to achieve a precision of 82.11% and a recall of 80.7%. However, ClassNet was only trained for MPI inspection of crankshafts where 130 images were gathered across 4 crankshafts [13]. This model will likely not perform well for the steel casting industry since most foundries produce a wide variety of parts with a large array of different geometries. Additionally, there is no publicly available dataset for MPI images mainly due to privacy restrictions.

Given the need for a system that can be applied to a larger variety of parts in the steel casting industry, this paper proposed a novel approach to reduce the escape rates of defects. Due to the lack of publicly available MPI images, the authors of this paper visited nine different steel foundries and gathered 300 images. An additional 10 images came from publicly available images on the internet. The challenges faced in developing a publicly available dataset will be highlighted in the Conclusions section. This paper proposes an entropy-driven grid labeling approach using the Faster R-CNN framework with a ResNet-101-FPN backbone. The goal of this method is to reduce escape and overkill rates of defects in wet MPI of steel casting and to demonstrate the benefits of a computer-assisted inspection approach in wet MPI.

Methods

In this section, the novel annotation approach of the object will be highlighted and discussed. The proposed method uses entropy-driven grid-style bounding box annotations. Compared to the standard box-level annotations, annotating with smaller grids reduces the

amount of background labeled as a crack and is less labor-intensive and time-consuming than pixel-level annotations. The specific grid size for each image is determined by maximizing Shannon's entropy. This allows for optimal grid sizes to be used based on the defect size in each image. This is important because grid sizes that are too small will have a negative impact on the model's performance since the characteristics of the crack are lost while grid sizes that are too large will incorrectly label large areas of the background as defect. Grid-style labeling is an effective method for cracks specifically because the crack can be segmented without losing the context of the object.

Grid-style labeling

Grids have been used extensively in past research to classify pavement cracks. The typical process involves processing images into grids of different scales i.e. 10 by 10 pixels, 20 by 20 pixels, etc. and cropping each grid out to be fed into the neural network as either containing the crack or background [14], [15], [16], [17]. For pavement cracks, this approach makes sense since cracks are occurring on a flat surface which reduces the need for spatial context. However, castings are known for their complex geometry and the geometry impacts where cracks tend to show up in a part. For example, sharp corners are more prone to developing cracks when undergoing heat treatment. To preserve the spatial context of the image, this study chose not to crop the images into small patches using grids but rather use the grids to do patch-level segmentation. This means that a high-resolution image (1024 by 1024 pixels) is fed into the model but instead of using a single bounding box to represent each crack, the box-level annotation is split up into multiple smaller bounding boxes. Figure 5-1 illustrates the difference in the areas of the images labeled as ground truth in the traditional method using a single bounding box per crack method versus the proposed grid-style labeling approach.

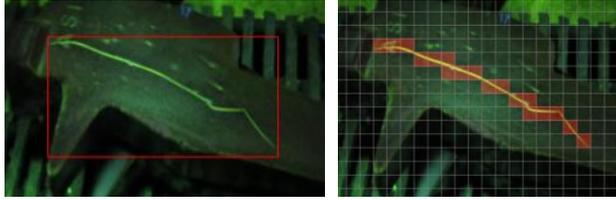


Figure 5-1. Left: One bounding box per crack; Right: Grid style labeling (25 by 25 pixels)

The advantages of using this approach specifically for identifying cracks in wet MPI, are that the spatial information is preserved, and it reduces the amount of background regions labeled as crack. Using the grid style labeling approach, one important factor that needs to be considered is the grid size. If the grid size is too large, the accuracy will suffer because of the increase in the background region being labeled as crack. If the grid size is too small, the features of what makes a crack identifiable will be lost. Additionally, the grid size selection impacts labeling time. Smaller grid sizes will result in longer annotation times. Therefore, it is critical to select an appropriate grid size for each image. This study experimented with 3 different grid sizes 10 by 10 pixels, 25 by 25 pixels, and 50 by 50 pixels. To further improve the accuracy of the labeling aspect, a dynamic approach to determining grid size was also tested. The idea here is that each crack may have different characteristics, i.e., size, shape, etc. which means that there exists a different optimal grid size based on the crack characteristics. To account for the various characteristics of cracks, this study used Shannon's entropy as an objective approach to determine the grid size for each image dynamically which will be discussed in more detail in the next sub-section.

Shannon's entropy

Entropy is used to measure the level of uncertainty of an information source [18]. On a high level, an image that contains all black pixels will have a low entropy compared to an image with varying levels of intensities. The formula for Shannon's entropy is as defined in Equation 1,

Equation 1. Shannon's entropy [19]

$$\text{Shannon's Entropy} = - \sum_{i=1}^n P(x_i) * \log_2 p(x_i)$$

where $P(x_i)$ is the fraction of objects in the i -th case and n being the total number of elements. More specifically, Shannon's entropy measures the variability of elements within the set. To provide an intuitive example, Figure 5-2 illustrates the graphs of two different images. Based on the shape of the distributions, the graph on the right which has a more compact distribution with a peak intensity of 127 on the grayscale will have lower entropy compared to the graph on the left.

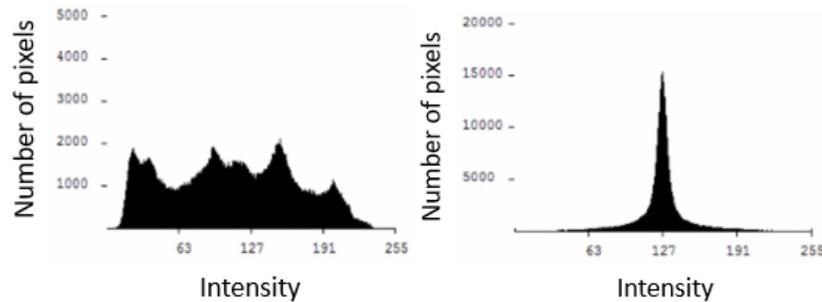


Figure 5-2. Two graphs represent the grayscale intensities of two different images. The graph on the right is more compact which means it will have a lower Shannon's entropy compared to the image on the left.

This measure to represent information in an image was leveraged for our application. To provide a visual interpretation of Shannon's entropy in wet MPI images, the authors mapped an example of the distribution of Shannon's entropy by overlaying a color-coded grid based on the percentile range of the entropy's value as shown in Figure 5-3. Red represents the 0 to 50th percentile range, orange represents the 50th to 75th percentile range, blue represents the 75th to 90th percentile range, and white represents the 90th to 100th percentile range. Overall, it can be observed that the regions with black pixels in the four corners of the image generally have the

lowest entropy. Additionally, the corner regions or cracks where intensities tend to vary have a higher entropy value.

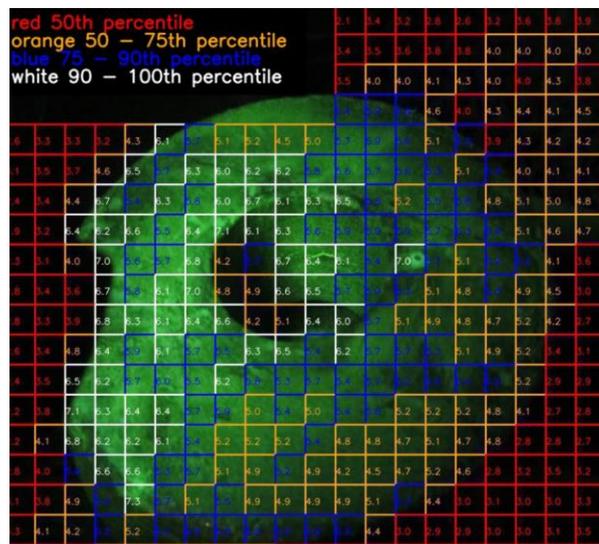


Figure 5-3. MPI image color-coded to represent various percentile ranges for Shannon's entropy for a grid size of 25 pixels

Although it is possible to compute Shannon's entropy for multiple grid sizes for the whole image, it is computationally expensive and would require additional processing steps to filter out the regions containing the crack. Therefore, this study proposes a simpler and faster approach to incorporate Shannon's entropy as a method for dynamic grid size selection. Figure 5-4 shows the proposed process:

1. The deep learning model is leveraged to provide the proposed region containing the crack
2. The entropy of multiple grid sizes is applied to this region (grid size 10 to 50 in steps of 5)
3. The entropy is plotted against grid size
4. The optimal grid size is selected based on maximizing entropy
5. The grid containing the crack is automatically labeled based on the trained model
6. The user can refine the grid selection if any mistakes were made by the model
7. Repeat steps 1 to 6 until all images are labeled

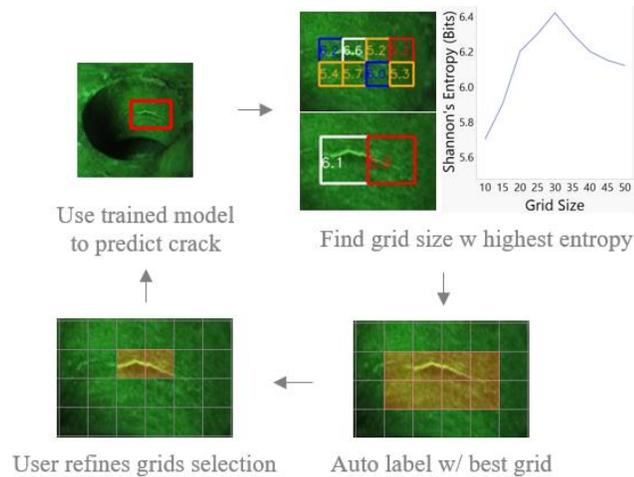


Figure 5-4. Process flow for dynamic selection of grid size using Shannon's entropy

The quantifiable benefits on the model's performance will be highlighted in the Experiments and Results section.

Experiments and Results

This section presents the dataset used in the experiments, the evaluation metrics, the hyper-parameters, and the model's performance. All experiments were conducted using Tensorflow [20] on an NVIDIA RTX 5000 GPU under a Windows 10 64-bit system. This study used Faster R-CNN with a ResNet-101 backbone which was fine-tuned from a pre-trained model on Microsoft COCO data set [21]. The Faster R-CNN is made up of two different modules which are a region proposal network (RPN) and a Fast R-CNN detector [22]. The RPN network provides localization of the cracks to the Fast R-CNN detector which then extracts the features and classifies the objects [23]. Additionally, YouOnlyLookOnce (YOLO) v3 with a Darknet-53 backbone was also tested to provide a comparison in performance between a two-stage method and a one-stage method. The Faster R-CNN model performed better than YOLO and was further optimized using entropy-driven grid-style labeling and data augmentation to improve the performance of the model.

Dataset

A dataset of 310 images that were used in this study were mainly gathered by visiting nine different steel foundries. Each foundry ran wet MPI on a variety of parts, which resulted in a diverse set of images. Additionally, each foundry had a different workspace leading to variation in lighting and background in the images acquired. For example, some foundries performed wet MPI in a dark room while others used blackout curtains to reduce ambient light in the test area. The differences in the lighting condition resulted in varying intensities of green in the image. The variation present in the images captured across the different foundries is illustrated in Figure 5-5.

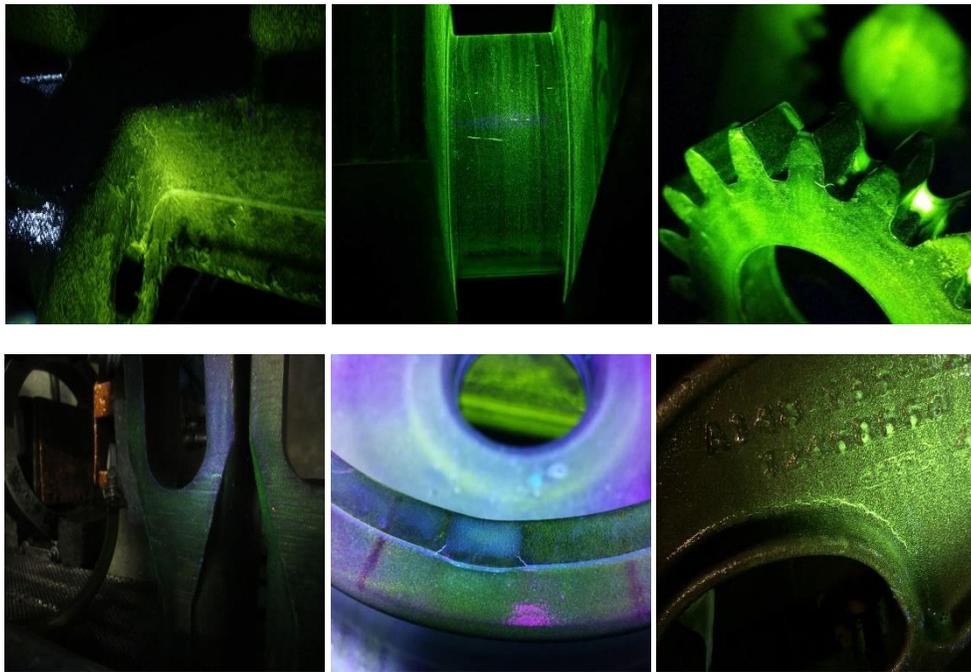


Figure 5-5. These images show the high variability in part geometry, illumination, crack characteristics, and image quality

Evaluation metric

Mean Average Precision (mAP) is one of the metrics most commonly used by CNN-based classifiers to evaluate a model's performance [24]. However, for manufacturing applications, the measurement error is typically quantified by calculating escape and overkill

rates. Hence, this paper used escape and overkill rates to evaluate the model. Overkill or type I error is when an area that does not have a defect is marked as having a defect. Escape or type II error is when an area containing a defect is not identified. Overkill and escape errors lead to increased costs and lower quality of parts.

Training details

Before evaluating grid-style labeling, this study compared the state-of-the-art one-stage approach, YOLO, against a two-stage approach, Faster R-CNN. One-stage approach models such as YOLO are known for their speed, which makes this the go-to method for applications that require real-time object detection [25]. However, one drawback of this method is the model tends to be less precise and has poorer performance on smaller objects due to the lower resolution output feature map [26], [27]. Two-stage methods such as Faster R-CNN tend to require more computational load but output more accurate predictions [23], [28]. This study evaluated the two models and found that the Faster R-CNN model performed better than the YOLO v3 model with 14% less escape and 8% less overkill. Hence, the authors used the Faster R-CNN model to be evaluated further in this study. The proposed method used Faster R-CNN model that was trained on the MS COCO dataset [29] as a starting point and fine-tuned the output layer of the model based on images of cracks from wet MPI. The dataset was split with 80 percent of the images used for training the model, 10 percent as validation to tune the model's hyper parameters, and the remaining 10 percent was used as the test set. Stochastic gradient descent (SGD) optimizer with a learning rate of 0.001 and momentum of 0.5 was used. The learning rate was set to gradually decay to 0.0001 during training. A mini batch size of 16 was used in this model and the number of epochs used to train the model was 15.

Design of experiments

As a baseline, the model was trained using ground truths that were annotated by the traditional method of using one bounding box per crack as shown in the left image of Figure 5-6. Subsequently, smaller bounding boxes were used which reduced the amount of background in the image labeled as crack as shown in the middle image of Figure 5-6. However, two clicks were required to create each bounding box, which was time-consuming. Additionally, the bounding boxes that were created using the traditional methods were irregular, making it hard to compare the impact of bounding box size on the model's performance. Hence, grid-style labeling was introduced to reduce the labeling time and as an easy way to control the size of bounding boxes (see right image of Figure 5-6).

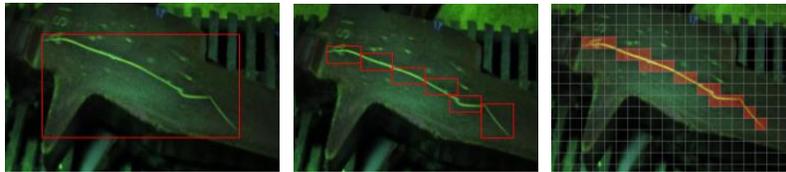


Figure 5-6. Left: One bounding box per crack; Middle: Smaller bounding boxes; Right: Grid style labeling (25 by 25 pixels)

With the grid style labeling method, three different sizes were tested, 10 by 10 pixel, 25 by 25 pixel, and 50 by 50 pixel. The size used was denoted by GSL (10), GSL (25), and GSL (50) where GSL stands for grid-style labeling and the number in the parenthesis represented the grid size used. The results showed that a grid size of 25 by 25 pixels performed the best when compared to using a single bounding box per crack, smaller bounding boxes, GSL (10), and GSL (50). Data augmentation was used to increase the sample size in the training and test datasets.

Using GSL (25), the training dataset was doubled by randomly applying two of the seven types of augmenters to each image in the training set (see Figure 5-7). With this approach, the overkill rate saw an incremental improvement of around 8% whereas no significant improvement

in the escape rate was observed. Tripling the training dataset with a similar approach resulted in poorer performance in both the overkill and escape rates which means that model is overfitting to the training data. Hence, the approach of doubling the training dataset by randomly applying two of the seven types of augmenters to each image in the training set was used (see Figure 5-8).

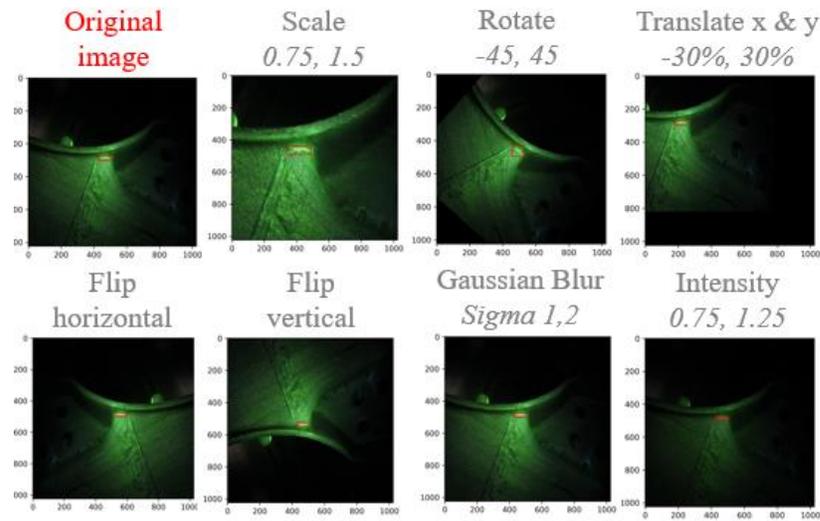


Figure 5-7. Type of augmenters applied

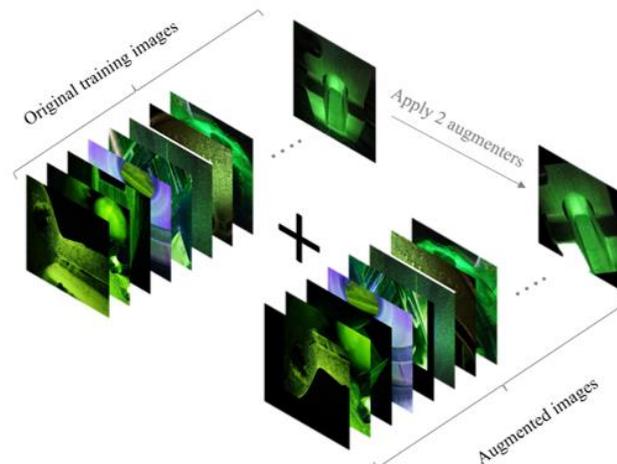


Figure 5-8. Data augmentation on the training set

Data augmentation was also applied to the test set to increase the probability of detecting the cracks in the image. After testing various combinations of augmenters, the approach that

produced the best results was creating 8 augmented images of each image in the test set and running the prediction on the original and augmented images. The final prediction is a combination of the predictions from the 9 images, which are then overlaid on the original image (see Figure 5-9). The 8 augmented images were a result of using the following augmenters 1) flip horizontal, 2) flip vertical, 3) rotate, 4) scale down, 5) scale up, 6) reduce the intensity, 7) increase intensity, and 8) gaussian blur. By adding augmentation in the training set and test set, the escape rate was reduced to 15%, however, the overkill increased to 28%. The results make intuitive sense because augmenting the test set provides more opportunity for the model to find the cracks but also increases the chances of the model predicting a false positive. Since the proposed implementation strategy is a hybrid approach with the predictions of the model being validated by a human, the more important metric was the escape rate since a missed defect could result in the part failing in the field and any overkill predictions can be later dismissed by the human operator.

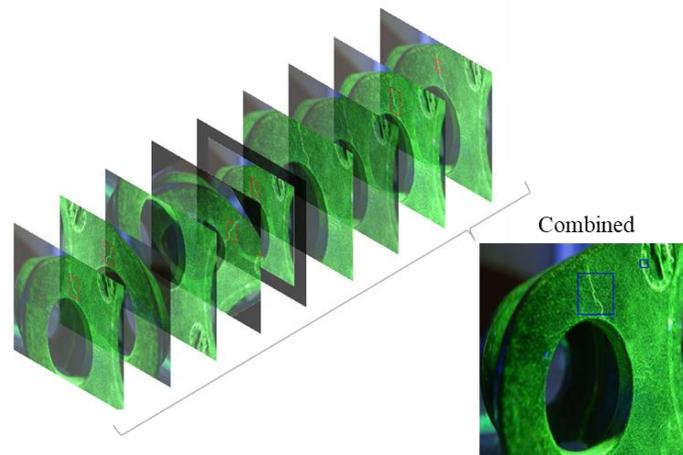


Figure 5-9. Data augmentation on the test set

The data augmentation strategy for the training and test dataset was then paired with an entropy-driven grid labeling approach (see Chapter 5 under Methods). By combining the

augmentation and entropy-driven labeling approach, the escape rate was reduced to 13%, however, the overkill increased to 32%. As previously mentioned, since the escape rate was considered the more important metric based on the implementation strategy, the entropy-driven labeling approach paired with the augmentation of the training and test set was found to perform the best. The optimized deep learning model has an escape rate of 13% which is better than the current human inspection method at 70% (See Figure 5-11).

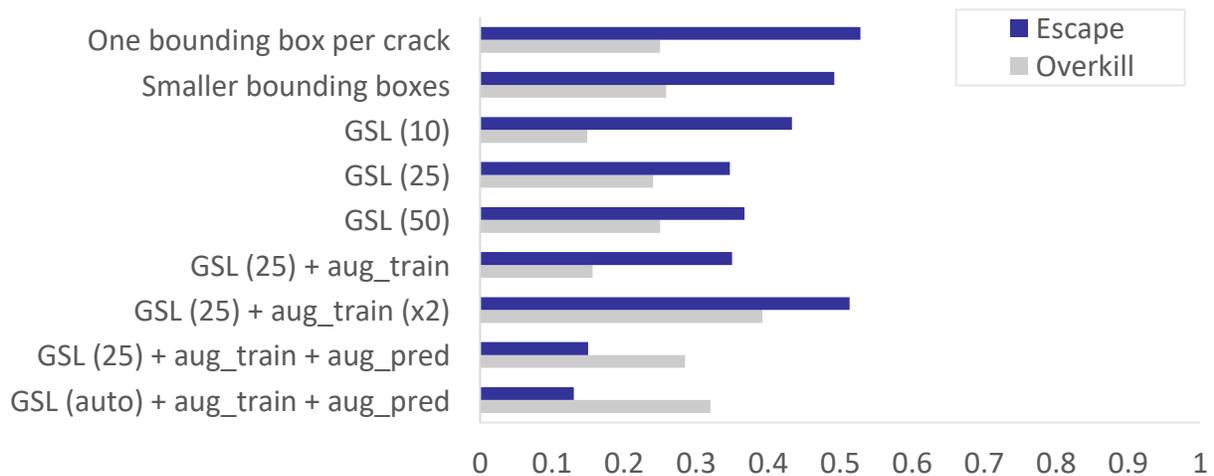


Figure 5-10. Bar graph of grid-style labeling experiments and data augmentation

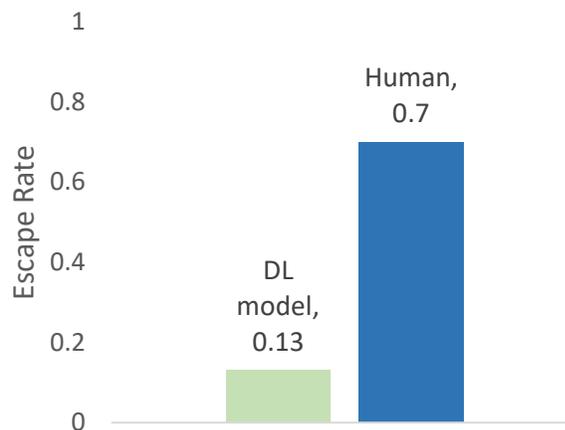


Figure 5-11. Deep learning model's performance versus human inspection

Proposed Implementation

Software

One major challenge observed in the process of gathering MPI images to train the model is the confidentiality agreement between the foundries and the buyers, which prevented the foundries from sharing images. This made it difficult to obtain the large quantities of images that are necessary to develop a good model. Although the current model can achieve an escape rate of 13%, further improvements are possible with the availability of more training images. Taking the confidentiality issue into consideration, the proposed implementation strategy would be to provide the current model trained with images from the nine different foundries and to allow foundries to build on this dataset by adding their images to the training set and train the model locally. This would allow them to not only leverage the dataset that was collected through this research but also to further build upon and improve the model by including examples from their specific foundry. Using this approach, the escape and overkill rates can be improved while minimizing privacy concerns.

The programming language used to make the software tool to label the image, train the model, and run predictions on images was Python. Familiarity with the language and ease of software implementation with Windows systems were the primary reasons for using this programming language. Additionally, Python has an extensive library for machine learning specific applications, which simplifies the development process and saves development time. Figure 5-12 shows the starting session of the Inspector Aid software tool with four buttons: label, train, evaluate model, and run live prediction. Element 1 in Figure 5-12 is the “Label” button where a graphical user interface (GUI) (see Figure 5-13) will appear as a new window that will allow the user to select a folder of images to label. Once a folder is selected, the images can be labeled in the GUI as shown in Figure 5-14, where each image will show up in the picture box

section. Elements 1 to 3 in Figure 5-14 allow the user to optimize the labeling accuracy, and Element 4 in Figure 5-14 allows the user to label the next image in the folder. The labeled images are then plotted in a new window in sets of four. Element 2 in Figure 5-12 is the “Train” button, where the labeled images will be used to train the model. Element 3 in Figure 5-12 is the “Evaluate Model” button where the trained model will be evaluated on the test set and the escape and overkill rate will be calculated. Lastly, Element 4 in Figure 5-12 is the “Run Live Prediction” button, which will allow the user to perform live prediction with a camera system based on the latest trained model.

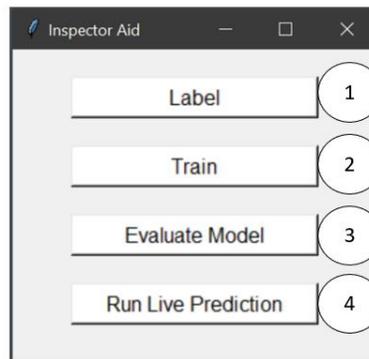


Figure 5-12. A standard Inspector Aid software tool session is shown, with each element of the user interface labeled

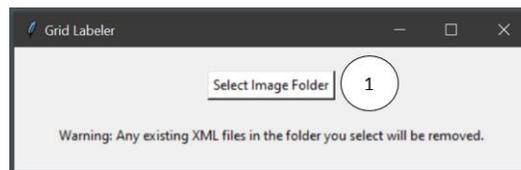


Figure 5-13. The new session is shown, where the user can select the folder of images that they want to label using the grid-style method

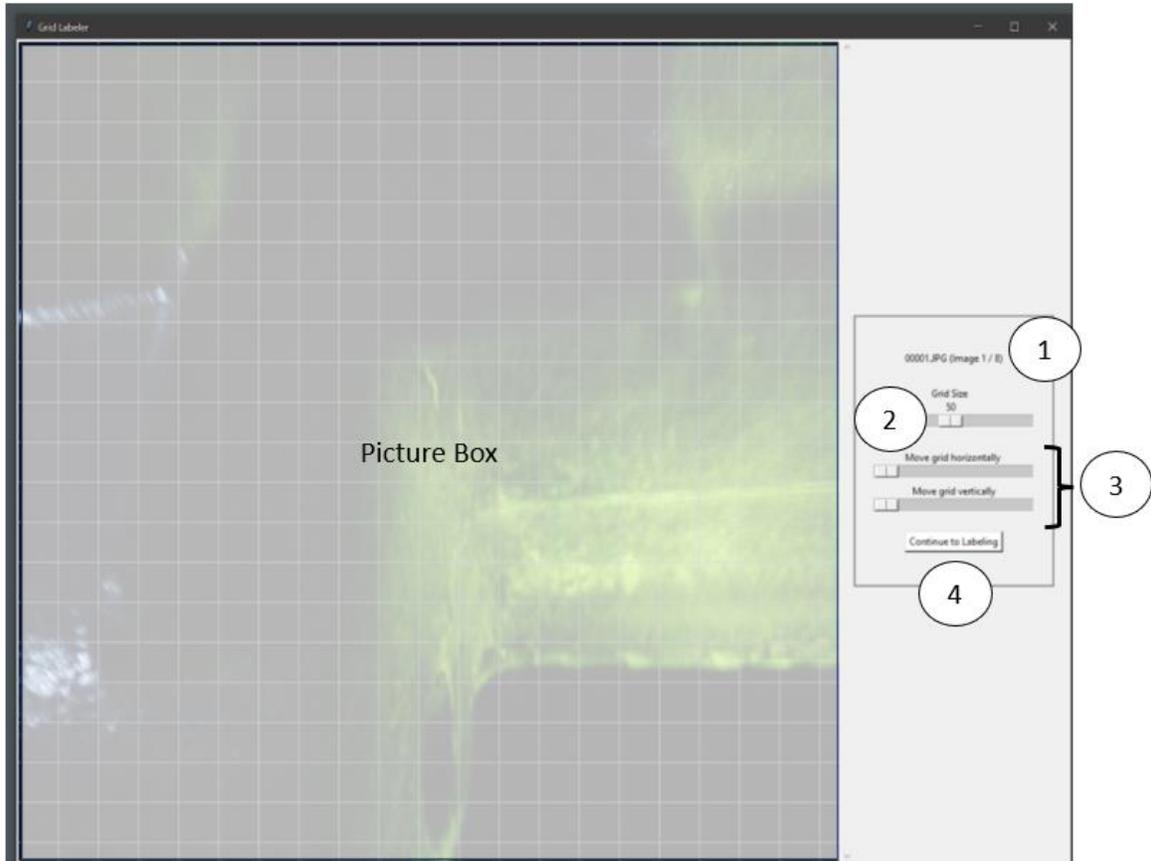


Figure 5-14. Once the folder is selected, each image is loaded to the picture box to be labeled

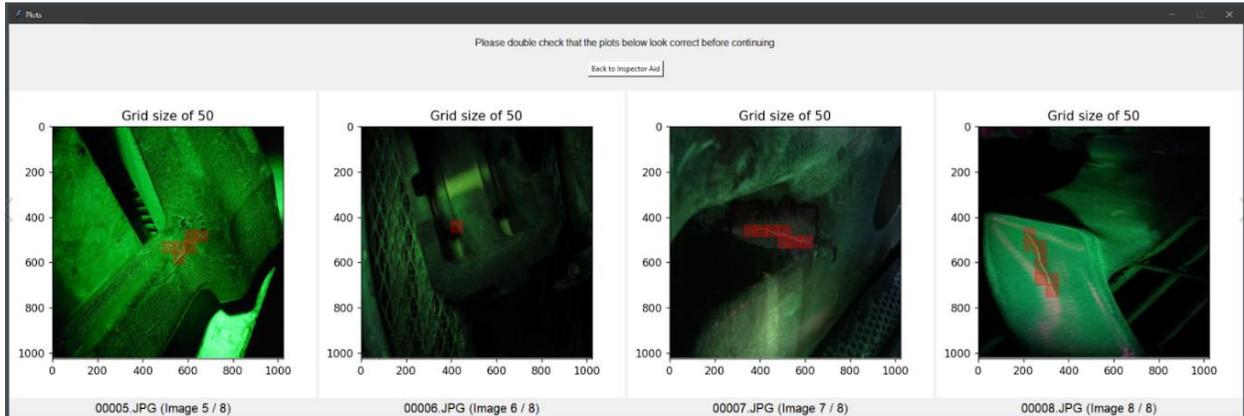


Figure 5-15. The plot of the labeled images in sets of four

Hardware

The specific equipment used in this research are as follows: 1) horizontal stationary magnetic particle machine (MD3-2060, Magnaflux®, Illinois), 2) Bench to ultraviolet (UV) light

mount (623365, Magnaflux®, Illinois), 3) UV light (C4 Magnum-GO, Rel®, Michigan), 4) Industrial Camera (Ace2 a2A1920-160uc, Basler®, Germany), and 5) 6mm f/1.8 Lens (C125-0618-5M, Basler®, Germany). The labeled setup is shown in Figure 5-16. The necessary fixtures to integrate the lighting and camera that were not available off the shelf were designed and 3D printed by the author to complete the hardware setup.

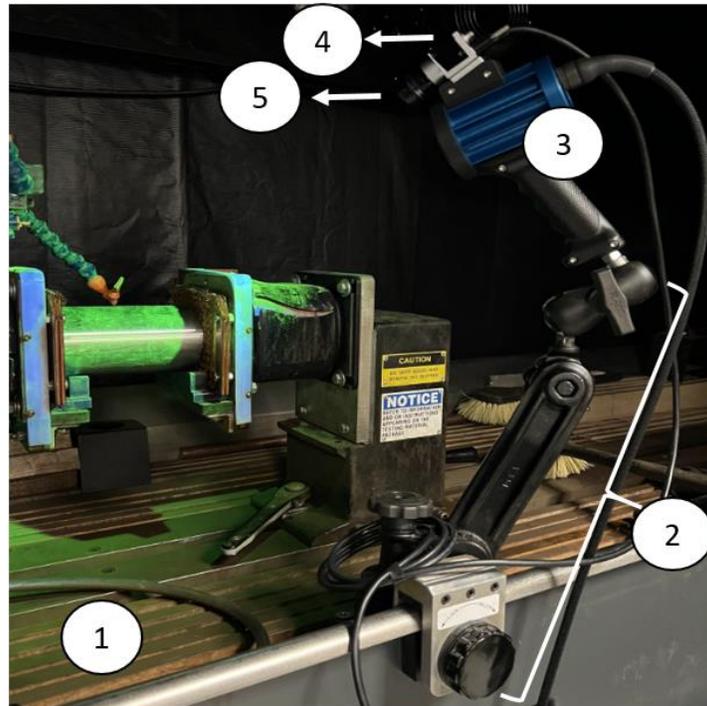


Figure 5-16. Live prediction hardware setup

Conclusions

The visual inspection step in the wet MPI process is crucial to ensure defects are identified and fixed prior to parts being used in the field. The findings in Chapter 3 have shown that the variability in the visual inspection process of wet MPI resulted in about a 70% defect escape rate. Because of the poor performance of human visual inspection, an alternative method to aid the visual inspection process was developed. The proposed method used deep learning to predict cracks in wet MPI based on images trained from a wide variety of parts across nine

different foundries. This resulted in an escape rate of 13% and an overkill rate of 32%. The results show a major improvement from the 70% escape rate of the current human visual inspection. Additionally, the hardware needed to enable real-time prediction at foundries is relatively inexpensive and simple to set up. Using the proposed software implementation in this paper, foundries have the option to further improve the model by providing additional training images for specific parts tested at their foundry to improve the model on their local system, eliminating confidentiality concerns.

Future research could involve testing different hardware integration technologies, e.g. augmented reality (AR) glasses, projectors, mobile phones, or robotics to improve the effectiveness of the inspector aid. This is important in wet MPI specifically because the darkroom environment means that the human eye will need some time to fully adapt to the changing light conditions. Creating a system that takes this into account will improve the effectiveness of the inspection. Additionally, further improvements to the accuracy of the model could be explored and benchmarked against the baseline model created in this study to further improve the performance of the artificially intelligent solution for wet MPI testing. Overall, this paper shows significant gains in reducing the escape rate of defects which supports the need for more research in the creation and implementation of a vision system in wet MPI.

References

- [1] S. M. Y. Lau, D. Eisenmann, and F. E. Peters, "Development of an Image Analysis Protocol to Define Noise in Wet Magnetic Particle Inspection," *Int. J. Met.*, vol. 15, pp. 1317–1325, Jan. 2021.
- [2] M. Blair, T. Stevens, and B. Linskey, Eds., *Steel Castings Handbook*, 6th ed. ASM International, 1995.
- [3] D. J. Eisenmann, D. D. Enyart, D. Kosaka, and C. Lo, "Fundamental Engineering Studies of Magnetic Particle Inspection and Impact on Standards and Industrial Practice," 2014.

- [4] J. E. See, C. G. Drury, A. Speed, A. Williams, and N. Khalandi, "The role of visual inspection in the 21st century," in *Proceedings of the Human Factors and Ergonomics Society*, 2017.
- [5] B. L. Luk and A. H. S. Chan, "Human Factors and Ergonomics in Dye Penetrant and Magnetic Particles Nondestructive Inspection Methods," *Adv. Ind. Eng. Oper. Res.*, vol. 0948, no. August, pp. 127–142, 2008.
- [6] C. G. Drury and J. G. Fox, "The imperfect inspector," in *Human reliability in quality control*, 1975.
- [7] A. Roth, Emily, Bisantz, "Chapter 14 Cognitive Work Analysis," in *Designing for People*
- [8] J. Schneider, K. Kluth, and H. Strasser, "Development and evaluation of specific lighting scenarios and ergonomic approaches to optimise visual inspection tasks," *Theor. Issues Ergon. Sci.*, 2011.
- [9] N. Rawashdeh, J. Abu-Khalaf, W. Khraisat, and S. Al-Hourani, "A visual inspection system of glass ampoule packaging defects: effect of lighting configurations," *Int. J. Comput. Integr. Manuf.*, 2018.
- [10] L. Campbell, R. Connor, J. Whitehead, and G. Washer, "Human factors affecting visual inspection of fatigue cracking in steel bridges," 2020.
- [11] J. E. See, "Visual Inspection: A Review of the Literature," *Sandia Rep.*, 2012.
- [12] J. e. Liu and F. P. An, "Image Classification Algorithm Based on Deep Learning-Kernel Function," *Sci. Program.*, 2020.
- [13] K. Tout, A. Meguenani, J. P. Urban, and C. Cudel, "Automated vision system for magnetic particle inspection of crankshafts using convolutional neural networks," *Int. J. Adv. Manuf. Technol.*, vol. 112, no. 11–12, pp. 3307–3326, 2021.
- [14] H. Fujita, M. Itagaki, K. Ichikawa, Y. K. Hooi, K. Kawahara, and A. Sarlan, "Fine-tuned Surface Object Detection Applying Pre-trained Mask R-CNN Models," 2020 *Int. Conf. Comput. Intell. ICCI 2020*, no. October, pp. 17–22, 2020.
- [15] H. Zeng, L. Li, Z. Cao, and L. Zhang, "Reliable and efficient image cropping: A grid anchor based approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019.
- [16] L. Celona, G. Ciocca, and P. Napoletano, "A grid anchor based cropping approach exploiting image aesthetics, geometric composition, and semantics," *Expert Syst. Appl.*, 2021.
- [17] X. Wang and Z. Hu, "Grid-based pavement crack analysis using deep learning," 2017 4th *Int. Conf. Transp. Inf. Safety, ICTIS 2017 - Proc.*, pp. 917–924, 2017.

- [18] A. Renyi, "On measures of entropy and information," in Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1961.
- [19] N. S. M. Raja, S. Arunmozhi, H. Lin, N. Dey, and V. Rajinikanth, "A Study on Segmentation of Leukocyte Image With Shannon's Entropy," 2018.
- [20] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016, 2016.
- [21] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2015.
- [22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., 2017.
- [23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," IEEE Trans. Pattern Anal. Mach. Intell., 2020.
- [24] P. Henderson and V. Ferrari, "End-to-end training of object class detectors for mean average precision," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017.
- [25] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A Review of Yolo Algorithm Developments," Procedia Comput. Sci., 2022.
- [26] Z. Cao, T. Liao, W. Song, Z. Chen, and C. Li, "Detecting the shuttlecock for a badminton robot: A YOLO based approach," Expert Syst. Appl., 2021.
- [27] W. Chen, H. Huang, S. Peng, C. Zhou, and C. Zhang, "YOLO-face: a real-time face detector," Vis. Comput.
- [28] Y. Xue and Y. Li, "A Fast Detection Method via Region-Based Fully Convolutional Neural Networks for Shield Tunnel Lining Defects," Comput. Civ. Infrastruct. Eng., vol. 33, no. 8, pp. 638–654, 2018.
- [29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," Comput. Vis. Pattern Recognit., 2016.

Appendix: Statement of Authorship

Conceptualization - Sharon Lau; methodology - Sharon Lau; software - Sharon Lau and Aron Mitchell; validation - Sharon Lau; formal analysis - Sharon Lau; investigation - Sharon

Lau; data curation - Sharon Lau; writing (original draft preparation) - Sharon Lau; writing (review and editing) - Sharon Lau, David Eisenmann, and Frank Peters.

CHAPTER 6. GENERAL CONCLUSION

Summary

The goal of this research was to improve the quality of steel castings through identifying measurement error in the wet magnetic particle inspection (MPI) process and developing solutions to improve the precision of the test method. Since there were no recent data evaluating the variability in wet MPI for steel foundries, the first step taken in this research was to conduct a percent match gauge repeatability and reproducibility (R&R) study. The gauge R&R study was conducted across four steel foundries where the average percent match repeatability and reproducibility were found to be 27% and 29%, respectively. The poor precision of the gauge R&R study further solidified the need for this research.

Several factors were identified that could lead to poor gauge R&R results. In Chapter 4, the effect of geometry on the magnetic field strength was first investigated. One advantage of metal castings is the ability to produce complex geometries. However, geometry is known to influence the magnetic field strength. Having adequate magnetic field strength in each region of the part is important to be able to attract sufficient magnetic particles to the area of the defect to illuminate the defect. To understand the impact of geometry on the magnetic field strength, flat and curved geometries were evaluated using a gauss meter. This study showed that concave regions tend to have lower magnetic field strength compared to flat and convex geometries. This is a concern because defects typically show up in concave regions. To increase the probability of locating defects in concave regions, a gauss meter should be used to ensure sufficient magnetic field strength is present in the concave regions and not just rely on the quantitative quality indicators (QQIs) that are placed on flat regions.

Additionally, Chapter 4 investigated the impact of surface roughness, orientation, current type, and magnetic field strength on the visibility of the defect. The noise area percentage (NAP) method was used to quantify level of difficulty in visually identifying a defect. The NAP method compares the green intensity of the defect with its surrounding area. Higher NAP values equate to higher difficulty in identifying a defect because the NAP values represent the percentage of pixels in the surrounding area that are brighter than the defect. Overall, the surface roughness and the orientation between the defect and magnetic field lines were the two factors that had a significant impact on the visibility of the defect for both surface-breaking and sub-surface defects.

Geometry is known to impact the direction of the magnetic field lines. Since the orientation between the defect and the magnetic field lines was found to be an important factor, further investigation into the effect of various geometries on the magnetic field direction was conducted. Four setups were tested: 1) flat geometry with flush contact points, 2) curved geometry with flush contact points 3) flat geometry with irregular contact points, and 4) curved geometry with irregular contact points. The results show that only for the flat geometry with flush contact points was the actual magnetic field direction the same as the expected magnetic field direction. This is important because if an operator is looking for a specific defect in an area, the magnetic field direction induced may not be the same as what the operator expects, which may lead to missed defects.

The current human visual inspection has an escape rate of around 70%. The poor performance of manual visual inspection indicates a need for an artificially intelligent vision system to aid the human inspector to find defects more reliably. Hence, in Chapter 5 an inspector aid was developed using deep learning to help the human inspector by providing predictions of

possible defects. A method for labeling using an entropy-driven grid-style labeling approach was used to optimize the accuracy and speed of the annotations. Additionally, data augmentation was used to increase the size of training and test images, further improving the model. Overall, the deep learning model was able to achieve an escape rate of 13% which is 57% better than the current manual visual inspection. An industrial camera, ultraviolet (UV) light, and mounts were integrated to complete the inspector aid system. The optics were optimized to work for the majority of casting sizes used on the wet horizontal benches in steel foundries.

Recommendation and Future Research

The criteria for defect detection for a specific part is typically agreed upon by the buyer and foundry. For example, any cracks larger 0.25 inch could be deemed unacceptable. However, the buyer and foundry should consider the minimum detectable defect size for the wet magnetic particle inspection (MPI) process when determining the criteria. Generally, the criteria for defect detection in metal castings is set by the criticality of the parts for safety and past experience with the part. There is less emphasis on identifying the smallest defect size that can be reliably located when the setting defect detection criteria in steel castings. In aerospace applications, the capability of the wet MPI method undergoes more scrutiny. A review article on the topic found twenty relevant articles summarizing the capability studies from 1960 to 2011 [1]. The review found that the smallest crack that the wet MPI process could reliably find for most implementations is 2.6 mm.

However, most of the studies conducted were based on a machined rectangular sample. Based on the work in chapter 4, surface roughness and geometry are known to impact the visibility of the defects. Since the past work in determining the minimum crack size that can be reliably detected did not include surface roughness and geometry as a factor, more research should be done to determine a more appropriate value for steel castings. However, conducting

probability of detection (POD) studies is not a feasible solution as can be inferred by the quantity of review articles covering this topic with only twenty relevant articles over the span of 51 years. POD studies are time consuming and expensive. As an alternative, the approach used in chapter 4 of this dissertation under the section “Impact of surface roughness, orientation, current type and magnetic field strength on the effectiveness of wet MPI” could be used to evaluate the capability of the wet MPI process given specific parameters. Research in this area will lead to increased awareness to set reasonable criteria for defect detection in wet MPI leading to a reduction in inspection time and improvements in quality control by identifying defects that the wet MPI process is capable of finding.

Additional factors that could impact defect visibility should be considered. Factors such as the defect characteristics by varying the depth, width, length, and type of the defects (e.g., linear, non-linear, subsurface, etc.) will show up differently in the wet MPI process. Along with investigating new types of defects, the orientation between the defect and the magnetic field direction, surface roughness, magnetic field strength, current type, and magnetization type should be considered. The experimental design should be carefully considered to optimize the tradeoff between time and number of combinations tested.

To further improve the capability of wet MPI, chapter 5 introduced an inspector aid which is an artificially intelligent solution to help the inspector identify defects. In chapter 5, the inspector aid developed was able to reduce the escape rate from 70% to 13%. However, the escape rate of 13% is assuming the inspector aid was used without the intervention of the human operator. The proposed implementation of the inspector aid in the steel foundries would be verified by the human inspector. It can be expected that the escape rate would be less than 13% since the human inspector will have a chance to catch any defects missed by the deep learning model. To verify

the true escape rate with the collaboration between the model and human inspector, a stress test needs to be done by setting up the inspector aid at a steel foundry for a couple of weeks.

To improve the performance of the inspector aid, the hardware setup can be further optimized. The setup proposed in Chapter 5 was optimized based on a fixed working distance because an existing off the shelf bench mount was used to fixture the lighting and optics. A custom fixture can be made to allow for more flexibility in the working distance which will allow for different sizes and positioning of parts to be captured. Additionally, one downside to the down-selected model, Faster R-CNN, is its slow prediction time. Because of this, the minimum requirement for the computational hardware would need to be set to ensure the solution is able to run in real-time. Further work into creating a lighter weight architecture with the same accuracy is also necessary to reduce the cost to setup this solution. Once, the inspector aid solution can be successfully deployed and scaled up to multiple foundries, the learnings from the implementation of the inspector aid in wet MPI can be transferred to dry MPI and florescent penetrant inspection (FPI).